HOW RUSSIAN TWITTER BOTS WEAPONIZE SOCIAL MEDIA TO INFLUENCE & DISINFORM

SafeGuardCyber

INTRODUCTION

This study originated with a simple question: What proportion of Twitter's most influential users have followings comprised of a high number of bots, and how does that composition make these users more vulnerable? For example, a high-profile CEO gaining a large number of followers in a short time might feel that is cause for pride, when in fact she should be more worried that those new followers are more like troops massing on a border.

Soon after posing this question, it became clear that the more interesting subjects weren't legitimate Twitter users, but the bots themselves. How did they behave individually versus collectively? What was the *modus operandi*? The subject of social media manipulation has been a prominent line of questioning since the 2016 Presidential Election. Information warfare may seem bloodless, but like conventional conflicts, it is waged on multiple fronts. Shifting our focus to the bots, we were able to glean deeper insights into bot tactics and operational organization.

For the purpose of this study, we limited ourselves to bots known to be connected to Russian influence and disinformation operations. In the course of our research, it became clear that the Russian disinformation effort is systematic and widespread, afflicting every digital touch point in the average American's life. And while we monitor bots on other channels, like Facebook and YouTube, this particular study is limited to Twitter. Our study led us to understand that Russian bots on Twitter:

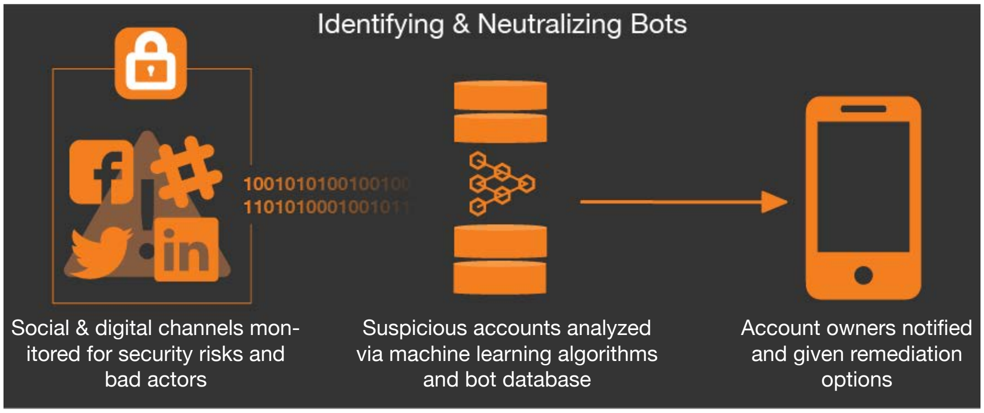
- Are deployed primarily to re-shape popular perceptions of events and news
- Operate within clearly identifiable content themes
- Extend their reach by hijacking or piggy-backing off organically created hashtags
- Are purpose-built to connect with one another to create amplification nodes

In addition to our findings, this report will look more closely at two specific cases that illustrate the tactics at play.

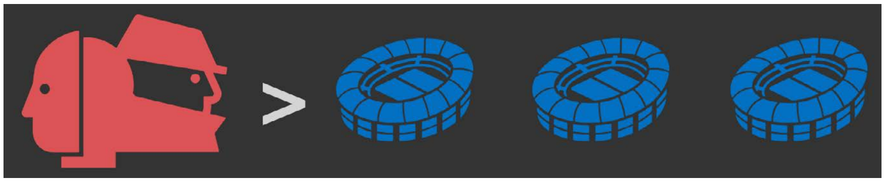
Lastly, alerts without action are useless. During the course of this study, our research team was accompanied by our product development teams to make the information applicable. As a result, the SafeGuard Cyber platform now has the capability to proactively detect and define bad actors from the moment they engage with our clients' brand channels or protected accounts. This process includes confirming known bots and identifying new ones, based on behavior patterns observed in our study. More to the point, our clients are prepared and equipped with an immediate recourse to neutralize these risks.

METHODOLOGY

Over the years we have identified and confirmed over 320,000 bots. To accomplish this, we used advanced algorithms that examined many aspects of bot behavior, content, and meta data. Suspected bots were then confirmed through crowd-sourced efforts. From there, our engineering teams applied machine-learning algorithms to determine the behaviors and characteristics that could reliably be used to positively identify bots. Our database grew, and the number of signatures analyzed also grew as our understanding of bots deepened.


Identifying & Neutralizing Bots
100101010010010
110101000100101
Social & digital channels monitored for security risks and bad actors
Suspicious accounts analyzed via machine learning algorithms and bot database
Account owners notified and given remediation options

This process of learning and adapting provided the framework for our bad actor detection capability. Suspect accounts are flagged by the SafeGuard Cyber platform. The identity is checked against the database. Should the account not be verified immediately, the account is run through a series of algorithms. If the account is then identified as a bot, it is added to our database for future checks.



Our database of Russian bots contains more than 320,000 accounts, more than three times the capacity of Michigan Stadium, the largest US stadium.

Our database acted as a petri dish from which to study our culture of Twitter bots. For this study, we subjected these bots to more robust scrutiny and detailed analysis by observing the content they shared via text analytics to infer their intent and behavior. We looked at volume of messaging, date, geo-location, and cross-referenced the data against the chronologies of real events and news.

WHAT WE LEARNED

Despite the nomenclature, bots are not a uniform army of automatons blanketing Twitter with the same tweets. These bot operations are far more sophisticated than the psy-ops of yesteryear.

The nature of social media affords Russian bot operators the paradoxical benefits of individualized specificity and generalized scale. Bots are designed to resemble real individuals to gain credibility among target populations. This means that individual bots do not post broadly about any and everything. Instead, they specialize. Some are designed to look like supporters of President Trump, while others appear to be left-of-center sympathizers to Russia. This level of differentiation is spread across Twitter at scale – hundreds of thousands of known bots – giving the impression of a diverse set of users sharing views on a wide range of issues.

What bots talk about

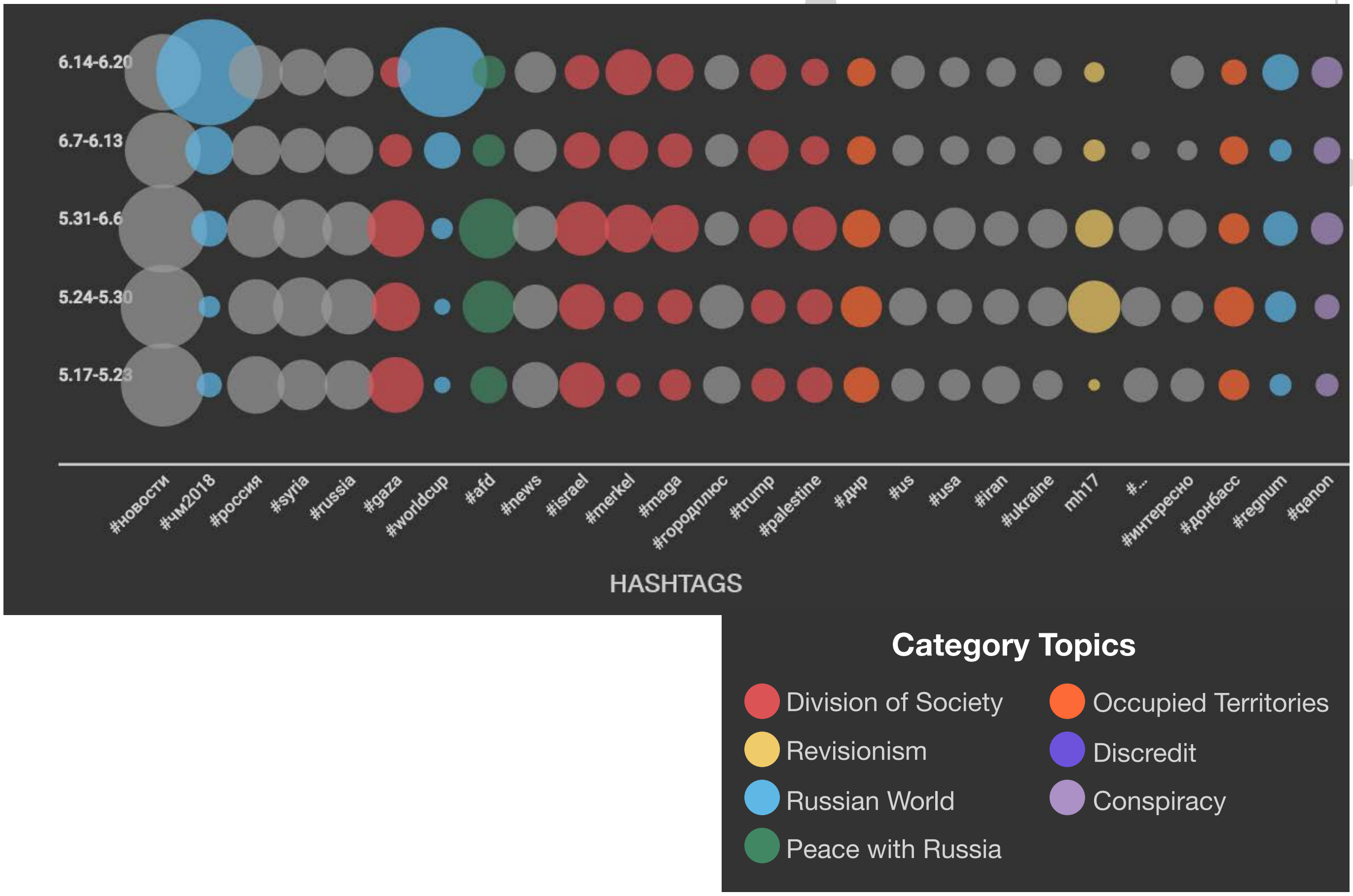
By focusing on our bot database, away from the larger Twitter population, our research team was able to discern clear thematic categories for tweet messaging. Each topic category has a specific purpose with respect to re-contextualizing conversations in a way that benefits the Russian government. Below is a table showing how themes are defined in relation to their goals, including examples of events and conversations where bots were deployed.

TOPIC CATEGORY	OBJECTIVE	EXAMPLE CONVERSATION
Division of Society	Destabilize, foment division	Brexit, Catalonian Independence, Charlottesville
Revisionism	Re-cast events in sympathetic light	MH-17, KGB legacy
Russian World	Tout Russian culture, language, etc.	World Cup, “invincible” missiles
Peace with Russia	Treaties, International agreements	NordStream II pipeline, anti-sanctions
Occupied Voices	Re-cast occupied territory in positive light	Crimea, Georgia
Discredit	Undermine verified facts	Ukraine as “fascist” state
Conspiracy	Surface & legitimize fringe ideas	QAnon, “globalist” agenda

Again, a military analogy proves both useful and limiting. Bots can be grouped into “battalions” by topic category. But, it’s clear that some of these categories overlap in certain cases. For example, “Conspiracy” tweets, it could be argued, are designed to discredit verifiable narratives or sow division. Compounding the difficulty in discerning truth from propaganda is the fact that bots are “irregular” forces, and as such are spread throughout the general Twitter population, in multiple countries. In this way, bots are being deployed to shape domestic perceptions in Russia, or to foment divisive conversations in the US or Germany.

WHAT WE LEARNED

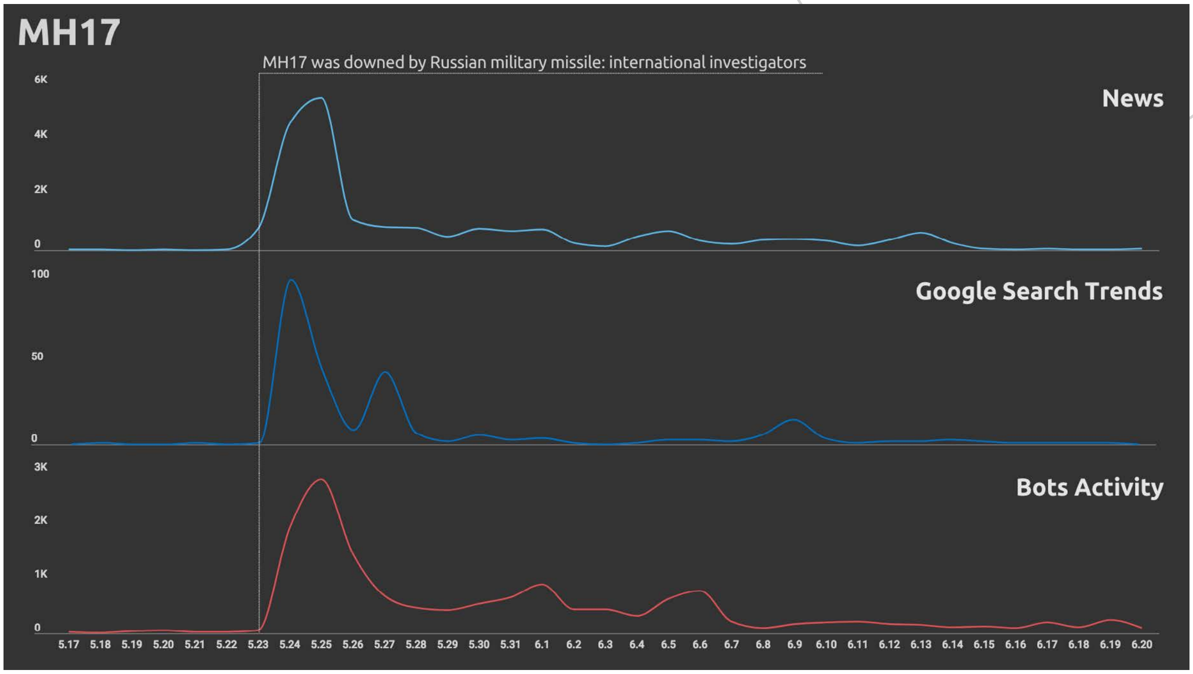
However, with insight into topic categories, we were able to see patterns among disinformation campaigns, depending on geography, topic, and time period. In the chart below, for example, you can see that the “Russian World” theme appears in greatest volume during and affiliated with World Cup content. Conversely, “Revisionist History” content was deployed around the May 2018 findings from international investigators about the downing of MH17, building to greatest volume during the week the findings were released.



When bots speak up

Bot activity is always-on, but not uniformly. At times, some bot “battalions” appear to go dormant, tweeting only enough to maintain the charade of an active user. Similarly, we observed how groups of bots were “activated” at appropriate moments, based on intent.

The bot battalions are most frequently mobilized to culture jam events or hijack the public perception and interpretation of news. Revisiting the release of the official report on MH17, it’s clear how bots were used when we correlated mentions from bots in our database against Google Trends search data and mentions from official news accounts.



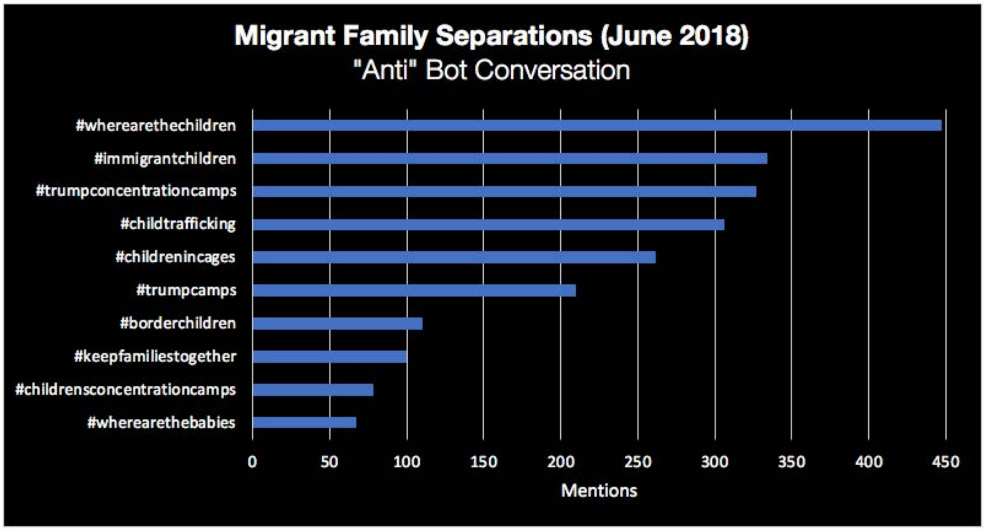
Bot activity (red line) kicks into high gear just following the news announcements. Conversation volume is maintained at a higher level once the news cycle ends. The spike is an attempt to intercept the news and the higher volume after the fact represents the continuing campaign to shape perception.

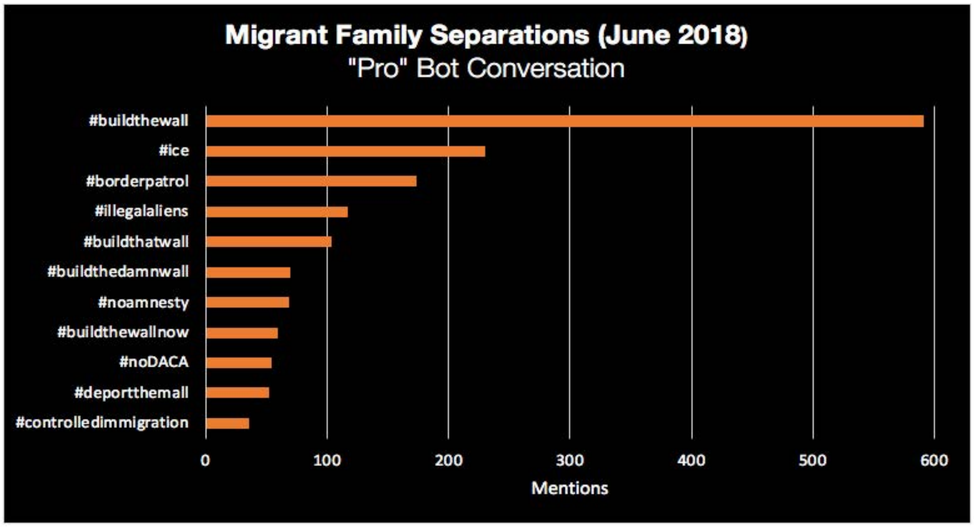
The challenge for Russian operators, of course, is having to wage this type of campaign across multiple languages and for dozens of events throughout the year. Hence, any one campaign – such as this MH17 example – cannot be sustained for long. Bots launch a full assault while a topic is trending, but must spare the ammo once the news cycle changes.

How bots speak about issues

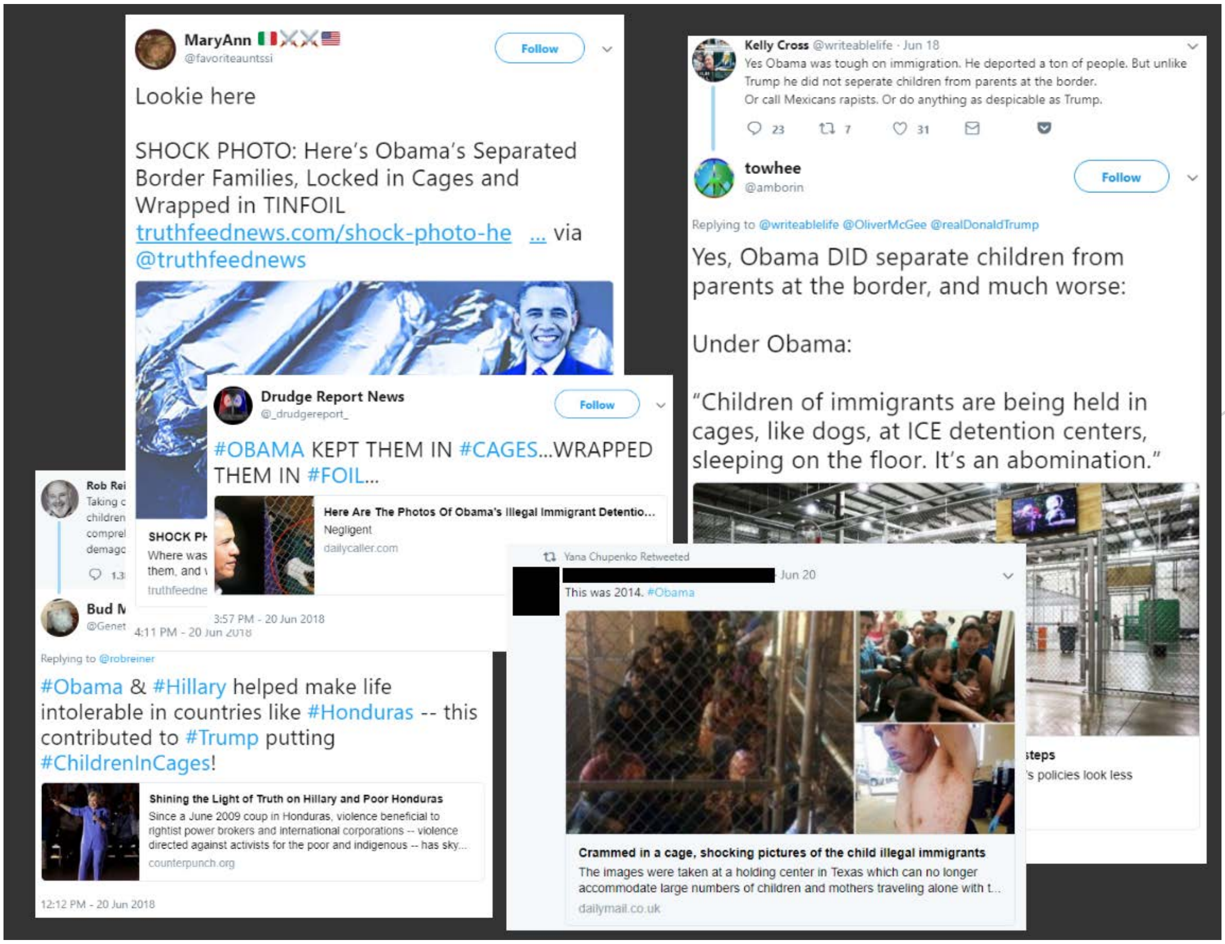
Bot operators are careful to speak in the language of their target audiences. When engaging a topic, they use carefully chosen hashtags to propel their content into the right conversations. It's not immediately clear whether poorly performing tags are the result of trial and error or a deliberate smoke screen, like mimicking misspellings. In any case, bots will seize on trending or organically created tags.

More to the point, bots deploy these hashtags frequently to champion or spread disinformation on both sides of an issue. As an example, take the June debate around the Trump administration's "zero tolerance" border policy which led to migrant children and parents being separated. Bots from our database were activated to stoke fervor on both sides of the issue.





Bots will also hijack hashtags to distort them. In the case of migrant separations, the audience protesting separations had used #childrenincages. During the news cycle, pro-Trump bots – some of which were designed to resemble real media or personalities – repurposed the hashtag to spread anti-Obama disinformation.



MaryAnn 🇮🇹❌❌🇺🇸
@favoriteauntssi
Follow

Lookie here

SHOCK PHOTO: Here's Obama's Separated Border Families, Locked in Cages and Wrapped in TINFOIL truthfeednews.com/shock-photo-he ... via @truthfeednews

SHOCK PH
Where was
them, and
truthfeedne

4:11 PM - 20 Jun 2018

Drudge Report News
@_drudgereport_
Follow

#OBAMA KEPT THEM IN #CAGES...WRAPPED THEM IN #FOIL...

Here Are The Photos Of Obama's Illegal Immigrant Detentio...
Negligent
dailycaller.com

3:57 PM - 20 Jun 2018

Rob Rei
Taking c
children
comprel
demagc
1.3

Bud M
@Genet

Replying to @robreiner

#Obama & #Hillary helped make life intolerable in countries like #Honduras -- this contributed to #Trump putting #ChildrenInCages!

Shining the Light of Truth on Hillary and Poor Honduras
Since a June 2009 coup in Honduras, violence beneficial to rightist power brokers and international corporations -- violence directed against activists for the poor and indigenous -- has sky...
counterpunch.org

12:12 PM - 20 Jun 2018

Kelly Cross @writeablelife · Jun 18
Yes Obama was tough on immigration. He deported a ton of people. But unlike Trump he did not seperate children from parents at the border.
Or call Mexicans rapists. Or do anything as despicable as Trump.
23 7 31

towhee
@amborin
Follow

Replying to @writeablelife @OliverMcGee @realDonaldTrump

Yes, Obama DID separate children from parents at the border, and much worse:

Under Obama:

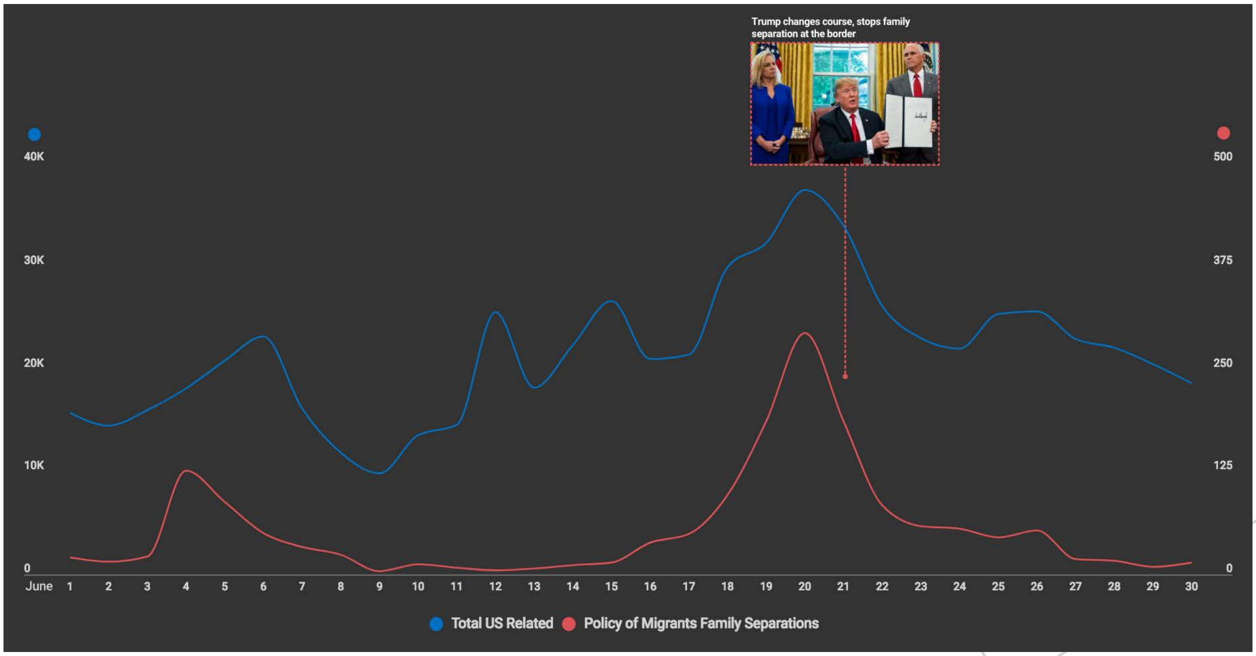
"Children of immigrants are being held in cages, like dogs, at ICE detention centers, sleeping on the floor. It's an abomination."

steps
's policies look less

Yana Chupenko Retweeted
[redacted] [redacted] · Jun 20
This was 2014. #Obama

Crammed in a cage, shocking pictures of the child illegal immigrants
The images were taken at a holding center in Texas which can no longer accommodate large numbers of children and mothers traveling alone with t...
dailymail.co.uk

And, when we correlate total bot activity against messaging around the family separations we see a similar chronological pattern as with MH17.



In the chart above, the spike in volume is clear. However, it's worth pointing out that bot messaging around migrant families at the height of activity was only 294 mentions out of 37,364 total US-related mentions among bots, or 0.79%. This seemingly low percentage, however, illustrates the efficiency of bot campaigns. The bot operators need only seed the conversation with the desired talking points or misleading stories. From there, they can rely on retweets and repurposing from ardent users to do the rest. Moreover, individual bots frequently follow other bots, creating "amplification nodes" throughout the Twitterverse. These connections help bots maintain a veneer of verisimilitude while accelerating misleading information.

In one case, our platform identified a Twitter account harassing one of our clients as a bot. The bot had, within two weeks of starting an account, only 279 followers. From there, our Threat Chain analysis revealed 84% of those 279 followers were in our bot database. Further, 70% of the retweets of this particular bot's content came from other bots. This small amplification node churned content at a rate that was sufficient to begin attracting legitimate users. At the time of writing, four weeks after the bot account was started, the bot's follower count has grown 4.5x to over 1,500 users.

CONCLUSION

From a starting point of curiosity, our study revealed a far more complex and nefarious network of Twitter bots. Understanding bot behavior is critical to containing and combating these operations, which have wide-ranging implications for civil society but also for private enterprise. To name only a few ways, bot armies can be deployed against companies to:

- Manipulate brand reputation through disinformation
- Gain employees' trust through social engineering
- Launch account takeover attacks
- Impersonate brand accounts to serve phishing sites to users

This study focuses on Russian bot operators' influence operations on Twitter. However, as stated above, information warfare is waged on multiple fronts. Our research team has documented coordinated campaigns across multiple digital channels and pathways. Twitter is but one facet of social media operations to influence conversations and perceptions. But similar tactics are used to ensure content deemed friendly to Russian government interests shows up for search. For example, black hat SEO tactics are used to establish and boost the rankings of propaganda sites, much like the Iranian operation identified earlier this year by FireEye. Bots are also used to game algorithms on other search channels such as up-voting videos on YouTube (arguably the world's second largest search engine), and the newer practice of black hat social SEO on Facebook, which we documented at the beginning of the year is linked to overt hacking and phishing attacks.

This investigation illustrates the complexity of the Russian bot operation on Twitter, and how quickly it can be mobilized to influence public perception. To date, the operation has retained a socio-political focus, but it could easily be turned against private corporations. Alternatively, the technology and tactics could be employed by others motivated more by financial gains than geopolitical ideologies.

SafeGuard Cyber

AMERICAS

410A E. Main Street
Charlottesville, VA
22901

+1 (434) 207 4265

sales@safeguardcyber.com

ASIA-PACIFIC

PO Box 523
Leichardt NSW 2040
Australia

+61 (437) 276 739

APACsales@safeguardcyber.com