Squashing cyberbullying: New approach is fast, accurate

June 11 2018

11:34 am Tue, 22 May

EDIT ⚙

| Wi-Fi | Location | Sound | Screen rotation | Bluetooth |

BatCave-R

Outdoor

Possible Bullying!!
On Instagram!!6 instances
11:34 AM

Cable charging
Charging, 22%

Connected as a media device
Touch for other USB options.

AT&T   ☰ Clear
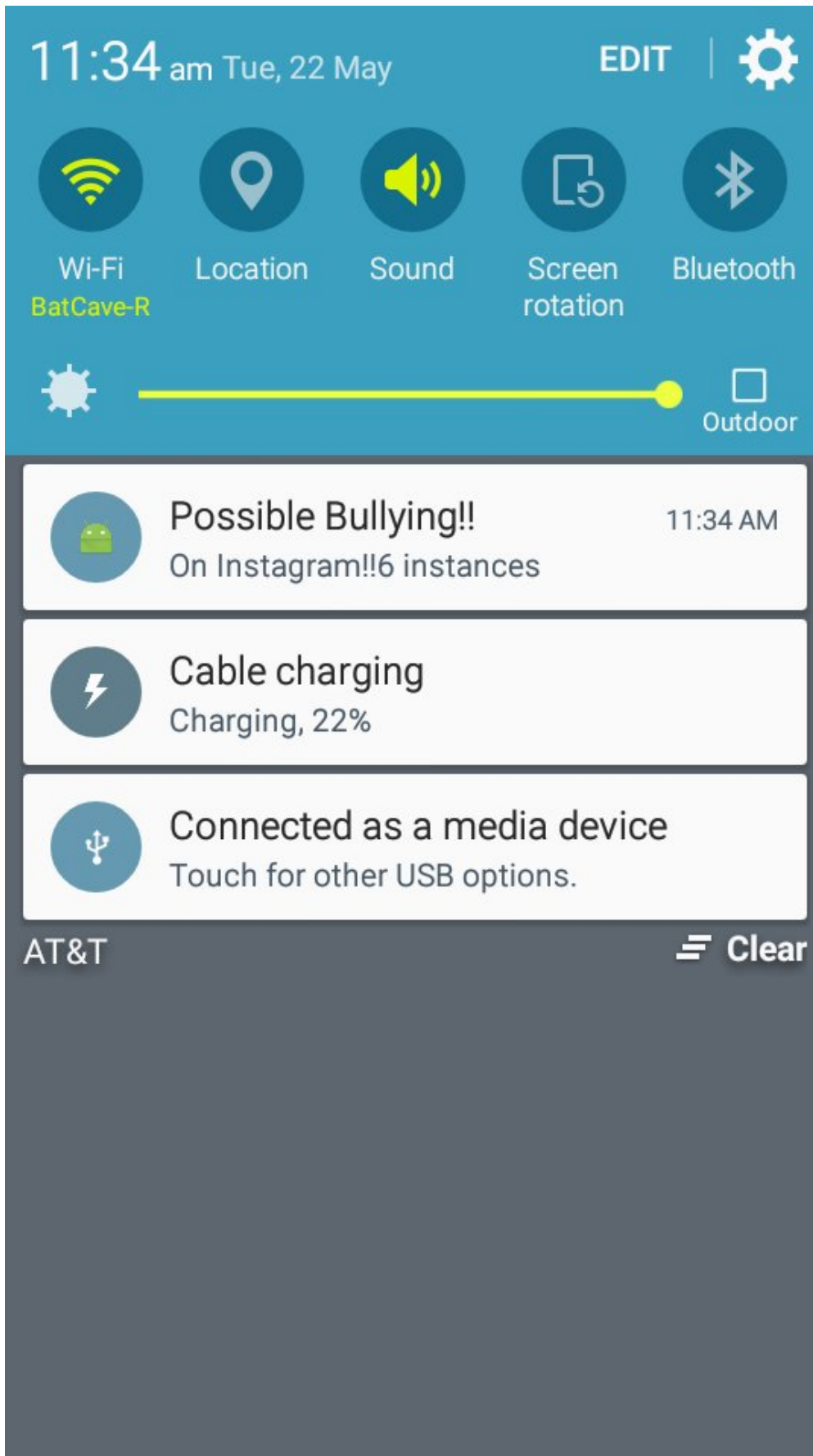
A new, free app for Android phones sends parents alerts when their children are being bullies on Instagram. Credit: Rahat Ibn Rafiq

Researchers at the University of Colorado Boulder have designed a new technique for spotting nasty personal attacks on social media networks like Instagram.

The new approach, developed by CU Boulder'sCyberSafety Research Center, combines several different computing tools to scan massive amounts of social media data, sending alerts to parents or network administrators that abuse has occurred.

It's rocket fast, too: In recently published research, the group reported that their approach uses five times less computing resources than existing tools. That's efficient enough to monitor a network the size of Instagram for a modest investment in server power, said study co-author Richard Han.

"The response of the social media networks to fake news has recently started to uptick, even though it took grave consequences to reach that point," said Han, an associate professor in the Department of Computer Science. "The response needs to be just as strong for cyberbullying."

The group also released a free app for Android phones that allows parents to receive alerts when their kids are the objects of bullying on Instagram. Study co-author Shivakant Mishra explained that the app, called BullyAlert, can learn from and adapt to what parents consider bullying.

"As parent, I know that a lot of times we are not in full knowledge of what our children are doing on their social networks," said Mishra, a professor in computer science. "An app like this that informs us when something problematic is happening is invaluable."

To build their toolbox, the CyberSafety researchers first employed real humans to teach a computer program how to separate benign online comments from abuse.

Next, the researchers designed a system that works a bit like hospital triage. When a user uploads a new post, the group's tools make a quick scan of the comments. If those comments look questionable, then that post gets high priority to receive further checks. But if the comments all seem charitable, then the system bumps the post to the bottom of its queue.

"Our goal is to focus on the most vulnerable sessions," Han said. "We still continue to monitor all of the sessions, but we monitor more frequently those sessions that we think are more problematic."

And it works: The researchers tested their approach on real-world data from Vine, a now-defunct video-sharing platform, and Instagram. Han explained that the team picked those networks because they make their data publicly available.

In research presented at a conference in April, the group calculated that their toolset could monitor traffic on Vine and Instagram in real-time, detecting cyberbullying behavior with 70 percent accuracy. What's more, the approach could also send up warning flags within two hours after the onset of abuse—a performance unmatched by currently available software.

Rahat Ibn Rafiq, a CU Boulder graduate student in computer science, is

the lead author of the new study. Other co-authors include CU Boulder Associate Professor Qin (Christine) Lv and Homa Hosseinmardi of the University of Southern California.

Provided by University of Colorado at Boulder