

# 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

I'm now at...



PRINCETON  
UNIVERSITY



CENTER FOR  
INFORMATION  
TECHNOLOGY  
POLICY  
PRINCETON UNIVERSITY

● LABOR DAY DEALS: Up to 75% Off What You Want N

● Century 21 Department Stores <service@shop.c21stores.com>

To [REDACTED] Today at 8:17 AM

This message contains blocked images. Show Images Change this setting

This message contains blocked images. Show Images Change this setting

+ SAVE 15% With A New C21STATUS Credit Card Once Approved [Shop Now](#)  
Having trouble viewing this email? [View in Browser](#)

Free Shipping on orders \$75+

Century 21 Women Shoes Handbags and Accessories Gift Cards Men Kids Home  
Department Stores

Apply Now

Compose

Archive Move Delete Spam More

YAHOO! MAIL

All Steven Englehardt, search your mailbox Search Mail Search Web Home Steven

Add Gmail, Outlook, AOL and more

Inbox (205)

Drafts

Sent

Archive

Spam (19)

Trash

Smart Views

Important

Unread

Starred

People

Social

Shopping

Travel

Finance

Folders

Recent

YAHOO! MAIL

Inbox Fw: CLEARANCE: Up to 60...   

Get Messages | Write | Chat | Address Book | Tag | Quick Filter | 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

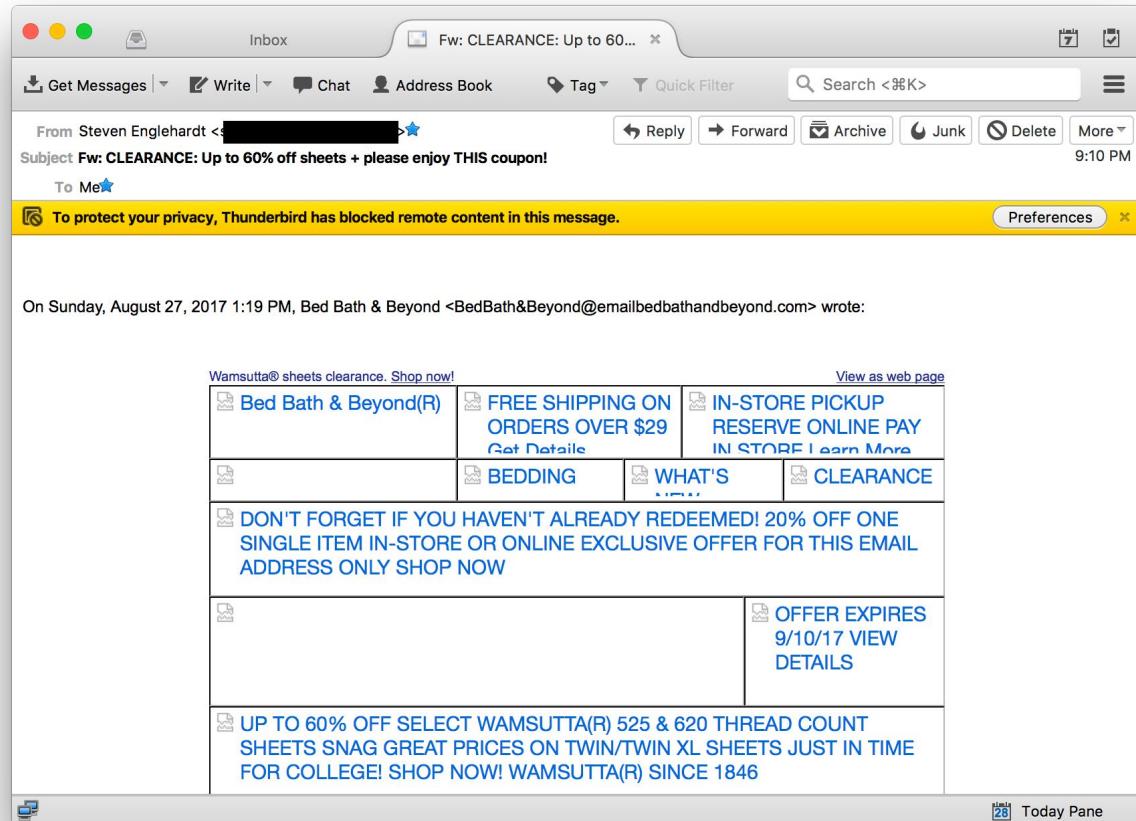
On Sunday, August 27, 2017 1:19 PM, Bed Bath & Beyond <BedBath&Beyond@emailbedbathandbeyond.com> wrote:

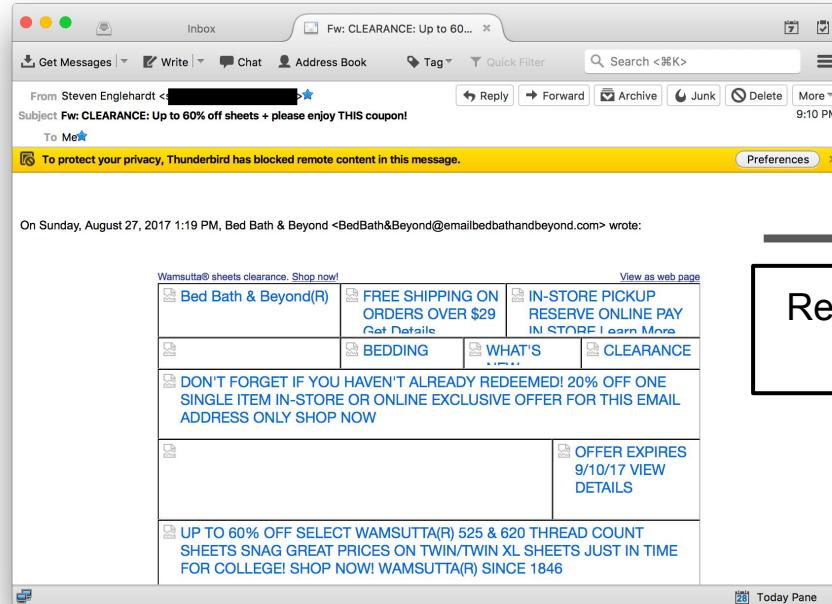
Wamsutta® sheets clearance. Shop now! View as web page

 Bed Bath & Beyond(R)	 FREE SHIPPING ON ORDERS OVER \$29 Get Details	 IN-STORE PICKUP RESERVE ONLINE PAY IN STORE <a href="#">Learn More</a>
	 BEDDING  WHAT'S NEW  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 <a href="#">VIEW DETAILS</a>
 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		

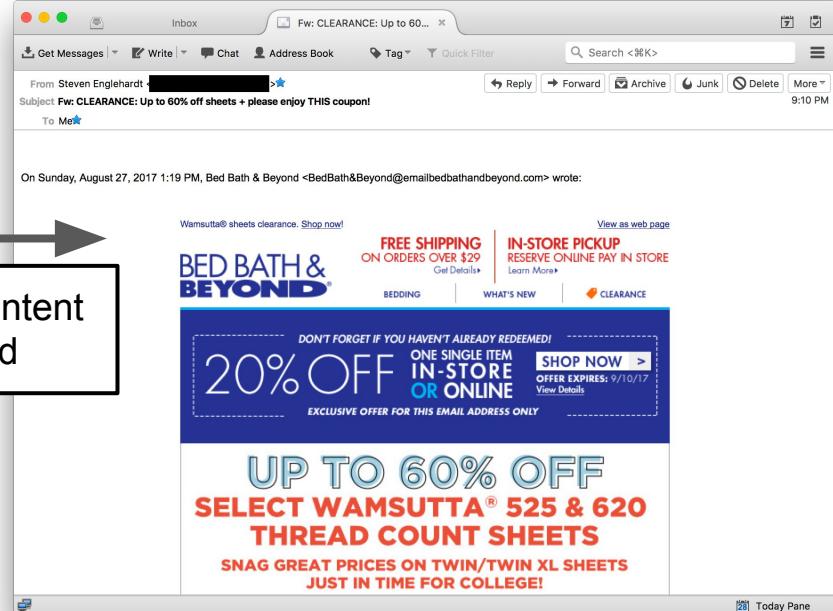
28 Today Pane

# Many emails are completely unreadable without remote content!





Remote content enabled



# What are the privacy implications?

Inbox Remote Content in Messag... ×

# Thunderbird

EDITING TOOLS

EXPLORE MORE TOPICS

TIPS AND TRICKS

INSTALL, MIGRATE AND UPDATE

EMAILS

NEWS FEEDS (RSS), BLOGS AND SOCIAL

CONTACTS

CALENDAR

CUSTOMIZE CONTROLS, OPTIONS AND ADD-ONS

PRIVACY AND SECURITY SETTINGS

Was this article helpful?

## Remote Content in Messages

Email messages can contain remote content such as images or stylesheets. To protect your privacy, Thunderbird does not load remote content automatically, but instead shows a notification bar to indicate that it blocked remote content.

Table of Contents

- [What is remote content and why is it blocked?](#)
- [How does Thunderbird protect my privacy?](#)
  - [Display remote content for a particular message](#)
  - [Display remote content by default](#)

What is remote content and why is it blocked?

Remote content are parts of a message (such as images) that are included in the message itself, but are downloaded from another location. Remote content is a privacy concern because it allows the message sender to know:

- each time you view the message
- rough details about what application and what platform you are using
- your current geographic location (a rough approximation by IP address)
- that your email address is actually used ("active")

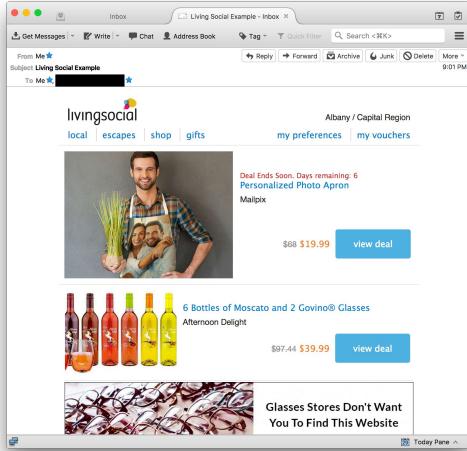
This is almost the same information websites commonly collect about you during normal web surfing. While browsing has the same privacy implications, what makes remote content in messages different is that it is targeted at you initially, so you can assume your access is directly linkable to your email address. This is also why spam often contains remote images (also known as "web bugs") which allow the spammer to mark your address as valid if the image is ever loaded.

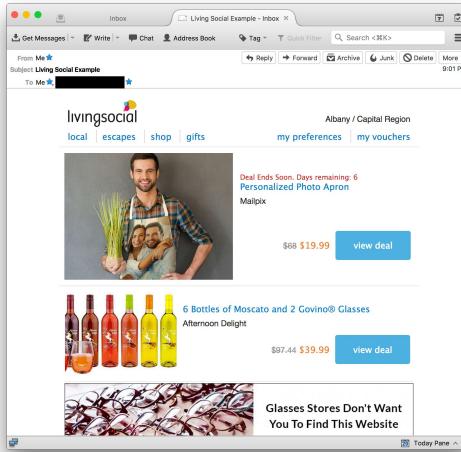
Today Pane

Remote content is a privacy concern because it allows the message sender to know:

- each time you view the message
- rough details about what application and what platform you are using
- your current geographic location (a rough approximation by IP address)
- that your email address is actually used ("active")

# 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)

**LiveIntent** (liadm.com)

**Oracle** (nexac.com)

**Acxiom** (rlcdn.com, pippio.com, acxiom-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 Apps** (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

**LiveIntent** (licasd.com)

**Google** (2mdn.net)

**Akamai** (akamai.net)

# A user's email address is the perfect identifier!

- It's unique
- It rarely changes
- It's the same across devices
- Consumers freely provide it to stores
- There's a lot of associated data

## PII-based tracking

```
UUID = {  
    MD5(bob@example.com),  
    SHA1(bob@example.com),  
    SHA256(bob@example.com)  
}
```



# Why hashed email addresses? User privacy!

## 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.

# Maybe hashing isn't so effective at protecting users...

The screenshot shows the Datafinder homepage with a dark header featuring the logo and navigation links: Targeted Lists, Data Tools, Predictive Analytics, and Pricing. The main content area has a purple background with a faint grid of numbers and symbols. It features the headline "Recover Encrypted Email Addresses" and a sub-headline about Versium's Email Decryption service starting at \$0.04 per email. Below this is a descriptive paragraph about recovering encrypted email addresses using common protocols, with a success rate of over 70%. At the bottom is a yellow call-to-action button with the text "Create Your Account Today »".

**Datafinder**  
Automated Data Intelligence

Targeted Lists   Data Tools   Predictive Analytics   Pricing

## 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.

Create Your Account Today »

## More on this:

<https://freedom-to-tinker.com/2018/04/09/four-cents-to-deanonymize-companies-reverse-hashed-email-addresses/>

# Methods

# Challenge: Measurements require the automated submission of PII to sites

## Mailing list sign-ups

Email Address

Birthdate

Your Country / Territory

State

Zip Code

Your Gender

By checking this box you agree to the TaylorSwift.com [Terms of Use](#) and [Privacy Policy](#).

**Subscribe**

## Login Forms

Sign in

Email address:

Password:

[I forgot my password.](#)

**SIGN IN**      **Cancel**

# Measuring email tracking at scale

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

Email, please

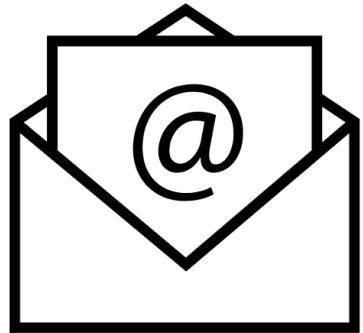
Confirm your email

**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

<https://github.com/citp/OpenWPM>



Email  
Tracking ≈ Web  
Tracking - Javascript

# 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)

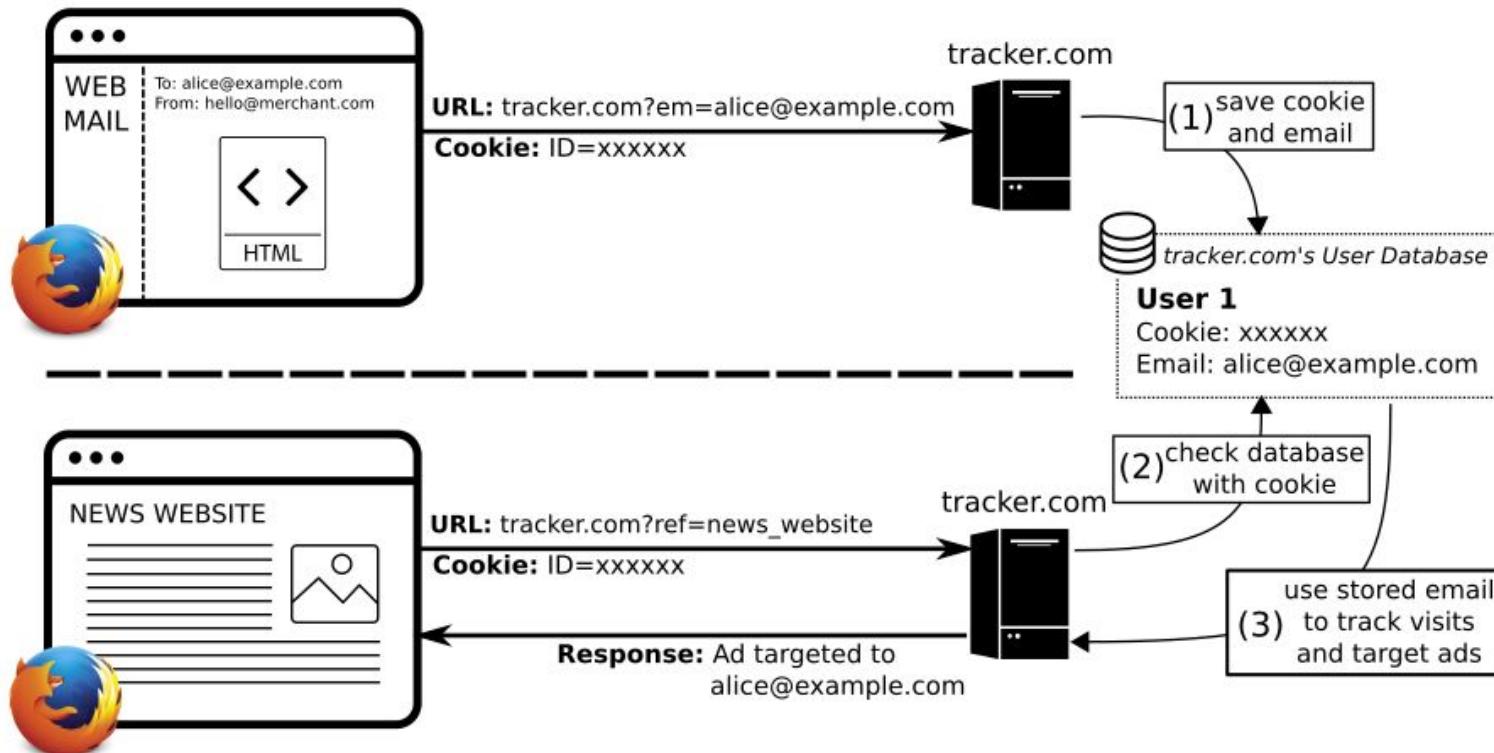
Leak	# of Senders	# of Recipients
MD5	100	38
SHA1	64	19
SHA256	69	13
Plaintext Domain	55	2
Plaintext Address	77	54
URL Encoded Address	6	8
SHA1 of MD5*	1	1
SHA256 of MD5*	1	1
MD5 of MD5*	1	1
SHA384	1	1

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

# A sample leak: 12 redirects in a single image tag

Row	Request URL
0	http://inbox.washingtonexaminer.com/imp?s=...&e=<EMAIL>&p=0
1	http://p.liadm.com/imp?...&m=<MD5>&sh=<SHA1>&sh2=<SHA256>&dom=<EMAIL_DOMAIN>
2	http://x.bidswitch.net-sync?ssp=liveintent&bidder_id=5298&licd=3357&x=EGF.M...
3	http://x.bidswitch.net-ul_cb-sync?ssp=liveintent&bidder_id=5298&licd=3357&x=EGF.M...
4	http://p.adsymptotic.com/d/px/?_pid=12688&_psign=d3e69...&bidswitch_ssp_id=liveintent&_redirect=...
5	http://p.adsymptotic.com/d/px/?_pid=12688&_psign=d3e69...&bidswit...&_redirect=...&_expected_cookie=...
6	http://x.bidswitch.net-sync?dsp_id=126&user_id=84f3...&ssp=liveintent
7	http://i.liadm.com/s/19751?bidder_id=5298&licd=3357&bidder_uuid=<UUID_1>
8	http://cm.g.doubleclick.net/pixel?google_nid=liveintent_dbm&google_cm&google_sc
9	http://cm.g.doubleclick.net/pixel?google_nid=liveintent_dbm&google_cm=&google_sc=&google_tc=
10	http://p.liadm.com/match_g?bidder_id=24314&bidder_uuid=<UUID_2>&google_cver=1
11	http://x.bidswitch.net-sync?ssp=liveintent&bidder_id=5298&licd=
12	http://pool.udsp.iponweb.net-sync?ssp=bidswitch&bidswitch_ssp_id=liveintent

# Trackers can correlate email and web tracking

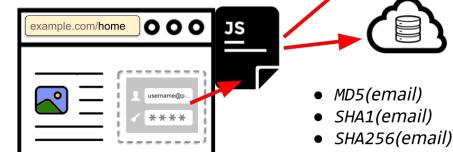
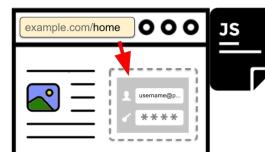
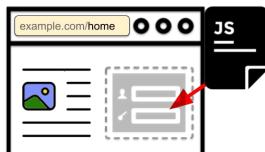


# Ongoing research: trackers also harvest email addresses from the web

User submits a login or registration form, clicks "Save" to store the credentials.



User visits a non-login page on the same site; this time the third party script is present



1. Third-party script injects an invisible login form

2. Login manager fills in user's email and password

3. The script reads the email address from the form and sends it to third-party servers

# Is identity-based marketing the future?

Recipient Organization	# of Senders
LiveIntent	68
Acxiom	46
Litmus Software	28
Conversant Media	26
Neustar	24
apxlv.com	18
54.211.147.17	18
Trancos	17
WPP	17
54.82.61.160	16

The top email collectors all sell “identity-based” marketing. Allowing advertisers to reach individuals on any device and connect with individual purchase data and other offline data.

# Defenses

Mail Client	Platform	Proxies Content	Blocks Images	Blocks Referrers	Blocks Cookies	Ext. Support
Gmail	Web	Yes	No*	L: Yes, I: Yes†	Yes†	Yes
Yahoo! Mail	Web	No	Yes	L: Yes, I: No	No	Yes
Outlook Web App	Web	No	Yes	No	No	Yes
Outlook.com	Web	No	No*	No	No	Yes
Yandex Mail	Web	Yes	No*	L: Yes, I: Yes†	Yes†	Yes
GMX	Web	No	No*	No	No	Yes
Zimbra	Web	No	Yes	No	No	Yes
163.com	Web	No	No*	No	No	Yes
Sina	Web	No	No	No	No	Yes
Apple Mail	iOS	No	No*	Yes	Yes	No
Gmail	iOS	Yes	No	Yes	Yes	No
Gmail	Android	Yes	No	Yes	Yes	No
Apple Mail	Desktop	No	No*	Yes	Yes	No
Windows Mail	Desktop	No	No*	Yes	No	No
Outlook 2016	Desktop	No	Yes	Yes	No	No
Thunderbird	Desktop	No	Yes	Yes	Optional (Default: No)	Yes

**Table 12.** A survey of the privacy impacting features of email clients. We explore whether the client proxies image requests, blocks images by default, blocks referrer headers from being sent (with image requests “I:” and with link clicks “L:”), blocks external resources from settings cookies, and whether or not the client supports request blocking extensions — either through the browser (for web clients) or directly (in the case of Thunderbird).

\*Images are only blocked for messages considered spam.

† Blocking occurs as a result of proxied content.

# Tracking defenses are incomplete

- **Block cookies**
  - Prevents PII leaks from being connected to tracking cookies
  - Doesn't prevent linkage of PII to IP address / passive fingerprint
- **Proxy image requests**
  - Prevents linkage of PII to cookies, IP, and fingerprint
  - Doesn't prevent targeted advertising / data collection
- **Block images**
  - Prevents tracking, but many emails are unreadable

# Ad blockers help, but don't fully protect users



Filtering requests with EasyList + EasyPrivacy

- Nearly half of the recipients of leaked email addresses are blocked (from 99 to 51)
- The number of senders leaking email addresses drops from 19% to 7%

...they also aren't available on all platforms

# Our proposal: Filtering at the service provider level

```
<html>
  <p>Hello there!</p>
  
  
  <p>Buy our products</p>
</html>
```

Performs almost as well as client-side filtering; misses redirects

# Our proposal: Filtering at the service provider level

## Server-side filtering using blocklists

```
<html>
  <p>Hello there!</p>
  
  
  <p>Buy our products</p>
</html>
```



```
<html>
  <p>Hello there!</p>
  
  <p>Buy our products</p>
</html>
```

Performs almost as well as client-side filtering; misses redirects

# Takeaways

# Takeaways

1. The line between email tracking/marketing and web tracking is blurry
2. We need better email tracking defenses.
  - a. Is measurement + filtering the only path forward?
3. A budding industry is building around tracking with hashed identifiers
  - a. Hard to block and control. Is policy the only solution?

Data + Code: [https://github.com/citp/email\\_tracking](https://github.com/citp/email_tracking)

Paper: <https://senglehardt.com/publications>