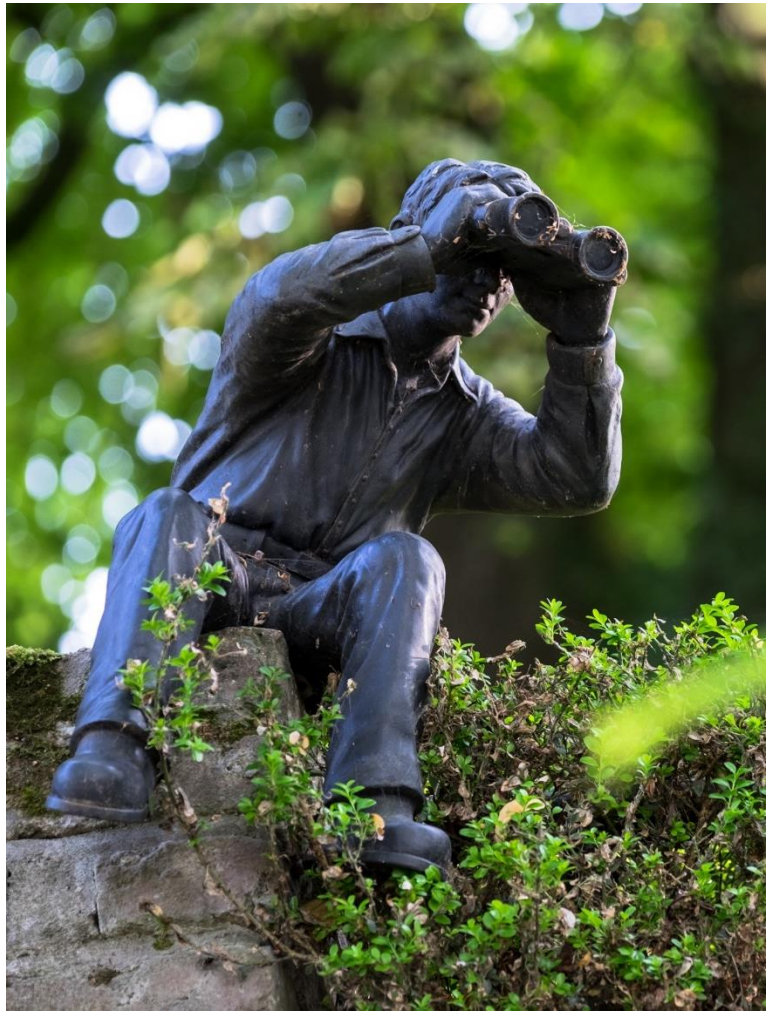
PrivacyMail: Towards Transparency in Email Tracking

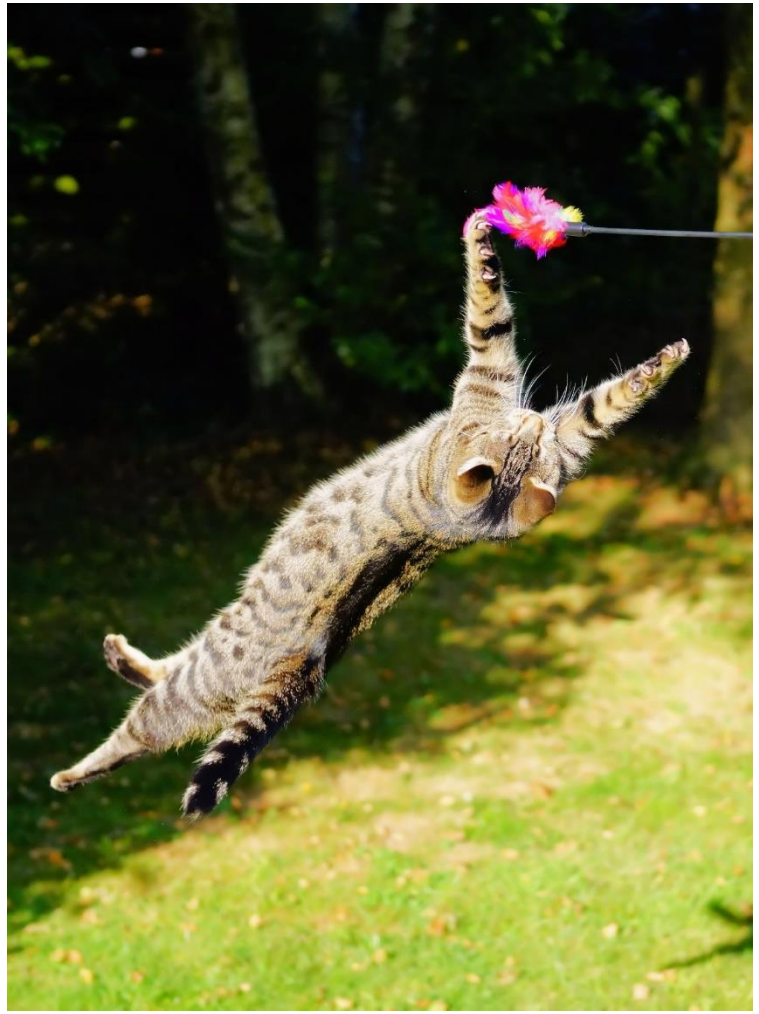
Max Maass, Stephan Schwär, Matthias Hollick – Secure Mobile Networking Lab, TU Darmstadt

Oh no, another online tracking talk?!



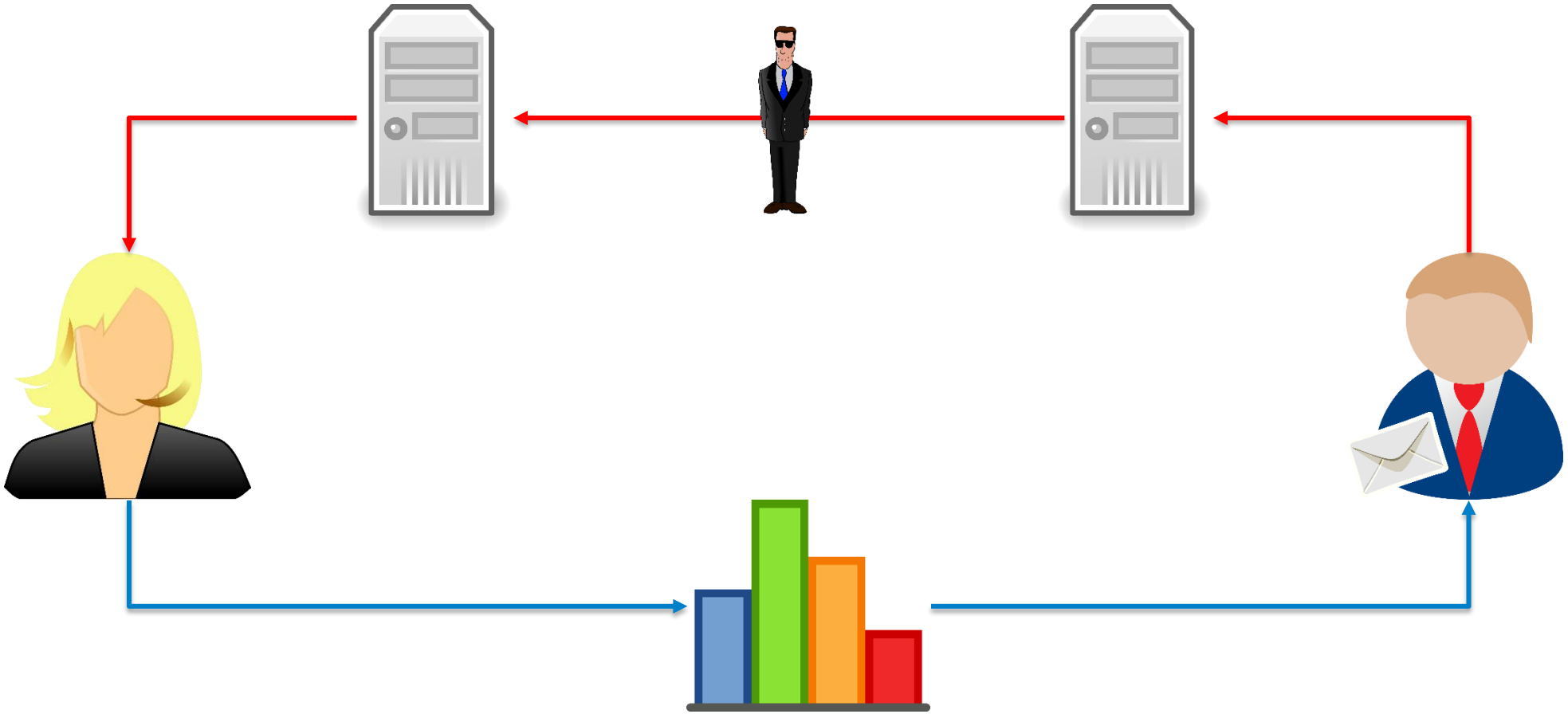
Slides & Paper: https://maass.xyz/talk/gpn2019/

Three reasons to keep listening



Slides & Paper: https://maass.xyz/talk/gpn2019/

Wait, Email tracking?



Slides & Paper: https://maass.xyz/talk/gpn2019/

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

Slides & Paper: https://maass.xyz/talk/gpn2019/

What's the big deal?

Tracking highly prevalent: between 24% [3] and 85% [1] of Emails contain tracking.

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].

Slides & Paper: https://maass.xyz/talk/gpn2019/

How can we detect it?

TECHNISCHE UNIVERSITÄT DARMSTADT

Static Analysis

Used in: [2, 3]

Dynamic Analysis

Used in: [1]

Slides & Paper: https://maass.xyz/talk/gpn2019/

Prior Work

Three studies on **Email privacy**:

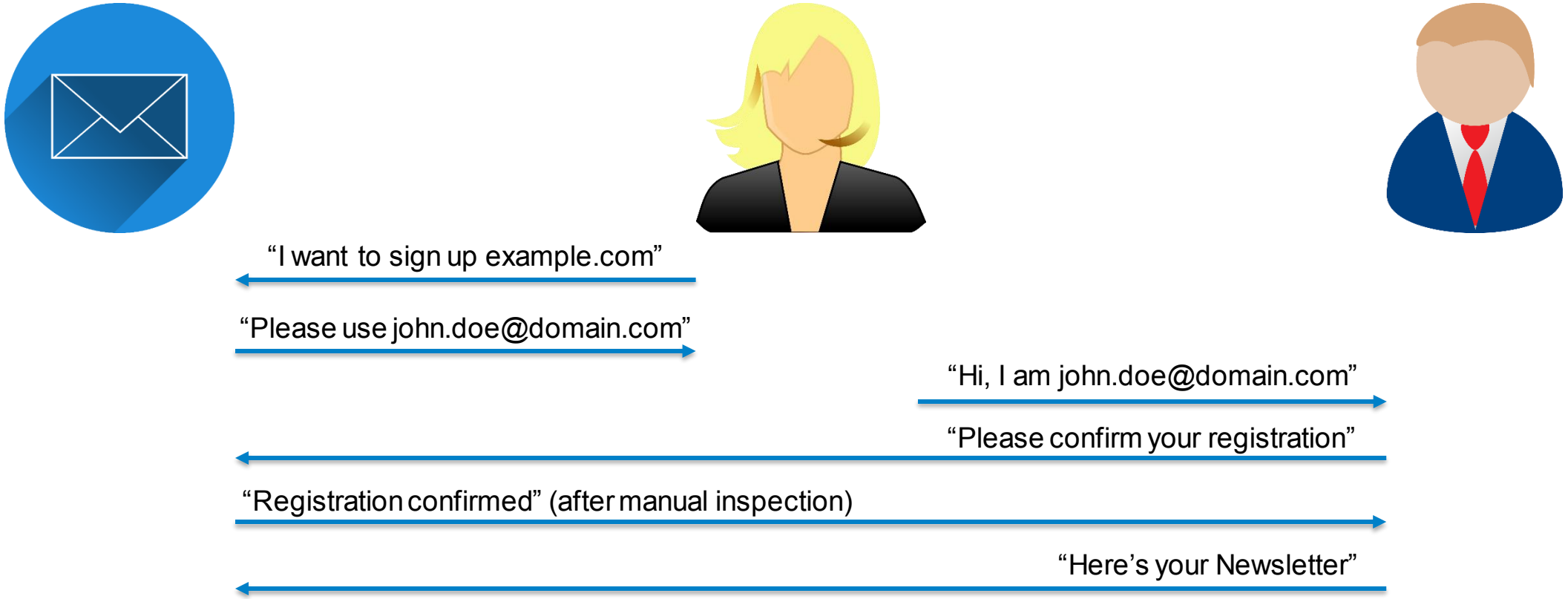
- Englehardt et al. [1]: Dynamic analysis of newsletters of popular websites. Find wide-spread tracking, information leakage. Also evaluate defensive measures.
- Xu et al. [2]: Static analysis of their own Email accounts and newsletters from top websites. Evaluated privacy risks. Also performed study 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]

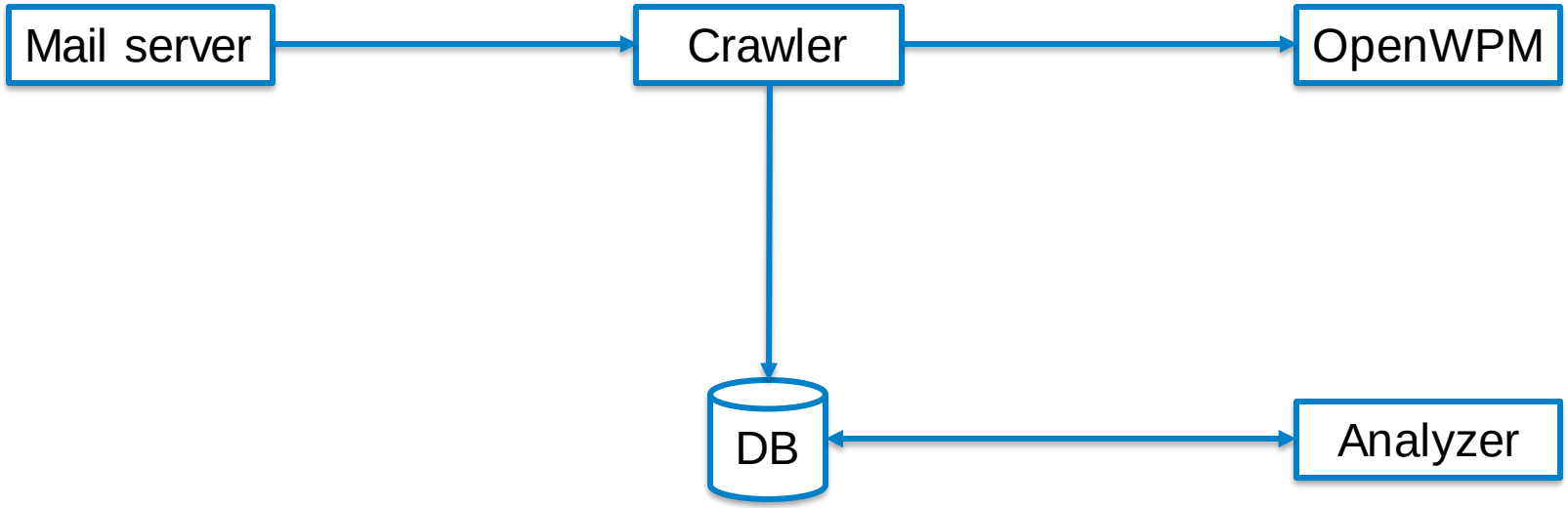
Slides & Paper: https://maass.xyz/talk/gpn2019/

Registering a Service



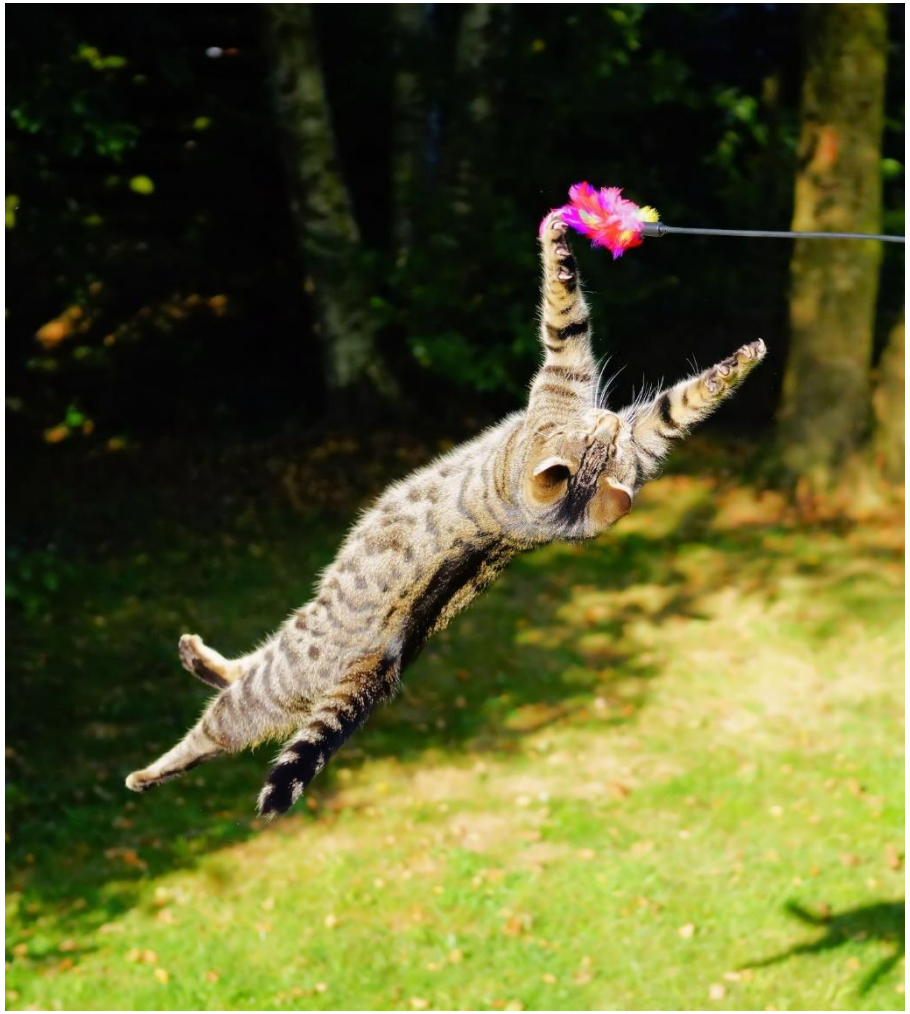
Slides & Paper: https://maass.xyz/talk/gpn2019/

Dynamic Analysis



Slides & Paper: https://maass.xyz/talk/gpn2019/

Live Demo



Slides & Paper: https://maass.xyz/talk/gpn2019/

Results – Third Party Prevalence

	# of services	Percentage of total
Embeds on view	112	82 %
Embeds on click	104	76 %
Embeds either	116	85 %

Results from 136 newsletters, 10 208 Emails analyzed

Slides & Paper: https://maass.xyz/talk/gpn2019/

Results – Third Party Prevalence

Third Party	Embed Count	Type
mailchimp.com	16	Tracker
googleapis.com	12	CDN
gstatic.com	12	CDN
list-manage.com	11	Tracker
srv2.de	10	Tracker
ioam.de	8	Tracker
cloudfront.net	6	CDN
amazonaws.com	6	CDN
exactag.com	4	Tracker
mojn.com	4	Tracker

Results from 136 newsletters, 10 208 Emails analyzed

Slides & Paper: https://maass.xyz/talk/gpn2019/

Results – Cookie Syncing

TECHNISCHE UNIVERSITÄT DARMSTADT

1. http://li.fastcompany.com/imp?[...]&e=**<plaintext email address>**&p=20182

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$_1$>**

5. http://sync.mathtag.com/sync/img?mt_exid=36&redir=http%3A%2F%2Fi.liadm.com%2Fs%2Fe%2F33013%2F0%2F**<MD5$_3$>**%3Fmpid%3D7156%26muid%3D%5BMM_UUID%5D&licd=27296&previous_uuid=**<MD5$_3$>**

6. http://sync.mathtag.com/sync/img?mt_exid=36&redir=http%3A%2F%2Fi.liadm.com%2Fs%2Fe%2F33013%2F0%2F**<MD5$_3$>**%3Fmpid%3D7156%26muid%3D%5BMM_UUID%5D&licd=27296&previous_uuid=**<MD5$_3$>**&mm_bnc&mm_bct

7. http://i.liadm.com/s/e/33013/0/**<MD5$_3$>**?mpid=7156&muid=**<UUID$_2$>**

Slides & Paper: https://maass.xyz/talk/gpn2019/

Results – Email Address Disclosure

Leak Algorithm	# Services	Examples
MD5	9	Expedia.de, asgoodasnew.com
URLencode	7	spd.de, humblebundle.com
SHA-256	6	Ticketmaster.de, lidl.de
Plaintext	5	spd.de, suedkurier.de
Base64	3	Expedia.de, booking.com
SHA-1	2	Fastcompany.com

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

Slides & Paper: https://maass.xyz/talk/gpn2019/

Decoding hashed Emails is hard, right?

datafinder.com/products/email-recovery

acxiom

ENTITY REPRESENTATION LINK (ERL) CONSTRUCTION

SHA1 is the current algorithm supported for generating the hash for ERLs. ERLs are constructed by first constructing the ER as described in the Entity Representation section of this document, then executing an SHA1 algorithm against that ER string to get the 40 byte representation of the SHA1 hash.

For example, the ER of "charlesjohnson2test@acxiom.com" would result in an ERL (SHA1 value) of "b2ff8b257933796a2e105593c8f570f73784f45c".

We also support MD5 and SHA256 for email only lookups.

ENTITY REPRESENTATION LINK SUPPORTED PERMUTATIONS

Email

PhoneNumber

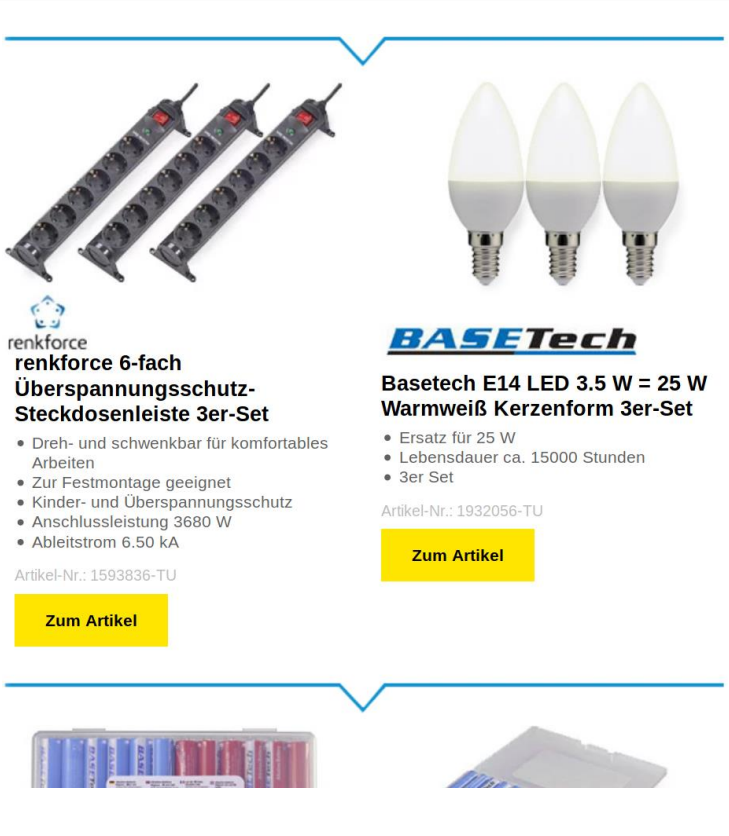
developer.myacxiom.com/code/api/endpoints/hashed-entity

Slides & Paper: https://maass.xyz/talk/gpn2019/

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



Slides & Paper: https://maass.xyz/talk/gpn2019/

Lessons Learned from Prior Email Privacy Research

TECHNISCHE UNIVERSITÄT DARMSTADT

- 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]

Slides & Paper: https://maass.xyz/talk/gpn2019/

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

Slides & Paper: https://maass.xyz/talk/gpn2019/

Literature and Acknowledgements

[1] Englehardt, S., Han, J., Narayanan, A.: **I never signed up for this! Privacy implications of email tracking**. Proc. Priv. Enhancing Technol. (2018)

[2] Xu, H., Hao, S., Sari, A., Wang, H.: **Privacy Risk Assessment on Email Tracking**. In: IEEE INFOCOM. (2018).

[3] Hu, H., Peng, P., Wang, G.: **Characterizing Pixel Tracking through the Lens of Disposable Email Services**. In: IEEE Security & Privacy. (2019).

[4] Maass, M., Wichmann, P., Pridöhl, H., Herrmann, D.: **PrivacyScore: Improving Privacy and Security via Crowd-sourced Benchmarks of Websites**. Lect. Notes Comput. Sci. 10518 LNCS (2017).

[5] Maass, M., Walter, N., Herrmann, D., Hollick, M.: **On the Difficulties of Incentivizing Online Privacy through Transparency: A Qualitative Survey of the German Health Insurance Market**. In: 14. Internationale Tagung Wirtschaftsinformatik (2019).

[6] Andersdotter, A., Jensen-Urstad, A.: **Evaluating Websites and Their Adherence to Data Protection Principles: Tools and Experiences**. In: IFIP Advances in Information and Communication Technology. (2016).

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)