IEEE Talks Big Data: Ritu Chadha

Dr. Ritu Chadha is an IEEE Senior member and serves as Executive Director at Vencore Labs (dba Applied Communication Sciences), where she oversees the application of data analytics to diverse domains, including healthcare, cyber security, and wireless networks. In this Q&A she discusses cyber security of wireless networks and the creation of testbed environments for testing cyber security algorithms.

Question: You work on something described as "detecting and mitigating attacks on the control plane of wireless ad hoc networks." Would you explain what that means?

Chadha: Let's take wireless ad hoc networks first. Think about first responders arriving at a natural disaster or even the scene of a terrorist attack. They arrive on location and have to communicate securely in areas where communication infrastructure doesn't already exist or has been destroyed. And these responders need to communicate in a way that cannot be easily disrupted.

The control plane comprises the protocols used by ad hoc network devices to set up their communications. Let's say you're using IEEE 802.11 technologies – commercial Wi-Fi – which can be used in an ad hoc networking situation where you don't have typical access points. You just need your devices to be able to directly communicate with each other using Wi-Fi, but without a central access point.

In this example, we can consider the ad hoc network a mesh network. Not all the devices would necessarily be in range of each other. So each device does not only talk directly to other devices within range, but can also route traffic from one device to another and you can route traffic to go across several of these nodes to get to their destination. In order for, say, my phone to talk directly to your phone, we would need software that routes traffic from one phone to the other – that's the routing protocol, which is part of the control plane.

In addition to a routing protocol, we'd have a MAC (Medium Access Control) protocol, which is a lower layer in the protocol stack. We'd need multiple protocols cooperating to get the data from one point to another and all of those protocols are potential vulnerabilities, attack vectors.

Question: That brings us to the "detecting and mitigating attacks" part. Would you explain how you go about accomplishing that piece?

Chadha: In our work, we're assuming that some of our devices might be captured by a person trying to attack our network and, therefore, they have inside access to our devices. Say they

capture our devices. We assume they're able to change the software on the device to exploit the vulnerabilities in the protocol to disrupt communications by dropping traffic or advertising false links in the mesh of devices.

To prevent such a scenario, we're analyzing the vulnerabilities in the protocols and developing techniques to identify such misbehavior. We have the capability of identifying authorized participants in this network. Though all the devices might be authorized and authenticated, one of them might have turned rogue because it was captured by an attacker. Or we might see an insider attack. Then we develop ways to keep the devices and their mesh functioning, even in the presence of such misbehavior. People's lives may depend on it.

Question: Does the solution to this challenge take the form of additional software?

Chadha: Yes. We've tried to make enhancements that enable us to detect misbehaviors. In the example I've created, where one of the phones drops all traffic that it's supposed to forward, we rely on the fact that the other phones in the network can sense what's happening. So software can react by noting that a phone – a node in the mesh – has been compromised and reroute communication to avoid forwarding data or communication through that device. The caveat is if that compromised device is the only way to get from Point A to Point B, there's nothing we can do.

Question: How would you describe how big data and/or data analytics are involved in this scenario?

Chadha: In order to develop models of what expected device behavior looks like versus a misbehaving device, we need to analyze the data. It's not "big data" per se, but it definitely requires analytics.

Question: You also work on something described as "innovative cyber experimentation capabilities." Again, would you explain what that means and provide an example?

Chadha: Our work in this area is more research-oriented. It's really aimed at enabling people who are performing basic research in cyber security to perform experimentation. Basically, we are providing an environment for cyber experimentations.

Let's say you're a researcher with an algorithm that you claim can detect specific kinds of attacks against, say, a specific operating system or a specific vulnerability in that OS. We provide an experimentation environment in which you can set up a mock network with instances of that vulnerability and other real world-like elements. Then you could deploy your algorithm and see how well it works. Of course, it's not the real world. It's a controlled

experimentation environment in which we provide the ability to rapidly set up these kinds of experiments and evaluate technologies.

Question: Would you describe the data analytics angle for this example?

Chadha: I would characterize this experimentation test bed as being a source of huge amounts of data that then have to be analyzed. For example, when we set up an experiment of the type I described where we insert certain vulnerabilities in the network and then we run various detection capabilities, there's a lot of data being generated because we capture every single packet that's sent across the network.

If you have even a dozen hosts generating e-mail traffic and Web traffic, that's a huge amount of data to capture. People who want to analyze that data can run their favorite modeling methodologies or machine-learning algorithms or whatever analytical technique they want to run on that data.

Question: Does this approach have wide application across many domains or industry verticals?

Chadha: I think so. I can envision going beyond our current focus on basic research and providing experimentation capabilities in more commercialized settings.

Question: Are there any public outputs that readers could consult to learn more?

Chadha: We recently got a paper accepted at the MILCOM conference where we published a description of five datasets that are available on the Web for people to run their favorite analytics on. We have these five datasets that capture different kinds of cyber scenarios with different attacks going on, and the datasets can be downloaded from our web site so that people can use them for running their machine learning algorithms or analytics technologies on and compare results. One thing the cyber community likes to have is standard datasets so they can say, "Okay, we all use the same datasets and this is how well my technology did. This is how well somebody else's did." It makes results comparable when you have published datasets that everybody can leverage. So the test bed provides a basis for an apples-to-apples comparison between different analytical algorithms.

Readers can see for themselves by going to CyberVAN (Cyber Virtual Assured Network).

Further, you'll hear some of these topic areas explored at IEEE Future Directions' flagship Technology Time Machine conference in San Diego, CA on 20-21 October 2016.