ROBOTROLLING

Executive Summary

Inauthentic activity in the Russian-language information space this quarter targeted NATO military exercises, peaking on 8 June with the commencement of BALTOPS in the Baltic Sea. Conversations amplified by both automated and anonymous accounts portrayed NATO's summer exercises as exceptionally aggressive and demonstrative of the alliance's ambitions to expand into Belarus and Ukraine.

Throughout this quarter, we observed a significant increase in English-language bot activity. This flurry of activity was triggered by public statements suggesting the US move nuclear weapons, troops, and military equipment from Germany to Poland for permanent deployment.

While the volume of Russian-language automated activity on both Twitter and VK shrunk this quarter, the percentage of conversations amplified by bot voices increased. State-backed media outlets Sputnik and RT were popular sources of information among both English- and Russian-language audiences.

The latter two sections of this report concentrate on the deleted account dataset published by Twitter in June 2020. First, we compare our list of previously identified bot accounts with those published by Twitter, which is visualised in Figure 5. Second, we evaluate Twitter's capacity for combatting inauthentic activity on its platform, concluding that rates of identification, suspension, and deletion of accounts are unnecessarily slow. ◼

The Big Picture

This edition of Robotrolling continues to track the online manipulation of information regarding the NATO presence in Poland, Estonia, Latvia, and Lithuania on the social media platforms Twitter and VK. Our analysis focuses on the activities of automated accounts (bots) and coordinated, anonymous human accounts (trolls). This issue identifies the key trends that emerged in the Russian- and English-language information spaces during the period 1 May to 31 July 2020.

Throughout this summer, activity referencing NATO's presence in Poland and the Baltics returned to previously-recorded levels. A total of 10 120 messages discussing the Enhanced Forward Presence (eFP) were shared on Twitter, nearly 2 000 more than were collected in the previous quarter, 1 February to 30 April 2020. This increase in message volume is the result of a dramatic surge in both authentic and inauthentic English-langauge engagement. The total number of unique users participating in English-language conversations, and the volume of tweets they produced, increased significantly.

While English-language activity soared this quarter, Russian-language messaging slightly waned. The overall volume of tweets in the Russian-language sphere diminished by 16%, while bot tweets saw a moderate decrease of 5%. Despite the reduction in volume of automated messages, the share of conversations conducted by bots increased modestly from 34% to 38%. As bot activity increased, faceless and identifiably human engagement shrunk during the summer months.

Automated activity on VK commanded a marginally larger portion of conversations this quarter, as the percentage of online conversations contributed by bots rose to 41%. However, this increase was not driven by a higher number of bot users engaging in discussions about the eFP, but because the pool of total unique users shrunk by nearly a quarter.

Messages published by both bot and anonymous accounts, which we refer to as inauthentic activity, exhibited nearly identical patterns in engagement with the NATO presence.

This summer, the focus of Russian-language bot activity on both Twitter and VK was relatively equally spread among the Baltic countries and Poland. On Russian-language Twitter, anonymous user activity decreased across the board. In contrast, 60% of bot conversations in the English-language information space were about Poland. ◼

RU **26%**

EN **9%**

RU **26%**
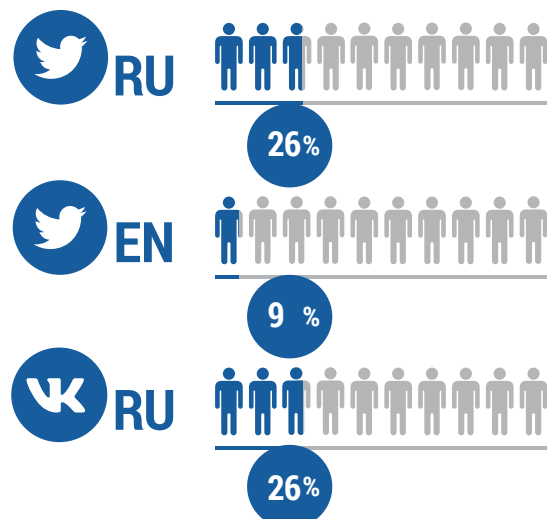
Country Overview

Russian-language discussions about NATO's presence in Poland and the Baltics peaked on 8 June on both Twitter and VK. This spike in inauthentic activity, driven primarily by bot accounts, coincided with the commencement of BALTOPS, a NATO military exercise in the Baltic Sea involving personnel from 19 NATO members and partners. On both platforms, bot activity remained somewhat constant throughout May and July, with a cluster of spikes in June corresponding with military exercises.

On English-language Twitter, bot activity intensified in tandem with recent developments in Polish affairs, particularly regarding US-Polish relations. Inauthentic activity remained low throughout May and began to increase in June before peaking on 14 July. The main spike this quarter on June 13 came from largely authentic accounts reacting to Polish president Duda's re-election.

Estonia

Estonia was the second-most mentioned country by English- and Russian-language automated accounts on both Twitter and VK, but received the most anonymous attention on Twitter. This rise in bot engagement is attributed to meetings between members of the Estonian Defence Ministry and NATO officials. In early June, Permanent Secretary of the Ministry of Defence Kristjan Prikk met with David Kettler, NATO Assistant Secretary General for Intelligence and Security. Bots highlighted that they discussed the "so-called" Russian threat faced by Estonia. The following month, Estonian Defense Minister Jüri Luik met with French military and NATO leadership. Automated users focussed on Luik's statement that Russia's missile system threatens the entire NATO alliance.

Latvia

This quarter, Latvia was the top target of Russian-language bots, but received much less attention on VK and among English-language bot users. Bots amplified the complementary narratives that Latvia is occupied by NATO and that Latvia perpetuates the 'falsehood' of oppression under the USSR. In early July, bots circulated a post claiming that the Latvian government is restricting ethnic Russians' freedoms by attempting to erase the Russian language from schools, media, and politics. Later that month, bots shared a report that Latvia's ambassador to Russia is urging Moscow to appoint a permanent representative to NATO to bolster relations.

Lithuania

Inauthentic activity directed at Lithuania on Twitter decreased this summer compared to the previous quarter. However, Lithuania remained the most-referred-to country by bots on VK. The dominant theme of bot messaging was the idea that Lithuania relies on the 'myth' of a Russian threat to justify its military spending. On 22 June, bots circulated a statement made by Raimundas Karoblis, the Minister of Defence, during a meeting with other defence ministers of NATO countries in which he emphasised the threat of Russia's missile capacity. At the end of July, France delivered military equipment and personnel to the NATO battalion in Rukla, which was criticised by Sputnik as a 'mobile circus tent'.

Poland

Mirroring our observations from the previous quarter, Poland received the bulk of English-language bot engagement and the fewest Russian-language bot mentions. On 15 May, the US ambassador to Poland suggested moving nuclear weapons from Germany to Poland. Shortly thereafter, bots circulated an RT article written by former American diplomat Stephen Pifer that warned against causing another Cuban missile crisis-style escalation between NATO and Russia. On 25 June, bot users shared the Russian Foreign Ministry reaction to rumours that the US will move 2 000 troops from Germany to Poland for permanent deployment. ■
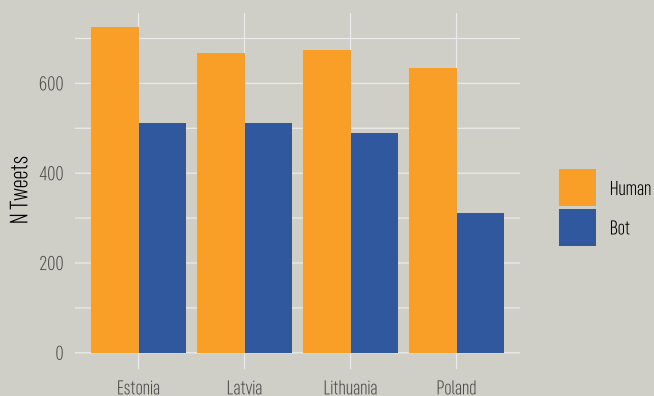

Figure 2: Country comparison of Russian language bot activity.
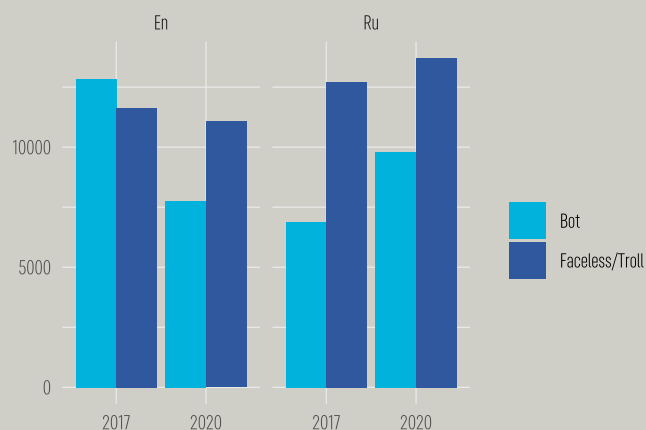

Figure 3: The average (median) number of posts by accounts which were subsequently deleted from the platform.

Themes

This quarter, increased levels of inauthentic activity coincided with NATO military exercises and reports of plans to alter NATO's current structure in Europe. The underlying message connecting these spikes in activity portrays the alliance as pursuing a policy of Russophobic aggression and coercive expansion. We observed bot accounts amplifying this content, including articles published by Kremlin-backed RT and Sputnik, as well as fringe outlets NewsFront and InfoWars. Bot-driven discourse targeting the Baltics and Poland continued to criticise the decision to hold joint military exercises amid the COVID-19 pandemic, accusing their leadership of prioritising NATO interests over public health.

Military exercises this summer were characterised as particularly provocative. In early June, Russian-language bots circulated claims that the BALTOPS exercise - along with other exercises in the Baltic Sea - was being used as a tool to force Sweden and Finland to join the NATO alliance. Inauthentic Russian-language activity peaked between 7 and 9 June, as bots kept their efforts focused on BALTOPS-related coverage. On 9 June, robotic users shared a transcript of a Sputnik Radio interview with Yuri Shvytkin, Deputy Chairman of the Duma Defence Committee, in which he accused NATO of exacerbating regional tension by conducting exercises near the Russian border. A NewsFront article published on 11 June reported that three NATO vessels were anchored near an unfinished section of the Nord Stream 2 pipeline, alleging that the purpose of their deployment could be sabotage.

Throughout this quarter, we identified bots engaging with the narrative that NATO is attempting to infiltrate Russia's traditional sphere of influence. In late June, they circulated claims that NATO deployed troops along the Polish border with Belarus ahead of the election in preparation for a "democratic intervention". Simultaneously, inauthentic accounts shared an article accusing NATO of using the conflict in the Donbass as a training ground for Baltic snipers. Another murky story from Ukraine was circulated by anonymous accounts on 16 July, reporting that several "saboteurs" in possession of NATO-grade equipment were killed in the Donbass. Although the details of this story are contested, its popularity illustrates how the theme of clandestine interference fuels bot activity.

While the bulk of inauthentic activity clustered around military and diplomatic events, a segment of the conversation was dedicated to bolstering Russia's image of military superiority. As BALTOPS was underway, bots circulated Russian media articles describing an editorial piece published in Forbes. The editorial denounced the plan to reduce US forces in Germany as weakening NATO's position in the region. The content of this article was appropriated by a range of news outlets—including some resources linked to Prigozhin's St. Petersburg troll farm—and systematically shared to falsely promote the claim that Forbes predicted Russia would be victorious in a hypothetical conflict. ■
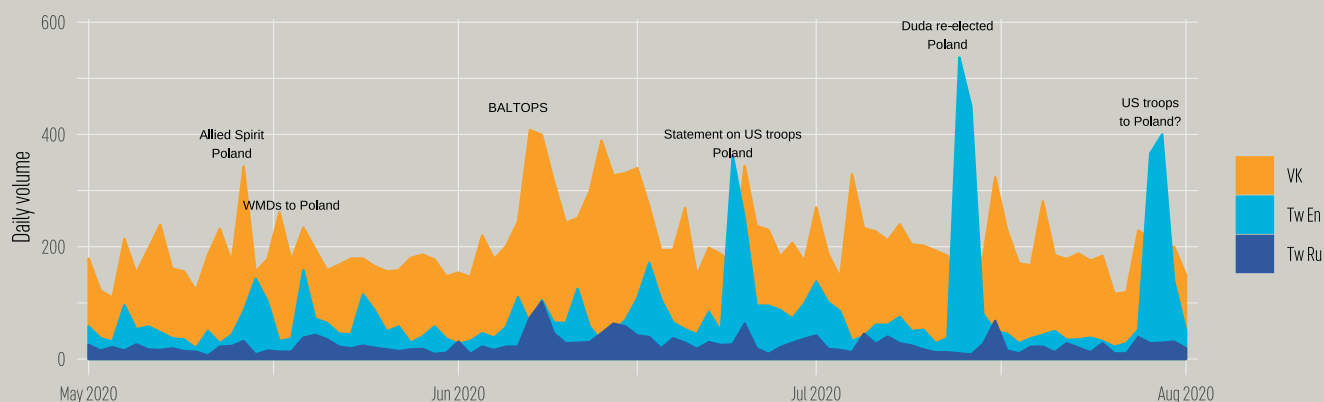

Figure 4: Timeline showing spikes in activity about the NATO presence on English and Russian-language Twitter, as well as VK.

Robo-topics

In this section, we analyse the extent to which Twitter has improved its ability to combat fake activity on its platform. In our first Robotrolling report, we estimated that the majority of users posting in Russian about NATO's presence in the Baltics and Poland originated from bot accounts. Having revised the algorithm, we believe that the percentage of automated Russian-language accounts posting during the first half of 2017 was 56%. As of August 2020, only 30% of these accounts have been deleted by Twitter. The majority of the remaining bots are currently dormant; the owner has either lost access to the account or has abandoned it.

The process of fake content removal is consistently slow. The (now dormant) bot accounts we flagged were only challenged in 2018, a year after we first identified them. As of August 2020, only 12% of the bot accounts active in 2019 have been deleted. The percentage for 2020 is smaller still.

In 2020 alone, 38 separate automated accounts with over 1 million tweets mentioned the NATO presence. To date, only five have been removed. The bar chart in Figure 3 shows the median number of posts from deleted accounts. It reveals that, in the four years we've been monitoring, there has been no significant increase in the speed with which accounts are removed after breaking Twitter's terms of service.

The current rate of deletion is severely lagging. For both English and Russian-language automated accounts, the average point of deletion tends to be in the range of 5 000 to 10 000 posts. Consequently, fake accounts posting about the NATO presence in the Baltics and Poland have the potential to amplify thousands of potentially malicious posts before action is taken.

Figure 5 features two representations of deleted account activity. Figure 5a shows the deleted accounts together with the main accounts they clustered around, whereas 5b zooms in on accounts that were closely connected through interaction with Current Policy, the central node in the state-backed network removed by Twitter in June 2020.

Blue nodes are active accounts and yellow nodes are deleted accounts, while orange nodes were included in Twitter's list of accounts used for information operations. The closer accounts are positioned together, the more they engaged with one another. The constellation shows pro-Kremlin and nationalist accounts at the top, whereas the central areas are occupied by major news outlets and independent or opposition figures, such as Alexei Navalny. ■
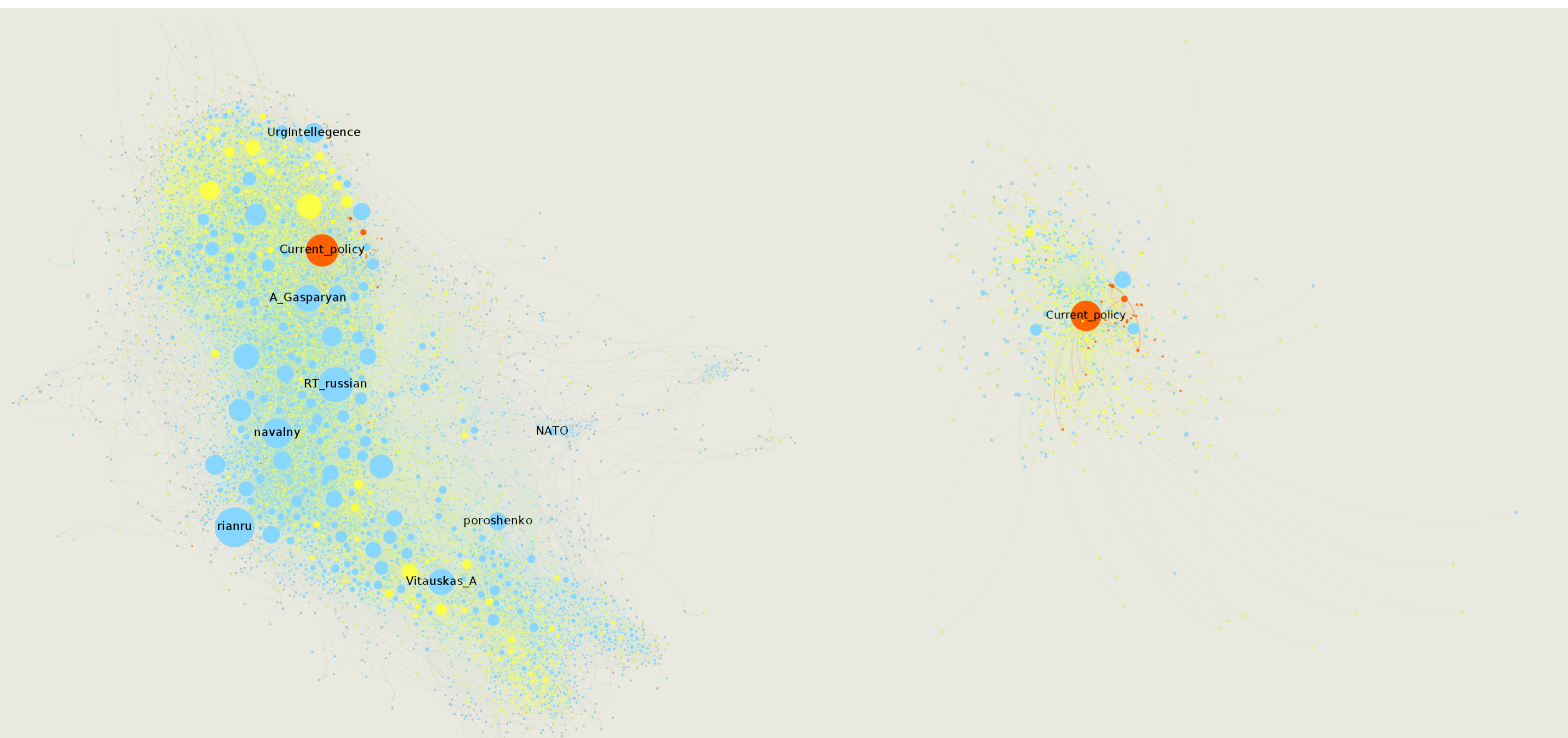

Figure 5 : Network diagram of social media accounts most closely with deleted accounts. Blue nodes are still active, yellow nodes are deleted, while orange ones were included in Twitter's list of accounts used for information operations. Nodes size in figure 5a is scaled by the number of connected accounts. Figure 5b zooms in on activity around Current Politics. Here node size is scaled to the number of posts mentioning NATO in the Baltics and Poland.

In Depth: Twitter and State-Backed Bots

In June 2020, Twitter released information about a cluster of fake accounts that they had removed. Many of the suspended accounts were already familiar to us. Along with reinforcing the accuracy of our bot-detection algorithm, this dataset provides a glimpse into Twitter's method of fake activity removal. While they were active, the primary purpose of these Twitter users was to artificially boost the visibility of particular news stories, a phenomenon identified and reported in Robotrolling as early as 2017.

Below is the statement from Twitter:

Aided in part by useful information sharing from external researchers and our peer companies, we investigated accounts associated with Current Policy, a media website engaging in state-backed political propaganda within Russia. A network of accounts related to this media operation was suspended for violations of our platform manipulation policy, specifically cross-posting and amplifying content in an inauthentic, coordinated manner for political ends.

Twitter's rationale is intriguing; were the automated accounts engaging in 'amplifying content in an inauthentic, coordinated manner' banned because Twitter is committed to removing bots from its platform, or because this coordination has been attributed to promoting 'state-backed political propaganda'?

If Twitter is serious about addressing the proliferation of bots on its platform, the answer must be because the activity itself is not tolerated. Yet evidence suggests that automated accounts not primarily spreading pro-Kremlin content were frequently not removed. Take Figure 5, where all the users coloured in blue were identified by our algorithm as having a bot likelihood in excess of 99%. Despite being frozen since 2018, they have not been removed by Twitter.

We were able to cross-tabulate messages about NATO in the publicly released but anonymised dataset with those in our collection to identify which accounts had been removed. Amongst the accounts released were many that we have previously observed discussing the NATO presence in Poland and the Baltics. The accounts we identified were most active during 2017, before Twitter increased its suspension efforts.

In April 2017, the number of tweets from the most active accounts dropped from 100 posts per day to no more than 50. This drop coincided with Twitter's intensified scrutiny of Russian bots, and demonstrates how such accounts attempted to evade detection by reducing their output.

The most active accounts were primarily using the same automation service, dlvr.it, which pushes tweets about newly published articles matching particular criteria. This blatant example of coordinated inauthentic activity should have been automatically detected by Twitter many years ago. Instead, many of these accounts remained active until December 2019.

Taken together, these data offer a number of striking insights: pro-Kremlin propaganda is regularly disseminated through automated accounts, and these accounts may be coordinated. While the underlying logics may be clear to the analyst, they are often sufficiently distinct as to evade automatic detection. Above all, these data reveal how adaptive online manipulators are to changes implemented by social media platforms. ■