Argonne applies machine learning to cybersecurity threats

November 13 2019, by Savannah Mitchem

Cyber threat analysis requires high-speed supercomputers, such as Theta at Argonne's Leadership Computing Facility, a DOE Office of Science User Facility. Credit: Argonne National Laboratory

It is indisputable that technology is now a fundamental and inextricable

part of our everyday existence—for most people, our employment, transportation, healthcare, education, and other quality of life measures are fully reliant on technology. Our dependence has created an urgent need for dynamic cybersecurity that protects U.S. government, research and industry assets in the face of technology advances and ever more sophisticated adversaries.

The U.S. Department of Energy's (DOE) Argonne National Laboratory is helping lead the way in researching and developing proactive cybersecurity, including measures that leverage machine learning, to help protect data and critical infrastructure from cyberattacks.

Machine learning is a category of artificial intelligence that involves training machines to continually learn from and identify patterns in data sets.

"Applying machine learning approaches to cybersecurity efforts makes sense due to the large amount of data involved," said Nate Evans, program lead for cybersecurity research in the Strategic Security Sciences (SSS) Division. "It is not efficient for humans to mine data for these patterns using traditional algorithms."

Argonne computer scientists develop machine learning algorithms using large data sets— comprising log data from different devices, network traffic information, and instances of malicious behavior—that enable the algorithms to recognize specific patterns of events that lead to attacks. When such patterns are identified, a response team investigates instances matching those patterns.

Following an attack, the response team patches the vulnerability in the laboratory's intrusion protection systems. Forensic analysis can then lead to changes that prevent similar future attacks.

"We are looking for ways to stop attacks before they happen," said Evans. "We're not only concerned with protecting our own lab, we're also developing methods to protect other national labs, and the country as a whole, from potential cyberattacks."

The machine learning approach allows a computer to serve as the cyber threat hunter, mining large data volumes while humans are freed to focus on the highest risk threats.

With huge amounts of data generated not only from Argonne but also by other national labs and elsewhere in DOE, analysis requires high-speed supercomputers, such as Theta at Argonne's Leadership Computing Facility, a DOE Office of Science User Facility.

"We're talking billions and billions of records per day," Evans said, "and the computer is identifying where there may be unusual or malicious traffic."

Researchers are working to toward testing their machine learning methods on private sector data as well, Evans said. Such future studies could yield knowledge transferable to the banking industry and other critical U.S. infrastructure, he said.

Teaching computers our language

Argonne scientists are using artificial intelligence to fight cybersecurity threats on many fronts. Computer scientist Sandeep Madireddy of Argonne's Mathematics and Computer Science (MCS) Division conducts research to facilitate the safe use of computer applications—word processors, spreadsheets, Web browsers, and the like. Machine learning techniques can be a powerful tool in combating cyberattacks that exploit security vulnerabilities in these ubiquitous programs.

Machine learning handles structured and unstructured data. Structured data is arranged in formal patterns such as tables that can easily feed into a model. Unstructured data often takes the form of text, a much more nuanced and complex data form.

"For unstructured data," said Madireddy, "our researchers build algorithms that extract information from data logs in text format using approaches like natural language processing, inspired by methods used in the commercial world for understanding text."

With natural language processing, sequences of letters serve as the input to the machine learning model. The algorithms then rely on ever-improving statistical language models to develop associations between terms and make predictions about the legitimacy of certain communications.

"We are trying to mine similarities among these texts, identify meaningful repeating patterns, and classify them as good or bad in terms of cybersecurity," said Madireddy. "We want to extract the anomalies."

For example, natural language processing can aid in distinguishing legitimate from phishing communications, in order to prevent a security breach through email applications.

Additionally, Argonne researchers are developing methods to mine time series data—data collected at successive, known time intervals—to provide another means of detecting cyberattacks. When a system is attacked, there is often a sudden behavioral change in the time series data received by the system. So-called change-point detection algorithms can use historical and current data to pinpoint the exact time such a drastic change occurred.

"This notifies us about some sort of anomalous behavior so that we can

take a closer look," said Madireddy.

Maintaining security and functionality

In addition to its cybersecurity research programs, Argonne is home to a Cybersecurity Program Office (CSPO) utilizing machine learning to protect the laboratory's digital information. For example, computer scientists in CSPO are developing machine learning algorithms to create a more flexible password protection protocol.

"We want to prevent false positives with regard to threat detection, so when someone logs in, we are steering away from the rigid protocol of allowing three tries before they are locked out," said Deputy Chief Information Security Officer Matt Kwiatkowski. "Instead, we can train computers to learn patterns of how people log in to our networks, such as their location and the time at which they are logging in, to make the protocol more flexible for employees, while also keeping the network secure."

The Cybersecurity Program Office is also developing machine learning algorithms as a cost-saving measure. For example, institutions typically pay third-party services to categorize different websites as informational, governmental, or social media. The team is trying to use machine learning to recognize patterns in website features in order to categorize them on their own.

Argonne's cybersecurity research, along with the organization's strong information security culture, keep the laboratory on the cutting edge, Kwiatkowski said.

"Our employees recognize the need for security, and they take it seriously," he said. "If a security measure hinders their work, we try to find creative ways to maintain security and functionality. It's about being

responsive to our people, being adaptable and always exploring new ways of doing things as the world of cybersecurity continues to rapidly evolve."

Every DOE national laboratory has an operational cybersecurity arm that focuses on protecting itself from cyberattacks. Some of the other laboratories focus on analyzing current threats and where they come from, while others focus on protecting the nation's power infrastructure. Argonne is one of the national laboratories that also has a robust cybersecurity research arm.

Provided by Argonne National Laboratory