

# **Collecting, categorizing and analyzing Emails from top Free Disposable Email Address Services**

**CSE508: Network Security**


By Shubham Jindal


SBU ID: 112129688


# DEA Services


- Unique public email address
- Disposable if compromised (spam, phishing)
- Free or paid
- Uses: Online registration/middle layer mailbox


← → ↻ <https://www.mailinator.com/v3/index.jsp?zone=public&query=mike#/#inboxpane> ☆ 📧 🍌 🟢 🟢 📧 1 🔍


 **Mailinator**


 Public Inboxes


 Public SMS Inbox

 Private Domains ▾

 Upgrade Pricing

 Request Trial

 API Documentation

 FAQ

Monetized by Viglink

Copyright 2018 Manubrain, LLC

public inbox: mike

Domain: **mailinator.com** ▾

mike Go!

From	Subject	Received
<input type="checkbox"/> SwimOutlet.com	Final Call for Up to 70% Off Black Friday Doorbusters!	moments ago
<input type="checkbox"/> NORDSTROM RACK	We *get* you	5 minutes ago
<input type="checkbox"/> Ms CHIANG Lai Yuen JP	RE:	13 minutes ago
<input type="checkbox"/> Zazzle Inc.	These Deals are Going Strong, But Not For Long! 📧	15 minutes ago
<input type="checkbox"/> mike@zippymail.info	Threat to your security! mike@zippymail.info has been compromised.	25 minutes ago
<input type="checkbox"/> mike@bobmail.info	Threat to your security! mike@bobmail.info has been compromised.	27 minutes ago
<input type="checkbox"/> mike@safetymail.info	Threat to your security! mike@safetymail.info has been compromised.	27 minutes ago
<input type="checkbox"/> epssljgjpj@pdndjuikglfosf	緊急特売会場	32 minutes ago
<input type="checkbox"/> Comptoir de l'Homme	📧 -25% sur TOUS LES PRODUITS	32 minutes ago
<input type="checkbox"/> no sender	[no subject]	33 minutes ago
<input type="checkbox"/> Nineties Hub	Here Are The Best Offers Of The Weekend!	36 minutes ago
<input type="checkbox"/> Bluewire Media	For your friend (and you)	38 minutes ago

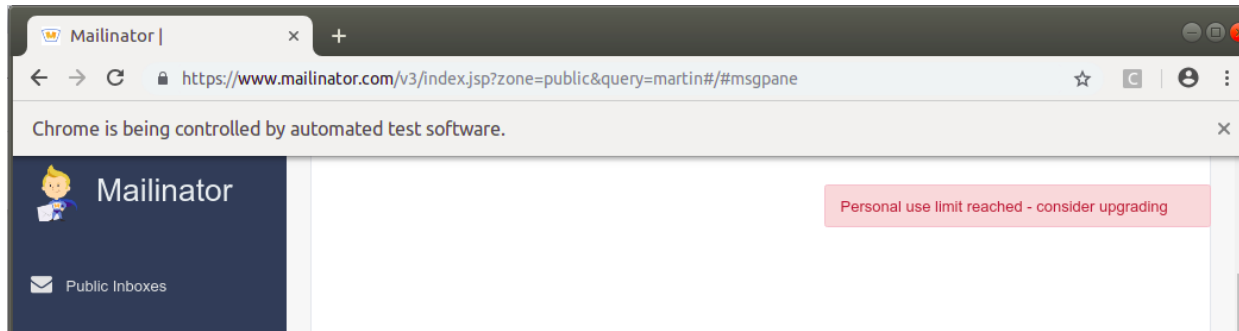
mike@mailinator.com

# Problem

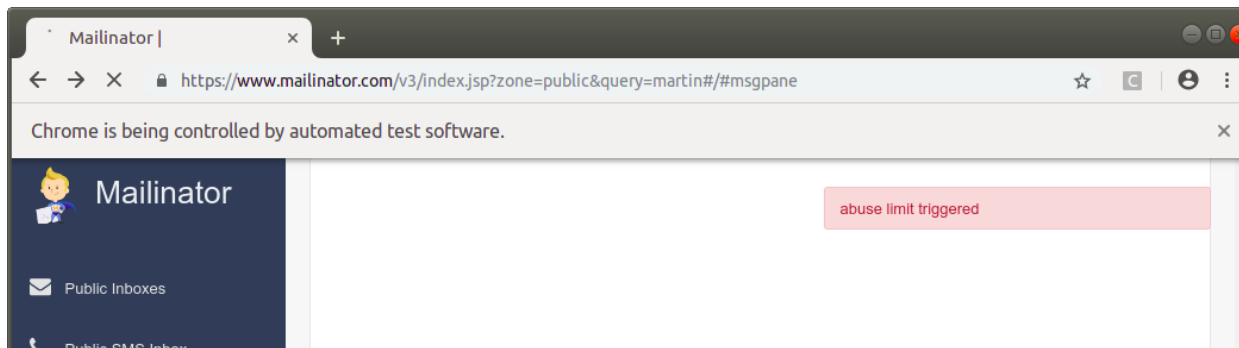
- Analyze the kinds of mails present in DEA mailboxes.
- Requirements:
  - Collection of mails
  - Categorization of mails
  - Draw insights

# Challenges

- **C1:** pages took time to load
- **C2:** mailinator has request limit and is protected by cloudFlare, number one service for protection against DOS and DDOS attacks.
- **C3:** mailinator blocks TOR proxy
- **C4:** Access lost after crossing limits.
- **C5:** Mailinator empties the email data at a certain point of time every day.
- **C6:** every run of the mailinator started fetching the emails from the start.



- Usage limit
- Abuse limit
- Blocks IP after abuse limit is crossed



# Results

- Collection: Selenium + python + cron + timing adjustments
- Classification: python + ML(scikit learn)
- 100 common names
- Top Google bad words:1703
- Collection results:

Total Mailboxes checked	:	101
Total Mails	:	2808

# Classification Result and Insights

cluster = 0



cluster = 1



cluster = 2



SimpleClassifier:

Profane Mail :539/2808  
 Pharma Mails :15  
 Profanity :19%  
 Pharma :0.5%  
 DEA Used :{'yopmail.com': 5}

MLClassifier:

- Top categories :  
Profanity/promotion/offer
- Some mails in Mandarin and Russian
- 26 extortion mails by hackers with BTC Wallet id.
- Presence of vulnerable urls



# Future Work

- Crawling links in mail.
- Analyzing source email addresses, time stamp.
- Image and OCR for mails
- Translating emails from Language Unicode to English (Google translator api - 1000000 words/ \$20).
- Include scripts for yopmail, tempr.email.

Project link:

<https://github.com/sjindal94/DEAemailscraper>