META

Removing Bad Actors on Facebook

July 31, 2018

LISTEN TO ARTICLE

Today we removed 32 Pages and accounts from Facebook and Instagram because they were involved in coordinated inauthentic behavior. This kind of behavior is not allowed on Facebook because we don't want people or organizations creating networks of accounts to mislead others about who they are, or what they're doing.

We're still in the very early stages of our investigation and don't have all the facts — including who may be behind this. But we are sharing what we know today given the connection between these bad actors and protests that are planned in Washington next week. We will update this post with more details when we have them, or if the facts we have change.

It's clear that whoever set up these accounts went to much greater lengths to obscure their true identities than the Russian-based Internet Research Agency (IRA) has in the past. We believe this could be partly due to changes we've made over the last year to make this kind of abuse much harder. But security is not something that's ever done. We face determined, well-funded adversaries who will never give up and are constantly changing tactics. It's an arms race and we need to constantly improve too. It's why we're investing heavily in more people and better technology to prevent bad actors misusing Facebook — as well as working much more closely with law enforcement and other tech companies to better understand the threats we face.

July 31, 2018

What We've Found So Far

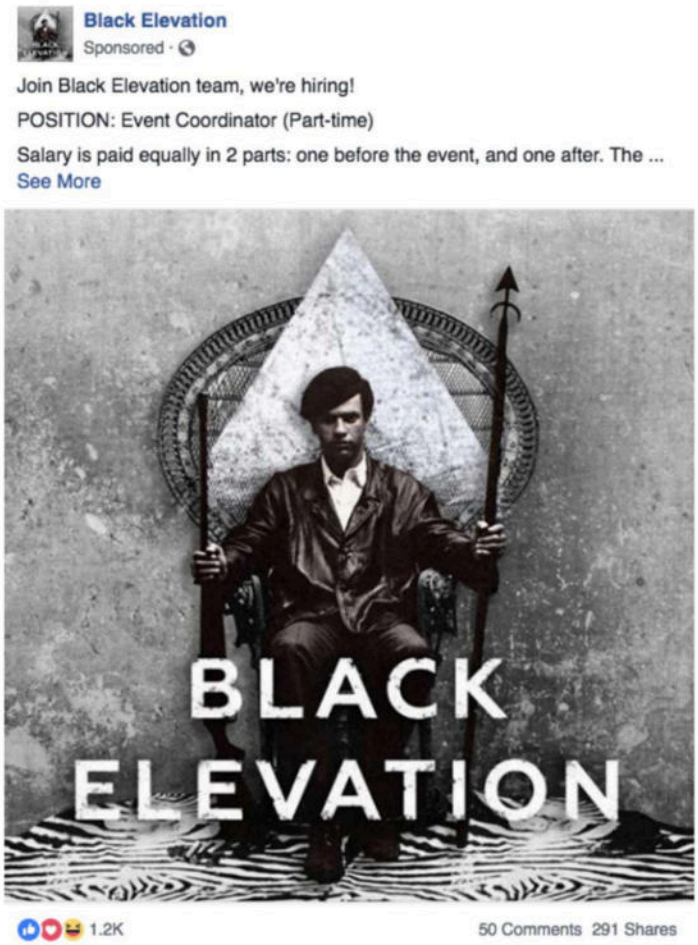
By Nathaniel Gleicher, Head of Cybersecurity Policy

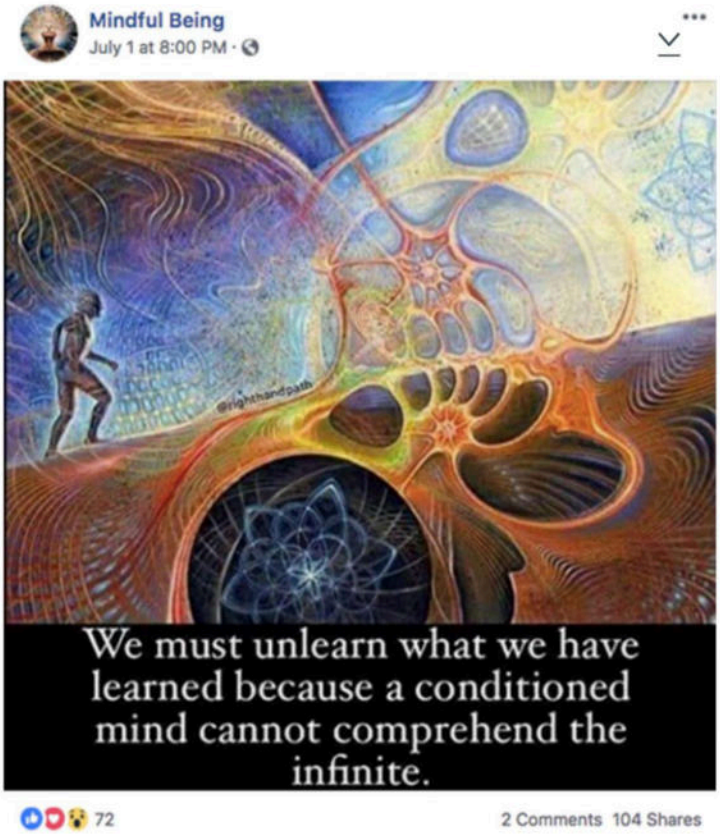
About two weeks ago we identified the first of eight Pages and 17 profiles on Facebook, as well as seven Instagram accounts, that violate our ban on coordinated inauthentic behavior. We removed all of them this morning once we'd completed our initial investigation and shared the information with US law

enforcement agencies, Congress, other technology companies, and the Atlantic Council's Digital Forensic Research Lab, a research organization that helps us identify and analyze abuse on Facebook.

- In total, more than 290,000 accounts followed at least one of these Pages, the earliest of which was created in March 2017. The latest was created in May 2018.
- The most followed Facebook Pages were "Aztlan Warriors," "Black Elevation," "Mindful Being," and "Resisters." The remaining Pages had between zero and 10 followers, and the Instagram accounts had zero followers.
- There were more than 9,500 organic posts created by these accounts on Facebook, and one piece of content on Instagram.
- They ran about 150 ads for approximately $11,000 on Facebook and Instagram, paid for in US and Canadian dollars. The first ad was created in April 2017, and the last was created in June 2018.
- The Pages created about 30 events since May 2017. About half had fewer than 100 accounts interested in attending. The largest had approximately 4,700 accounts interested in attending, and 1,400 users said that they would attend.

We are still reviewing all of the content and ads from these Pages. In the meantime here are some examples of the content and ads posted by these Pages.





These bad actors have been more careful to cover their tracks, in part due to the actions we've taken to prevent abuse over the past year. For example they used VPNs and internet phone services, and paid third parties to run ads on their behalf. As we've told law enforcement and Congress, we still don't have firm evidence to say with certainty who's behind this effort**.** Some of the activity is consistent with what we saw from the IRA before and after the 2016 elections. And we've found evidence of some connections between these accounts and IRA accounts we disabled last year, which is covered below. But there are differences, too. For example, while IP addresses are easy to spoof, the IRA accounts we disabled last year sometimes used Russian IP addresses. We haven't seen those here.

We found this activity as part of our ongoing efforts to identify coordinated inauthentic behavior. Given these bad actors are now working harder to obscure their identities, we need to find every small mistake they make. It's why we're following up on thousands of leads, including information from law enforcement and lessons we learned from last year's IRA investigation. The IRA engaged with many legitimate Pages, so these leads sometimes turn up nothing. However, one of these

leads did turn up something. One of the IRA accounts we disabled in 2017 shared a Facebook Event hosted by the "Resisters" Page. This Page also previously had an IRA account as one of its admins for only seven minutes. These discoveries helped us uncover the other inauthentic accounts we disabled today.

The "Resisters" Page also created a Facebook Event for a protest on August 10 to 12 and enlisted support from real people. The Event – "No Unite the Right 2 – DC" – was scheduled to protest an August "Unite the Right" event in Washington. Inauthentic admins of the "Resisters" Page connected with admins from five legitimate Pages to co-host the event. These legitimate Pages unwittingly helped build interest in "No Unite Right 2 – DC" and posted information about transportation, materials, and locations so people could get to the protests.

We disabled the event earlier today and have reached out to the admins of the five other Pages to update them on what happened. This afternoon, we'll begin informing the approximately 2,600 users interested in the event, and the more than 600 users who said they'd attend, about what happened.

We don't have all the facts, but we'll work closely with others as we continue our investigation. We hope to get new information from law enforcement and other companies so we can better understand what happened — and we'll share any additional findings with law enforcement and Congress. However, we may never be able to identify the source with the same level of confidence we had in naming the IRA last year. See Alex Stamos' post below on why attribution can be really hard.

We're seeing real benefits from working with outside experts. Partners like the Atlantic Council have provided invaluable help in identifying bad actors and analyzing their behavior across the internet. Based on leads from the recent US Department of Justice indictment, the Atlantic Council identified a Facebook group with roughly 4,000 members. It was created by Russian government actors but had been dormant since we disabled the group's admins last year. Groups typically persist on Facebook even when their admins are disabled, but we chose to remove this group to protect the privacy of its members in advance of a report that the Atlantic Council plans to publish as soon as it concludes its analysis. It will follow this report in the coming weeks with an analysis of the Pages, accounts and profiles we disabled today.

July 31, 2018

How Much Can Companies Know About Who's Behind Cyber Threats?

By Alex Stamos, Chief Security Officer

Deciding when and how to publicly link suspicious activity to a specific organization, government, or individual is a challenge that governments and many companies face. Last year, we said the Russia-based Internet Research Agency (IRA) was behind much of the abuse we found around the 2016 election. But today we're shutting down 32 Pages and accounts engaged in coordinated inauthentic behavior without saying that a specific group or country is responsible.

The process of attributing observed activity to particular threat actors has been much debated by academics and within the intelligence community. All modern intelligence agencies use their own internal guidelines to help them consistently communicate their findings to policymakers and the public. Companies, by comparison, operate with relatively limited information from outside sources — though as we get more involved in detecting and investigating this kind of misuse, we also need clear and consistent ways to confront and communicate these issues head on.

Determining Who is Behind an Action

The first challenge is figuring out the type of entity to which we are attributing responsibility. This is harder than it might sound. It is standard for both traditional security attacks and information operations to be conducted using commercial infrastructure or computers belonging to innocent people that have been compromised. As a result, simple techniques like blaming the owner of an IP address that was used to register a malicious account usually aren't sufficient to accurately determine who's responsible.

Instead, we try to:

- Link suspicious activity to the individual or group with *primary operational responsibility* for the malicious action. We can then potentially associate multiple campaigns to one set of actors, study how they abuse our systems, and take appropriate countermeasures.

- Tie a specific actor to a *real-world sponsor*. This could include a political organization, a nation-state, or a non-political entity.

The relationship between malicious actors and real-world sponsors can be difficult to determine in practice, especially for activity sponsored by nation-states. In his seminal paper on the topic, Jason Healey described a spectrum to measure the degree of state responsibility for cyber attacks. This included 10 discrete steps ranging from "state-prohibited," where a state actively stops attacks originating from their territory, to "state-integrated," where the attackers serve as fully integrated resources of the national government.

This framework is helpful when looking at the two major organized attempts to interfere in the 2016 US election on Facebook that we have found to date. One set of actors used hacking techniques to steal information from email accounts — and then contacted journalists using social media to encourage them to publish stories about the stolen data. Based on our investigation and information provided by the US government, we concluded that this work was the responsibility of groups tied to the GRU, or Russian military intelligence. The recent Special Counsel indictment of GRU officers supports our assessment in this case, and we would consider these actions to be "state-integrated" on Healey's spectrum.

The other major organized effort did not include traditional cyber attacks but was instead designed to sow division using social media. Based on our own investigations, we assessed with high confidence that this group was part of the IRA. There has been a public debate about the relationship between the IRA and the Russian government — though most seem to conclude this activity is between "state-encouraged" and "state-ordered" using Healey's definitions.

Four Methods of Attribution

Academics have written about a variety of methods for attributing activity to cyber actors, but for our purposes we simplify these methods into an attribution model with four general categories. And while all of these are appropriate for government organizations, we do not believe some of them should be used by companies:

- **Political Motivations**: In this model, inferred political motivations are measured against the known political goals of a nation-state. Providing public

attribution based on political evidence is especially challenging for companies because we don't have the information needed to make this kind of evaluation. For example, we lack the analytical capabilities, signals intelligence, and human sources available to the intelligence community. As a result, we don't believe it is appropriate for Facebook to give public comment on the political motivations of nation-states.

- **Coordination**: Sometimes we will observe signs of coordination between threat actors even when the evidence indicates that they are operating separate technical infrastructure. We have to be careful, though, because coincidences can happen. Collaboration that requires sharing of secrets, such as the possession of stolen data before it has been publicly disclosed, should be treated as much stronger evidence than open interactions in public forums.

- **Tools, Techniques and Procedures (TTPs)**: By looking at how a threat group performs their actions to achieve a goal — including reconnaissance, planning, exploitation, command and control, and exfiltration or distribution of information — it is often possible to infer a linkage between a specific incident and a known threat actor. We believe there is value in providing our assessment of how TTPs compare with previous events, but we don't plan to rely solely upon TTPs to provide any direct attribution.

- **Technical Forensics**: By studying the specific indicators of compromise (IOCs) left behind in an incident, it's sometimes possible to trace activity back to a known or new organized actor. Sometimes these IOCs point to a specific group using shared software or infrastructure, or to a specific geographic location. In situations where we have high confidence in our technical forensics, we provide our best attribution publicly and report the specific information to the appropriate government authorities. This is especially true when these forensics are compatible with independently gathered information from one of our private or public partners.

Applying the Framework to Our New Discovery

Here is how we use this framework to discuss attribution of the accounts and Pages we removed today:

- As mentioned, we will not provide an assessment of the political motivations of the group behind this activity.

- We have found evidence of connections between these accounts and previously identified IRA accounts. For example, in one instance a known IRA account was an administrator on a Facebook Page controlled by this group. These are important details, but on their own insufficient to support a firm determination, as we have also seen examples of authentic political groups interacting with IRA content in the past.

- Some of the tools, techniques and procedures of this actor are consistent with those we saw from the IRA in 2016 and 2017. But we don't believe this evidence is strong enough to provide public attribution to the IRA. The TTPs of the IRA have been widely discussed and disseminated, including by Facebook, and it's possible that a separate actor could be copying their techniques.

- Our technical forensics are insufficient to provide high confidence attribution at this time. We have proactively reported our technical findings to US law enforcement because they have much more information than we do, and may in time be in a position to provide public attribution.

Given all this, we are not going to attribute this activity to any one group right now. This set of actors has better operational security and does more to conceal their identities than the IRA did around the 2016 election, which is to be expected. We were able to tie previous abuse to the IRA partly because of several unique aspects of their behavior that allowed us to connect a large number of seemingly unrelated accounts. After we named the IRA, we expected the organization to evolve. The set of actors we see now might be the IRA with improved capabilities, or it could be a separate group. This is one of the fundamental limitations of attribution: offensive organizations improve their techniques once they have been uncovered, and it is wishful thinking to believe that we will always be able to identify persistent actors with high confidence.

The lack of firm attribution in this case or others does not suggest a lack of action. We have invested heavily in people and technology to detect inauthentic attempts to influence political discourse, and enforcing our policies doesn't require us to confidently attribute the identity of those who violate them or their potential links to foreign actors. We recognize the importance of sharing our best assessment of attribution with the public, and despite the challenges we intend to continue our work to find and stop this behavior, and to publish our results responsibly.

Categories:
Data and Privacy, Election Integrity, Integrity and Security, Meta, Safety and Expression

Tags:
Community Standards and Enforcement, Coordinated Inauthentic Behavior, Elections, Press Call, Security News

DOWNLOAD ALL IMAGES

Share this article

RELATED NEWS

FACEBOOK

Cracking Down on Spammy Content on Facebook

April 24, 2025

RECENT NEWS

META

How Meta Is Preparing for the 2026 US Midterm Elections

February 19, 2026

WhatsApp Adds New Features to Simplify Storage, Switch Accounts, and More

Meta Partners With Arm to Develop New Class of Data Center Silicon

Follow Meta Newsroom

Boosting Your Support and Safety on Meta's Apps With AI

Creator Fast Track: A New Way to Quickly Grow Your Audience and Earn Money on Facebook

Meta

Meta Store

Community

Our actions

About us

Site terms and policies

App support

Community Standards

Data Policy

Terms

Cookie policy

United States

(English)