**Using the Facts to Protect Enterprise Networks: CERT's NetSA Team**
**Transcript**

## Part 1: Providing Insight for Network Defenders and Decision Makers

**Julia Allen:** Welcome to CERT's Podcast Series: Security for Business Leaders. The CERT program is part of the Software Engineering Institute , a federally-funded research and development center at Carnegie Mellon University in Pittsburgh , Pennsylvania. You can find out more about us at cert. org. Shownotes for today's conversation are available at the podcast website.

My name is Julia Allen. I'm a senior researcher at CERT , working on operational resilience and software assurance. Today I'm pleased to welcome my colleague , Tim Shimeall , also a senior researcher with CERT's Network Situational Awareness team.

Today Tim and I will be discussing practical approaches for determining what's happening on your networks and the networks to which you and your organization are connected.

So welcome Tim , really glad to have you with us today.

**Tim Shimeall:** Glad to be with you.

**Julia Allen:** Okay , so I think it'd be helpful to give our listeners a little bit of background. So how long has the CERT Network Situational Awareness team been in business , and just generally what's the scope of your efforts?

**Tim Shimeall:** The Network Situational Awareness Team , or NetSA , has been going since about 2001 , early 2001 , with a major reorganization in 2004 to its present shape. We do analysis geared towards providing business leaders and network defenders with a quantitative understanding of the activity on their network , that can be relevant to making decisions as to how to defend your network , or how to respond to potential threats to your network. And the focus of our efforts is really not quite so much on building tools as it is on providing methods and providing insight to network defenders and to their decision makers.

**Julia Allen:** So would it be fair to say that you take a gestalt or a , probably in some cases , a very detailed look at traffic or behaviors or patterns , and then try to deduce what might be happening based on those patterns and behaviors?

**Tim Shimeall:** Very much. We often work less with packet captures and packet content based stuff , and more with higher level abstractions of network activity. This is largely because we're working on very , very large networks , on the order of a CIDR/8 block. But the techniques that we use have been used on smaller networks.

One of the principle abstractions that we work with is network flow data , which is aggregated header information , without the packet content. And it provides essentially records of all of the communication going across a network , including communication that might be intercepted by defenses along the way , or information or communication that would be ignored by the destination.
So we can determine broad scale what is happening in a fairly ubiquitous fashion across the network.

**Julia Allen:** Excellent. So who would you say is the primary audience for the work products or tools , techniques , methods that your team recommends?

**Tim Shimeall:** The direct audience is probably network defenders , or people that are interested in doing analysis of network activity. This includes CERT's traditional audience of incident response personnel , but it also includes people that are interested in traffic balancing , load balancing , network engineering , service inventory -- service inventory and service control.

And then the secondary audience , or the indirect audience , is the decision makers of these groups. Because what we're trying to do is give the tools that let the frontline personnel inform the decision makers in an effective manner.

**Julia Allen:** Right , so that they're better educated about where to invest in their protection strategies, right?

**Tim Shimeall:** Better educated, better able to assess how well the defense measures are actually working against their threats , and have a generally better understanding of how their network is actually operating.

**Julia Allen:** Well I know you have developed , as you've been at this for quite a few years now , some key principles and strategies that the NetSA team has identified for gaining better what we sometimes call situational awareness, or a better understanding of what's going on the internet and on your network. So could you tell us a little bit about these key principles and strategies?

**Tim Shimeall:** My favorite summary of network situational awareness is it's really a three-step process. The first step is to know your network. And a lot of organizations have a model of their network built up by the network engineering process. But that tends to be biased towards the intended behavior of the network as opposed to accidental behavior that could occur across the network. What you really want to do for defensive decision making, for security decision making , is really know what's, all of what's going on on your network, and provide some useful summary information of what's going on.

The second phase is to know the internet. And this involves knowing what services you are providing to the internet, what services you are receiving from the internet, what potential threats are out there that could be affecting you, what potential changes in technology are going on that could be affecting you, and how all of that interacts with the way your network actually works. Know how the two fit together; know how the trends on the internet interact with the trends- with the behavior of your local network.

**Julia Allen:** Right. And I suspect on the 'know your internet' part, that's probably one of the most challenging areas, just in terms of keeping up to date, right? - with everything that's happening; the new technologies as well as the new attack strategies , right?

**Tim Shimeall:** It is. But the 'know the internet' is very much scoped by your points of presence. So that involves being able to monitor the traffic across your internet points of presence, the traffic that's being dealt with at your ISP or at your POPs, and try and deal with, try and produce useful understanding of what behavior is there. You're worrying about what's the internet as it means to you.

**Julia Allen:** Right. So in other words you're interpreting things that are happening based on -- you called them points of presence -- but all the ways in which you specifically interact, correct?

**Tim Shimeall:** Correct.

## Part 2: Analyzing Aggregated Headers , Network Flows , and Communication Patterns

**Julia Allen:** Okay. So you started us down this path a little bit in your opening remarks. But how do you, from the NetSA team's point of view, how do you go about determining what is happening on the internet as it affects you or on your own network? Can you give us some ideas of the analysis that you perform and how that shows up?

**Tim Shimeall:** We do a lot of analysis on aggregated header information, on network flow information. And this is largely looking at what services are being provided, both from and to the internet, what are the communication partners out there on the internet, what are new emerging groups of communication patterns.

So for dealing with something like say phishing; phishing's very, very difficult to detect on a message by message basis, because they've gotten very clever at faking messages. But one feature about a lot of the phishing sites is they're sent from very transient locations, which is very different from normal email communication. Most of the business email that you receive are from people that you've communicated with before. So if you're tracking the normal set of email communication partners that you've got, and then looking at variations off of the normal, particularly very transient variations off the normal, you can get a sense for where phishing sites might actually occur. The more you know about the distant end – Are these cable modem pools? Are these dynamic address pools? ; things like that -- the more refinements you can put in place to understanding where some of these attacks might actually occur from.

Most of the time that an individual is using a network, they are interacting with what we call the locality set. They're interacting with groups of hosts that they've already communicated with. And then you look for variations off of that locality set for indications of possible malicious or lax changes of behavior.

**Julia Allen:** Do you take – just to define this a little bit better -- do you take header, traffic header sets across geographies, across time? Is there a sample size that seems to work the best or does it depend based on what you're looking for?

**Tim Shimeall:** Well organizations can instrument their own networks and start to collect this information. Network flow information is actually fairly tractable because it can be collected in a very, very efficient manner and stored in a very, very efficient manner. We literally can store communication that might involve gigabytes moving across the internet, or moving across the local network, and record the fact that the traffic has occurred in as little as 30 bytes of information.

**Julia Allen:** Wow.

**Tim Shimeall:** And this, because we've got these very small records, we can collect across broad infrastructures and we can keep around the records for long periods of time. We actually have sponsors that keep six months to a year of traffic records online, and that are able to do large scale analysis across their infrastructure for this time period. And that gives them a lot of precision in looking for emerging trends, in matching against patterns of activity.

**Julia Allen:** That's incredible because I always think of organizations being inundated with network data. But it sounds like, based on your flow analysis strategies and approaches and looking at headers, you can really consolidate down to the essential information.

**Tim Shimeall:** That's correct. And we actually have collection infrastructures that we've built because at the time when we started nobody else was out there building tools like this. And we have a fairly mature tools suite that we actually have built to enable this kind of analysis. The tool suite is known a the System for Internet Level Knowledge, or SiLK , and it's available via the NetSA Tools website.

**Julia Allen:** So you gave an example about phishing but I think it would help a little bit if you could walk us through maybe a typical scenario or two of looking at a sample size and making some inferences or doing some analysis to detect maybe a distributed denial-of- service attack ramp-up or maybe some low-level port scans to gain intelligence. Are there some kind of typical scenarios that you see that you could just briefly walk us through?

**Tim Shimeall:** Well one way of spotting the ramp-up of distributed denial-of-service attacks is to note the number of new hosts that are starting to communicate with your network, particularly hosts that are communicating on unusual ports or protocols. Because often the DDoS tools will randomize ports and things like that to try and bypass security controls.

By spotting the fact that we've got a bunch of new hosts that are out there trying to initiate communication with us, and, in fact, that the number of new hosts is rapidly passing the number of conventional hosts, that's a good indicator that something's going on that's very unusual and could potentially be for a DDoS hitting your network. And again, it relies upon having some broad-scale knowledge of who do we normally communicate with, what's the relative rate of communication. In some cases you can spot this ramp-up before you start to see your bandwidth being severely impacted. And that gives you opportunities to do things like communicate with your ISP and say, "Hey , can we get some rate throttling in place? So at least we preserve some connectivity and don't kill our border routers" and things like that.

Another example (rather than talking about scanning), another example that we often do is looking at how much network bandwidth is being used by various services or by various hosts internally and be able to spot bandwidth hogs -- hosts that are taking more than their share of activity or services that potentially may not be mission relevant to our organization but are consuming large amounts of our organization's infrastructure. Do you have an unusual amount of download activity? Do you have activity that indicates "Hey , we've got a lot of people that are listening to internet radio in the workforce, " and that's actually to the point where it starts to impact the throughput of our network. But being able to spot those kinds of activity across the network, and spot points of contention across the network, is something that's fairly straightforward to do with netflow, just simply by balancing, by looking at what ports, what protocols are getting most of the traffic, and what hosts are emitting and consuming most of that traffic.

## Part 3: Methods and Tools; Making the Business Case

**Julia Allen:** So then do you actually produce -- and we're going to be talking about the methods and tools as well -- but do you actually then produce some kind of a visual depiction of bandwidth usage over time by, as you said, by service or by user group or by some sub- network?

**Tim Shimeall:** Yeah, we can easily produce trend graphs, which are time series graphs related to network utilization. We can easily subdivide that according to common ports and protocols. We also can produce scatter graphs that identify most common producers, most common consumers, and present the data very visually so that decision makers have a fairly actionable set of activity. And this has led to some of our partners saying, "Well we don't want this. And we want to deprecate our response time with respect to that," in order to better balance their network utilization.

Scanning -- we have some fairly advanced scan detection methods that we've used, largely because we want to be able to separate out scanned traffic from non-scanned traffic in terms of determining what behavior is being exhibited by the network. Also because scans are often a precursor for other sorts of attacks, and be able to note when there's been a change in scanning behavior on your network.

Unfortunately modern scan rates are such that no one has the available resources to respond to each scan. What you try and do is manage the response en masse across your network , and be able to note when things are being scanned for and are there particular focuses of a scan.

**Julia Allen:** You know you raise a good point, which is we know in security you can't secure everything at the same level. It doesn't make good business sense to protect everything at the same level because some services, processes, business assets are clearly more valuable than others. Do you find the same kind of thinking applies when you're looking at network situational awareness where you really can use some of the methods and approaches you've described to hone in on the areas that are most important?

**Tim Shimeall:** Very much, and you very much need to keep the most important in focus. Because it's not unusual to spot weird patterns that may or may not be relevant to the defense of your network. Are there servers that are out there that are misconfigured and the misconfiguration causes them to emit bursts of behavior? That's not a good thing but it's not necessarily a security problem.

On the other hand, someone trying to flood out particular hosts and act in particular ways is a security issue. And you would want to be able to make rapid distinctions between the two. You can spend a lot of time analyzing weird behavior and really not get much additional security information out of it. If you keep in mind what your mission is in terms of improving security, that gives you a lens through which you can observe your traffic and focus in on those aspects that'll help you out the most.

**Julia Allen:** So what are some of the other methods and tools that the NetSA team has developed that are generally available for our listeners to take advantage of?

**Tim Shimeall:** We do have some visualization tools which are available, particularly the RAVE engine that's available on the site. We've got a collection of methods that we've incorporated using our tools to drive forward analyses for various sorts of behaviors. So if you want to have a prototype of how do you do a bandwidth study, we've got an example of that. Do you want to know how to measure activity across a given set of ports or for a given set of hosts? We have tips that tell you how to do that.

But in broad, we're really trying to look at what makes sense to decision makers, and what allows decision makers to make evidence-based decisions rather than decisions that are based on fear or uncertainty. And by making decisions from measurement, they're making decisions that are often more balanced and more immediately relevant.

**Julia Allen:** You know, that's a good point as we come to our close. So if you could put yourself in the mind of a network administrator or some primary member of your target audience having to present a business case for using this work, using these tools and methods, having a better understanding, and trying to present that up the chain of command to get funding for this kind of effort -- what are some of the arguments that you might make up to a CIO (chief information officer) or a chief risk officer?

**Tim Shimeall:** One, it allows the CIO to have really a very balanced picture of their network; to state how their network is behaving as it is as opposed to how it's intended to be. And that is often very, very useful.

Secondly, it allows for very rapid recognition, independent of small and easily changed characteristics of various parts of attacks. So we can spot a denial-of-service ramp-up without really worrying about what is the exact semantics of this denial-of-service. We can spot scanning behavior and changes in scanning behavior without really worrying about what are the packet characteristics of the scan? And so that provides a degree of broad applicability that really isn't available with conventional IDS (intrusion detection system) approaches.

The other aspect of it is you can collect this very economically. Because the network footprint -- because it's a passive collection rather than active collection, and because the network footprint of the collector is so small, it's very cost efficient and very bandwidth efficient to collect some of this information.

**Julia Allen:** Excellent. Well Tim this has really been a fantastic introduction and presentation on the kind of resources that your team is producing. Are there some places where our listeners can learn more, and some ways they can take advantage of this work?

**Tim Shimeall:** Certainly. One, we are always interested in people that are interested in improving the state of their network situational awareness. And so they can easily communicate via the CERT website and via the contact methods on the CERT website. There's also a lot of analysis studies that are present in the Network Situational Awareness portion of the CERT website.

We also have an annual conference that we run called FloCon. And the next one will be New Orleans this coming January. And certainly people can register and come and attend and hear what's going on across the community rather than just CERT's people. We have people from around the world that come and produce with that conference.

So there's a lot of different ways that people can build their awareness, either online or in person.

**Julia Allen:** Well fantastic. I thank you very , very much for your time, and perhaps I can convince you to do an update with us in the future.

**Tim Shimeall:** I would look forward to that.