A better kind of cybersecurity strategy

December 10 2020, by Peter Dizikes

The multilateral nature of cybersecurity today makes it markedly different than conventional security, according to a new study co-authored by an MIT professor. Credits: Jose-Luis Olivares, MIT

During the opening ceremonies of the 2018 Winter Olympics, held in PyeongChang, South Korea, Russian hackers launched a cyberattack that disrupted television and internet systems at the games. The incident was

resolved quickly, but because Russia used North Korean IP addresses for the attack, the source of the disruption was unclear in the event's immediate aftermath.

There is a lesson in that attack, and others like it, at a time when hostilities between countries increasingly occur online. In contrast to conventional national security thinking, such skirmishes call for a new strategic outlook, according to a new paper co-authored by an MIT professor.

The core of the matter involves deterrence and retaliation. In conventional warfare, deterrence usually consists of potential retaliatory military strikes against enemies. But in cybersecurity, this is more complicated. If identifying cyberattackers is difficult, then retaliating too quickly or too often, on the basis of limited information such as the location of certain IP addresses, can be counterproductive. Indeed, it can embolden other countries to launch their own attacks, by leading them to think they will not be blamed.

"If one country becomes more aggressive, then the equilibrium response is that all countries are going to end up becoming more aggressive," says Alexander Wolitzky, an MIT economist who specializes in game theory. "If after every cyberattack my first instinct is to retaliate against Russia and China, this gives North Korea and Iran impunity to engage in cyberattacks."

But Wolitzky and his colleagues do think there is a viable new approach, involving a more judicious and well-informed use of selective retaliation.

"Imperfect attribution makes deterrence multilateral," Wolitzky says. "You have to think about everybody's incentives together. Focusing your attention on the most likely culprits could be a big mistake."

The paper, "Deterrence with Imperfect Attribution," appears in the latest issue of the *American Political Science Review*. In addition to Wolitzky, the authors are Sandeep Baliga, the John L. and Helen Kellogg Professor of Managerial Economics and Decision Sciences at Northwestern University's Kellogg School of Management; and Ethan Bueno de Mesquita, the Sydney Stein Professor and deputy dean of the Harris School of Public Policy at the University of Chicago.

The study is a joint project, in which Baliga added to the research team by contacting Wolitzky, whose own work applies game theory to a wide variety of situations, including war, international affairs, network behavior, labor relations, and even technology adoption.

"In some sense this is a canonical kind of question for game theorists to think about," Wolitzky says, noting that the development of game theory as an intellectual field stems from the study of nuclear deterrence during the Cold War. "We were interested in what's different about cyberdeterrence, in contrast to conventional or nuclear deterrence. And of course there are a lot of differences, but one thing that we settled on pretty early is this attribution problem." In their paper, the authors note that, as former U.S. Deputy Secretary of Defense William Lynn once put it, "Whereas a missile comes with a return address, a computer virus generally does not."

In some cases, countries are not even aware of major cyberattacks against them; Iran only belatedly realized it had been attacked by the Stuxnet worm over a period of years, damaging centrifuges being used in the country's nuclear weapons program.

In the paper, the scholars largely examined scenarios where countries are aware of cyberattacks against them but have imperfect information about the attacks and attackers. After modeling these events extensively, the researchers determined that the multilateral nature of cybersecurity

today makes it markedly different than conventional security. There is a much higher chance in multilateral conditions that retaliation can backfire, generating additional attacks from multiple sources.

"You don't necessarily want to commit to be more aggressive after every signal," Wolitzky says.

What does work, however, is simultaneously improving detection of attacks and gathering more information about the identity of the attackers, so that a country can pinpoint the other nations they could meaningfully retaliate against.

But even gathering more information to inform strategic decisions is a tricky process, as the scholars show. Detecting more attacks while being unable to identify the attackers does not clarify specific decisions, for instance. And gathering more information but having "too much certainty in attribution" can lead a country straight back into the problem of lashing out against some states, even as others are continuing to plan and commit attacks.

"The optimal doctrine in this case in some sense will commit you to retaliate more after the clearest signals, the most unambiguous signals," Wolitzky says. "If you blindly commit yourself more to retaliate after every attack, you increase the risk you're going to be retaliating after false alarms."

Wolitzky points out that the paper's model can apply to issues beyond cybersecurity. The problem of stopping pollution can have the same dynamics. If, for instance, numerous firms are polluting a river, singling just one out for punishment can embolden the others to continue.

Still, the authors do hope the paper will generate discussion in the foreign-policy community, with cyberattacks continuing to be a

significant source of national security concern.

"People thought the possibility of failing to detect or attribute a cyberattack mattered, but there hadn't [necessarily] been a recognition of the multilateral implications of this," Wolitzky says. "I do think there is interest in thinking about the applications of that."

More information: Sandeep Baliga et al. Deterrence with Imperfect Attribution. *American Political Science Review* (2020) DOI: doi.org/10.1017/S0003055420000362

Provided by Massachusetts Institute of Technology