

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

hanghu@vt.edu →



Hang Hu

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

Leaked email address can lead to real-life scandal



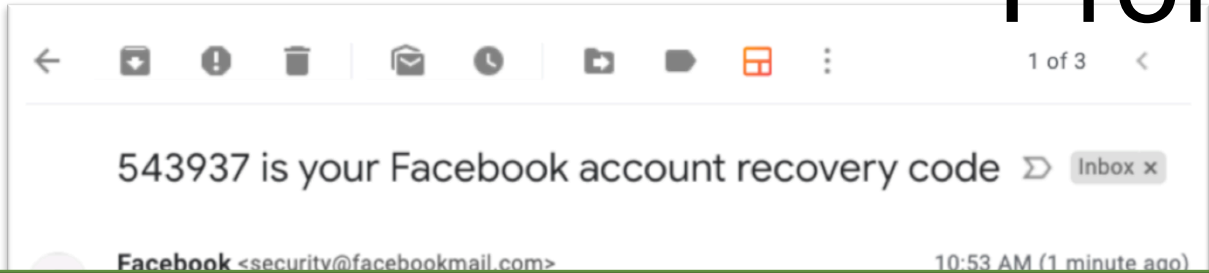
07/12/2015: Ashley Madison hacked

Website to look for an affair

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

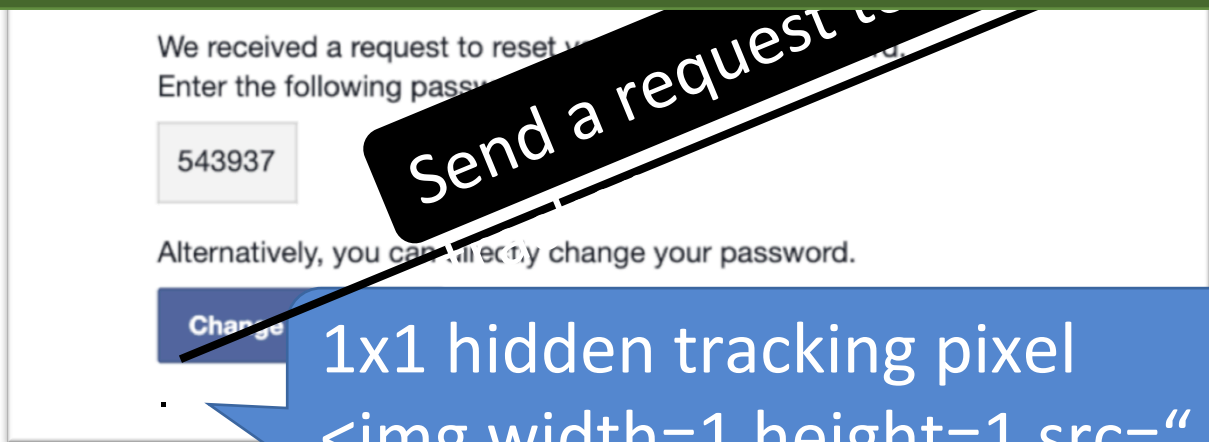
08/27/2015: Leaked users face blackmail threats

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



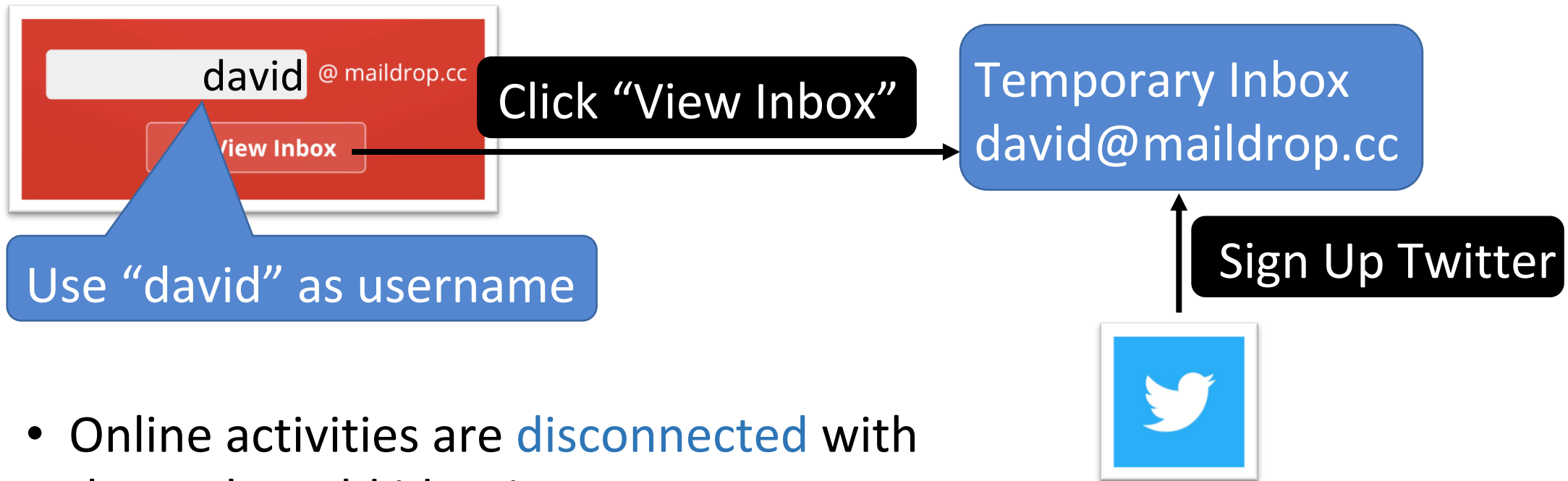
1x1 hidden tracking pixel

``

Tracker → The device

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

david@maildrop.cc



Triggered by me when I use

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

Quora Digest <english-digest-noreply@quora.com>

May 08 11:24am

A couple and their 2 daughters went to the movies. Each daughter has 4 brothers. Each...?

Drop <info@i.massdrop.com>

May 08 10:48am

Ravnica Allegiance Guild Kit (Set of 5), Stuffed Fables Board Game, Dragon Shield Nest 100+ De...

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?

This is what they say

This is what they actually do

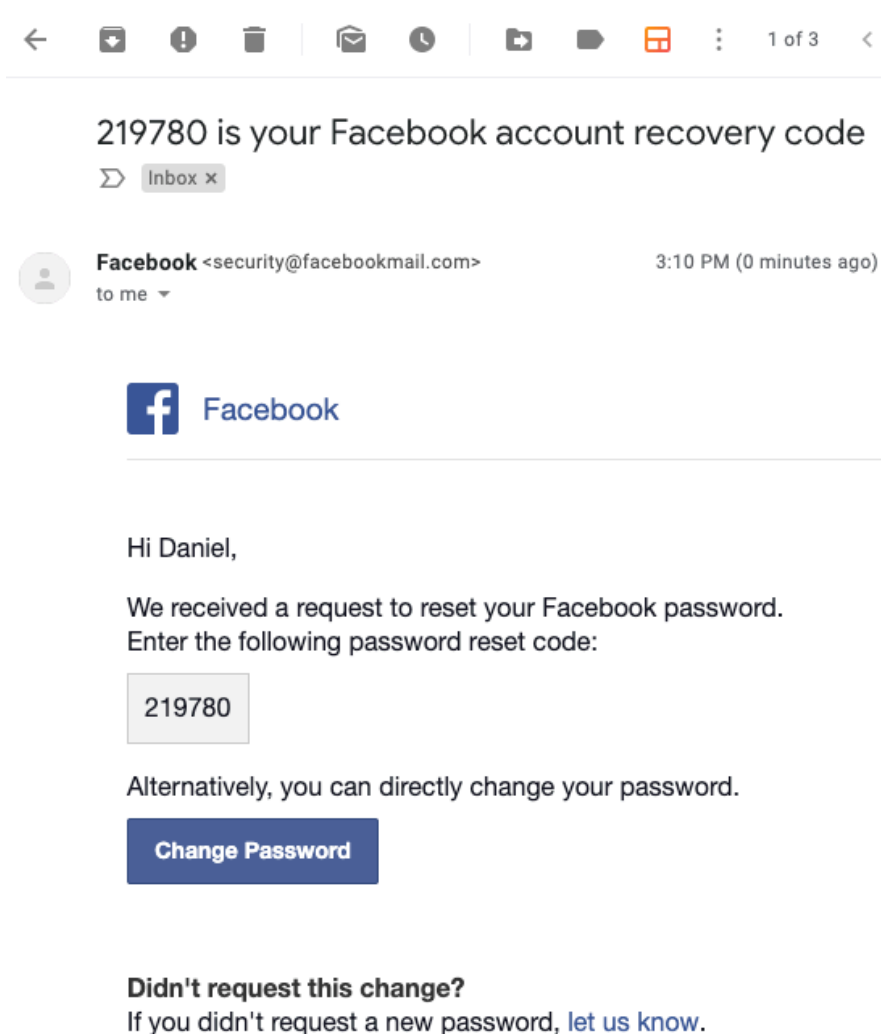
Website	Claimed Time
Guerrillamail.com	1 hour

Disposable email services don't delete emails as quickly as promised

Mailfall.com	25 mins
Mailsac.com	Dynamic



What Are the Risky Usages?



	# Detected in Data
Number	1,399
Security Number (SSN)	926
Identification Number (EIN)	701

online accounts under this email can be hijacked

312) Registration
,715) Password Reset
,802) Authentication Code

} 86K

512,361) All unsolicited emails, newsletters, ads

and notifications

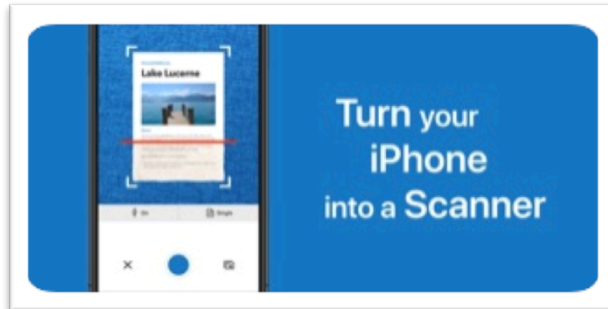
Risky Usage: Case Study

The logo for HealthCare.gov, featuring the word "Health" in black, "Care" in blue, and ".gov" in grey.

4000+ emails from healthcare.gov
Account carries sensitive information



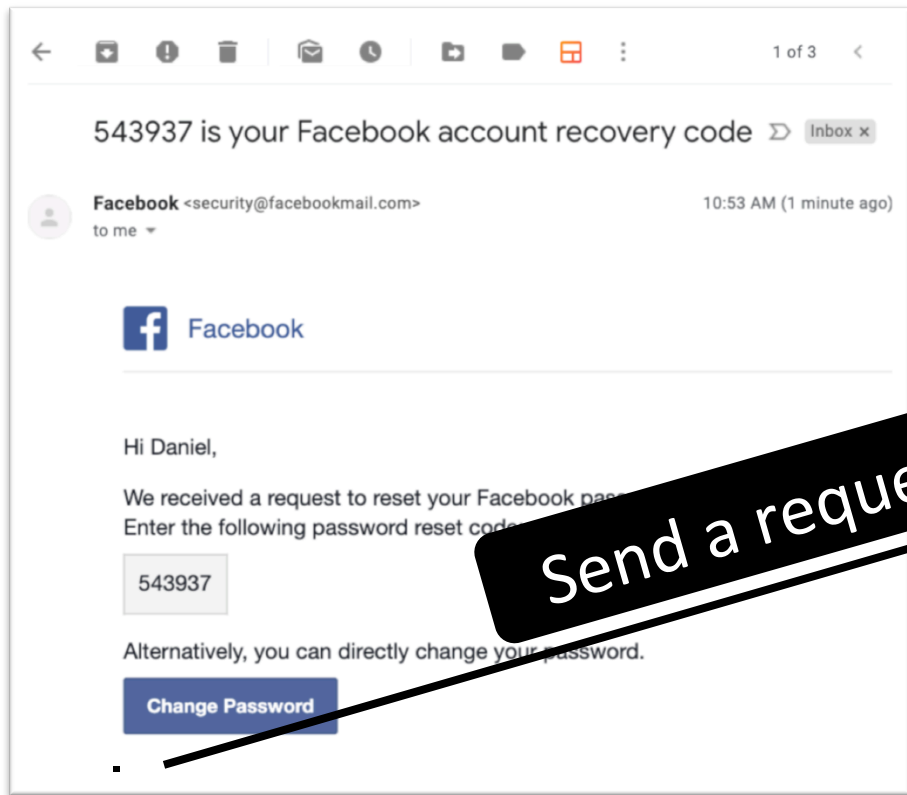
Emails from af.mil
Contain SSN and date of birth
Password reset is available



Receive all scanned PDF documents
(signed contract or other sensitive docs)

Use Real-world Dataset to Study Email Tracking

Sender: Facebook

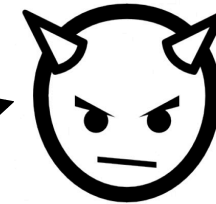


Send a request to the tracker

Tracker

First-party Tracking

If tracker is facebook.com



Tracker

If tracker is
google.com

Third-party Tracking

Tracking Detection

The URL contains an identifier of the receiver

1. The ID is the **email address** of the receiver

Or

2. The ID is the **hash of the email address** of the receiver

32 hash functions

33,824 combinations of hash

The

Tracking Detection (Cont.): Handling Evasion

Hidden Tracker	# Emails	# Direct Trackers
DoubleClick.net	96,430	164
Adsrvr.org	48,858	130
Rlcdn.com	42,745	132
Pippio.com	41,140	59
Liadm.com	29,643	252

`` → Beacon → C.com → A tracking pixel

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

#	5,007,000			
Trackers	N/A			

Popular Services Are More Likely To Track You

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

	Sender Count	% Tracking
Popular Senders	2,052 (1%)	46.9%
Non-popular Senders	208,321 (99%)	5.2%



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

[2] [IEEE S&P'12] Third-party web tracking: Policy and technology

[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

	Web	Mobile
Gmail	Proxy	Proxy
Outlook	Non-block	Non-block
Yahoo	Proxy	Proxy
iCloud	Non-block	Non-block

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

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