'I need my girlfriend off TikTok': How hackers game abuse-reporting systems

December 8 2021, by Brian Contreras, Los Angeles Times

Credit: Unsplash/CC0 Public Domain

One hundred and forty-seven dollar signs fill the opening lines of the computer program. Rendered in an icy blue against a matte black background, each "$" has been carefully placed so that, all together, they

spell out a name: "H4xton."

It's a signature of sorts, and not a subtle one. Actual code doesn't show up until a third of the way down the screen.

The purpose of that code: to send a surge of content violation reports to the moderators of the wildly popular short-form video app TikTok, with the intent of getting videos removed and their creators banned.

It's a practice called "mass reporting," and for would-be TikTok celebrities, it's the sort of thing that keeps you up at night.

As with many social media platforms, TikTok relies on users to report content they think violates the platform's rules. With a few quick taps, TikTokers can flag videos as falling into specific categories of prohibited content—misleading information, hate speech, pornography—and send them to the company for review. Given the immense scale of content that gets posted to the app, this crowdsourcing is an important weapon in TikTok's content moderation arsenal.

Mass reporting simply scales that process up. Rather than one person reporting a post to TikTok, multiple people all report it in concert or—as programs such as H4xton's purport to do—a single person uses automated scripts to send multiple reports.

H4xton, who described himself as a 14-year-old from Denmark, said he saw his "TikTok Reportation Bot" as a force for good. "I want to eliminate those who spread false information or … made fun of others," he said, citing QAnon and anti-vax conspiracy theories. (He declined to share his real name, saying he was concerned about being doxxed, or having personal information spread online; The Times was unable to independently confirm his identity.)

But the practice has become something of a boogeyman on TikTok, where having a video removed can mean losing a chance to go viral, build a brand or catch the eye of corporate sponsors. It's an especially frightening prospect because many TikTokers believe that mass reporting is effective even against posts that don't actually break the rules. If a video gets too many reports, they worry, TikTok will remove it, regardless of whether those reports were fair.

It's a very 2021 thing to fear. The policing of user-generated internet content has emerged as a hot-button issue in the age of social-mediated connectivity, pitting free speech proponents against those who seek to protect [internet users](#) from digital toxicity. Spurred by concerns about misinformation and extremism—as well as events such as the Jan. 6 insurrection—many Democrats have called for social media companies to moderate user content more aggressively. Republicans have responded with cries of censorship and threats to punish internet companies that restrict expression.

Mass reporting tools exist for other social media platforms too. But TikTok's popularity and [growth rate](#)—it was the most downloaded app in the world last year—raise the stakes of what happens there for influencers and other power-users.

When The Times spoke this summer with a number of Black TikTokers about their struggles on the app, several expressed suspicion that organized mass reporting campaigns had targeted them for their race and political outspokenness, resulting in posts being taken down which didn't seem to violate any site policies. Other users—from transgender and Jewish TikTokers to gossip blogger Perez Hilton and mega-influencer Bella Poarch—have similarly speculated that they've been restricted from using TikTok, or had their content removed from it, after bad actors co-opted the platform's reporting system.

"TikTok has so much traffic, I just wonder if it gets to a certain threshold of people reporting [a video] that they just take it down," said Jacob Coyne, 29, a TikToker focused on making Christian content who's struggled with video takedowns he thinks stem from mass reporting campaigns.

H4xton posted his mass reporting script on GitHub, a popular website for hosting computer code—but that's not the only place such tools can be found. On YouTube, videos set to up-tempo electronica walk curious viewers through where to find and how to run mass reporting software. Hacking and piracy forums with names such as Leak Zone, ELeaks and RaidForums offer similar access. Under download links for mass reporting scripts, anonymous users leave comments including "I need my girlfriend off of TikTok" and "I really want to see my local classmates banned."

The opacity of most social media content moderation makes it hard to know how big of a problem mass reporting actually is.

Sarah Roberts, an assistant professor of information studies at UCLA, said that social media users experience content moderation as a complicated, dynamic, often opaque web of policies that makes it "difficult to understand or accurately assess" what they did wrong.

"Although users have things like Terms of Service and Community Guidelines, how those actually are implemented in their granularity—in an operational setting by content moderators—is often considered proprietary information," Roberts said. "So when [content moderation] happens, in the absence of a clear explanation, a user might feel that there are circumstances conspiring against them."

"The creepiest part," she said, "is that in some cases that might be true."

Such cases include instances of "brigading," or coordinated campaigns of harassment in the form of hostile replies or downvotes. Forums such as the notoriously toxic 8chan have historically served as home bases for such efforts. Prominent politicians including Donald Trump and Ted Cruz have also, without evidence, accused Twitter of "shadowbanning," or suppressing the reach of certain users' accounts without telling them.

TikTok has downplayed the risk that mass reporting poses to users and says it has systems in place to prevent the tactic from succeeding. A statement the company put out in July said that although certain categories of content are moderated by algorithms, human moderators review reported posts. Last year, the company said it had more than 10,000 employees working on trust and safety efforts.

The company has also said that mass reporting "does not lead to an automatic removal or to a greater likelihood of removal" by platform moderators.

Some of the programmers behind automated mass reporting tools affirm this. H4xton—who spoke with The Times over a mix of online messaging apps—said that his Reportation Bot can only get TikToks taken down that legitimately violate the platform's rules. It can speed up a moderation process that might otherwise take days, he said, but "won't work if there is not anything wrong with the video."

Filza Omran, a 22-year-old Saudi coder who identified himself as the author of another mass reporting script posted on GitHub, said that if his tool was used to mass-report a video that didn't break any of TikTok's rules, the most he thinks would happen would be that the reported account would get briefly blocked from posting new videos. Within minutes, Omran said over the messaging app Telegram, TikTok would confirm that the reported video hadn't broken any rules and restore the user's full access.

But other people involved in this shadow economy make more sweeping claims. One of the scripts circulated on hacker forums comes with the description: "Quick little bot I made. Mass reports an account til it gets banned which takes about an hour."

A user The Times found in the comments section below a different mass reporting tool, who identified himself as an 18-year-old Hungarian named Dénes Zarfa Szú, said that he's personally used mass reporting tools "to mass report bully posts" and accounts peddling sexual content. He said the limiting factor on those tools' efficacy has been how popular a post was, not whether that post broke any rules.

"You can take down almost anything," Szú said in an email, as long as it's not "insanely popular."

And a 20-year-old programmer from Kurdistan who goes by the screen name Mohamed Linux due to privacy concerns said that a mass reporting tool he made could get videos deleted even if they didn't break any rules.

These are difficult claims to prove without back-end access to TikTok's moderation system—and Linux, who discussed his work via Telegram, said his program no longer works because TikTok fixed a bug he'd been exploiting. (The Times found Linux's code on GitHub, although Linux said it had been leaked there and that he normally sells it to private buyers for $50.)

Yet the lack of clarity around how well mass reporting works hasn't stopped it from capturing the imaginations of TikTokers, many of whom lack better answers as to why their videos keep disappearing. In the comments section below a recent statement that TikTok made acknowledging concerns about mass reporting, swarms of users—some of them with millions of followers —complained that mass reporting had

led to their posts and accounts getting banned for unfair or altogether fabricated reasons.

Among those critics was Allen Polyakov, a gamer and TikTok creator affiliated with the esports organization Luminosity Gaming, who wrote that the platform had "taken down many posts and streams of mine because I've been mass reported." Elaborating on those complaints later, he told The Times that mass reporting became a big issue for him only after he began getting popular on TikTok.

"Around summer of last year, I started seeing that a lot of my videos were getting taken down," said Polyakov, 27. But he couldn't figure out why certain videos had been removed: "I would post a video of me playing Fortnite and it would get taken down" after being falsely flagged for containing nudity or sexual activity.

The seemingly nonsensical nature of the takedowns led him to think trolls were mass-reporting his posts. It wasn't pure speculation either: he said people have come into his live-streams and bragged about successfully mass reporting his content, needling him with taunts of "We got your video taken down" and "How does it feel to lose a viral video?"

Polyakov made clear that he loves TikTok. "It's changed my life and given me so many opportunities," he said. But the platform seems to follow a "guilty 'til proven innocent" ethos, he said, which errs on the side of removing videos that receive lots of reports, and then leaves it up to creators to appeal those decisions after the fact.

Those appeals can take a few days, he said, which might as well be a millennium given TikTok's fast-moving culture. "I would win most of my appeals—but because it's already down for 48 to 72 hours, the trend might have went away; the relevance of that video might have went away."

As with many goods and services that exist on the periphery of polite society, there's no guarantee that mass-reporting tools will work. Complaints about broken links and useless programs are common on the hacker forums where such software is posted.

But technical reviews of several mass-reporting tools posted on GitHub—including those written by H4xton, Omran and Linux—suggest that this cottage industry is not entirely smoke and mirrors.

Francesco Bailo, a lecturer in digital and social media at the University of Technology Sydney, said that what these tools "claim to do is not technically complicated."

"Do they work? Possibly they worked when they were first written," Bailo said in an email. But the programs "don't seem to be actively maintained," which is essential given that TikTok is probably "monitoring and contrasting this kind of activity" in a sort of coding arms race.

Patrik Wikstrom, a communication professor at the Queensland University of Technology, was similarly circumspect.

"They might work, but they most likely need a significant amount of hand-holding to do the job well," Wikstrom said via email. Because TikTok doesn't want content reports to be sent from anywhere but the confines of the company's own app, he said, mass reporting requires some technical trickery: "I suspect they need a lot of manual work not to get kicked out."

But however unreliable mass-reporting tools are—and however successful TikTok is in separating their complaints from more legitimate ones—influencers including Coyne and Polyakov insist that the problem is one the company needs to start taking more seriously.

"This is literally the only platform that I've ever had any issues" on, Polyakov said. " I can post any [video](#) that I have on TikTok anywhere else, and it won't be an issue."

"Might you get some kids being assholes in the comments?" he said. "Yeah—but they don't have the ability to take down your account."

©2021 Los Angeles Times.
Distributed by Tribune Content Agency, LLC.