FEBRUARY 2023

DETAILED REPORT

Quarterly Adversarial Threat Report

Ben Nimmo, Global Threat Intelligence Lead
Margarita Franklin, Director of Public Affairs, Security
David Agranovich, Director, Threat Disruption
Lindsay Hundley, Security Policy Manager
Mike Torrey, Security Engineer

Meta

TABLE OF CONTENTS

PURPOSE OF THIS REPORT

Our public threat reporting began over five years ago when we first shared our findings about a coordinated network of fake accounts from Russia engaged in misleading people about who is behind it and what they are doing. Since then, global threats have significantly evolved, and we have expanded our ability to respond to a wider range of adversarial behaviors – from coordinated inauthentic behavior (CIB) to cyber espionage to mass reporting and other adversarial behaviors. As part of our quarterly adversarial threat reports, we now publish information about the networks we take down and our broader threat research to make it easier for people to see progress we're making in one place. We welcome ideas from the security community to help us make these reports more informative so we can adjust as we learn from feedback. In this quarterly report, we included three networks removed for coordinated inauthentic behavior and mass reporting, in addition to an update on the adversarial trends we've identified in the year since Russia began its full-scale war against Ukraine.

For a quantitative view into our Community Standards' enforcement in the fourth quarter of 2022, including content-based actions we've taken at scale and our broader integrity work, please visit our Transparency Center.

What Is Coordinated Inauthentic Behavior (Cib)?

We view CIB as coordinated efforts to manipulate public debate for a strategic goal, in which fake accounts are central to the operation. In each case, people coordinate with one another and use fake accounts to mislead others about who they are and what they are doing. When we investigate and remove these operations, we focus on behavior rather than content — no matter who's behind them, what they post or whether they're foreign or domestic.

Continuous CIB enforcement: We monitor for efforts to come back by the networks we previously removed. Using both automated and manual detection, we continuously remove accounts and Pages connected to networks we took down in the past.

What Is Mass Reporting Or Coordinated Abusive Reporting?

Under our Inauthentic Behavior policies, we remove mass reporting activity when we find adversarial networks that coordinate to abuse our reporting systems to get accounts or content incorrectly taken down from our platform, typically with the intention of silencing others.

SUMMARY OF OUR FINDINGS

- Our Q4 2022 threat report provides a view into the risks we see across multiple policy violations including Coordinated Inauthentic Behavior (CIB) and mass reporting.
- While Russian-origin attempts at **covert activity (CIB) related to Russia's war in Ukraine** have sharply increased, **overt efforts by Russian state-controlled media** have reportedly decreased over the last 12 months on our platform. We saw state-controlled media shifting to other platforms and using new domains to try to escape the additional transparency on (and demotions against) links to their websites. During the same period, covert influence operations have adopted a brute-force, "smash-and-grab" approach of high-volume but very low-quality campaigns across the internet. Notably, the two largest covert operations focused on the war in Ukraine that we disrupted were linked to private actors, including those associated with the sanctioned Russian individual Yevgeny Prigozhin, continuing a number of global trends we've called out in our threat reporting. These actors can provide plausible deniability to their customers, but they also have an interest in exaggerating their own effectiveness, engaging in client-facing perception hacking to burnish their credentials with those who might be paying them. It is critical to analyze the impact of these deceptive efforts (or lack of it) based on evidence, not on the actors' own claims, while continuously strengthening our whole-of-society defenses across the internet. More here.
- In our previous threat reporting, we called out the **rise of domestic influence operations**, which are particularly concerning when they combine deceptive techniques with the real-world power of a state. The three CIB networks we removed last quarter — **in Serbia, Cuba, and Bolivia** — continued this trend and were in some way linked to governments or ruling parties in their respective countries. Each targeted domestic populations to praise the government and criticize the opposition.
- We took action against a **CIB network in Serbia** linked to employees of the Serbian Progressive Party, known as its Internet Team, and state employees from around Serbia. They targeted domestic audiences across many internet services, including Facebook, Instagram, Twitter, YouTube, in addition to local news media to create a perception of widespread and authentic grassroots support for Serbian President Aleksandar Vučić and the Serbian Progressive party. More here.
- We also took down a **CIB operation in Cuba** that targeted primarily domestic audiences in that country and also the Cuban diaspora abroad. Our investigation linked this network to the Cuban government. The people behind it operated across many internet services, including Facebook, Instagram, Telegram, Twitter, YouTube and Picta, a Cuban social network, in an effort to create the perception of widespread support for the Cuban government. More here.

- Finally, we removed what we call a **blended operation** — coordinated adversarial activities that violated multiple policies at once – **in Bolivia** linked to the current government and Movimiento al Socialismo (MAS party), including individuals claiming to be part of a group known as “Guerreros Digitales” (“digital warriors”). It engaged in both coordinated inauthentic behavior and mass reporting (or coordinated abusive reporting) domestically in support of the Bolivian government and to criticize and attempt to silence the opposition. This operation ran across many internet services, including Facebook, Instagram, Twitter, YouTube, TikTok, Spotify, Telegram, and websites associated with its own “news media” brands. More here.

01

Analysis: Russian influence operations related to war in Ukraine

Since Russia invaded Ukraine almost a year ago, our teams have been on high alert to identify emerging threats, respond as quickly as we can, and share the steps we take to protect people's ability to stay safe and have access to accurate information.

This war has written a new chapter in our industry's collective understanding of influence operations, both overt and covert. While we've seen some of these elements around the world, this is the first time we've observed attempts at covert influence operations deployed at this scale alongside a military invasion and subsequent land warfare between two states. In response to military aggression of this magnitude, it's also the first time we've taken the unprecedented step of reducing the distribution of state media outlets.

Looking back at the past 12 months, there were a few notable changes. Russian state media outlets have significantly reduced their activity on our platforms and pivoted elsewhere, while covert influence operations have shifted to a brute-force, "smash-and-grab" approach of high volume but very low quality across the internet. We've used different enforcements against these distinct efforts, and for both areas – covert and overt influence operations – our actions have substantially impacted their activities.

State media: reduced posting, reduced engagement

As part of our response to the Russian government's invasion of Ukraine, we applied new, unprecedented enforcements to Russian state-controlled media. These included preventing these outlets from running ads globally, demonetizing their Pages and IG accounts, demoting their content in people's feeds, and adding in-product nudges that ask people to confirm they want to share or navigate to off-platform content from these outlets. We took these measures globally – across all languages – in addition to the usual transparency labels we apply to state-controlled media. Even outside of crises, we believe that people should know if the news they read is coming from a state-controlled publication.

The most recent research by Graphika found a substantial drop in the activity of these Pages themselves around the world. It also noted a substantial drop in people's engagement with the content these entities posted, which aligns with earlier open-source reporting. This behavior appears around the world, not only in places where Russian state media face additional government restrictions. Although they are still active on our platforms, a number of Russian state-controlled media have shared posts urging followers to find them on other services instead. We've also seen public reporting that some of these outlets increased their activity on other platforms.

We also detected and blocked some attempts to evade our enforcement by these state-controlled outlets. For example, we saw state-controlled media using new domains to try to evade the additional transparency on, and demotions against, links to their websites. We also found and labeled some newly created Pages associated with these outlets. However, these attempts were relatively infrequent.

The decrease in activity on our platform and the reported increase elsewhere, and the relatively low volume of recidivist attempts are likely related to the nature of media entities, which typically invest in brand building and growing their audiences over a sustained period, and attempt to occupy the information environment long term. In response to the transparency and other measures we put in place in early 2022, these Russian-origin outlets appear to have chosen to focus their investment elsewhere in hopes for a higher return.

Overall, our approach in taking these steps during these unprecedented times of war has been to balance enabling speech but not reach for Russian state-controlled media on our platforms. We've taken steps to limit the reach of these entities for people who might unintentionally encounter this content in their newsfeed, but preserve the ability for determined users to find these outlets and view their content in a way where it can be fact-checked and viewed alongside counter-speech.

Covert influence operations: smash and grab

The covert influence operations from Russia that we've disrupted since the war began put very little effort into building discernible brands or personas on our services. They instead appeared to put effort in branding elsewhere, including Telegram channels and their own websites. On our platforms, these networks resembled brute-force, "smash-and-grab" attempts to use a large number of low-quality accounts all at once, in the hope that at least a few might survive and escape detection.

Below, we share new analysis that looks back at previously unreported responses by these operations to our takedowns.

Historically, operations that originated in Russia have targeted Ukraine more than any other country.

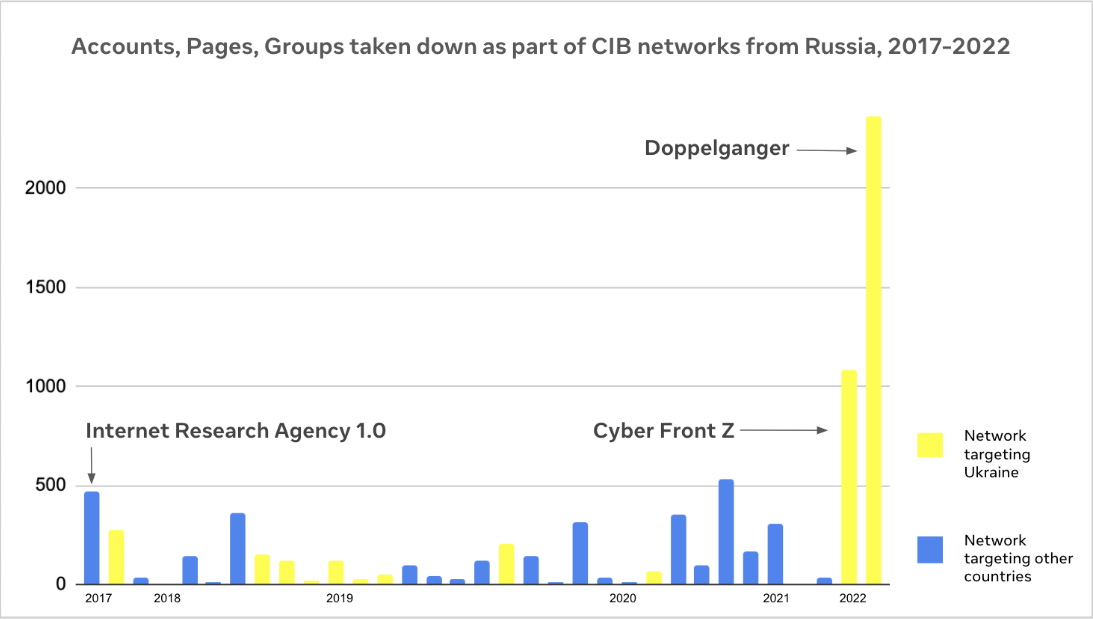

Image: Chart showing the number of accounts, Pages and Groups in Meta's takedowns of CIB networks originating in Russia, 2017-2022. Graphs in yellow show networks that were partly or wholly focused on Ukraine.

Before we disrupted these networks and publicly reported our findings in August and September, they operated at a high tempo across social media and the internet more broadly, but with very low quality and low overall return on their activity. In fact, this activity bore a closer resemblance to a spammers' playbook than the more stealthy and deceptive Russian influence operations we've disrupted in the past. The elements of sophistication in these latest covert operations were related to off-platform activity where they attempted to find and exploit vulnerabilities in the broader information ecosystem across the internet.

For example, the "Doppelganger" network actively exploited gaps in domain registration systems by setting up web addresses that were almost identical to those of major news outlets (such as bild[.]pics instead of bild[.]de), and creating sophisticated websites that spoofed those outlets' appearance. According to public reporting by EU Disinfolab, this network also restricted access to some of its domains, so that only people in the target country could view them. This is likely in an attempt to make the deceptive inauthentic networks harder to detect. Both Doppelganger and Cyber Front Z targeted many social media platforms, including Telegram, TikTok, Twitter, YouTube, LinkedIn, VKontakte and Odnoklassniki.

In response to our takedowns, each of these unconnected networks tried aggressively to rebuild their operations across the internet – much more so than has been usual for covert influence operations from Russia. Our teams have continued to detect and enforce against thousands of attempts to create fake accounts. After our initial enforcement against the spoofed domains, we've blocked hundreds more of them from being shared on our platforms, and shared them publicly so that other platforms and researchers can study them and take action, as appropriate.

The uniquely aggressive persistence of these operations, together with their high volume and low quality, may point to the operators' lack of preparation, the nature of influence operations in high-intensity war times, and the nature of the entities behind them. *First*, these operations appeared to hastily respond to the state’s military actions with no time to build credibility and influence in promoting pro-Russia narratives. *Second*, they ran an attempt to flood the information environment in support of the war. *Third*, both were likely paid to keep running these efforts aimed at defending Russia’s war and weakening global support for Ukraine, no matter how low the quality or impact of their work.

As we’ve called out in the past, private actors like these can provide plausible deniability to the sponsors to conduct their operations, but they also have an interest in exaggerating their own effectiveness, engaging in customer-facing perception hacking to burnish their credentials with those who might be paying them.

These operations’ adversarial adaptations have provided us with useful insights that we’ve leveraged to strengthen our automated detection. Our efforts have led to them expending operational resources and time to keep on trying to evade our enforcements rather than landing their narratives in front of authentic audiences, resulting in a low return on their investment.

Notably, these covert operations attempted to target gaps in defenses across the wider internet: from the fragmented domain registration ecosystem allowing them to spoof reputable media organizations to using custom-built or other social media tools to coordinate deception campaigns. This means that no single defender team can tackle these threats in isolation. It’s vital to keep building a whole-of-society response to influence operations, and it’s why we publish threat indicators associated with covert activity, where appropriate, to enable further awareness and analysis across our industry and society at large.

02

Serbia

We removed 5,374 Facebook accounts, 12 Groups and 100 accounts on Instagram for violating our policy against coordinated inauthentic behavior. This network originated in Serbia and primarily targeted domestic audiences in that country.

This activity ran across the internet and targeted social media services including Facebook, Instagram, Twitter, YouTube, in addition to local news media to create a perception of widespread and authentic grassroots support for Serbian President Aleksandar Vučić and the Serbian Progressive party. Unlike traditional troll farms we've disrupted over the years in Albania, Nicaragua and Russia which were run by co-located operators working in a single office or set of offices, this operation instead relied on individuals running their efforts from across Serbia.

On our services, the people behind this activity used fake accounts — some of which were detected and disabled by our automated systems prior to our investigation — to manage Groups, mass comment and like content to make it appear more popular than it was. They also posted original memes criticizing the opposition figures in Serbia. This network typically posted in Serbian about news and political events in the region, tailoring comments to each post they amplified. Outside of social media platforms, this operation also actively commented directly on news media websites, in support of the government and its actions.

They appeared to have initially coordinated this cross-internet operation and its posting efforts through a stand-alone web application that allowed operators to comment and post on various news websites and log into social media. We investigated and blocked the app in 2020 after reviewing public reporting by investigative journalists at Balkan Insight, but the operation likely continued to use the app for posting on other platforms and news websites.

Also in 2020, after investigating a tip from our peers at Twitter, we took down a cluster of fake accounts for violating our policies against spam and inauthentic behavior. Based on the insights we gained from these earlier investigations combined with further threat research, we were able to tie them together and uncover the broader scope of this activity across the internet and attribute this

CIB network to employees of the Serbian Progressive Party, known as its Internet Team, and state employees from around Serbia.

- *Presence on Facebook and Instagram:* 5,374 Facebook accounts, 12 Groups and 1,060 Instagram accounts.
- *Followers:* At least 350 accounts joined one or more of these Groups and at least 26,000 accounts followed one or more of these Instagram accounts.
- *Advertising:* At least $150,000 in spending for ads on Facebook and Instagram, paid for in US dollars and euros.

03

Cuba

We removed 363 Facebook accounts, 270 Pages, 229 Groups and 72 accounts on Instagram for violating our policy against coordinated inauthentic behavior. This network originated in Cuba and primarily targeted domestic audiences in that country and also the Cuban diaspora abroad.

The people behind this activity relied on fake accounts – some of which were detected and disabled by our automated systems prior to this investigation – to manage Pages and Groups, post and amplify content, comment on other people's posts, and drive traffic to off-platform websites. Some of these accounts used profile photos likely generated using machine learning techniques like GAN (generative adversarial networks).

The network appeared to have pursued two main efforts to create the perception of widespread support for the Cuban government across many internet platforms, including Facebook, Instagram, Telegram, Twitter, YouTube and Picta, a Cuban social network. *First*, it used basic fake accounts to share and like pro-government content. *Second*, it created a number of more elaborate fictitious personal brands that featured distinctive logos, profile photos, visual styles and hashtags. The operation used these brands to post criticisms of government opponents and call on people to report them with no success. Some of the memes created by the network included photos of its targets, many of which referred to the government critics as "worms".

The individuals behind this operation posted Spanish-language videos, audio clips, articles, photos and memes that criticized members of the opposition and those who had criticized the government, including members of the Cuban diaspora in the United States and elsewhere.

We found this activity as part of our internal investigation into suspected coordinated inauthentic behavior in the region. Although the people behind this operation attempted to conceal their identities and coordination, our investigation found links to the Cuban government and its various entities.

- *Presence on Facebook and Instagram:* 363 Facebook accounts, 270 Pages, 229 Groups and 72 Instagram accounts
- *Followers: A*t least 650,000 accounts followed one or more of these Pages, around 510,000 accounts joined one or more of these Groups and at least 8,000 accounts followed one or more of these Instagram accounts

- *Advertising:* About $100 in spending for ads on Facebook and Instagram, paid for mostly in US dollars.

04

Bolivia

We removed 1,041 Facebook accounts, 450 Pages, 14 Groups and 130 accounts on Instagram for violating our policies against coordinated inauthentic behavior and mass reporting under our inauthentic behavior policies. This network originated in Bolivia and focused on domestic audiences in that country.

The people behind this network engaged in what we call a blended operation — coordinated adversarial activities that violated multiple policies at once. As part of CIB efforts, they used fake accounts to manage Pages posing as independent news media outlets, drive people to off-platform websites, and post and amplify content to make it appear more popular than it was. As part of coordinated abusive reporting, this network used some of the fake accounts to file large numbers of false reports, including against the Pages of news organizations and members of the opposition in an attempt to get them removed and silence them.

This operation ran across many internet services, including Facebook, Instagram, Twitter, YouTube, TikTok, Spotify, Telegram, and websites associated with its own “news media” brands. The individuals behind this effort were publicly reported to be working from a “bunker” office in Santa Cruz, Bolivia. There, among other locations, they coordinated their efforts to use fake accounts to post in support of the Bolivian government and to criticize and harass the opposition. They primarily posted Spanish-language pro-government commentary, memes and links to this operation’s off-platform websites. In a likely attempt to evade detection, this network interspersed political content with non-political posts about local news and events.

We found this activity as part of our internal investigation into suspected mass reporting in the region. Our investigation benefitted from open-source reporting on the portion of this network’s activity. Although the people behind this operation attempted to conceal their identities and coordination, our investigation found links to the current Bolivian government and Movimiento al Socialismo or MAS party, including individuals claiming to be part of a group known as “Guerreros Digitales” (“digital warriors”). We banned this group from our services.

- *Presence on Facebook and Instagram:* 1,041 Facebook accounts, 450 Pages, 14 Groups and 130 Instagram accounts.

- *Followers:* At least 2.3 million accounts followed one or more of these Pages, around 57,000 accounts joined one or more of these Groups and at least 23,000 accounts followed one or more of these Instagram accounts.
- *Advertising:* At least $1.1 million in spending for ads on Facebook and Instagram, paid for mostly in Bolivian Boliviano.