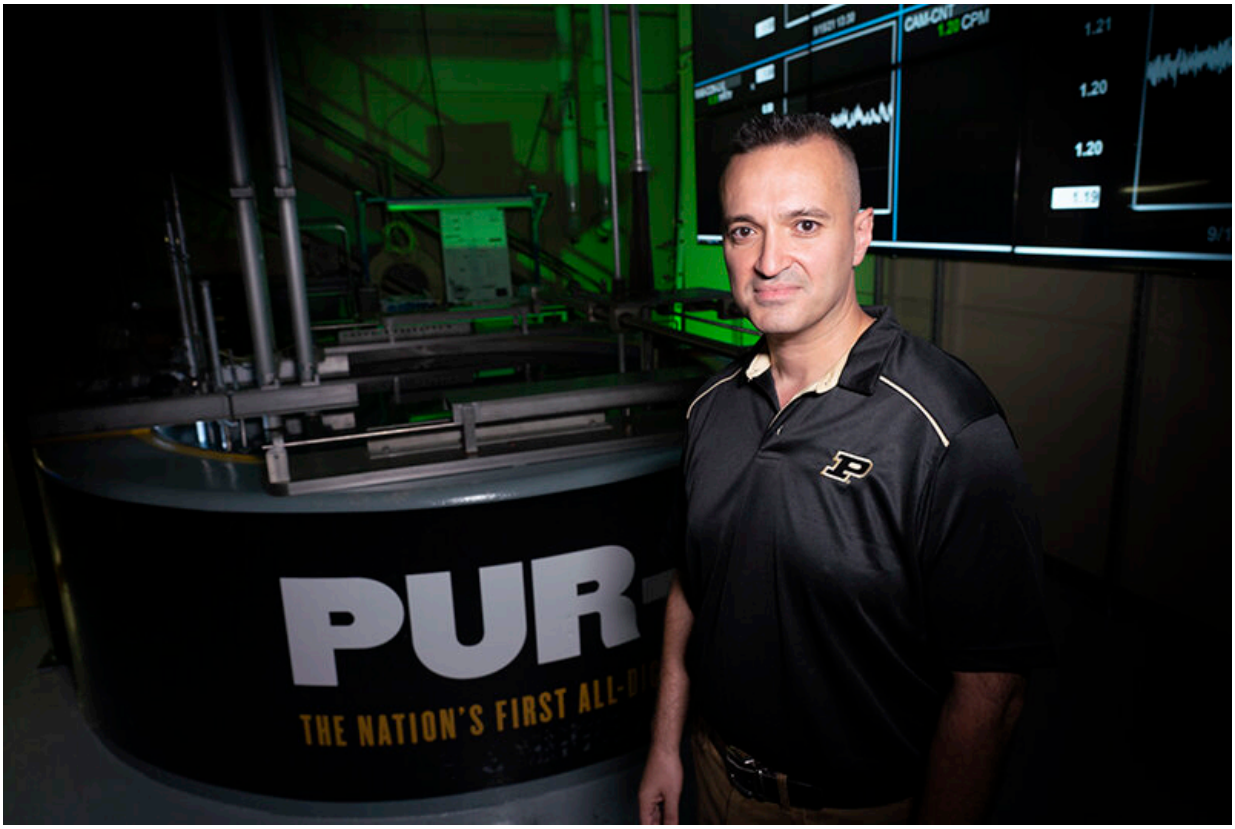
Researchers create 'self-aware' algorithm to ward off hacking attempts

October 7 2021, by Rob Mitchum



Equipping computer models with “covert cognizance” could protect electric grids, manufacturing facilities and nuclear power plants from hackers, says Hany Abdel-Khalik, a Purdue associate professor of nuclear engineering. Credit: Purdue University / Vincent Walter

It sounds like a scene from a spy thriller. An attacker gets through the IT

defenses of a nuclear power plant and feeds it fake, realistic data, tricking its computer systems and personnel into thinking operations are normal. The attacker then disrupts the function of key plant machinery, causing it to misperform or break down. By the time system operators realize they've been duped, it's too late, with catastrophic results.

The scenario isn't fictional; it happened in 2010, when the Stuxnet virus was used to damage nuclear centrifuges in Iran. And as ransomware and other cyberattacks around the world increase, system operators worry more about these sophisticated "false data injection" strikes. In the wrong hands, the computer models and data analytics—based on artificial intelligence—that ensure smooth operation of today's electric grids, manufacturing facilities, and power plants could be turned against themselves.

Purdue University's Hany Abdel-Khalik has come up with a powerful response: To make the computer models that run these cyberphysical systems both self-aware and self-healing. Using the background noise within these systems' data streams, Abdel-Khalik and his students embed invisible, ever-changing, one-time-use signals that turn passive components into active watchers. Even if an attacker is armed with a perfect duplicate of a system's model, any attempt to introduce falsified data will be immediately detected and rejected by the system itself, requiring no human response.

"We call it covert cognizance," said Abdel-Khalik, an associate professor of nuclear engineering and researcher with Purdue's Center for Education and Research in Information Assurance and Security (CERIAS). "Imagine having a bunch of bees hovering around you. Once you move a little bit, the whole network of bees responds, so it has that butterfly effect. Here, if someone sticks their finger in the data, the whole system will know that there was an intrusion, and it will be able to correct the modified data."

Trust through self-awareness

Abdel-Khalik will be the first to say that he is a nuclear engineer, not a computer scientist. But today, critical infrastructure systems in energy, water, and manufacturing all use advanced computational techniques, including machine learning, predictive analytics, and artificial intelligence. Employees use these models to monitor readings from their machinery and verify that they are within normal ranges. From studying the efficiency of reactor systems and how they respond to equipment failures and other disruptions, Abdel-Khalik grew familiar with the "digital twins" employed by these facilities: Duplicate simulations of data-monitoring models that help system operators determine when true errors arise.

But gradually he became interested in intentional rather than accidental failures, particularly what could happen when a malicious attacker has a digital twin of their own to work with. It's not a far-fetched situation, as the simulators used to control nuclear reactors and other critical infrastructure can be easily acquired. There's also the perennial risk that someone inside a system, with access to the control model and its digital twin, could attempt a sneak attack.

"Traditionally, your defense is as good as your knowledge of the model. If they know your model pretty well, then your defense can be breached," said Yeni Li, a recent graduate from the group, whose Ph.D. research focused on the detection of such attacks using model-based methods.

Abdel-Khalik said, "Any type of system right now that is based on the control looking at information and making a decision is vulnerable to these types of attacks. If you have access to the data, and then you change the information, then whoever's making the decision is going to be basing their decision on fake data."

To thwart this strategy, Abdel-Khalik and Arvind Sundaram, a third-year graduate student in nuclear engineering, found a way to hide signals in the unobservable "noise space" of the system. Control models juggle thousands of different data variables, but only a fraction of them are actually used in the core calculations that affect the model's outputs and predictions. By slightly altering these nonessential variables, their algorithm produces a signal so that individual components of a system can verify the authenticity of the data coming in and react accordingly.

"When you have components that are loosely coupled with each other, the system really isn't aware of the other components or even of itself," Sundaram said. "It just responds to its inputs. When you're making it self-aware, you build an anomaly detection model within itself. If something is wrong, it needs to not just detect that, but also operate in a way that doesn't respect the malicious input that's come in."

For added security, these signals are generated by the random noise of the system hardware, for example, fluctuations in temperature or power consumption. An attacker holding a digital twin of a facility's model could not anticipate or re-create these perpetually shifting data signatures, and even someone with internal access would not be able to crack the code.

"Anytime you develop a security solution, you can trust it, but you still have to give somebody the keys," Abdel-Khalik said. "If that person turns on you, then all bets are off. Here, we're saying that the added perturbations are based on the noise of the system itself. So there's no way I would know what the noise of the system is, even as an insider. It's being recorded automatically and added to the signal."

Though the papers published by the team members so far have focused on using their paradigm in nuclear reactors, the researchers see potential for applications across industries—any system that uses a control loop

and sensors, Sundaram said. The same methods could be used also for objectives beyond cybersecurity, such as self-healing anomaly detection that could prevent costly shutdowns, and a new form of cryptography that would enable the secure sharing of data from critical systems with outside researchers.

Cyber gets physical

As nuclear engineers, Abdel-Khalik and Sundaram benefit from the expertise and resources of CERIAS to find entry points into the worlds of cybersecurity and computer science. Abdel-Khalik credits Elisa Bertino, the Samuel D. Conte Professor of Computer Science and CERIAS research director, with the original spark that led to creating the covert cognizance algorithm, and thanks the center for exposing him to new partnerships and opportunities.

Founded in 1998, CERIAS is one of the oldest and largest research centers in the world concentrating on cybersecurity. Its mission, says managing director Joel Rasmus, has always been interdisciplinary, and today the center works with researchers from 18 departments and eight colleges at Purdue. Abdel-Khalik's research is a perfect example of this diverse network.

"When most people think about cybersecurity, they only think about computer science," Rasmus said. "Here's a nuclear engineering faculty member who's doing unbelievably great cyber and cyberphysical security work. We've been able to link him with computer scientists at Purdue who understand this problem, but yet don't understand anything about nuclear engineering or the power grid, so they're able to collaborate with him."

Abdel-Khalik and Sundaram have begun to explore the commercial possibilities of covert cognizance through a startup company. That

startup, Covert Defenses LLC, has recently engaged with Entanglement Inc., an early-stage deep tech company, to develop a go-to-market strategy.

In parallel, the team will be working to develop a software toolkit that can be integrated with the cyberphysical test beds at CERIAS and the Pacific Northwest National Laboratory, where sensors and actuators coupled to software provide a simulation of large-scale industrial systems.

"We can provide additional applications for the technologies that he's developing, since this is an idea that can help nearly every cyberphysical domain, such as advanced manufacturing or transportation," Rasmus said. "We want to make sure that the research that we're doing actually helps move the world forward, that it helps solve actual real-world problems."

More information: Arvind Sundaram et al, Covert Cognizance: A Novel Predictive Modeling Paradigm, *Nuclear Technology* (2021). DOI: 10.1080/00295450.2020.1812349

Matthias Eckhart et al, Digital Twins for Cyber-Physical Systems Security: State of the Art and Outlook, *Security and Quality in Cyber-Physical Systems Engineering* (2019). DOI: 10.1007/978-3-030-25312-7_14

Yeni Li et al, Data trustworthiness signatures for nuclear reactor dynamics simulation, *Progress in Nuclear Energy* (2021). DOI: 10.1016/j.pnucene.2020.103612

Arvind Sundaram et al, Validation of Covert Cognizance Active Defenses, *Nuclear Science and Engineering* (2021). DOI: 10.1080/00295639.2021.1897731

Provided by Purdue University