PrivacyMail: Towards Transparency in Email Tracking

Max Maass, Stephan Schwär, Matthias Hollick – Secure Mobile Networking Lab, TU Darmstadt
Time to check Emails?
Three reasons to keep listening
Wait, Email tracking?
Wait, **Email** tracking?

### Tracking views
- Remote images
- Remote style sheets

### Tracking interactions
- Personalized links

### Linking identities
- Email is used on multiple devices
- Allows linking the advertising profiles
What’s the big deal?


The website knows [2]:
• If you opened the eMail
• When you opened the eMail
• Which device you used
  • Potentially linking it to other profiles online
• Which software you used
• Where you were (IP-based geolocation)

This data can also be shared with others using HTTP redirects for cookie syncing [1].
How can we detect it?

- **Static Analysis**

  Used in: [2, 3]

- **Dynamic Analysis**

  Used in: [1]
Three studies on **Email privacy**:

- Englehardt et al. [1]: Dynamic analysis of newsletters of popular websites. Find wide-spread tracking, information leakage. Also evaluate defensive measures.
- Xu et al. [2]: Static analysis of their own Email accounts and newsletters from top websites. Evaluated privacy risks. Also performed study about user acceptance of tracking.
- Hu et al. [3]: Static analysis of large corpus collected from disposable Email services. Also studies risks of using disposable Email systems.

Similar systems for **web privacy**:

- PrivacyScore.org [4, 5]
- Webbkoll.dataskydd.net [6]
Registering a Service

“I want to sign up example.com”

“Please use john.doe@domain.com”

“Hi, I am john.doe@domain.com”

“Please confirm your registration”

“Registration confirmed” (after manual inspection)

“Here’s your Newsletter”
Dynamic Analysis
Live Demo
Results – Third Party Prevalence

<table>
<thead>
<tr>
<th></th>
<th># of services</th>
<th>Percentage of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embeds on view</td>
<td>112</td>
<td>82 %</td>
</tr>
<tr>
<td>Embeds on click</td>
<td>104</td>
<td>76 %</td>
</tr>
<tr>
<td>Embeds either</td>
<td>116</td>
<td>85 %</td>
</tr>
</tbody>
</table>

Results from 136 newsletters, 10 208 Emails analyzed
Results – Third Party Prevalence

<table>
<thead>
<tr>
<th>Third Party</th>
<th>Embed Count</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>mailchimp.com</td>
<td>16</td>
<td>Tracker</td>
</tr>
<tr>
<td>googleapis.com</td>
<td>12</td>
<td>CDN</td>
</tr>
<tr>
<td>gstatic.com</td>
<td>12</td>
<td>CDN</td>
</tr>
<tr>
<td>list-manage.com</td>
<td>11</td>
<td>Tracker</td>
</tr>
<tr>
<td>srv2.de</td>
<td>10</td>
<td>Tracker</td>
</tr>
<tr>
<td>ioam.de</td>
<td>8</td>
<td>Tracker</td>
</tr>
<tr>
<td>cloudfront.net</td>
<td>6</td>
<td>CDN</td>
</tr>
<tr>
<td>amazonaws.com</td>
<td>6</td>
<td>CDN</td>
</tr>
<tr>
<td>exactag.com</td>
<td>4</td>
<td>Tracker</td>
</tr>
<tr>
<td>mojn.com</td>
<td>4</td>
<td>Tracker</td>
</tr>
</tbody>
</table>

Results from 136 newsletters, 10 208 Emails analyzed
Results – Cookie Syncing


2. http://p.liadm.com/imp?[...]&m=<MD5_1>&sh=<SHA1>&sh2=<SHA256>[...]&dom=<plaintext email domain>

3. http://i.liadm.com/s/h/33013?m=<MD5_1>&sh1=<SHA1>&sh2=<SHA256>[...]

4. http://i.liadm.com/s/h/33013?sh2=<SHA256>[...]&m=<MD5_1>&[...]&sh1=<SHA1>&previous_uuid=<UUID_1>


7. http://i.liadm.com/s/e/33013/0/<MD5_2>?mpid=7156&muid=<UUID_2>
## Results – Email Address Disclosure

<table>
<thead>
<tr>
<th>Leak Algorithm</th>
<th># Services</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD5</td>
<td>9</td>
<td>Expedia.de, asgoodasnew.com</td>
</tr>
<tr>
<td>URLencode</td>
<td>7</td>
<td>spd.de, humblebundle.com</td>
</tr>
<tr>
<td>SHA-256</td>
<td>6</td>
<td>Ticketmaster.de, lidl.de</td>
</tr>
<tr>
<td>Plaintext</td>
<td>5</td>
<td>spd.de, suedkurier.de</td>
</tr>
<tr>
<td>Base64</td>
<td>3</td>
<td>Expedia.de, booking.com</td>
</tr>
<tr>
<td>SHA-1</td>
<td>2</td>
<td>Fastcompany.com</td>
</tr>
</tbody>
</table>

**Leak Algorithm** | **# 3Ps**
--- | ---
MD5 | 15
URLencode | 12
SHA-256 | 10
Plaintext | 8
Base64 | 3
SHA-1 | 2

Results from 136 newsletters, 10 208 Emails analyzed
Decoding hashed Emails is hard, right?

Recover Encrypted Email Addresses
Versium’s Email Decryption, starting at $0.04 per email or $0.08 with consumer data append
Recover email addresses that have been encrypted using the most common hashing and encryption protocols, with more than a 70% success rate.

datafinder.com/products/email-recovery

developer.myacxiom.com/code/api/endpoints/hashed-entity
Results – A/B Testing

A/B testing detected by comparing related eMails (time, title, …)

3 sites use A/B testing, all of them online shops
Lessons Learned from Prior Email Privacy Research

- Lack of awareness in the general population [2]
- Useful defense mechanisms are missing
  - “Asking nicely” probably won’t work
  - Ad-blocking lists have bad coverage for Email tracking [1]
  - “Just use plaintext mail only” works for experts, but does not work for entire populations
  - “Don’t load remote content” defends against view-tracking, but not click-tracking
- We attempt transparency for online tracking [4], but had mixed success rate in the past [5]
Conclusion

- Web tracking is only part of the online privacy picture
- Email tracking should be considered a threat
- We provide a transparency system
- Feel free to use it: https://PrivacyMail.info/

- Problems? Ideas? Pull Requests?
  https://github.com/PrivacyMail/PrivacyMail
- Want access to the data? Contact me!
  mmaass [at] seemoo.tu-darmstadt.de
Literature and Acknowledgements


Part of this research was funded by the DFG as part of subproject C.1 within the RTG 2050 “Privacy and Trust for Mobile Users”. Image source: pixabay.com (public domain images)
Meet the Trackers – LiveIntent

Leverage proven, full-funnel features and strategies needed to power customer acquisition or retention.

“We’ve been able to generate a significant number of conversions with LiveIntent by driving quality site traffic and reaching our site visitors with relevant ads after they’ve left the site.”

Emily Larson Hausner
Nordstrom Rack, LiveIntent partner since 2017

Dynamic Campaign Optimization
Engage high-quality prospects using machine learning technology that considers everything from demographic to product offering.

CRM-Based Targeting and Suppression
Increase relevancy and efficiency by using your first-party CRM data to target or suppress existing customers and find new customers that look and act like them.

Behavioral and Product Remarketing
Convert in-market consumers with personalized, relevant creatives informed by intent, like category browsing or cart abandonment.

Source: liveintent.com/advertiser-solutions/
Meet the Trackers - MediaMath

Engage your customers across all devices
Manage campaigns across every channel from a single workflow so you can tell a consistent story to your customers that is relevant, efficient and targeted to achieve the goals your business needs.

See how:
OMNICHANNEL CAMPAIGN MANAGEMENT →
CONNECTEDID →

Find your best customers
Identify your existing and future customers with the greatest value, target them across every device and reach them on the premium ad inventory where they will see your message.

See how:
LOOKALIKE AUDIENCES →
AUDIENCE SCORING →
CONNECTEDID →
MEDIA MARKETPLACE →

Address the unique needs of your business
Unlock the full power of programmatic with professional services that can extend our offerings with custom solutions, education to bring your team up to speed if programmatic is new territory, and dedicated account management with experts that can help you transform your business with programmatic marketing.

See how:
PROFESSIONAL SERVICES →
NEW MARKETING INSTITUTE →

Measure the impact of your marketing
Apply the power of machine learning to understanding the true incremental impact of your marketing and use these observations to improve the performance of your current campaigns to maximize your marketing spend.

See how:
LIFT MEASUREMENT →
BRAIN INSIGHTS →
AUDIENCE INSIGHTS →

Use your data better
Your first-party data is your most valuable asset. Onboard, segment, activate and understand your audiences in a single platform without data-loss or latency.

See how:
DATA MANAGEMENT PLATFORM →
AUDIENCE ONBOARDING →
ADAPTIVE SEGMENTS →
AUDIENCE INSIGHTS →

Access premium ad inventory
Get in front of your best customers by securing the highest quality ad inventory in every channel. Over the past decade, MediaMath has built an ecosystem of the highest quality publishers and exchanges where you can get your message in front of your customers.

See how:
SUPPLY →
OMNICHANNEL CAMPAIGN MANAGEMENT →

Source: mediamath.com/brands/