Using Network Flow Data to Profile Your Network and Reduce Vulnerabilities
Transcript

Part 1: What a Network Profile Can Tell You About Your Network

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. Show notes for today's conversation are available at the podcast website.

My name is Julia Allen. I'm a principal researcher at CERT, working on operational resilience. I'm pleased to welcome Austin Whisnant and Sid Faber, members of CERT's Network Situational Awareness Team.

Today Austin, Sid, and I will be talking about how organizations can use network flow data -- and we'll tell you what that is -- to identify the top public facing assets and services with most network traffic on your network. And the reason this is important is because a potential attacker can observe and take advantage of this information. This conversation is based on an August 2012 report that Austin and Sid authored, which is titled *Network Profiling Using Flow*.

So with no further ado, welcome to the podcast series Austin.

Austin Whisnant: Thanks Julia.

Julia Allen: And Sid, we're really glad to have you with us today.

Sid Faber: Hello Julia. Thank you.

Julia Allen: So Austin, why don't you get us started? Can you say a little bit, just to set the stage? Tell us what a network profile is and why it's useful to an organization.

Austin Whisnant: Sure. So a network profile is basically a list of all of the assets on a network and what their associated purpose is. And that's beneficial for system administrators to know what they have on their network and what they need to defend. It's beneficial to business administrators for purchasing and allocating resources. And it's also useful for the operators to know what's going on, on their network.

Julia Allen: Okay. And so often we advise our customers and the people that we interact with to build an asset inventory. So could a network profile be used to help generate such an inventory?

Austin Whisnant: Sure yes, that's exactly what it could be used for. And people can make it as detailed as they need to or just leave it as basically a list of what their assets are.

Julia Allen: So if I have such a profile, how might I use it to respond to an incident or help with a security assessment?

Austin Whisnant: Well if you have the profile then you know what assets are supposed to be doing on your network. You know which machines are servers and which ones are clients and those kinds of things. So you can see when something is not doing what it's supposed to be doing.

Julia Allen: So you can actually observe the behavior of a particular asset. Like say there's some services that are real high priority, you could use the profile to help you watch those servers?

Austin Whisnant: Well, the profile is more of a baseline, so you know what they're supposed to be doing. And other network monitoring equipment will tell you what it's actually doing, and then you can compare what it's currently doing to what your profile says it should be doing.

Julia Allen: Great, great. Well that really helps. So Sid, we're always worried about the bad guys, right? So how could -- if a potential attacker had access to this type of information, how could they use it and turn it to their advantage?

Sid Faber: So I think one of the most important things to point out about the profiling activity is you have to assume that the attacker actually does have all this information at their hands already. One of the things we know about attackers is that they're always scanning our networks and determining what's on the network. And essentially with the profile, we're doing that ahead of time. We're taking advantage of knowing what services are facing the outside world and learning everything that the attacker might know. So that way we're essentially one step ahead of them.

One of the surprising things that we've noticed too is that as we look at our network profiles, we more often than not will find misconfigurations. We'll find things that are hanging out off of your internet that you didn't realize were there. And then once they get brought to light, it's a pretty simple issue to tighten things down and make sure that your network is behaving the way that you had designed it, not the way it kind of evolved over time.

Julia Allen: So by doing that, you can at least -- theoretically you can eliminate maybe some weaknesses or vulnerabilities or points of entry that you didn't know about before, right?

Sid Faber: Absolutely. We've found in some of the places where we've done network profiling things like there's a machine that was directly connected to the internet for a research or evaluation purpose that later on had been connected to the internal network at the same time. And when we picked that up on the profile, the system administrator was surprised that the machine was still there and then turned right around and shut that machine off from the internet.

That was essentially a way that an attacker could bypass the corporate firewall and get right into the inside network but -- and it wasn't a standard configuration, wasn't something that was expected. But by simply watching the traffic, they were able to pick up on it and just simply shut that down.

Part 2: Seven Steps for Building a Network Profile

Julia Allen: Great, great. So Austin let's turn to your report a little bit. It's quite comprehensive and certainly well worth a read for listeners who are interested in this topic. But by way of introduction, in your report you describe seven steps for actually starting to create and build a network profile.

And I was hoping you could introduce those briefly here, and then you and Sid and I can dig into them in a little bit more detail. So could you just summarize those top level steps?

Austin Whisnant: Yes sure. What I'd like to point out real quickly is that this report uses “flow” specifically, network flow data, rather than packet capture or some other programs for inventorying networks. So we're focusing on flow in this report.

But we go through seven steps. And we start with basically just gathering information that you might already have about your network, like network maps or purchase orders for equipment, things like that. Sometimes that stuff can be incorrect or out of date. So we just want to gather that first. Then we select an initial data set that we want to look through; maybe a day's worth of traffic or a week's work of traffic, depending on the size of the network.

And then we look at, or we determine which address space we're concerned about. It may be the whole network; it may be something else. We look at cataloging the common services, so things like Web, DNS, maybe VPN or FTP. And then we move on to cataloging other services that might be more specific to your network. Maybe you have a special SQL protocol or something like that.

And then we catalog whatever machines are left over. There may be a handful of IP addresses that haven't been profiled yet. And basically the last step is just maintaining the profile and determining how it's useful to you and how to use it on your network.

Julia Allen: Before we get into a little more detail here, can you say a little bit about the resources and the time that you've observed that it takes to put this process in place and really make it go?

Austin Whisnant: Yes. So it depends a lot on the size of the network and how dynamic the network is. So if it's a small network that's not changing very often, then you may be able to finish this in a few days or a week. And then you won't have to update it for another six months or something like that.

But if it's a very large-scale network that has a lot of changing components, then it might, it'll take you a long time to get through it; maybe a few weeks to do the initial profile. But you'll also want to be concerned about updating it fairly often; since your assets are changing often you'll want to keep up with that.

Julia Allen: So do you sometimes advise your users of this approach, and in your own experience, to start with maybe a small subnet to get your feet wet and then increase? Or do you think it's better to take on a larger scope initially? Or is it the proverbial "it depends"?

Austin Whisnant: Well if you don't have the resources to do a huge dataset, then it's fine to start with a subnet; or maybe profile subnets separately and then just aggregate them all together when you're done. That's a perfectly acceptable way to do it.

Julia Allen: Great. So let's dig into some of the steps a little bit, just to make sure our listeners get some of the -- put some meat on the bones, if you will.

So your first two steps -- so you talked a little bit about the kind of information, network information, you need to gather and putting an initial dataset together. So could you say a little bit more about that?

Austin Whisnant: When you first gather all of the information you may have, oftentimes network maps are out of date, lists of servers are out of date. And going back to what Sid was saying earlier, you may have something on your network that you think is doing something. You may think it's a web server but it actually turns out that it's serving email traffic or something like that. And that's something that the profile will pull out.

So when you finish your profile you can go back to those lists, that initial information that you gathered, and update it and see if everything's correct, if everything's doing what it's supposed to be doing. Then as for the initial dataset, like I said it depends on the size of the network. But sometimes you may want to pull just your outgoing traffic or pull like a specific subnet; just whatever you're capable of doing with your resources in the time that you have to do this.

Julia Allen: And what about -- I know when we were preparing for the podcast you talked about the criticality of sensor placement. Can you say a little bit about that?

Austin Whisnant: Yes. So depending on where your sensors are placed, you may get different results. In this profile we're focusing on what your network would look like to an outsider. So we

want to place the sensors around the perimeter. You want to make sure that you're getting all of your perimeter, your entrances, covered. And that way you're not missing any traffic, you're not duplicating any traffic, things like that.

Sid Faber: Julia, if I could slide another comment in there...

Julia Allen: Please.

Sid Faber: ...about sensor placement. Interesting thing about flow is that for a lot of organizations flow is native; it's built in. We started working with flow because routers already generate flow. So most of your common routers, all you have to do is change a configuration option, and as long as the router's not overburdened and it has room to do a little bit of extra processing, it can create the flow and then send that flow into a collection engine.

We use our own tools here at CERT. And we'll talk a little bit more I think later on in the podcast about what other resources are available. But there's also commercial products that might already be within your organization that'll collect the flow. So it should be pretty easy to at least get kick-started on generating flow and sending that to some central place where you can do some analysis.

Julia Allen: Oh that's a great point Sid because I was -- one of the questions I was going to ask you or Austin is number of sensors. Is there an optimal number of sensors or does there become a threshold beyond which it becomes too burdensome both to collect and to receive and process all that information? But if the routers already have flow built in, that seems to be a no-brainer, right?

Sid Faber: Yes, yes. Obviously there's limitations on what your routers can do, and it's better if you can have a dedicated sensor. But we see a lot of tools that are starting to come on the market that'll collect flow directly from the routers with what they have.

So if you have that, you're great. And the best place to collect them we find is actually at your internet border, your internet choke points. Most organizations know pretty well where their internet presence is at, mainly because they have to pay the bill for it. So whatever you're paying the bill for, for your internet presence, that's a good router to start with.

Julia Allen: Great, great. Okay so Austin moving on, your next step is about identifying the active monitored address space; in other words what IP addresses and what hosts and what assets and services you actually want to monitor for. So could you say a little bit about how to identify that address space?

Austin Whisnant: Yes. Well what we started with in the paper was looking at all of the hosts on your network that have active TCP connections. And we separated out TCP traffic from other traffic, just so that we could be sure that there weren't any ghosts -- what we call ghost's traffic in there; things that don't make full connections and scanners and things like that. And once we pull out the active TCP connections, then we move on to the other protocols: UDP, ICMP, whatever else might be on your network, and we pull those separately.

And then we just combine those two together. Basically you have a list of IP addresses that are active on your network. And then you want to look at information that you gathered previously and see if there aren't any other assets that might not be included in there. You may have things like backup gateways or something like that, that typically don't have any traffic but you may still want to include in your profile. So you just want to aggregate all those together and have a list of IP addresses.

Julia Allen: Great. And once you have that, once you know that basically sets your scope -- you've got your sensors in place; you know what data you want to collect -- then what about the

cataloging? When you talked about cataloging common and other services and leftover assets -- that's easy to say and I'm sure quite difficult to do. So how do you move forward on those set of steps?

Austin Whisnant: Yes. So by cataloging services we mean just pick a service such as Web, Web traffic, and find the assets that are communicating on Web traffic. So when we have a client server architecture, we split up the servers and the clients, and we'll start by looking for Web servers first.

Then the profile goes through all the details and the port numbers and protocols and all of the technical details that you need to do that. But basically you just look for a specific port with outgoing traffic, and if the asset is communicating mainly on that port, like it's mainly communicating through Web traffic, then you might catalog that as a Web server. And same with DNS and VPN and those more common services.

Julia Allen: And what about any other services or leftover assets?

Austin Whisnant: So once you go through all of the main ones that are common to most networks, you may have -- you may notice that there are some IP addresses that you had found when we talked about identifying the active hosts. You may have some of those that you haven't profiled yet.

So you want to go back and instead of looking for a specific service, you want to just look at the traffic that that address is producing and see what's the highest, what's the highest service or what's the highest service that it's using, or the highest two maybe, and catalog that particular asset that way.

Julia Allen: And as you're doing this, how are you capturing all this information? I know in a minute I'm going to ask Sid about tools and maybe that's a better place to talk about it. But I'm trying to visualize what all this looks like. Can you give me an idea? I know it's hard when we're just using audio only and I know you've got examples in your report. But can you give me an idea about what some of this data looks like?

Austin Whisnant: Yes. It depends a little bit on which tool you're using and Sid will go through that a little bit. But basically what the report looks at is just a command line interface, a text interface, that shows you records, flow records, that have like the port numbers, the protocols, the IP addresses. It may have flags, byte counts, packet counts, all of those things. And you have tools at your disposal to sort through those and pull out unique ones and look at statistics and all kinds of stuff. It's a very powerful way to look at this huge amount of data.

Julia Allen: Yes, it sounds like a very rich dataset.

Austin Whisnant: It is, very much.

Part 3: Netflow Analysis Tools; Detect Anomalies by Tracking Trends

Julia Allen: So let's talk about that Sid. I'm kind of itching to get into tools. So you mentioned flow. But let's talk a little bit more broadly -- what we use and what we recommend others use for collecting and analyzing network traffic.

And also, as Austin said earlier, there's this whole issue of keeping the profile up to date and how often you need to do that and clearly tools will help with that. So can you walk us through tooling as it applies to this problem space?

Sid Faber: Sure, absolutely. Here at CERT our, the tool that we use is -- it's known as the System for Internet- Level Knowledge; shorthand for that is SiLK. And it's a flow analysis tool. And it

includes not just generating flow from packet captures but also storing it and analyzing it. And it's mainly a Linux/Unix toolset. It's open-source so it's free for download. So we use that. And it's geared towards collecting flow in very large environments, which is why there's not a lot of front-end graphical interface on it. There's a little bit of an overhead in getting started and understanding the commands to it.

There are other tools out there that are much more end-user driven that will gather flow data for you. I'm hoping that some of our listeners will actually encourage some of those vendors to do some more network profiling. I think I've seen some of that show up in the tools where we get an inventory of what our systems are. But to refine that a little bit more with some of the things that we talk about in the report I think would be of a lot of benefit for our listeners.

Julia Allen: So do you have examples of what some of those other tools might be?

Sid Faber: Sure. I know one of the ones that has a lot of footprint in the marketplace is Lancope. Cisco has a tool for flow analysis obviously. The name escapes me right now. I believe it's a part of a suite of tools that Cisco offers for network management and monitoring.

Julia Allen: Great, great. And what about -- so I'm thinking about this incredibly large data stream that's coming your way. And if you do this periodically, clearly you're trying to collect trends over time. Can you say a little bit about that aspect of automation and also keeping the profile up to date?

Sid Faber: Sure. Trending we've found is extremely valuable and watching it over time. First of all, watching your network traffic over just a week's worth of time, after doing the profiling activities, as Austin described.

One of the things we've found are things like you'll notice on Sunday evenings perhaps an FTP upload happens where you refresh content to a content distribution network (maybe an upload to Akamai) and it's a large spike. And you see a pop in your network traffic; it shows up every week and you recognize it.

We've seen other places where, for instance, something will happen over the weekend that was actually a user that came in over the weekend and did some stuff on the network that you wouldn't expect to be normal -- might be something worth looking into, maybe not, depending on network policy.

We see the ability to track things like software updates and patching. In one case we had an organization where we were monitoring their web servers and you would see web server traffic, normal web server traffic. But from time to time the web servers would act as web clients as though somebody were sitting there downloading. What it was was the servers would actually pull for updates and you would see those updates come down to the servers.

And as long as that's within your network policy, that would be normal. But if you have a standard way to manage your software updates then that would be an anomaly -- something that you'd want to look into; perhaps a misconfigured server.

So watching those trends over time becomes really valuable. And obviously as we watch that over more than just a week, but watch it over months, perhaps even up to a year, we can tell how our bandwidth is being consumed. We can tell adoption of new technologies such as streaming audio and video, and bring your own device and mobile devices. And you can profile how those things happen and potentially predict how your organization will be impacted by the adoption of new technologies.

Julia Allen: So clearly, we're doing this and recommending that our users and listeners do this to have a more accurate picture of what their public face is. But attackers have access to all of these tools as well, correct? So can they get, with appropriate cleverness, can they get a similar picture? And is that also one of the reasons why you want to do this so you know what your external facing signature or footprint is?

Sid Faber: So they can definitely have a view into what your network server, what your network footprint is. They can see what resources that you've published out to the internet that you know about. They can also see potentially the ones that you don't know about -- as we mentioned early on, the ones that they detect by scanning.

I'd like to also point out as we're talking about attackers, one of the things that we've started to cultivate with the idea of network profiling is a more proactive approach to network security. So what we want to get at is here you're getting a feel for how your network behaves so that you can detect those anomalies.

We've seen a lot of products come out about anomaly detection. But before you can find anomalies, you have to know what's normal. So some of the events that I mentioned earlier are anomalies. Whether or not they're security events, that's up to the manager or the policy; it depends on policy and so on.

But the idea of understanding what's normal on your network is a huge win; and when there is an event that goes on and you can pick that event out, you can determine that there's an anomaly going on and then you can trace it down to find that more advanced threat that's not detectable by your traditional signature-based detection methods.

Julia Allen: Excellent. And I know this is a companion to, as Austin said earlier, many other types of network monitoring that goes on in a normal day-to-day IT shop. But is there an ability with the tooling to actually generate alerts based on the trending that you described or is that really not appropriate for this type of analysis?

Sid Faber: So with the technology that's out there right now, there is some alerting that's built in, and there'll be more as the tools continue to mature and be updated. The pretty straightforward alerting is mostly driven by watch lists. So I think every organization has a list of IP addresses that they don't like or that are associated with malicious code and so on.

And so if you identify traffic to those devices, to those IP addresses, you could identify that; that could alert right away. But more often what we see is something like a daily workflow where there's a daily report that'll show up in someone's inbox. And they'll look at it and see: “Has network traffic been like I would expect it to be?” or “Are my web servers talking as I expect them to be?” “Do I have extraneous SSH traffic, or something along that, that's worth looking into?” More often it's a lot more about understanding the dynamics of network behavior than it is about a traditional: “Here's an alert that I have to run to ground.”

Julia Allen: Right, right. I was just going back to your comments about anomaly detection and having to be -- for that to be effective, you need to know what your known good state is, and therefore be able to measure against some threshold departures against that known state. So I was thinking well maybe we could turn this to our advantage in that respect.

Sid Faber: Absolutely. And I think another huge thing is to grow that knowledge base of your analysts; whether they're network analysts, security analysts, or depending on how the organization is set up. But somebody that really understands how the network is used from day to day.

Then that skillset, that becomes invaluable to the organization as you have to face a threat that you hadn't planned on. How do you handle an outage of a particular service, regardless of the root cause of it? There's somebody that understands the network dynamics and knows how to react to unique threats.

Julia Allen: Great, great. Well Austin before we come to our close, were there any other comments that you wanted to make about the process or your experience in building network profiles -- anything that we might have missed?

Austin Whisnant: I think Sid has said pretty much everything that needs to be. One of the main things that I took away from the first time I profiled an enterprise-sized network was the amount of stuff that you find that just shouldn't be going on, on the network.

Sid had mentioned some of those anomalies from other networks. And it turned out to be very useful to the organization to find all of those anomalies and to be able to lock their network down better than they had before.

Julia Allen: Excellent. Well thank you. So Sid, you get the last shot -- and Austin, if you care to -- do you have some places where our listeners can learn more about this body of work?

Sid Faber: Sure. I'd encourage our listeners to take a look at the SEI and the CERT websites. There's the report itself is posted on the website and that's open for all to use. There's also the Network Situational Awareness site, underneath CERT, has a lot of resources; and in particular the open source toolset, the SiLK toolset as I mentioned, is there.

So there's also a handful of other tools out there and they are in very active development. We expect some significant updates to those to come out within the next month or two. So keep an eye on that. There's also some mailing lists associated with it where our developers -- you can go to them if you have questions about how to install the toolsets or how to get up and running.

And finally the one thing I'd encourage everybody to take a look at is FloCon. That's our annual conference on network flow analysis. So that's again available off the CERT website. And that'll be held in January, the second week of January, in Albuquerque, New Mexico.

There's complimentary training there. In the first day of the session they'll train on how to use SiLK; as well as we're -- I believe we'll have another analysis toolset, Argus. Some of our listeners might be familiar with that. They'll be presenting there as well; so training on that. And then research and presentations on how people are actually using flow for analysis out in the field.

Julia Allen: Great. Well Austin, thank you so very much for your time and your preparation and for this great body of work. I really appreciate your time today. Thanks so much.

Austin Whisnant: Thanks for having us.

Julia Allen: And Sid, excellent conversation. I really learned a lot listening to both you and Austin. So thank you for your time today.

Sid Faber: Thank you very much.