

ROBOTROLLING
2018. ISSUE 3



ROBOTROLLING

PREPARED AND PUBLISHED BY THE
**NATO STRATEGIC COMMUNICATIONS
CENTRE OF EXCELLENCE**



Executive Summary

President Trump’s whirlwind tour of Europe in July provoked ferocious discussion about NATO on social media. Anonymous human-controlled English-language accounts, expressing positions in support of or in opposition to the US President, dominated online conversations. Compared to the levels observed in the Spring issue of Robotrolling, the volume of English-language messages has more than doubled.

The increasing proportion of anonymous accounts active during key political moments indicate that anonymity is being abused to cloak manipulation on social networks. We call on social media companies to keep investing in countering platform misuse.

The social media companies Reddit and Twitter have released lists of accounts identified as originating from the

notorious St Petersburg ‘troll factory’—the Internet Research Agency (IRA). In this issue, we present the first quantitative analysis comparing English- and Russian-language posts from these accounts. The IRA bombarded citizens in Russia and its neighbouring states with pro-Kremlin propaganda. For English, fake accounts posed as Trump supporters, and argued both sides of the Black Lives Matter controversy. Russian-language material closely echoed and amplified the narratives popularised by Russian state-media.

Amongst the accounts identified by Twitter, 26 also posted about NATO in the Baltics and Poland. Our algorithm correctly identified 24 of these as bot accounts. The other two accounts were anonymous human-controlled (troll) accounts. ■

The Big Picture

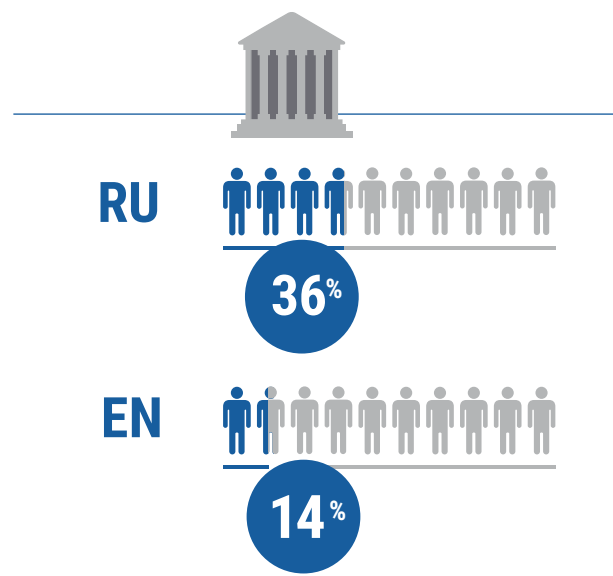
During the period 1 May – 31 July 2018, Russian-language bots created 49% of all Russian-language messages about NATO in the Baltic States and Poland. Of accounts posting in Russian, 36% were predominantly automated. In comparison, bots made up 14% of accounts tweeting in English and created 19% of all English-language messages for the quarter.

We analysed Twitter-mentions of NATO together with one or more of the host countries Estonia, Latvia, Lithuania, and Poland. The total number of posts considered was 14 600, of which 23% were in Russian. The number of active users was 9 100. Both the number of posts and the number of users more than doubled compared to the period February – April 2018.

In July 2018, the NATO summit coincided with an increase in the proportion of English-language posts about NATO originating both from automated (bot) accounts, and from anonymous human-controlled accounts.

For the first time, we observed that human-created Russian-language messages about each of the four states were in the majority. However, bots were still responsible for nearly 50% of all Russian-language content. The proportion of activity from anonymous or low-quality accounts was comparable for both English- and Russian-language spaces—approximately 35%. This percentage is significantly higher for English than expected simply as a result of Twitter banning automated accounts. As a result, we see the Russian-language space as currently dominated by inauthentic activity. As Twitter has moved to restrict automated users, anonymous, human-controlled accounts are becoming increasingly prevalent, both for the English and Russian-language spaces.

President Trump’s trip to Europe in July 2018 provoked unprecedented English-language Twitter activity about NATO in the Baltics and Poland. In the past, English-language spikes have been the result of bot activity, but in July 44% of English-language messages were traced back to anonymous, human-operated accounts compared with an average of 25% since March 2017. ■



Country Overview

The Russian-language space was dominated by commentary on events surrounding the military exercises in late May and early June. English-language content repeatedly spiked in July—at the time of the NATO summit in Brussels, and in anticipation of the meeting between President Trump and President Putin in Helsinki.

Russian-language bot activity first spiked on 28 May when a Polish Ministry of Defence proposal to fund a permanent US base in Poland was released. The most commentary, from bots and humans alike, was prompted by former Estonian President Ilves’s comments in early June, when he speculated about possible NATO reprisals in the event of an attack on Estonia. Together, these comments provoked the highest level of Russian-language Twitter activity in the last year. By contrast, July was largely free of bot messages.

Estonia

Russian-language bots echoed the negative coverage in Russian state-media about the Open Spirit exercises in Estonia in May, and Saber Strike in the Baltic States and Poland in June. Of particular note was the attention paid to President Ilves’s comments that an attack on Estonia would risk Russian territorial losses. In the English-language space, reports on Saber Strike gained modest traction, largely attributable to bot activity. During the exercises, bot activity exceeded human activity for the first time since the Russian military exercise Zapad in September 2017. As Figure 4 shows, the resignation of the US ambassador to Estonia on 29 June drew high levels of commentary, as did reports that Putin’s jet took an unauthorised shortcut across Estonian airspace en route to the meeting with Trump in Helsinki on 16 July.

Latvia

The main event concerning NATO in Latvia was Canadian Prime Minister Trudeau’s visit prior to the NATO summit in July. Coverage of the visit was largely positive. Compared to previous quarters, the number of English-language posts almost tripled to 3 500; the number of Russian-language posts remained stable at 1 600.

Lithuania

Lithuania attracted the highest percentage of Russian-language bot posts—49%. On 13 June the Lithuanian Ministry of Defence exposed a fake report about a boy supposedly killed during the Saber Strike exercises. At the NATO summit, President Dalia Grybauskaitė told reporters that President Trump had not threatened to quit the alliance. Messages about both statements were shared extensively by English-language users. Meanwhile, users posting in Russian highlighted an incident in which 13 American soldiers suffered minor injuries in a road accident.

Poland

English-language mentions of NATO in Poland more than doubled this quarter, and Russian-language messages increased by more than 50%. The increase was in large part driven by reports in May that the Ministry of Defence was preparing to fund the permanent stationing of a US armoured division in Poland. The high level of comments persisted throughout June and July, as partisan arguments about President Trump’s position on Russia and NATO drew on facts about Poland’s defence spending and the purchase of US Patriot missiles. ■

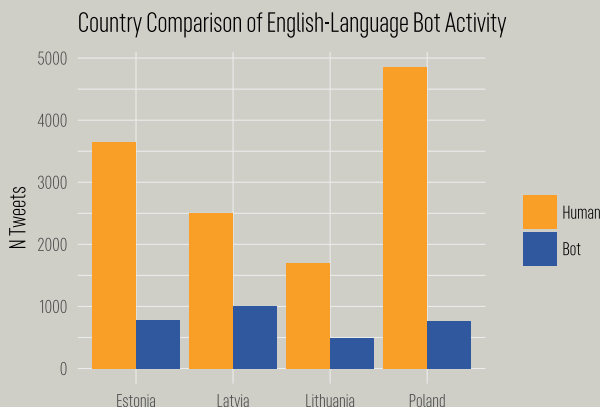


Figure 2: Distribution of English-language posts mentioning NATO and Estonia, Latvia, Lithuania, or Poland.

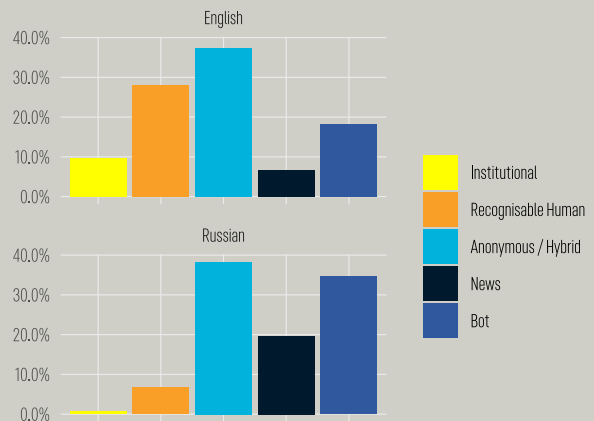


Figure 3: Percentage of Twitter posts by account type and language.

Themes

In July, hyper-partisan anonymous English-language accounts debated President Trump's position on NATO and Russia. As a result, the trajectories of English- and Russian-language content diverged. The number of Russian-language messages was 50% higher for June than for July; for English the July volume was twice that of June. A timeline of Twitter mentions can be seen in Figure 4. The peak of 1 500 messages—excluding retweets—during the NATO summit was twice as high as the previous peak during Trump's July 2017 trip to Poland.

The low levels of Russian-language content in July are not only striking, they are also consistent: the volume of comments on the popular Russian-language network VKontakte follow the same pattern. Russian-language social media users did not forget about the Alliance—at the time of the Brussels summit messages about NATO were high at twice the volume of messages posted in the period coinciding with the Skripal poisonings. But, Poland and the Baltics were not part of the conversation. Instead, messaging focussed on discord within the Alliance, on demands that its European members pay more for security, on threats that the US might withdraw, and on hopes that a diminished NATO would result in a new multi-polar order where Trump and Putin are players, but the European NATO states are not.

Only a handful of the accounts active in the Russian-language space represented institutions or belonged to recognisable individuals (see Figure 3). In addition to the messages coming from predominantly automated news accounts, roughly 35% of messages were from bot accounts, which, at first glance, seemed to belong to private citizens, but in fact were operated

automatically. The proportion of bot accounts is slowly decreasing for both in English and Russian. However, this decrease is offset by an increase in anonymous human activity.

This quarter, the rise in anonymous English-language activity has been especially noticeable. Twitter users expressing hyper-partisan views—on both sides of the political spectrum—were disproportionately likely to hide behind anonymous profiles. There are many legitimate reasons why citizens may choose to voice their political views anonymously. However, there is justified concern that individuals, groups, and shadowy digital marketing agencies operate multiple accounts to create false impressions of public opinion, or to promote sponsored messages.

In a sample of English-language Twitter posts from anonymous accounts, 4 in 10 came from ardent Trump supporters, 2 in 10 were passionately anti-Trump, and roughly 4 in 10 posts came from other miscellaneous users. In July, twice the proportion of anonymous English-language messaging originated from accounts posing as supporters or detractors of the President, compared to previous months. With the US mid-term elections approaching, the aggregate numbers suggest fake human activity is on the rise. The fact that NATO's deterrence efforts in Poland and the Baltics are being mobilised in this way is concerning.

Trump supporters cited the sale of US missiles to Poland as evidence that the President is being tough on Russia. His detractors argued the President's brinkmanship threatens to undermine the Alliance's message of unity and assurance, and its deterrence efforts in the Baltics and Poland. ■

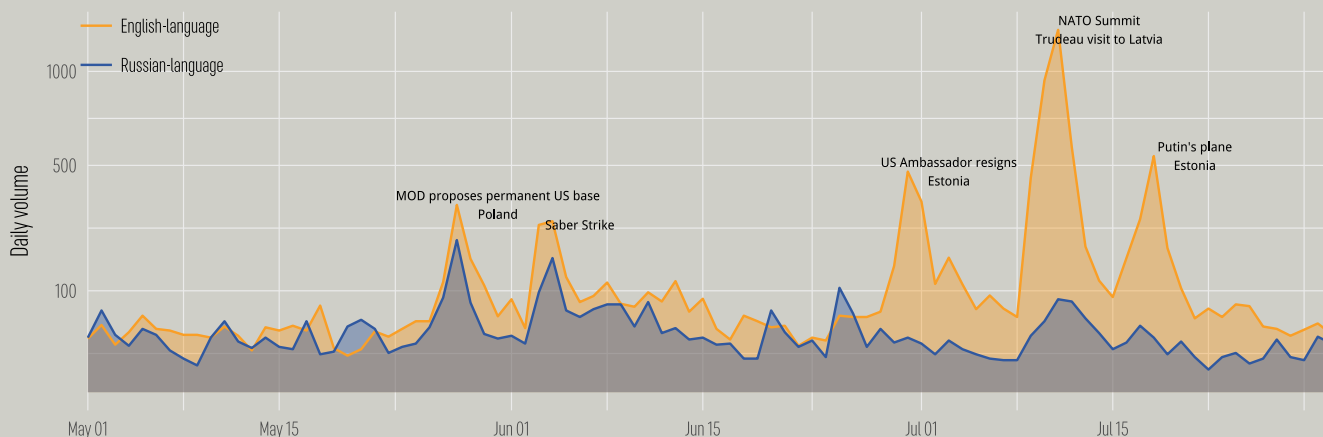


Figure 4: Timeline of all Russian-language mentions of NATO on Twitter and VKontakte for the period 1 September 2017 – 30 April 2018

Robo-topics

Public datasets released in April and July 2018 document the social media activities of the Russian St Petersburg ‘troll farm’ the Internet Research Agency (IRA) on Twitter¹ and Reddit². Our analysis is the first to compare Reddit conversations with English and Russian-language Twitter posts from identified IRA accounts. We found that the approach to different English-language environments was similar, but markedly different to Russian Twitter.

We used machine learning and natural language processing to identify the main topics of conversation in the three datasets. A simplified overview of these topics is shown in Figure 5, visualising the sharp difference in the messaging aimed at English and Russian-language audiences. The topics—represented by circular nodes—are scaled to reflect the number of posts about a particular topic, while the proximity and connections between topics shown map the frequency at which topics occur together.

Conversations on Reddit were dominated by issues relating to US politics and racial tensions. Messages in support of the Black Lives Matter campaign and decrying police brutality dominated, as the accounts attempted to widen existing divisions within communities, as increase intergroup tensions and conflict. The Reddit dataset also contained conversations about civil unrest, the Syrian conflict, international terrorism, and financial news, together with ‘other’ non-political content which appears designed to create more ‘realistic’ account histories.

English-language Twitter posts were concerned with the same topics as those on Reddit, but the proportion of messages about US politics was greater than for racial issues. These topics were also less closely linked. Political commentary included messaging in support of President Trump and Make America Great Again, but also in support of other candidates in the US primaries and at times critical of US Government. On the other hand, material about the Black Lives Matter movement picked up on broader issues likely to appeal to liberal and democratic-leaning audiences.

Within the Russian Twitter space, the conversation was dominated by Russian international relations, the conflict in Ukraine, and political posts relating to internal Russian politics. These topics do not feature prominently in English-language conversations. The only point of convergence across all three datasets is the conflict in Syria and the threat of international terrorism. Within Russian Twitter the largest category was content that we classified as ‘other’. This content related to the entire spectrum of Russian news—from sports to weather to entertainment and business. The broad character of these subjects reflects the fact that the Kremlin-supported trolls and bots are indiscriminately promoting news content from state media outlets. ■

¹ <https://github.com/fivethirtyeight/russian-troll-tweets/>

² <https://github.com/ALCC01/reddit-suspicious-accounts>

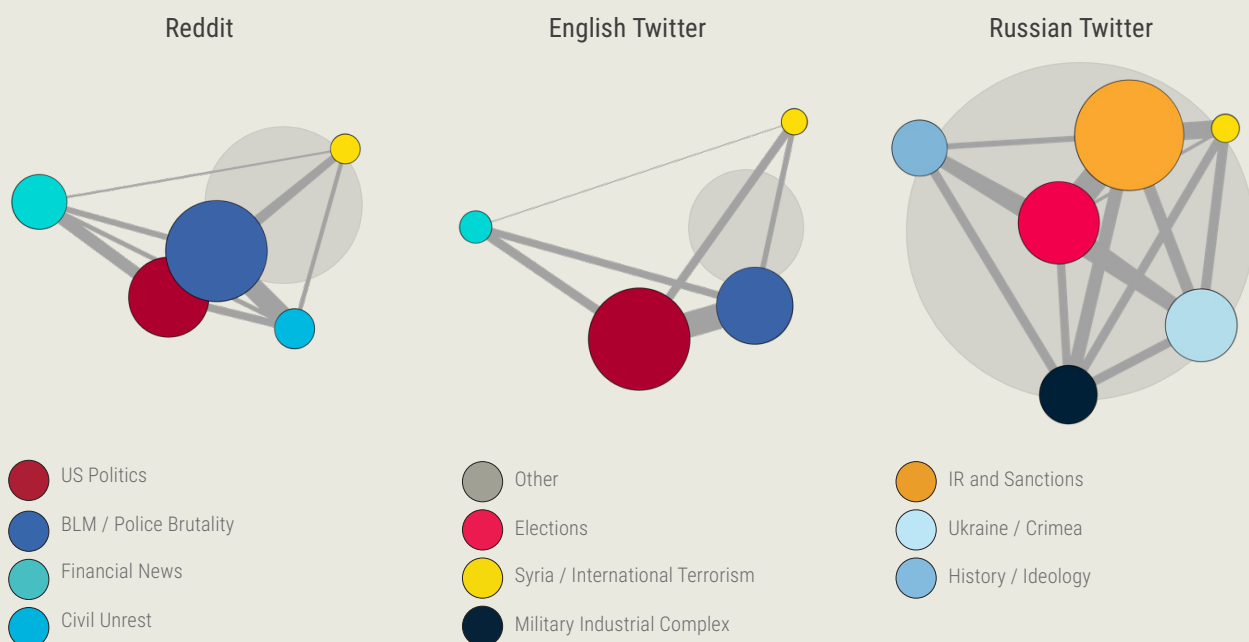


Figure 5: What do Kremlin trolls talk about? Russian-language content exhibits a scatter-gun approach, while English-language content narrowly targets charged political and social issues.

In Depth: Trolling Operations

Of the Russian-language accounts identified by Twitter, 26 users mentioned NATO in the Baltics and Poland. Of these, our classifier identified three as English-language bots, 21 as Russian-language bots (four of these were news-bots), and two as human-operated accounts posing as anonymous Russian citizens.

IRA accounts targeting messages at Russian-speaking audiences posted in support of the ruling party, of the regime's position on Syria and Ukraine and to exaggerate divisions and threats in the West. In English-language spaces the content was antagonistic, stoking tempers on both sides of the US political divide by disseminating content about the US election, US politics, and politically-charged social issues.

The IRA's activity showed sensitivity to breaking news stories and the ability to rapidly switch focus. Messages on Russian-language Twitter trended away from international questions in 2015 to become more internally focused. In particular, messages drawing on historical and ideological themes have emerged. By comparison, English-language Twitter posts have shown a shift towards US politics and away from racial issues.

On Reddit, the conversation has moved from US politics—the dominant topic in 2015—to topics of racial division. Activity regarding both topics tracks closely, highlighting how the two topics are discussed in close relation to each other on this platform, leveraging divisive social issues to antagonise and inflame political discussions.

Patterns in platform usage reveal that a higher level of customization and targeting has been invested in English-language activity, compared to activity in the Russian-language space.

Compared to Russian-language Twitter posts, English-language posts were four times as likely not to contain external links, twice as likely

to be directed at other users, and three times as likely to contain hashtags. Messages produced in English actively sought to engage with and build an audience within Twitter. Russian-language messages mainly distributed external news content and promoted messages copied from state-run media. Twitter was likely used to game other metrics, such as 'most shared' news story and video rankings.

Metric	En	Ru
Links to external sites	64%	90%
Mentions of other users	15%	6%
Hashtag use	46%	15%
Highly specialised accounts	34%	15%

Users active in English-language spaces exhibited greater focus in their topics of conversation. We labelled accounts commenting on a single topic more than 50% of the time as 'highly specialised'. On English-language Twitter we found 34% of accounts were highly specialised, compared to only 15% of Russian-language Twitter users. Thus, English-language content seems to be tailored to specific groups.

Journalists reported that the IRA operation was expensive.³ We read discrepancies between content produced for English and Russian as reflecting that greater effort went into manipulating the English-language conversation. Greater effort indicates the stakes were higher in English, the operation more sensitive, and the interventions therefore less crude. This may also indicate that it was easier to manipulate the Russian-language space, or that other platforms received greater attention. While social media companies may not be able to eliminate manipulation completely, they can make manipulation less attractive by making it increasingly difficult and expensive. ■

³ <https://www.nytimes.com/2015/06/07/magazine/the-agency.html>

Prepared by Dr. Rolf Fredheim and John Gallacher published by

**NATO STRATEGIC COMMUNICATIONS
CENTRE OF EXCELLENCE**

The NATO StratCom Centre of Excellence, based in Latvia, is a Multinational, Cross-sector Organization which provides Comprehensive analyses, Advice and Practical Support to the Alliance and Allied Nations.

www.stratcomcoe.org | [@stratcomcoe](https://twitter.com/stratcomcoe) | info@stratcomcoe.org