National Science Foundation
WHERE DISCOVERIES BEGIN

DISCOVERY

When hackers talk, this research team listens

Online conversations help fill critical gap in cybersecurity knowledge about attackers' motivations, possible targets

October 8, 2015

If you're a hacker, you gather as much data as you can on your targets, in search of something valuable.
If you're researcher Hsinchun Chen, you gather as much data as you can on the hackers.

Chen, a professor of management information systems at the University of Arizona, works in a little-explored, but hugely important area of cybersecurity: Exploring the motivations of hackers and other cyberattackers, and trying to predict how they might act, based on their behaviors.

With support from the National Science Foundation's (NSF) Social, Behavioral and Economic Sciences directorate and the Directorate for Computer and Information Science and Engineering under the Secure and Trustworthy Cyberspace (SaTC) program, Chen and his collaborators have generated findings that shed light on how hacker communities interact and share information--and even created actionable intelligence for criminal investigations by federal agencies.

But the research's goal is even more ambitious. Chen wants to develop models that might be able to take information on how hackers behave and use it to predict their next targets, as well as their methods for attack.

"The most important part isn't looking back and saying 'what have they done?'" Chen says. "It's looking forward and saying 'What are the emerging threats?' We're really trying to understand the intent of the people planning attacks. Instead of looking at the bullets, you're looking at the shooters."

The research holds significant promise for the social sciences, as well as information science. The team aims to develop and test theories about hacker cultures, based on their online interactions. That involves modeling the social attributes of hacker networks and investigating how their groups are organized.

Chen is hardly a stranger to this kind of work. For the past decade, he's worked on--and headed--NSF-funded research projects that examine other potentially threatening online communities, producing a long trail of papers and tools along the way.

He developed COPLINK, a software system used by more than 3,500 law enforcement networks nationwide to look for information on drug networks, border smuggling operations and other criminal activity. With an international group of terrorism research centers and security agencies, he helped create the Dark Web project, which has tapped into extremist communications and social networks to generate one of the world's largest databases of terrorist information.

Still, he said, tapping into hacker behavior has proved even more of a challenge.

"This community," he says, "is even more tightly knit."

'Honor Among Thieves'

How do you research hacker intent? By gathering all of the hacker community content possible.

Chen and his collaborators collect all of the "artifacts" they possibly can automatically--primarily from hacker forums and hundreds of text communication channels known as IRC chats, and millions of messages--from hackers around the world.

Through automated text mining that can search for everything from relevant terms and topics to "sentiment analysis," Chen and his collaborators are able to distill that chatter down to a much smaller body of communications that deal with top-tier, likely threats. That slimmed-down pool of data constitutes roughly five percent of the total collected, Chen says.

By studying those data, they've found hackers build social structures just like any other kind of community.

For instance, he says, "honor among thieves" applies to hackers, and as a community they punish any transgressions. Communities begin to distrust hackers that lose money, steal from partners-in-crime or make mistakes that harm their associates, leaving them isolated.

And there's more. Hackers work in groups and collaborate on projects, seeking counsel from trusted friends and leaning on one another's expertise. They have underground economies and methods for sharing data and selling stolen goods. They analyze others' work and post reviews. Top-tier cyberattackers each have some specialty and a preferred payment method whenever hacking-for-hire, Chen says.

Hackers work in groups and collaborate on projects

By being proactive about capturing artifacts from communications, the researchers can even see things missed by studies that focus on the damage wrought by hacks and other attacks. Instead of just seeing that a large number of credit card numbers has been compromised, for example, the researchers can observe what cyberattackers are using those cards for--even the ones that have yet to be reported as stolen.

"It takes a very different approach from previous cybersecurity research," Chen says. "You really want to understand the intent, the *modus operandi* of operators. Instead of just finding out about one operation at a time, you're looking at an entire source of information about ongoing activities."

New Tools

Chen estimates that about 20 to 30 percent of the research and analysis that his team generates can be transitioned into actionable intelligence for law enforcement agencies and the industry. The researchers have provided such information to agencies to help with ongoing investigations.

But, while he acknowledges that aspect of the job is "exciting," he adds "I'm a computer scientist--not a law enforcement agent."

There are other data analysis projects that look for pending threats. Chen says his research is focused on creating new tools that will allow scientists and their partners at federal agencies to analyze hacker chats and other data in ways that are faster, more efficient and better at making predictions about future threats based on hackers' intentions.

"I'm not interested in the hackers themselves," he says. "I'm interested in developing the best science that will help advance cybersecurity big data research."

Text mining, data mining, sentiment analysis and other automated analysis tools that incorporate artificial intelligence are very good at cutting down about 95 percent of the noise from massive sets of information gleaned from IRC chats and other sources, leaving researchers with the aforementioned 5 percent of top-tier threat information.

Chen wants to make those filters even better. Doing so requires following a cycle of research and development: building an analysis tool; using it on sets of information drawn from hacker communications; refining it; gathering more data; applying the tool to that larger set of information. Repeat.

It's a process with no end point and one that will require researchers to adapt to new hacker communications methods, shifting intentions in the malicious hacker community and an ever-expanding pool of data. But keeping up with emerging cyberthreats--and perhaps even getting ahead of them--requires the process continue.

"There's an overwhelming amount of data," Chen says. "You need ways to analyze those data and distill them into actionable intelligence."

-- Robert J. Margetta, (703) 292-2663 rmargett@nsf.gov

Investigators
Hsinchun Chen
Daniel Zeng
Thomas Holt
Salim Hariri
Ronald Breiger

Related Institutions/Organizations
University of Arizona

Related Awards
#1314631 SBE TTP: Medium: Securing Cyber Space: Understanding the Cyber Attackers and Attacks via Social Media Analytics

Total Grants
$1,301,944