ROBOTROLLING
2018. ISSUE 4

ROBOTROLLING

PREPARED AND PUBLISHED BY THE

NATO STRATEGIC COMMUNICATIONS CENTRE OF EXCELLENCE

Executive Summary

This issue of Robotrolling examines users suspended by Twitter. Contrary to expectation, most of the accounts were human-controlled accounts rather than bots. Since 2017, the speed at which Twitter suspended misbehaving users has by two measures almost doubled. However, removals of Russian-language accounts have been considerably slower than for English.

The speed of removal can be critical, for instance in the context of an election. The Latvian elections, conducted on 6 October 2018, passed with remarkably little Russian-language activity about the NATO presence in the country.

Our analyses show a movement in the past year away from automated manipulation to humans operating fake or disposable identities online. The figures published in this issue reflect the good work done to tackle bots, but show much work remains to tackle manipulation through fake human-controlled accounts.

Bots created 46% of Russian-language messaging about the NATO presence in the Baltics and Poland. More than 50% of Russian-language messaging about Estonia this quarter came from automated accounts.

Anonymous human-operated accounts posted 46% of all English-language messages about Poland, compared to 29% for the Baltic States. This discrepancy is both anomalous and persistent. Some of the messaging is probably artificial.

We continue to publish measures of fake social activity in the hope that quantifying the problem will focus minds on solving it. ■

The Big Picture

Robotrolling analyses social media manipulation about the NATO presence in the Baltic States and Poland. There are two principal types of manipulation: automated activity from robotic accounts, and messaging from fake human accounts, for instance from a so-called troll factory.

The total number of posts considered is 6 800, of which 38% were in Russian. The number of active users is 4 200. This quarter the number of posts and the number of active users reverted to the long term average following a doubling in activity in response to the July NATO summit.

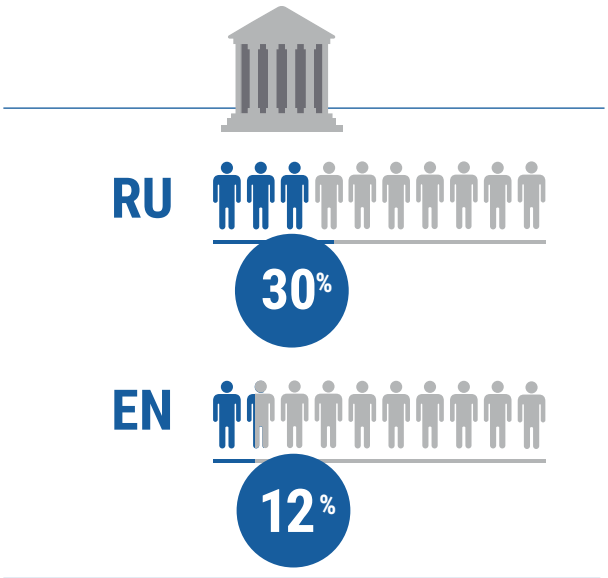
The level of bot activity this quarter is the lowest observed to date, both for Russian and English language content.

In the period 1 August – 31 October 2018, Russian-language bots created 46% of all Russian-language messages about NATO in the Baltic States and Poland. Of the accounts posting in Russian, 30% were predominantly automated. In comparison, bots made up 12% of accounts tweeting in English. These bots created 17% of all English-language messages this quarter.

Russian-language messaging this quarter overwhelmingly focused on mishaps during NATO exercises. By contrast, attention to the Latvian exercises in August and the elections in October received little attention.

We continue to expand our analytical toolkit. In this issue, we introduce analyses of deleted accounts. We have also expanded our model to estimate account types including news media and institutional accounts. The social network visualization in Figure 4 incorporates this complexity to give unique insight into the character of the Twitter conversation.

Social media manipulation continues to transition from being largely automated to increasingly being conducted manually through disposable accounts. It is hard to say whether individual anonymous accounts are operated out of so-called troll-factories. But, the anomalous and persistent high level of anonymous messaging by English-language accounts about NATO in Poland does indicate that some of the messaging is artificial. ■



Country Overview

Russian-language Twitter commentary about the NATO presence in the Baltics and Poland peaked in August amid reports that a Spanish jet had misfired while participating in NATO exercises in Estonia. English-language messaging peaked on 19 September during Polish President Andrzej Duda's visit to the US.

Estonia remains the focus of automated Russian-language accounts, whereas Poland has the highest proportion of comments from anonymous human users, both for Russian and English. In the English-language space, the volume of messages about Poland equals that of Estonia, Latvia, and Lithuania put together.

In the English-language space, activity from anonymous accounts increased in the run-up to the US mid-terms, especially after President Trump expressed a wish to withdraw from the Intermediate-Range Nuclear Forces (INF) treaty. In the lead-up to the large NATO Trident Juncture 2018 Twitter messaging remained subdued.

Estonia

The NATO presence in Estonia remains a key focus of robotic activity. On 8 September, a Spanish jet misfired while participating in a NATO exercise. Although the incident did not cause any damage, it did attract extensive commentary on Twitter. A few days later, the Mount Show, an online Russian-language political satire show, discussed the incident under the heading 'NATO attacks Estonia'. Messages about the show and the incident itself were primarily spread by human-operated accounts. In contrast, Russian-language bots focused on an interview given by Mikk Marran, Director General of the Estonian Foreign Intelligence Service, in which he said the Kremlin is continuously seeking to undermine EU and NATO unity.

Latvia

From August 20 to September 2, Latvian and allied troops conducted the exercise Namejs. RT's reporting, which centred on the scale of the NATO presence and supposed ethnic tensions, gained some traction in English-language Twitter messaging. The Latvian elections in early October coincided with increased activity from English-language bots. Russian-language messaging about NATO and Latvia was unusually low this quarter, at less than 300 messages.

Lithuania

In mid-September, Chancellor Angela Merkel visited the German troops stationed in Lithuania. On 7 October, a German soldier died in an accident during drills held in Lithuania. The incident drew more commentary in Russian than English. Roughly half the Russian-language messaging about the incident came from bot accounts.

Poland

Polish President Andrzej Duda visited the US in September. At the White House, he discussed the case for building an American military base in Poland to as a means to deter any Russian aggression. This base – Fort Trump – would be paid for by the Polish government. The New York Times' reporting on the visit sparked extensive commentary and sharing on Twitter. The level of bot activity was low.

The volume of English-language Twitter posts about NATO in Poland remains at a high level. Anonymous human-operated accounts posted 46% of all English-language messages, compared to 29% for the Baltic States.■

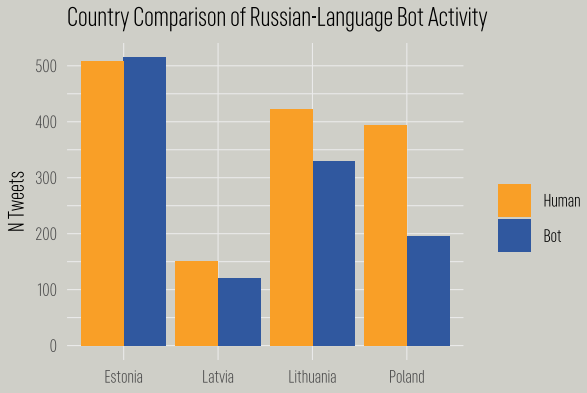


Figure 2: Country comparison of Russian-language bot activity for posts mentioning NATO and Estonia, Latvia, Lithuania, or Poland.

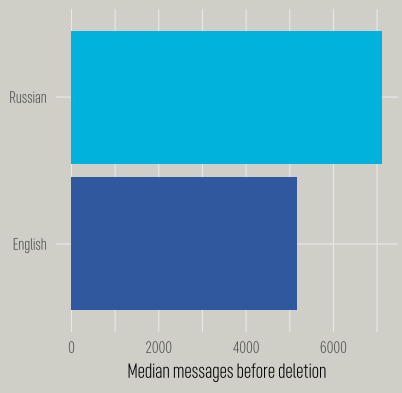


Figure 3: Deleted users - median number of messages posted before human-operated accounts were removed.

Themes

Our Robotrolling reports attempt to present as clear a view of online manipulation as possible through data and statistics. Yet the judgements about what constitutes bad behaviour is subjective, and the reader may disagree with our methodologies. In this section, we rely on Twitter's own judgement by analysing accounts banned or suspended from the platform as of 1 May 2018.

In the summer of 2017, Twitter reportedly made big changes to its platform to crack-down on bots and trolls. Since then the social network has boasted of challenging 9.9 million automated accounts per week (May 2018). In this section, we examine accounts deleted from the platform since this change, in the six months from 1 May to 31 October 2018. During this period, 2 900 unique tweets have been posted about our keywords by Twitter users who have since been removed from the platform. This figure corresponds to about 13% of all tweets during the period. Two thirds of the 900 users in our dataset that Twitter removed posted primarily in English.

Our analyses have consistently shown that the problem of manipulation is worse for Russian-language posts than for English, and that relatively speaking the rate of improvement is slower for Russian. It is therefore perhaps unsurprising that the bulk of deleted accounts operate in English.

According to Twitter, the platform will delete or suspend users for three main reasons: if the account is identified as 'spammy', if the account may have been hacked, or if the account exhibits abusive behaviour. Of these, spammy accounts make up the majority. Additionally, users may have disappeared because they chose to leave the platform. Spammy accounts are often fully or partially automated. They might direct traffic to external websites, promote various products or ideas, or game social metrics.

If we compare accounts removed in 2018 to 2017, Twitter has this year succeeded in taking quicker action. In 2017, accounts which were subsequently deleted on average posted at least 12 000 messages. For 2018 the figure stands at 6 500 messages, a dramatic improvement. Also, the proportion of accounts removed before reaching 1000 messages has improved from 13% to 22%. While these numbers suggest that users who break the platform's rules are still not swiftly dealt with, the progress is laudable.

We have observed thousands of spam bot accounts, which have not been removed from the site. Many of these have been inactive since 2017. In fact, the majority of deleted accounts were not bots. Instead, they tended to be disposable anonymous human-controlled accounts.

However, as Figure 3 shows, malicious account removal is slower for Russian-language activity. This is surprising, given the scale of the problem is larger for this space. The discrepancy is especially noticeable for accounts we believe are operated by humans, not bots. In the period January – October 2018, Russian-language accounts on average posted 7 000 times before deletion, compared to 5 000 times for English-language accounts. And the proportion of accounts removed is smaller, at 20%, compared to 29% for English.

Given these accounts are unlikely to be automated, they will mostly have been removed due to engaging in abusive behaviour. The discrepancy between languages strongly indicates that reporting and moderating procedures for Russian (and probably other foreign languages) are less effective. ■

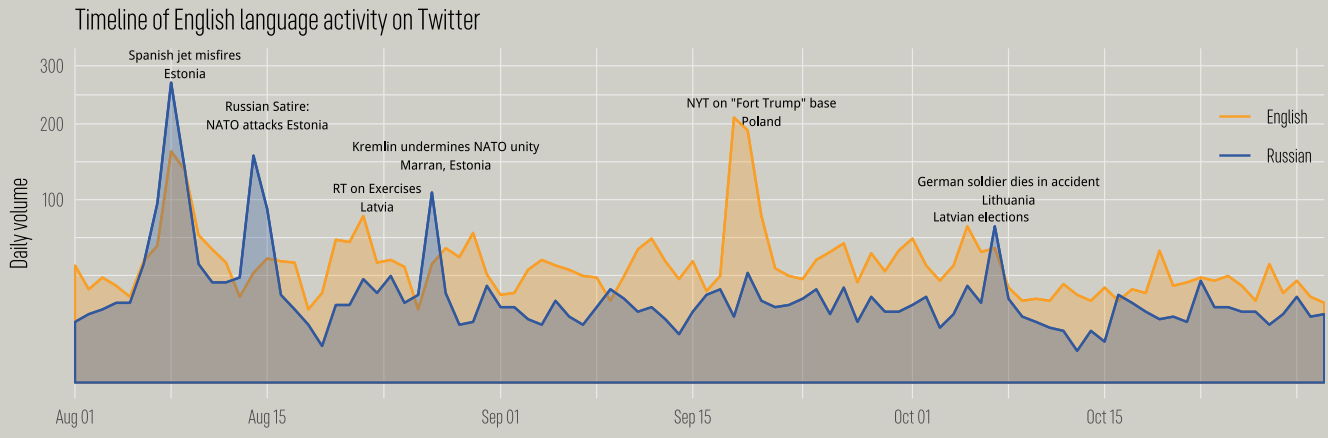


Figure 4: Timeline contrasting English and Russian-language Twitter posts for the period May – July 2018.

Robo-topics

The Russian-language Twitter space is dramatically different to the English-language space. It is dominated by activity from a hard-core of largely anonymous accounts and professional bloggers. Few institutions have a noticeable footprint.

Figure 5 below gives a unique perspective on the Twitter conversation. The main figure represents users as circles, scaled to the number of times they have mentioned NATO in the Baltics and Poland. The edges connect users to each other, based on whom they most frequently mention or retweet. The users are coloured according to account-type as estimated by our algorithm. Around the edge of the figure are disconnected nodes. These accounts have mentioned our keywords, but have not retweeted or mentioned other users within the dataset. A disproportionate number of these accounts are automated.

The figure shows two main groups of users – those operating in the Russian-language to the right, and those in English to the left. The difference in colouring should be immediately striking: note how little blue there is within the Russian-dominated area. This means there are few prominent institutional or recognisably human accounts. The English-language area is different: a large group of active institutional accounts cluster to the bottom of the figure. This group includes NATO. Moving upwards is a predominantly blue area dominated by recognisably human and news media accounts. Beyond this is a space centred on US President Donald Trump. This area includes a high proportion of anonymous accounts. In the centre, bridging the Russian and English spaces we can find the Kremlin-backed media outlets RT and Sputnik. Users interested in various types of conspiracy theories dominate this space.

The panes to the right slice this network according to various criteria. The first shows Russian-language users. The second highlights users active this quarter, and the third shows inactive users who posted in the last 12 months, but not in the last 3 months.

The pane splitting users into active and inactive reveals that anonymous users centred on Donald Trump has been quiet this period. Instead, the main activated communities are the entire Russian-language space, the area around Sputnik and RT, and the core NATO community.

Additionally, notice that:

- Virtually the entire core Russian-language community seen in the past 12 months also took part in the conversation about NATO in the Baltics this quarter. This core is highly active and persistently engaged in the NATO conversation.
- Visualised in this form, bots are relatively peripheral to the conversation.
- The disconnected accounts surrounding the Russian-language area create a golden semi-circle, as they are dominated by bots.
- RT and Sputnik are popular with Internet users interested in conspiracy theories. They form something resembling a bridge for pro-Kremlin messaging into the English-language space. ■

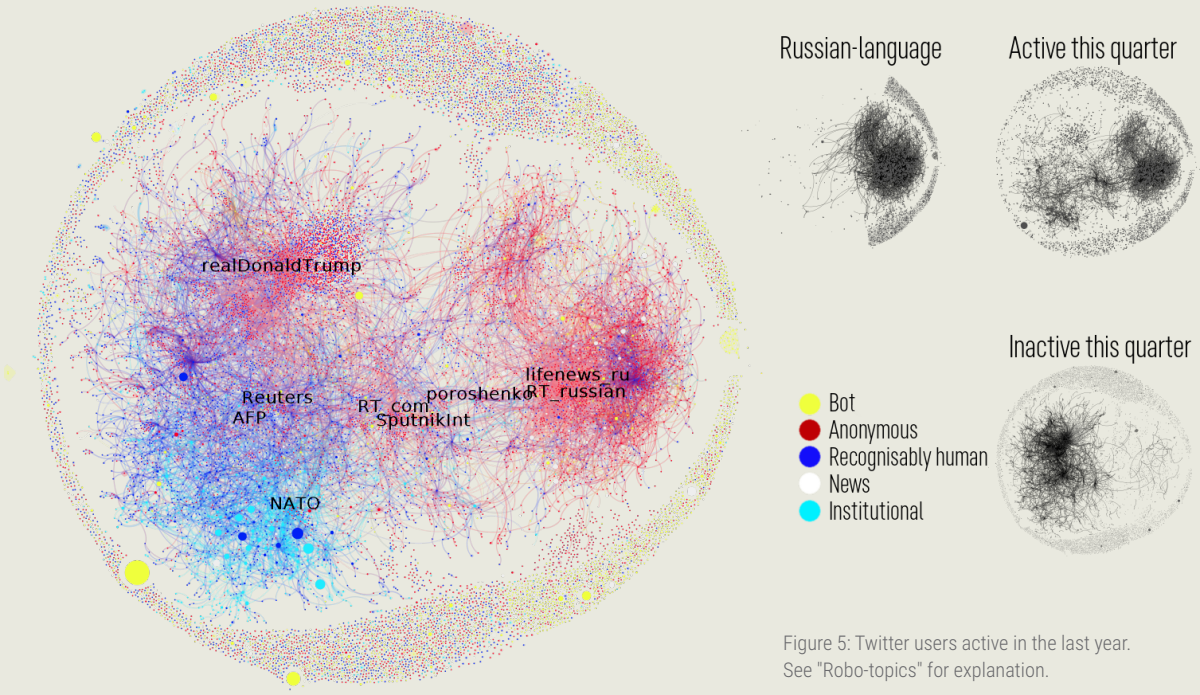


Figure 5: Twitter users active in the last year. See "Robo-topics" for explanation.

In Depth: Bots and Elections

The Latvian elections, conducted on 6 October 2018, passed with remarkably little Russian-language activity about the NATO presence in the country. In September and October the volume of messaging about Latvia, from bots and anonymous accounts alike, hit successive lows at a third of the long-term median number of 150 unique messages per month.

It is in itself surprising that messaging about Latvia and NATO should be at such low volumes, but the electoral calendar makes these figures even more remarkable. Though one must be careful not to read too much into aggregate figures, there is no evidence here of Kremlin-backed accounts spewing out content on Twitter with a view to affecting the elections. On the contrary, if anything it is possible care was taken to avoid actions that might be construed as interference.

Elsewhere this quarter, Sweden conducted parliamentary elections on 9 September. The Swedish Defence Research Agency (FOI) has published reports on bot activity in the run up to the vote.[1] The study is based on a machine learning approach similar to that underlying these Robotrolling reports, but the bot estimates are not directly comparable due to differences in model design and implementation. Nonetheless, a number of findings are worthy of note:

The researchers estimate that between 6 and 17% of accounts which tweeted about the Swedish elections were automated. These figures put the Swedish political conversation roughly on par with the English-language activity about NATO in the Baltics and Poland for 2018, and significantly less polluted by automation than the Russian-language space.

The authors note that of the political parties, the right-wing party the Sweden Democrats and the newly formed party Alternative for Sweden received the most support from automated accounts. Further, users expressing traditionalist, authoritarian, or nationalist views were more likely to be banned from the platform. The report makes no suggestion that the automated activity had foreign origins.

During the pre-election period, the volume of activity from political bots almost doubled as the election approached. This reinforces our finding that bot activity is dynamic and may peak at politically sensitive moments.

The researchers note that account deletions appeared to decline as the election approached, and speculated this was due to Twitter being slow to suspend accounts. This finding supports the view expressed in this report that action is required to speed up the moderating and account suspension process. Especially in an election context where attention is focused on a single day, delayed action is not good enough as manipulators may achieve their goals before being suspended.

The study is a good example of civil authorities working together with the research community to quantify the scale of online manipulation. Publishing figures such as these helps cut through the hyperbole about manipulation, anchoring the conversation in facts. Moreover, tech companies and investors alike measure performance against metrics such as monthly active users. We hope that publishing metrics about manipulation will encourage social media platforms to continue addressing the problem.

Overall, the news about the elections is good. It appears both the Swedish and Latvian elections passed with less social media manipulation than observers might have expected. ■

[1] Swedish Defence Research Agency, Swedish Election and Bots

Prepared by Dr. Rolf Fredheim published by
NATO STRATEGIC COMMUNICATIONS
CENTRE OF EXCELLENCE

www.stratcomcoe.org | @stratcomcoe | info@stratcomcoe.org