

# CSE 599B:

# Technology-Enabled Misinformation

Franziska (Franzi) Roesner

[franzi@cs.washington.edu](mailto:franzi@cs.washington.edu)

*Fall 2018*



PAUL G. ALLEN SCHOOL  
OF COMPUTER SCIENCE & ENGINEERING



UNIVERSITY of WASHINGTON

SECURITY AND PRIVACY  
RESEARCH LAB

The Switch

# Twitter is sweeping out fake accounts like never before, putting user growth at risk

---


Twitter suspended more than 70 million accounts in May and June, and the pace has continued in July


## Crackdown on 'bots' sweeps up people who tweet often


By SARA BURNETT   August 4, 2018





# BotSentinel.com


 Bot Sentinel


 Dashboard


 Trending Topics

 Trollbot Network

 Recent Tweets

 Tweet Archive


 Top 100 Trollbots

 All **41 684**


BotSentinel.com


0%


**Trollbot Rating: Normal**  
This report was created for @uwcse  
Report created: 2018-10-26 02:17:52




NormalTolerableProblematicAlarming

 Normal: 0% - 24%


 Tolerable: 25% - 49%

 Problematic: 50% - 74%

 Alarming: 75% - 100%

Our analysis has concluded **uwcse** exhibits normal tweet activity and is not a trollbot account.

ch...

 Check User

76 Fake News

0 added today.

968 Inactive

3 became inactive today.

# IN THE SENATE OF THE UNITED STATES

JUNE 25, 2018

Mrs. FEINSTEIN introduced the following bill; which was read twice and referred to the Committee on Commerce, Science, and Transportation

**A**

To protect the right of the American people to receive news and information by regulating the use of social media intended to impersonate human beings on social media.

**The Bot Disclosure and Accountability Act** would compel social media companies to institute policies that require users on their platform who operate automated software programs designed to mimic or impersonate human beings to disclose this fact on their account profiles. It would also require the platforms to develop "a process to identify, assess and verify" bot activity and take "reasonable" steps to prevent bots from impersonating human users online.

It would also ban the use of bot programs designed to impersonate humans by political campaigns, parties and authorized committees. It does not address the practice of campaigns or their affiliates **paying human trolls** to provide similar campaign amplification efforts for a candidate online.

1 *Be it enacted by the*  
2 *tives of the United States of America in Congress assembled,*

## 3 **SECTION 1. SHORT TITLE.**

4 This Act may be cited as the “Bot Disclosure and  
5 Accountability Act of 2018”.

# Freelance Abuse [Motoyama et al., USENIX Security 2011]

Category	Job Type	Description	Count	%
Legitimate [§A.1]	Web Design/Coding	Create, modify, or design a Web site	769	38.5
	Multimedia Related	Complete multimedia-related task (e.g., Flash)	265	13.2
	Private Jobs	Jobs designated for a particular worker	138	6.9
	Desktop/Mobile Applications	Create a desktop or mobile application	100	5.0
	Legitimate Miscellaneous	Miscellaneous jobs	177	8.8
Accounts [§A.2]	Account Registrations	Create accounts with no defined requirements	22	1.1
	Human CAPTCHA Solving	Requests for human CAPTCHA solving	19	0.9
	Verified Accounts	Create verified accounts (e.g. phone)	14	0.7
SEO [§A.3]	SEO Content Generation	Requests for SEO content (e.g., articles, blogs)	195	9.8
	Link Building (Grey Hat)	Get backlinks using grey hat methods	53	2.6
	Link Building (White Hat)	Get backlinks using no grey/black hat methods	20	1.0
	SEO Miscellaneous	Nonspecific SEO-related job postings	61	3.0
Spamming [§A.4]	Ad Posting	Post content for human consumption	25	1.2
	Bulk Mailing	Send bulk emails	8	0.4
OSN Linking [§A.5]	Create Social Networking Links	Get friends/subscribers/fans/followers/etc.	33	1.7
Misc [§A.6]	Abuse Tools	Tools used for abuse (e.g., CAPTCHA OCR)	41	2.1
	Clicks/CPA/Leads/Signups	Get clicks, emails, zip codes, signups, etc.	32	1.6
	Manual Data Extraction	Manually visit websites and scrape content	21	1.1
	Gather Email/Contact Lists	Research contact details for targeted people	17	0.9
	Academic Fraud	Write essays, code homework assignments, etc.	10	0.5
	Reviews/Astroturfing	Create positive reviews	1	0.1
	Other Malicious	Miscellaneous jobs with malicious intentions	35	1.8

# Fraudulent Accounts

[Thomas et al, USENIX Security 2013]

Price: **\$0.04** Median account price

Delivery: **1day** Median time before accounts arrive

Fraud: **13%** Accounts are resold, accessed after sale

- Prices from buyaccs.com

Web Service	Price per Thousand
Hotmail.com, resale*	\$2.00
Hotmail.com	\$4.00
Yahoo	\$6.00
Twitter	\$20.00
Google (PVA)**	\$100.00
Facebook (PVA)**	\$100.00

\* Resale indicates account was previously used in another activity

\*\* PVA indicates a phone verified account; challenge response text to cell phone





Tweets  
**4,064**

Following  
**995**

Followers  
**29.5K**

Likes  
**407**

**Elon Musk** ✓

@LibmanCompany

My Official Promotion Account

Joined October 2010

Tweet to

Message

638 Photos and videos



Tweets

Tweets & replies

Media

Pinned Tweet



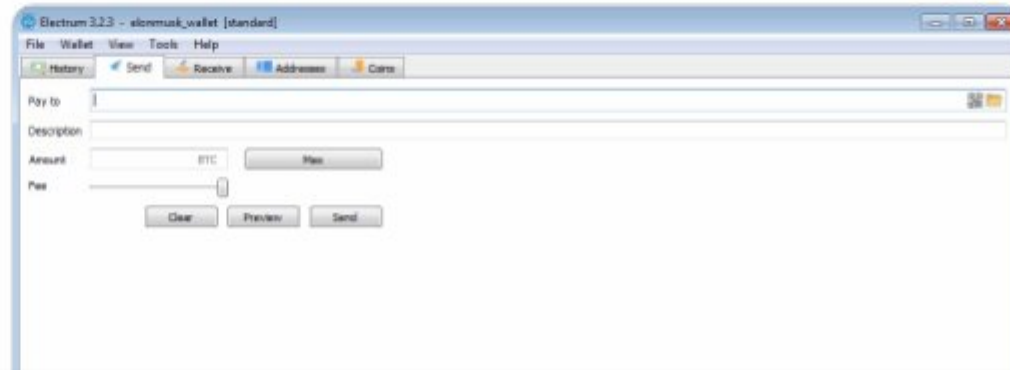
**Elon Musk** ✓ @LibmanCompany · 1h

I'm giving 5 000 Bitcoin (BTC) to my followers!

To identify your address, send from .1–3 BTC to the address below and get from 1-30 BTC back to your address!

Address BTC - 186w8LXTkss9EENGUKTNfrvo6utQ7codC

If you are late, your BTC will be sent back.



# Compromised Accounts

[Egele et al., NDSS 2013]

Our approach uses a composition of statistical modeling and anomaly detection to identify accounts that experience a **sudden change in behavior. ...**

We look for **groups of accounts that all experience similar changes within a short period of time**, assuming that these changes are the result of a malicious campaign that is unfolding.

	[5]	[3]	[4]	[6]	[7]	[17]	[18]	[19]	COMPA
<b>Network Features</b>									
Avg # conn. of neighbors						✓			
Avg messages of neighbors						✓			
Friends to Followers (F2F)	✓	✓			✓				
F2F of neighbors						✓			
Mutual links						✓	✓	✓	
User distance								✓	
<b>Single Message Features</b>									
Suspicious content	✓								
URL blacklist			✓						
<b>Friends features</b>									
Friend name entropy					✓				
Number of friends	✓				✓				
Profile age	✓								
<b>Stream Features</b>									
Activity per day	✓								
Applications used						✓			✓
Following Rate						✓			
Language									✓
Message length	✓								
Messages sent					✓				
Message similarity		✓	✓	✓	✓	✓			
Message timing		✓	✓						✓
Proximity									✓
Retweet ratio	✓								
Topics	✓								✓
URL entropy			✓						
URL ratio	✓	✓		✓	✓	✓			
URL repetition				✓					✓
User interaction	✓	✓		✓					✓

**Table 1.** Comparison of the features used by previous work



# Looking Ahead

- Defenses
  - False news detection
  - UI/UX interventions
- Projects
  - Checkpoint presentations and reports **next Friday**
  - Peer project workshopping the following week