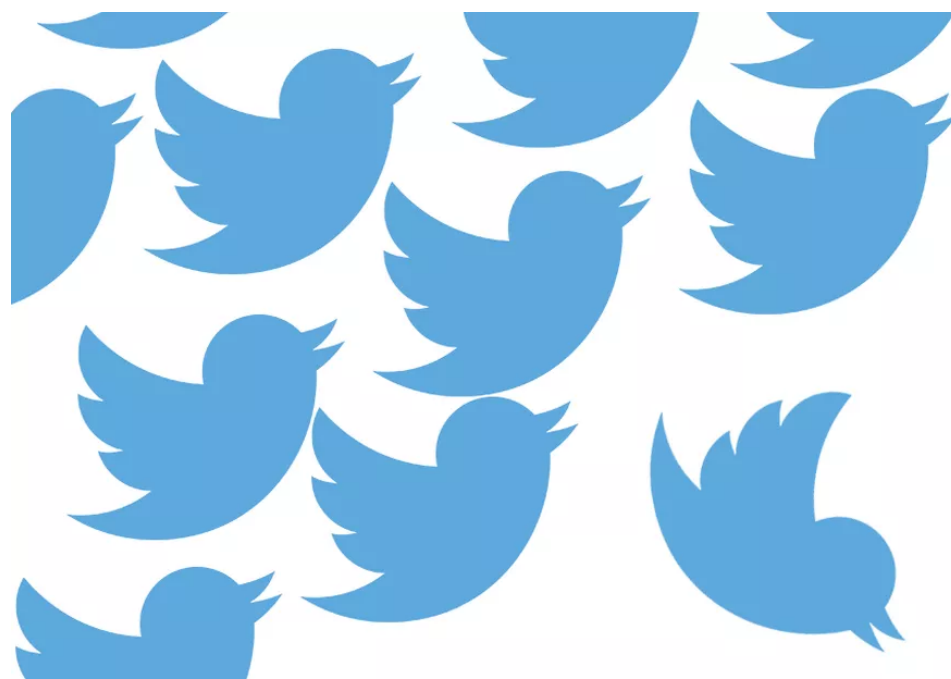


# One of Twitter's new anti-abuse measures is the oldest trick in the forum moderation book

*Tachy goes to coventry*

By Dieter Bohn | @backlon | Feb 16, 2017, 9:07am EST



Overnight, [BuzzFeed uncovered](#) one of the ways that Twitter is filtering out abuse on its platform in its [latest anti-harassment initiative](#). Users have begun getting notices that their tweets are on a kind of time-out. These users are being told that “only your followers can see your activity on Twitter for the amount of time shown below,” followed by a number of hours — the examples seen so far are all 12 hours. It starts when the user clicks a button to “Continue to Twitter.”

Allowing a user to continue to post to a forum but limiting who can actually see those posts is commonly known as a [shadow ban or a stealth ban](#). But it has other names, and it's one of the oldest moderation tricks in the forum book. In its earlier iterations, vBulletin forum software called it “[tachy goes to coventry](#).”

It's an effective tactic because the abusive user still feels as though he is spewing bile into the community, but nobody actually has to see it. Twitter's take on it is smart from a moderation standpoint because it's at least letting the user know it's happening, while classic shadow banning on forums would sometimes happen without letting the user know in a kind of reverse troll — which would inevitably lead to a backlash of [sockpuppet](#) accounts when the user figured it out. In the meantime, though, it's satisfying to know that when trolls think they're trolling, what they're actually doing is shouting into the void.

On smaller forums, time-outs can also be effective for some users — it's a minor punishment for an infraction and it gives angry users some time to cool off and see how the community interacts without their input. The ones who aren't actively malicious might see that it goes better when they're not trolling and end up coming back as more valuable members of the group.

>automatically get limited cause I said retard

I'm just considering leaving Twitter, fuck them [pic.twitter.com/2NZpOPmlo2](https://pic.twitter.com/2NZpOPmlo2)

— Drybones △ (@Drybones5) [February 14, 2017](#)

Twitter itself has been intentionally vague about what its precise policies and tools are for limiting the reach of harassers, because the company believes that they will “seek to use the information to game the system,” as [Casey Newton put it](#). That's another thing that often happens on smaller forum communities. After a ban citing a rule, the user will often try to litigate the precise rule that was used to ban them, and usually those arguments are not in good faith.

However (you knew there'd be a however), this is Twitter we're talking about, not a small forum community. So the things that work on forums — with users who often actually deserve the benefit of the doubt — are not guaranteed to work on Twitter. Something is better than nothing, of course, and the nuance in Twitter's variation on shadow banning seems to imply that the company is giving some thought to how it tackles abuse.

On the forums I've run, I've never used rule litigation as an excuse to hide the rules, and found that shadow banning often ends up being more trouble than it's worth. It's laudable

that Twitter is trying to be nuanced in its approach. But when it comes to banning trolls, nuance isn't the only thing you need. Sometimes you just have to break out the [banhammer](#).