Characterizing Pixel Tracking through the Lens of Disposable Email Services

Hang Hu, Peng Peng, Gang Wang

Computer Science
Virginia Tech
hanghu@vt.edu
Privacy #1: Online Activity → Real-world Identity

Email address is one of the most important online Personally Identifiable Information (PII)

Leaked email address can lead to real-life scandal

07/12/2015: Ashley Madison hacked
Website to look for an affair

08/18/2015: 32 million user email addresses released by hackers, many gov, mil and corporate addresses found

08/27/2015: Leaked users face blackmail threats
Privacy #2: Email Tracking → User Profiling

Using email tracking:
1. Business does user profiling for targeted ads (discrimination)
2. Phishers make more informed and flexible strategy

1x1 hidden tracking pixel
```
<img width=1 height=1 src="...">
```
Alternative: Disposable Email Services

• Instead of using real email address to register online services, use disposable email address for short-term usage

• Online activities are disconnected with the real-world identity
Research Questions

1. What do users use disposable email services for?
2. What are the potential risks for using disposable email services?

A measurement study
1. Chose 7 popular disposable email services
2. Monitored 70,000 disposable email inboxes
3. Collected 2.3 million emails from 210K sender domains

Use this large dataset of emails to study email tracking
Dataset: 7 Popular Disposable Email Services

Guerrillamail.com  Temp-mail.org  Mailsac.com  Mailfall.com  Maildrop.cc  Mailinator.com  Mailnesia.com

Processed 11 billion+ emails, with 100k+ emails/h going in

All emails @mailinator.com are public, readable, and discoverable by anyone at any time - and are automatically deleted after a few hours

Privacy Policy of Mailinator.com
Disposable Inboxes Are Publicly Shared

1. Disposable inbox is shared by multiple users
2. Popular usernames are used by more users thus receive more emails
Data Collection

- Get popular usernames from existing data breaches
- Use popular usernames to collect more email messages

10K Popular Usernames

7 Disposable Email Services

Online Services

We collected 2,332,544 messages from 210,373 sender domains during Oct. 2017 - Jan. 2018

Mail
“david”
...

Infer user activity from collected
# How Long Do They Keep Received Messages?

<table>
<thead>
<tr>
<th>Website</th>
<th>Claimed Time</th>
<th>Actual Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guerrillamail.com</td>
<td>1 hour</td>
<td>1 hour</td>
</tr>
<tr>
<td>Mailinator.com</td>
<td>A few hours</td>
<td>10.5 – 16.5 hours</td>
</tr>
<tr>
<td>Temp-mail.org</td>
<td>25 mins</td>
<td>3 hours</td>
</tr>
<tr>
<td>Maildrop.cc</td>
<td>Dynamic</td>
<td>24 hours</td>
</tr>
<tr>
<td>Mailnesia.com</td>
<td>Dynamic</td>
<td>12.6 – 13.1 hours</td>
</tr>
<tr>
<td>Mailfall.com</td>
<td>25 mins</td>
<td>&gt;= 30 days</td>
</tr>
<tr>
<td>Mailsac.com</td>
<td>Dynamic</td>
<td>19.9 – 20.7 days</td>
</tr>
</tbody>
</table>

This is what they say vs. This is what they actually do

Disposable email services don’t delete emails as quickly as promised.
What Are the Risky Usages?

<table>
<thead>
<tr>
<th>PII Type</th>
<th># Detected in Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit Card Number</td>
<td>1,399</td>
</tr>
<tr>
<td>Social Security Number (SSN)</td>
<td>926</td>
</tr>
<tr>
<td>Employer Identification Number (EIN)</td>
<td>701</td>
</tr>
</tbody>
</table>

- 3.7% (61,812) Registration
- 0.86% (14,715) Password Reset
- 0.75% (12,802) Authentication Code
- 94.8% (1,612,361) All unsolicited emails, newsletters, ads and notifications

Online accounts under this email can be hijacked via password reset.
Risky Usage: Case Study

4000+ emails from healthcare.gov
Account carries sensitive information

Emails from af.mil
Contain SSN and date of birth
Password reset is available

Receive all scanned PDF documents
(signed contract or other sensitive docs)
Use Real-world Dataset to Study Email Tracking

Sender: Facebook

First-party Tracking

If tracker is facebook.com

If tracker is google.com

Third-party Tracking

Send a request to the tracker
Tracking Detection

The `<img>` URL contains an identifier of the receiver

1. The ID is the email address of the receiver
   `<img src="https://xx.com?id=hanghu@vt.edu">`

2. The ID is the hash of the email address of the receiver
   `<img src="https://xx.com?id=MD5(hanghu@vt.edu)"`>

Or

The

32 hash functions
33,824 combinations of hash

`<img width=1 height=1 src="https://xx.com">`
Tracking Detection (Cont.): Handling Evasion

<table>
<thead>
<tr>
<th>Hidden Tracker</th>
<th># Emails</th>
<th># Direct Trackers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doubleclick.net</td>
<td>96,430</td>
<td>164</td>
</tr>
<tr>
<td>Adsrvr.org</td>
<td>48,858</td>
<td>130</td>
</tr>
<tr>
<td>Rlcdn.com</td>
<td>42,745</td>
<td>132</td>
</tr>
<tr>
<td>Pippio.com</td>
<td>41,140</td>
<td>59</td>
</tr>
<tr>
<td>Liadm.com</td>
<td>29,643</td>
<td>252</td>
</tr>
</tbody>
</table>

Top Hidden Trackers

Popular hidden trackers receive tracking information from a large number of direct trackers in real time.
Email Tracking Analysis

- How prevalent is email tracking?
- How prevalent is first-party and third-party tracking?

1. First-party tracking is more prevalent than third-party tracking
2. Overall only a small percentage (5.5%) of senders perform tracking

<table>
<thead>
<tr>
<th>Type</th>
<th>Total Tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>2,332,544</td>
</tr>
<tr>
<td>1st-party</td>
<td>573,244 (24.6%)</td>
</tr>
<tr>
<td>3rd-party</td>
<td>264,501</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># Senders</th>
<th>210,373</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st-party</td>
<td>11,688 (5.5%)</td>
</tr>
<tr>
<td>3rd-party</td>
<td>5,403</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># &lt;img&gt;</th>
<th>3,887,658</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st-party</td>
<td>509,419</td>
</tr>
<tr>
<td>3rd-party</td>
<td>179,223</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># Trackers</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st-party</td>
<td>13,563</td>
</tr>
<tr>
<td>3rd-party</td>
<td>5,381</td>
</tr>
</tbody>
</table>
Popular Services Are More Likely To Track You

We consider sender domains within Alexa top 10K as “popular” senders.

<table>
<thead>
<tr>
<th>Popular Services</th>
<th>Sender Count</th>
<th>% Tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Popular Senders</td>
<td>2,052 (1%)</td>
<td>46.9%</td>
</tr>
<tr>
<td>Non-popular Senders</td>
<td>208,321 (99%)</td>
<td>5.2%</td>
</tr>
</tbody>
</table>
Email Tracking VS. Web Tracking

Web tracking has been extensively studied [1, 2]
• Google is the top tracker, tracking 80% Alexa top 1 million websites
  Previously largest email tracking study [3]
• Emails from 902 senders

Email tracking:
1. **Is not** as prevalent as web tracking
   Only 5.5% of all sender domains are tracking receivers
2. **Is not** dominated by a single company
   Top 10 trackers cover only 31.8% of all senders who do tracking

[1] [EC’16] Understanding emerging threats to online advertising
[3] [PETS’18] I never signed up for this! Privacy implications of email tracking
Conclusion

• The first measurement study on disposable email services
  • Collected 2.3 million messages from 7 disposable email services
  • New understandings of what they are used for and risky usages

• Empirically analyzed email tracking activities
  • Prevalence of tracking activities
  • Evasive tracking methods

We hope our work can increase awareness of email tracking privacy concern and accelerate the defense and legislation deployment
Thank You
Dataset Bias

The dataset inevitably suffers from bias
Disposable email services aren’t representative of personal inboxes

Unique value of dataset from disposable email services
• Cover a wide range of online services (210,000+)
• Study email tracking from the perspective of online services instead of the perspective of email users
Email Tracking Countermeasure

- Email tracking blocker (like Adblocker)
- Image querying proxy
- Image pre-fetching + proxy
- Block all outgoing requests

<table>
<thead>
<tr>
<th></th>
<th>Web</th>
<th>Mobile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gmail</td>
<td>Proxy</td>
<td>Proxy</td>
</tr>
<tr>
<td>Outlook</td>
<td>Non-block</td>
<td>Non-block</td>
</tr>
<tr>
<td>Yahoo</td>
<td>Proxy</td>
<td>Proxy</td>
</tr>
<tr>
<td>iCloud</td>
<td>Non-block</td>
<td>Non-block</td>
</tr>
</tbody>
</table>
Disposable SMS Study

- Collected 386,327 messages from over 400 phone numbers in 28 countries [4]
- Evaluated security posture of benign services
- Characterized malicious behavior via SMS gateway

[4] [IEEE S&P’16] Sending out an SMS: Characterizing the Security of the SMS Ecosystem with Public Gateways
Ethical Considerations

• Study follows a prior study about disposable SMS messages [4]
• All messages collected are publicly available
• Removed all PII from collected messages
• Send emails to all inbox to offer an opportunity to opt out
• Didn’t access any account registered under disposable email addresses

[4] [IEEE S&P’16] Sending out an SMS: Characterizing the Security of the SMS Ecosystem with Public Gateways