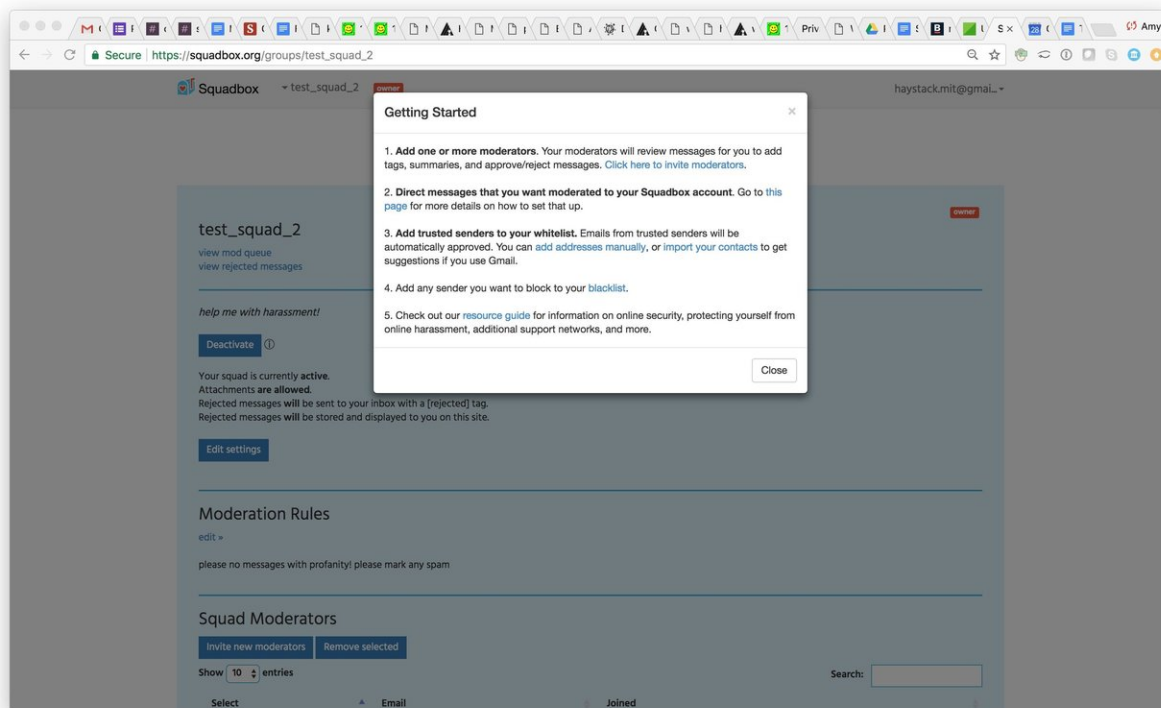


'Squadbox' uses 'friendsourcing' to better support targets of cyberbullying

April 5 2018



Squadbox interface. Credit: Massachusetts Institute of Technology

Harassment has become ubiquitous on social media and in the online world- Twitter has been under fire for how it handles harassment, YouTube's trending algorithm has occasionally promoted offensive videos, and reporting abuse on Instagram is still quite difficult.

While current tools for harassment such as blocking users or filtering trigger words partially helps, a lot of online mistreatment is so subtle that an algorithm alone might not pick up on various cues.

But what if people could actually leverage their friends to help moderate their accounts and shield them from abusive messages?

A team from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) use that approach with "Squadbox", a new crowdsourcing tool that enables people who have been the targets of harassment to coordinate "squads" of friends to filter messages and support them during attacks.

The team interviewed a range of scientists, activists, and Youtube personalities, and found that many people who are harassed on [email](#) rely on friends and family to shield themselves from abusive messages.

Senior author and MIT professor David Karger says that Squadbox aims to make this so-called "friend-sourcing" more efficient and less work for its moderators.

"If you just give moderators the keys to your inbox, how does the moderator know when there's an email to moderate, and which email has already been handled by other moderators?" says Karger. "Squadbox allows users to customize how incoming email is handled, divvying up the work to make sure there's no duplication of effort."

Karger wrote the new paper with PhD student Amy Zhang and former MIT student and current software engineer Kaitlin Mahar. The team will present the work in April at ACM's CHI Conference on Human Factors in Computing Systems in Montreal, Canada.

"Friend-sourced" moderation

With Squadbox, the "owner" of the squad can set up filters to automatically forward incoming content to its moderation pipeline. Once an email arrives, a moderator decides which emails are harassment, and which can be forwarded back to the person's inbox.

For example, let's say a journalist wants to have a public email address to receive news tips, but fears having one because of how often she gets harassed by people who disagree with her reporting. She could create a Squadbox account with two close colleagues as moderators, and then use her account anywhere she wants to share her email address. Any email she receives at that address goes through her squad first.

"Previous solutions depended entirely on automated techniques, or were overly dependent on social solutions like simply giving one's account information to a friend," says Clifford Lampe, a professor of information at the University of Michigan. "This line of work helps provide a map for one hybrid solution to harassment that augments human support with tools in a meaningful way."

While the team plans to extend the capabilities of Squadbox to work with [social media](#) platforms, they say that email is a particularly useful system for studying harassment.

"What's interesting about email is that it's not out in the open like Twitter or Facebook," says Karger. "That means harassers may feel less accountability because their comments aren't public."

Squadbox also lets users create "whitelists" and "blacklists" of senders whose emails will be automatically approved or rejected without moderation. Users can also deactivate and reactivate the system, read scores on messages' toxicity, and even respond to harassers.

"Harassment is an inherently disempowering experience, and giving

survivors options allows them to move back into control," says Emily May, co-founder and executive director of the global anti-[harassment](#) initiative Hollaback! "Squadbox is designed to not just remind people that they have community around them, but to activate that community on their behalf."

The team evaluated Squadbox on five pairs of friends, and found that using friends helped with fears of privacy and allowed for more tailored decisions for victims. However, users still worried that a friend might be sensitive to slow response times, and that the system could also be "spreading" the burden. The team says that having multiple moderators could potentially help this issue.

"Squadbox offers a tangible system of support for people who are harassed online by using other people," says May. "We need to put survivors back in control and create communities strong enough to declare: this isn't the internet we want. It's not the world we want. And we're not going to stand for it anymore."

Provided by Massachusetts Institute of Technology

Citation: 'Squadbox' uses 'friendsourcing' to better support targets of cyberbullying (2018, April 5) retrieved 27 April 2024 from

<https://techxplore.com/news/2018-04-squadbox-friendsourcing-cyberbullying.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
