Analyzing Internet Traffic for Better Cyber Situational Awareness
Transcript

Part 1: How Do We Know What's Happening on the Internet?

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 security governance and software assurance. Today I'm pleased to welcome Derek Gabbard, Co-founder and Technology Director for Lookingglass. Derek's firm specializes in internet traffic monitoring, analysis, and visualization for situational awareness, which we'll be talking about today. Just for information, Derek's also a former member of Carnegie Mellon's CERT program. So Derek and I are going to be kicking around how network data and traffic analysis can help listeners and organizations stay one step ahead of attackers. So welcome Derek. Glad to have you with us today.

Derek Gabbard: Thanks, Julia. Glad to be here.

Julia Allen: So according to some great information you've posted on your website from Cisco and one of your recent papers, I think it's no surprise to any of us that internet traffic is set to double every year for the foreseeable future, both in terms of volume and rate, maybe even more so in terms of rate. So given that, it's really difficult for any organization to know what's happening out there. What are some of the dimensions of that kind of volume and rate issue that you see?

Derek Gabbard: Well, the interesting thing about the volume and rate changing is sort of the reasons that it's changing. So there's the YouTubes and the viral video aspect of driving internet traffic. But – and those things are certainly interesting from an enterprise security perspective, mostly for the purposes of monitoring employee behavior and making sure they're on task, doing what they're supposed to do. But the movement toward cloud computing and outsourced services and a reliance upon the internet for business operations that has, I guess, kind of started maybe six or seven years ago to be really the wave of the future and now here we are. That's kind of lost in that traffic doubling over the next few years.

And so there are a lot of dimensions to it. First that it's a – there's a need to have an understanding of what the internet looks like, the way that an organization is connected to the network for the purposes of creating and defining and executing security posture. But then as you move into the cloud computing type of mindset and services and applications and capabilities are hosted other places, you're fighting in that sea of bandwidth for bandwidth, for routing, for service with a whole lot of other folks that are doing a variety of different things.

And so when you spin it back to the requirements for an internal network security or an external network security person or organization, it really comes down to risk identification, risk assessment, and being able to see and understand where your services and capabilities are. And how trends and issues that are happening on the internet – whether it's traffic like we're talking about or connectivity or peering relationships – how those things that are going on outside your borders are going to have the ability and potential ability to impact your operations, to impact your security, to impact your availability.

And so I think the place where we're moving to now is this need for situational awareness of the network that goes beyond the traditional borders of the past. So you used to be able to draw the perimeter around the network. And you could be very focused on network traffic inside that perimeter. You could be very focused on network security inside that perimeter, on creating the right kinds of access controls and so on.

But as we've taken services and applications and put them outside of that traditional perimeter with cloud computing and hosted services other places, it becomes important to the business continuity planners or the security folks to have an appreciation and an understanding for what's going on on those networks that are actually serving up those services to the organization.

I've talked to a lot of financial services organizations where my company's a part of, an affiliate member of, the Financial Services Information Sharing and Analysis Center. And in talking to those folks, they're telling us, "We have services that are hosted somewhere else. And I'm having a difficult time understanding what the dependencies are between me and them from an internet routing and typology perspective." And then when you add to that this huge amount of increase of traffic, they're starting to become concerned that availability, that bandwidth is going to be an issue and then obviously, that security will be an issue in that same vein as well.

Julia Allen: So when you talk about this idea – I mean I hear the term situational awareness a lot. Can you kind of net it out for us? When you're talking about an organization being situationally aware of, as you said, their connectivity and their external dependence to the internet, what does that mean and why is that something that you think business leaders need to be paying more attention to?

Derek Gabbard: Well, in the past, you really had pretty much all of the services and applications that you hosted or that were part of your mission – the things that you delivered for your organization – they were within your control or at least within your sphere of control. And nowadays, those things have been outsourced to the point – when you look at cloud computing, when you look at Amazon's hosted services, there are organizations that are starting to put components of their own internal operationally critical systems and services outside their network and into the cloud.

And so to really understand situational awareness or to get to situational awareness, I think it requires that you have a knowledge and appreciation of where and how those things are hosted, the connectivity to and from your network, to and from your

customers' networks or your suppliers' networks, or whoever needs to touch and gain access to that information or service that's being hosted somewhere else. And that you can kind of draw that out through the whole information supply chain so that you can see "these are the 20 organizations that need access. If I'm doing business with them, they need access to my services or applications that sit in the cloud. Here's where those things live the cloud, here's my network and how I'm attached to that." And to be able to really start to identify key weaknesses, critical connections between service providers, critical, potentially critical geographic regions. If you start to map things out to take in the cyber world and trying to lay it into the physical world so that you understand where you have dependencies upon systems or services or assets or infrastructure that sits in a specific region.

These are the things that I'm hearing from the financial services folks as well as from the federal government. We do a lot of work with them as well. And they're trying to understand what it looks like, what the nexus between physical infrastructure and cyber infrastructure, and the nexus between required availability and services to overall connectivity and to geo location and to tying it all the way out from end to end so that you can know who and how people are going to access the services and applications and systems that they need through you or from you.

Julia Allen: So would it fair to say that, again it's getting harder and harder to keep track of all of that because it extends beyond the direct control of the organization or even the direct control of their suppliers or their partners. But you're really doing this to insure that you have the proper risk mitigation strategies in place, that you're protecting the data that's transiting those services, that you understand, as you said, who you're critically dependent on. So from a service continuity point of view, that may be where you want to shore up and pay more attention in terms of oversight – those kinds of things?

Derek Gabbard: Yeah. I've always tried to be a fan of operationalizing the risk management. And so to be secure for the sake of being secure doesn't necessarily make sense. It makes sense in the context of your mission and your business and whatever it is that you're trying to accomplish. And so to take your mission and your business and whatever it is you're trying to accomplish and to be able to overlay that with this picture of the internet and this picture of the interdependencies that you have on your suppliers and all the things that you mentioned before – their providers, your upstream provider's upstream provider. So that you can start to see the cascading impacts that can happen as a result of, say, a regional issue like Hurricane Ike or Hurricane Katrina. Or a connectivity issue like two major ISPs (Internet Service Providers) de-peering from each other as happened last year. And to be able to, in advance of those type of events happening, to be able to identify and to mitigate the risks that are associated with any of those types of events. That to me is the basis or the starting point for situational awareness.

Part 2: Automation that Scales; Innovation through Visualization

Julia Allen: Okay, well that's great, that helps a lot. So turning our attention to starting to get our heads around actions we can take, I noticed that in one of your

papers you describe four key factors – sometimes they're called motivators – that can really help organizations become more aware in this cloud distributed arena, can become more aware of what attacks may be on the horizon or worse yet, maybe even happening right now. So perhaps you could start by introducing those key factors and then we'll walk through them in more detail.

Derek Gabbard: Sure. The areas that we called out of the future of network analysis is, it's called AIRE – the Automation, Innovation, Reaction, and Expansion – those four key categories. As a preface to where we're going with it, some of the founders of the company, other founders of the company and myself, we were trying to come up with what's going to be the next security tool. Five years before there was a firewall, people didn't maybe even think there was a need for one. And so now they're ubiquitous.

And so let's spin this thing forward 10 or 15 years and figure out what's the next security device like that that's going to become ubiquitous. And we really thought that it was going to have to be something along the lines of network analysis but network analysis that gives you meaningful results around risks and around operational capabilities.

And so given all those other – the things that we talked about as far as the distributed nature of services around bandwidth expansion, significant bandwidth expansion, those type of issues that were going on at the same as you're trying to develop this tool we said, "These are the areas that if you could get it right, if you could create significant automation, if it would be an innovative approach to analysis that provides for significant capabilities around reaction, and then, hopefully, pre-reaction.

And then it will scale and can expand with wherever the network traffic levels go to in the future, that that would be a good basis for building this, the next network analysis or network security tool." So those are the four – automation, innovation, reaction, and expansion.

Julia Allen: Okay, well let's walk through these because I think there are some real interesting insights and groundwork that you lay in the analysis and the approach that you suggest. So let's start with automation. So clearly if we even have a prayer – I mean humans are obsolete when it comes to processing these kinds of data flows and traffic in the mix. We need some type of fairly sophisticated automation to deal with the rate and volume of data. So what kinds of advances or what kinds of potential opportunities or current opportunities do you see in the automation area?

Derek Gabbard: Well, hopefully there will continue to be significant advances in processing power. But that is a given I think. So we wanted to go to and create this capability for basically taking all the different types of data sets that you're going to be dealing with when it comes to the rest of these things, the rest of the analysis requirements. And to build a backend data fusion capability that takes these individual data sets and creates relationships to them, not necessarily on the fly – it takes a little bit of human intervention. But then once that's done, creates the

capability for significant and relatively quick ingestion of additional or new network data as it pours in to this tool.

So basically, the way that we've approached and the way I've seen others approach automation is to do it once manually. So how would we do this? And some of the colleagues from my company now were analysts in government agencies in the past and they spent time manually doing a lot of this work to basically create relationships to and try to glean information from the data. And to take that sort of manual – we did it this way a few times and this sort of worked out – and to be able to turn that into an automated system process to do that same amount of relationship creation and data digestion that it would take an individual analyst literally a couple of weeks.

Once you have that done a couple times and really understand it – to create it so that it is essentially done in a matter of a couple seconds when the system does it because it does it the same way every time and quite quickly – then you're moving down the path toward automation.

Julia Allen: Right. And as you said, it has to be the kind of an approach that allows you to continually take in more and more increasing levels of data, volume of data, types of data, do synthesis, do analysis, and present some kind of a more abstracted or a more summary picture for humans to digest, because – and like you said, to have that be repeatable, right?

Derek Gabbard: Yeah. That's been the, I think, the area that has been the most beneficial of the work that we've been a part of to this point. And that's that we approach the problem from two different perspectives. One is from a very, very knowledgeable network analyst who knows the protocols inside and out and really just wants the system to crunch the data in a way that he would do, he or she would do, if he or she had the time.

And then on the flip side of that, we wanted to approach the output of this automation to be something that is stoplight. Basically management; capable of being easily understood and quickly digested by senior management in the organization. So red, yellow, green. So we show network traffic and the analysis of that traffic across these areas to be red, yellow, and green. And so there's a significant amount of processing and thresholding that has to go on below that. But if you can get it down to be able to be distilled to that easy of a stoplight picture for senior executives – that don't have the time or the energy to spend doing significant amounts of analysis. But at the same time, give the analyst the ability to do that detailed level of analysis on the same data with sort of the same results. One being really very quick and easy and very management focused and one being with a lot more significant technical detail output, then the automation part of what you're doing, if it serves both, is working well.

Julia Allen: Okay. So let's talk a little bit about innovation. Based on working in this field for a while, are you seeing some potential breakthrough ideas that will help us get greater innovation around this network and data analysis issue?

Derek Gabbard: Well, I would say that I am and I would say that a lot of it centers around visualization and presentation. And so that – the data sets are pretty much the data sets. They're going to change over time. There's been a lot of, I would say, very good work done in correlation – so correlation engines to take a vast amount of data and spit out one event with a lot of different data that drives that event. There's been some significant advances in the amount of data that packet capture devices or network forensics devices can handle at one time. And those things are all – they're incremental changes and they're very much needed.

But I think that the breakthrough that's waiting to happen and maybe is starting to happen, really centers around the ability for the tool or a set of tools to provide that quick and meaningful visually appealing and easily understood actionable information to a variety of different levels of user. And it's a really – it's kind of an interesting time to be in that particular part of the field, because there's – if you look at the tools – the tools that were built to do this analysis were, for the most part, built to do enterprise-level network analysis.

So they were looking at internal network protocols. They were looking at relatively finite amounts of data and relatively – even the biggest enterprises – and relatively saturated pipes. You're talking about decent amounts of data but nothing huge. And so now to take that and to put it onto the carrier class or internet backbone for analysis and visualization has really sort of turned things on their ear as far as the old tools. And has required a new set of visual front ends to present that – a lot of times the same information. But when you're looking at the federal government space, if you're sitting in DHS in US CERT and you're looking at Einstein data from potentially 100 agencies, that's a lot different as far as size and complexity to visualize than it was to look at a university's network or even a decent size enterprise network in the past.

And so there's been a lot of research lately, funded by the government and in the private sector, to really try to tackle that problem and take those vast amounts of data and to show it in innovative and enticing ways.

Part 3: Real-Time Reaction; Expanding to Track New Attacks

Julia Allen: Right. Well, it seems to me, just given the dimensions of scale and complexity of what we're talking about, that kind of visualization, presentation in a human, understandable, meaningful way kind of leads us to the next key factor, which is being able to analyze and react in real time, right? You've got to have some kind of an accurate, close to real time picture of what's happening right now before you can take action, right?

Derek Gabbard: Yeah. I mean, there's two reasons to collect this data. And one is forensic and historical. And you have your – you have the ability to analyze that at your leisure, depending on how you define "at your leisure." But it's certainly not time sensitive and time critical. And so there's that mentality, which fine tooth comb and really drive down into the data deep and find the needle in the needle stack as the common saying is now.

But the other side of it is to try to take real-time actionable data from that sea of information that US CERT or other large organizations are pulling in on a daily basis, or an hourly or every minute basis. And so if you got the automation down and if you have the visual presentation down, then you start to be able to get to, like you said, some semblance of the ability to quickly and accurately identify at least the places where the analyst needs to spend their critical time. This is something that needs immediate attention where they need to spend that time. As opposed to sort of the older way of "there's a ton of information and I just try to use my best judgment to figure out where to start looking." So the first two definitely drive, the automation and innovation definitely reaction.

Julia Allen: Sure, sure. I mean that just all seems to fit very well in terms of interdependencies. So let's go ahead and get to the last piece, which is maybe a little bit more forward looking. If you have that kind of foundation to draw from, what do you think are some examples where the attack community might be expanding into new areas that we could use this automation, innovation, reaction approach to help analysts kind of capitalize on that learning and maybe even have a chance of hopefully keeping up and maybe even get a little bit in front?

Derek Gabbard: Yeah, it's an interesting question, because in the constant struggle between the attackers and the defenders, it seems like the defenders are always playing catch up. And that's going to be a tough nut to crack. Because as you move into becoming more forward looking, you open yourself up to the attacker's changing methodology and changing attack patterns and changing attack vectors and whatnot. But the mindset behind creating this need for and definition of expansion is that what is considered baseline today will be nowhere near baseline in five years.

Julia Allen: Right. In fact, it probably could be – it would probably be obsolete, right?

Derek Gabbard: Right. I mean, at the very least, it will be woefully under gunned and at the worst, it would be completely obsolete. And so the mindset toward modularity and toward expandability is really what we're trying to call out here so that as new – as there's a requirement for a change in either the automation or the visualization, that you're not going back to square one and starting to try to basically create your first two – the automation and innovation, visualization primarily – from scratch. So that there's tweaks that happen to the system or to the capability that allow you to stay current. And current being for size and speed and scale. And current also being for when it comes to the way things have evolved from a security or other types of network operations requirements.

An example is there may be changes to the core routing protocols of the internet in the not too distant future, with secure BGP (Border Gateway Protocol) that's being funded by the U.S. government. If that happens, then that may substantially change tools that were developed to do BGP analysis. And so if you have to go back to the drawing board and start over, then you haven't really positioned the tool for expansion. Same thing for secure DNS (Domain Name System). Those protocols are going to change.

And then the attack vectors obviously have changed over time as well. And so something that gives you the ability to change both size and speed and scale and the guts of the analysis, the guts of what you're looking for, would be a homerun as far as a security tool goes. Because it would help you keep ahead of all of the different ways that the playing field can change, whether the field itself changes or the players change.

Julia Allen: Well, Derek, I know we've barely scratched the surface here, and I sure appreciate kind of the thought provoking nature of some of the challenges that you've laid out for us as a community. Do you have some places where our listeners can learn more on the subject?

Derek Gabbard: Well, we have some white papers at our website, which will be called out in the show notes, that talk directly about this type of capability. And some of our other sort of forward looking, what does a security analyst need to be thinking about today and tomorrow in a bunch of different areas.

And I would recommend, just like any security analyst would probably recommend, to stay plugged in with the communities like Black Hat and DEFCON to really understand the current status and to try to see where things are going. You can learn a lot just by listening and trying to keep up with the things that are interesting to the hacker or white hat, black hat, gray hat, whatever, communities whenever you can. It certainly has helped us as we've worked on our tools to try to get ahead of this stuff, to know what they think of it and where they're going as a community as well.

Julia Allen: Well Derek, I so appreciate your time today, your expertise, the hard and challenging work that you and your colleagues, both in your own firm and in the community at large are doing. So thanks so much for your time today.

Derek Gabbard: Thanks, Julia.