

# I never signed up for this!

Privacy implications of email tracking

**Steven Englehardt**

@s\_englehardt  
senglehardt.com

Joint work with:  
Jeffrey Han and  
Arvind Narayanan

Inbox Fw: CLEARANCE: Up to 60...

Get Messages Write Chat Address Book Tag QuickFilter Search <⌘K>

From Steven Englehardt <[REDACTED]>  
Subject Fw: CLEARANCE: Up to 60% off sheets, please  
To Me

To protect your privacy, Thunderbird has blocked remote content in this message.

To protect your privacy, Thunderbird has blocked remote content in this message. Preferences

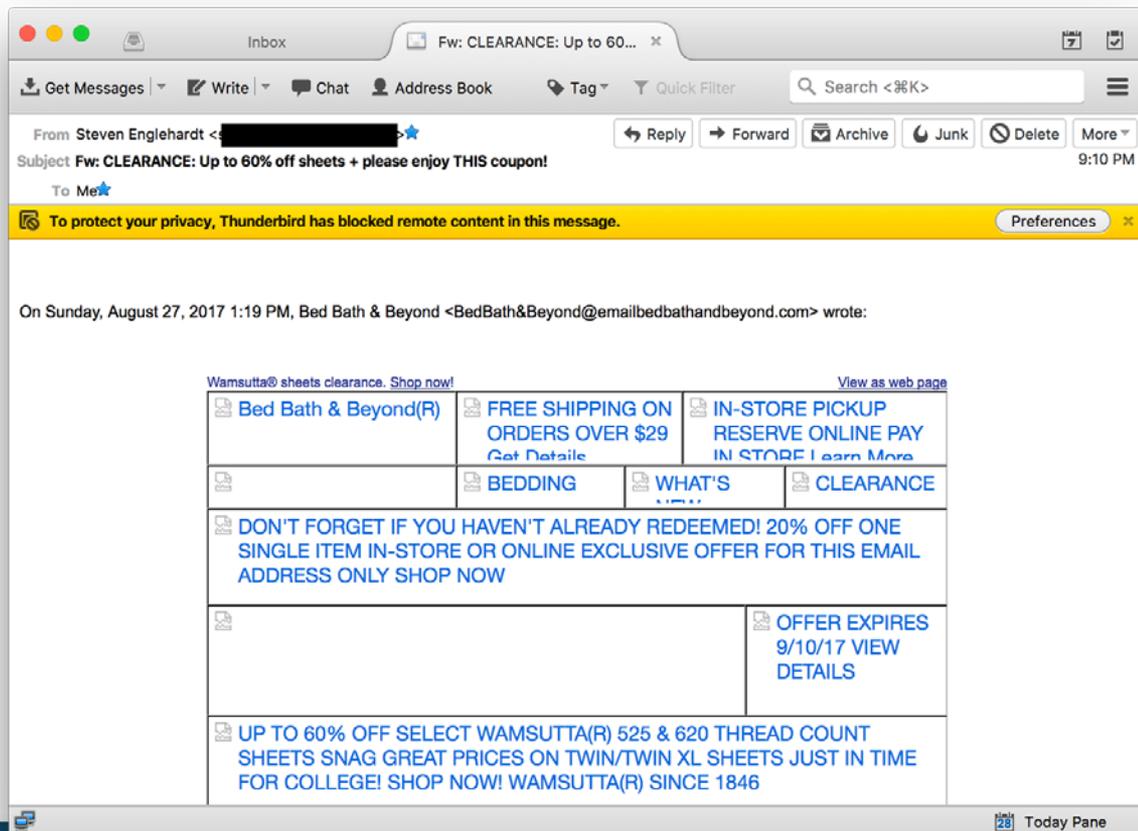
On Sunday, August 27, 2017 1:19 PM, Bed Bath & Beyond <BedBath&Beyond@emailbedbathandbeyond.com> wrote:

Wamsutta® sheets clearance. [Shop now!](#) [View as web page](#)

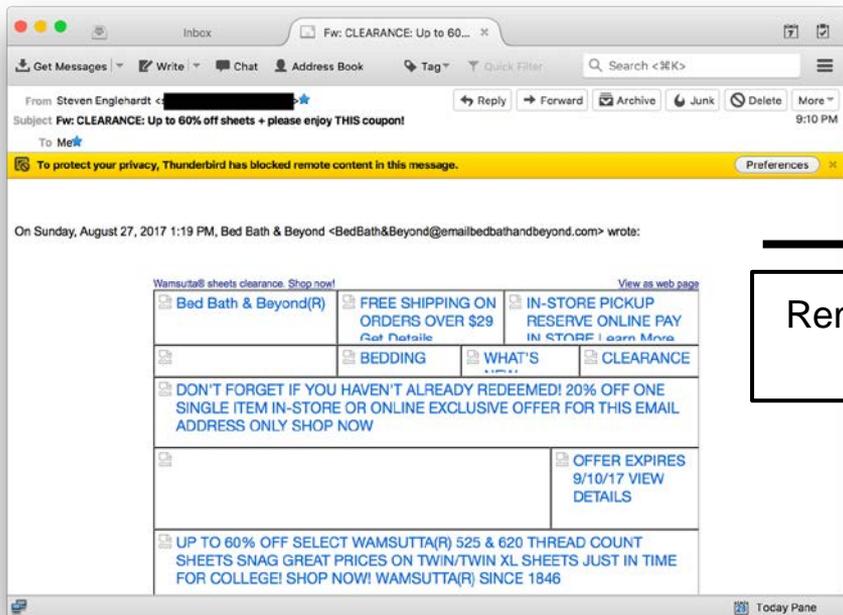
|  |  |  |
|--|--|--|
| Bed Bath & Beyond(R)   | FREE SHIPPING ON ORDERS OVER \$29<br><a href="#">Get Details</a> | IN-STORE PICKUP RESERVE ONLINE PAY IN STORE   <a href="#">Learn More</a> |
| BEDDING  | WHAT'S   | CLEARANCE  |
| DON'T FORGET IF YOU HAVEN'T ALREADY REDEEMED! 20% OFF ONE SINGLE ITEM IN-STORE OR ONLINE EXCLUSIVE OFFER FOR THIS EMAIL ADDRESS ONLY SHOP NOW                      |  |  |
| OFFER EXPIRES 9/10/17 VIEW DETAILS   |  |  |
| UP TO 60% OFF SELECT WAMSUTTA(R) 525 & 620 THREAD COUNT SHEETS SNAG GREAT PRICES ON TWIN/TWIN XL SHEETS JUST IN TIME FOR COLLEGE! SHOP NOW! WAMSUTTA(R) SINCE 1846 |  |  |

Today Pane

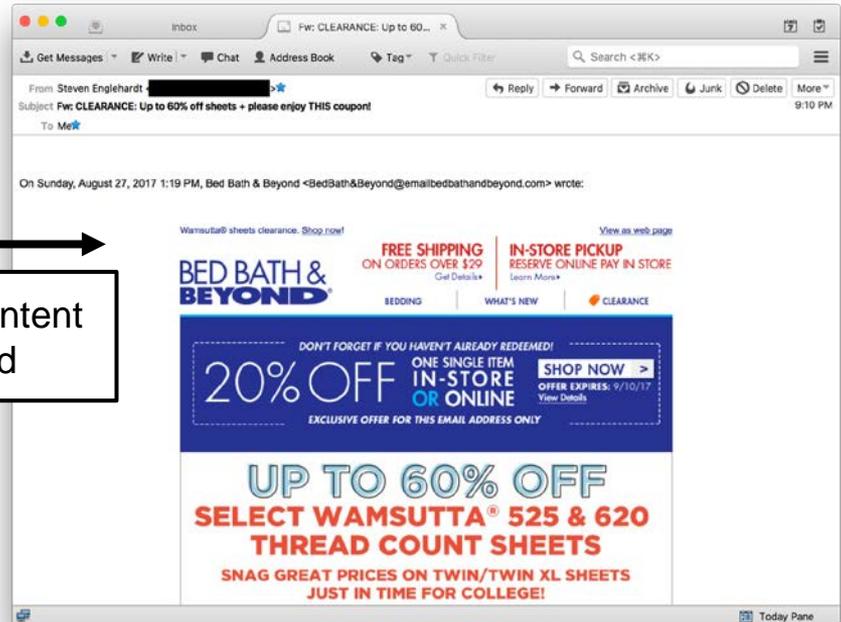
Who *doesn't*  
love reading  
email like this?



# What are the privacy implications?



Remote content enabled



Start Email Tracking Today | HubSpot Sales Software - Mozilla Firefox

Start Email Tracking Today x +

← → <https://www.hubspot.com/products/sales/email-tracking> Search

English Sales: 877 650 8270 Sign In Get started

HubSpot SALES HUB Software Pricing Resources Partners About

# Email Tracking

Know the second a lead opens an email, send a perfectly timed follow-up, and close deals faster than ever.

Priority

 **Adam Carpenter** Priority  
Opened email Your Biglytics Trial  
> 2 opens OPEN

Today

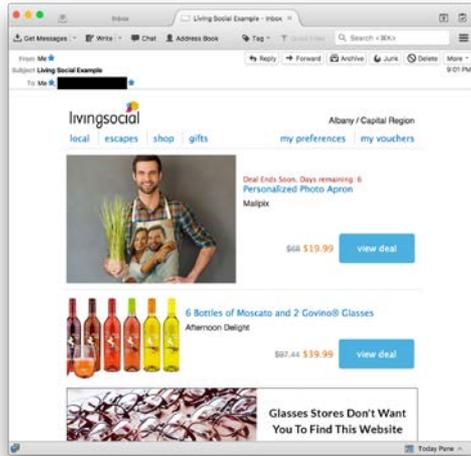
 **Adam Carpenter** 8:53 am  
Opened email Your Biglytics Trial  
> 2 opens OPEN

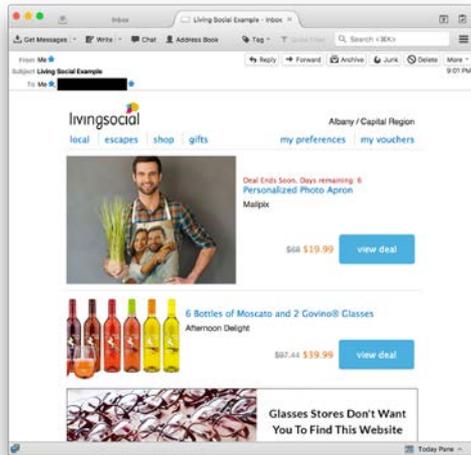
Yesterday

 **Someone** Yesterday 1:45 pm  
Opened email The Future of Big Data | LIVE Webinar  
> 1 open OPEN

PRIVACY

# Emails are tracked far beyond send tracking





Your device contacts 24 companies  
→ 20 can track you (if supported)  
→ 10 receive an MD5 hash of your email address

Receives MD5(email address) & Sets a Cookie

- American List Counsel** (alcmpn.com)
- LivelIntent** (liadm.com)
- Oracle** (nexac.com)
- Axiom** (rlcdn.com, pippio.com, axiom-online.com)
- Criteo** (criteo.com)
- Conversant Media** (dotomi.com)
- V12 Data** (v12group.com)
- VideoAmp** (videoamp.com)
- <Unknown>** (alocdn.com)

Sets a Cookie

- OpenX** (openx.net)
- comScore** (scorecardresearch.com, voicefive.com)
- Oracle** (bluekai.com)
- Google** (doubleclick.net)
- Realtime Targeting Aps** (mojn.com)

- MediaMath** (mathtag.com)
- TapAd** (tapad.com)
- IPONWEB** (bidswitch.net)
- AOL** (advertising.com)
- Centro** (sitescout.com)
- The Trade Desk** (adsrvr.org)
- Adobe** (demdex.net)

Receives MD5(email addr.)

- Criteo** (emailretargeting.com)
- Neustar** (agkn.com)

Receives Bare Request

- LivelIntent** (licasd.com)
- Google** (2mdn.net)
- Akamai** (akamai.net)



Email  
Tracking

≈

Web  
Tracking

-

Javascript

# Measuring email tracking at scale

**Sign up for email & get 25% off\***

Email, please

Confirm your email

**SIGN UP NOW**

\*Valid for first-time registrants only & applies to reg. price items only. [Privacy Policy](#)

1. Crawled 15,700 sites
2. Signed up for mailing lists
3. Received 13,000 emails from ~900 sites
4. Measured tracking with OpenWPM

# Our Findings

# Many of the top web trackers are in emails

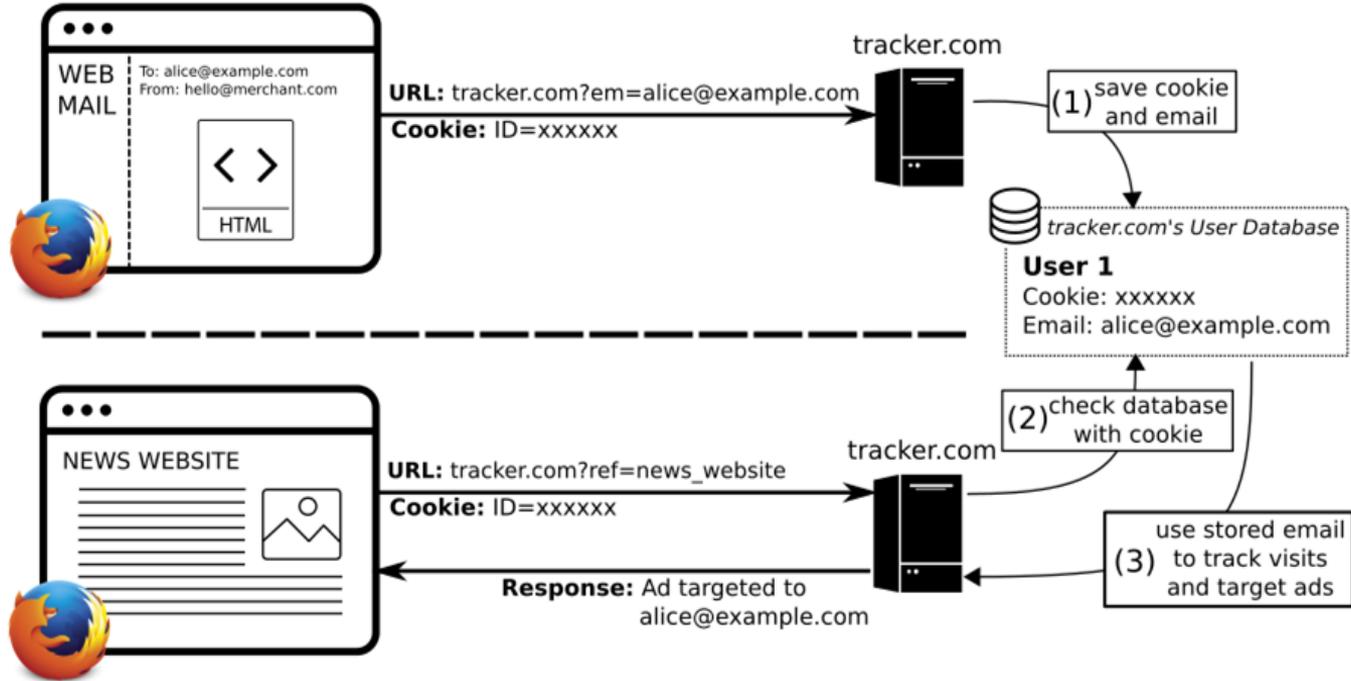
| Domain               | % of Emails | % of Top 1M |
|----------------------|-------------|-------------|
| doubleclick.net      | 22.2        | 47.5        |
| mathtag.com          | 14.2        | 7.9         |
| dotomi.com           | 12.7        | 3.5         |
| adnxs.com            | 12.2        | 13.2        |
| tapad.com            | 11.0        | 2.6         |
| liadm.com            | 11.0        | 0.4         |
| returnpath.net       | 11.0        | <0.1        |
| bidswitch.net        | 10.5        | 4.9         |
| fonts.googleapis.com | 10.2        | 39.4        |
| list-manage.com      | 10.1        | <0.1        |

85% of emails embed third parties (with an average of 5 per email)

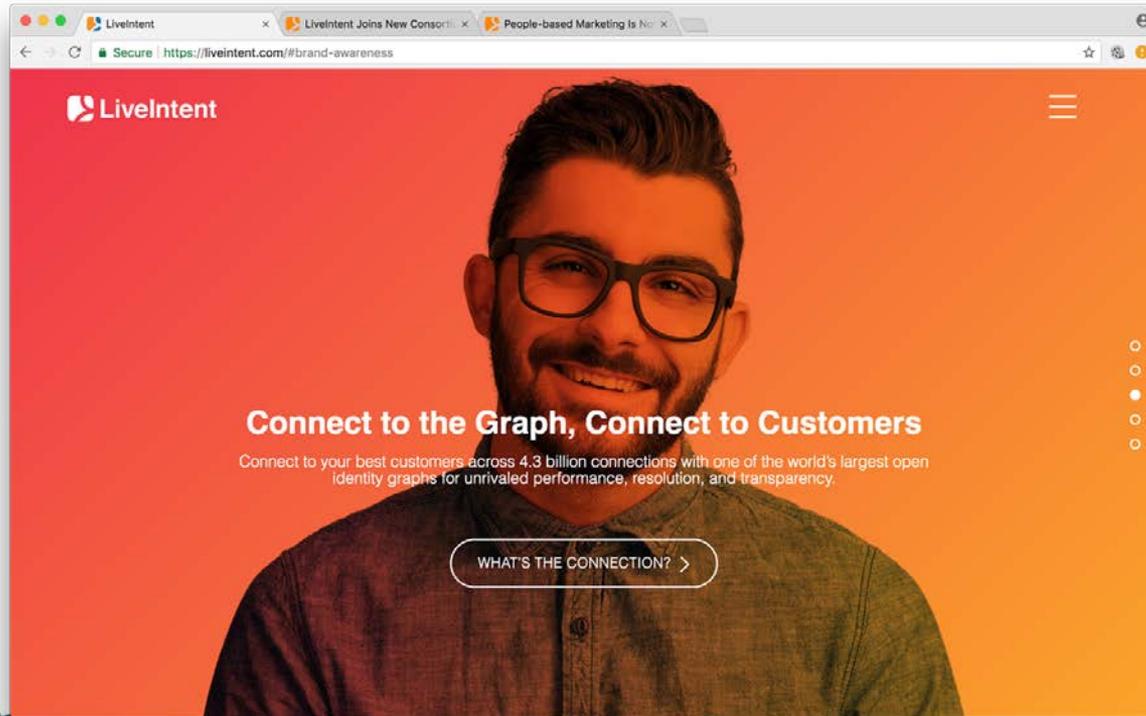
# 29% of emails ( from 19% of senders) leak the email address to third parties

| Leak                | # of Senders | # of Recipients |
|---------------------|--------------|-----------------|
| MD5                 | 100          | 38              |
| SHA1                | 64           | 19              |
| SHA256              | 69           | 13              |
| Plaintext Domain    | 55           | 2               |
| Plaintext Address   | 77           | 54              |
| URL Encoded Address | 6            | 8               |
| SHA1 of MD5*        | 1            | 1               |
| SHA256 of MD5*      | 1            | 1               |
| MD5 of MD5*         | 1            | 1               |
| SHA384              | 1            | 1               |

# Trackers can correlate email and web tracking



# “People-based” Marketing



# LiveIntent Blog Post

Source: <https://blog.liveintent.com/people-based-marketing-not-complicated/>

As an identifier, email is both deterministic and persistent. That is, when a consumer gives out a verified email, it usually belongs to only that consumer. That can't be said of all typical advertising identifiers. Cookies, for example, live on desktop browsers that are often shared with no way to distinguish who's using it. And whereas email is cross-device, cookies aren't.

# LiveIntent Privacy Policy

Source: <https://liveintent.com/services-privacy-policy>

LiveIntent may also receive non-personal information from online and offline sources, including the types described below, from our business partners

# LiveIntent Privacy Policy

Source: <https://liveintent.com/services-privacy-policy>

To de-identify this information, either we or our business partners [hash it].

# Criteo Privacy Policy

Source: <https://www.criteo.com/privacy/>

we use a double hashing method ... to ensure the non-reversibility of your information. A hash of your email corresponds to a series of characters that does not permit your identification.

# Does hashing protect user privacy?



Tracker Database

## Email Hash

b5184f3fb0fe35e4319b729f05017f6e

## Tracking Data

- <https://www.webmd.com/cancer/default.htm>
- <http://www.foxnews.com/>
- LivingSocial *Healthy Living* email campaign
- \$105 in *Personal Health* purchases from CVS
- \$55 purchase from Babies"R"Us

# Does hashing protect user privacy?



Tracker Database

## Email Hash

b5184f3fb0fe35e4319b729f05017f6e

## Tracking Data

- <https://www.webmd.com/cancer/default.htm>
- <http://www.foxnews.com/>
- LivingSocial *Healthy Living* email campaign
- \$105 in *Personal Health* purchases from CVS
- \$55 purchase from Babies"R"Us

Run a “re-identification attack” yourself! Open your terminal and enter:

```
Linux: echo -n ste@cs.princeton.edu | md5sum
```

```
MacOS: echo -n ste@cs.princeton.edu | md5
```

Easy

ste@cs.princeton.edu → b5184f3fb0fe35e4319b729f05017f6e

Hard

b5184f3fb0fe35e4319b729f05017f6e → ste@cs.princeton.edu

Easy

ste@cs.princeton.edu → b5184f3fb0fe35e4319b729f05017f6e

~~Hard~~

b5184f3fb0fe35e4319b729f05017f6e → ste@cs.princeton.edu

Easy (when you can guess the possible inputs)



16eaf6d2cef77e145db18804d2aa4fd56e



16eaf6d2cef77e145db18804d2aa4fd56e

jh34@alumni.princeton.edu →  
261495fd24d108b3c573527b3854af00  
ste@cs.princeton.edu →  
b5184f3fb0fe35e4319b729f05017f6e  
arvindn@cs.princeton.edu →  
16eaf6d2cef77e145db18804d2aa4fd5

# Email addresses aren't secrets!

Use email database leaks...



The screenshot shows a news article header with social media icons (M, 4i@, Twitter, Facebook) and buttons for 'Sign in' and 'Get started'. The main headline reads '1.4 Billion Clear Text Credentials Discovered in a Single Database'. Below the headline is a dark banner with a red diagonal watermark that says 'LARGEST CREDENTIAL BREACH EXPOSURE'. The background of the banner features a grid of red and white text, likely representing a list of leaked credentials.

...and just guess the rest.

GPU cloud computer: \$24.48 / hour  
→ 450 billion MD5 hashes / second

~4.7 billion email addresses total. If we generate a real address every 1 in 1 million guesses, **we can generate the entire space for less than \$75.**

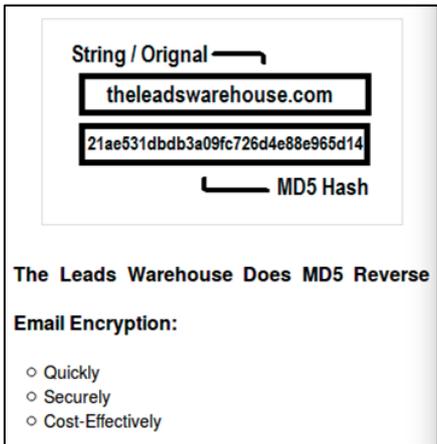
Past research recovered 45-70% of emails.

**More info:** <https://freedom-to-tinker.com/2017/09/28/i-never-signed-up-for-this-privacy-implications-of-email-tracking/>

**The pitfalls of hashing for privacy.** <https://www.comp.nus.edu.sg/~amrit/papers/pitfalls.pdf>

# Don't want to guess? Reverse hashes for \$0.04/email

theleadwarehouse.com



String / Original →

theleadwarehouse.com

21ae531dbdb3a09fc726d4e88e965d14

← MD5 Hash

**The Leads Warehouse Does MD5 Reverse**

**Email Encryption:**

- Quickly
- Securely
- Cost-Effectively

infutor.com



**Data Snapshot: MD5 and SHA1 Email Identification and Use Cases**

What is MD5 and SHA1?

MD5 and SHA1 are algorithms used to verify data integrity. Originally created for online security applications to verify data integrity, the MD5 (Message Digest 5) and SHA1 (Secure

datafinder.com



**Datafinder**  
Automated Data Intelligence

Login Signup

## 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.

# Takeaways

1. The line between email and web tracking is blurry
2. Email addresses are commonly leaked to trackers in emails
3. Claims of “de-identification” are suspect

## More Info

- **Full paper:**  
[https://senglehardt.com/papers/pets18\\_email\\_tracking.pdf](https://senglehardt.com/papers/pets18_email_tracking.pdf)
- **More on identity leaks:** <https://freedom-to-tinker.com/tag/noboundaries/>