Improving the Common Vulnerability Scoring System

Featuring Art Manion, Deana Shick, and Jonathan Spring

Welcome to the SEI Podcast Series, a production of the Carnegie Mellon University Software Engineering Institute. The SEI is a federally funded research and development center sponsored by the U.S. Department of Defense. A transcript of today's podcast is posted on the SEI website at sei.cmu.edu/podcasts.

Deana Shick: Hi, my name is Deana Shick. I am a threat analyst on the Vulnerability Analysis team within the CERT Coordination Center. Today, I am joined by two of my colleagues, Art Manion and Jonathan Spring. We are going to be talking to you today about improving the Common Vulnerability Scoring System, sometimes called CVSS for short.

Welcome, Art and Jono.

Art Manion: Thanks, Deana.

Jonathan Spring: Thank you.

Deana: Let's start off by just telling everybody by what we do, and I'll start. My name is Deana Shick. I've been a member of the technical staff here at CERT for the last five years. I started off doing threat intelligence analysis and have transitioned onto the Vulnerability Analysis team under Art Manion. I typically focus on vulnerability response processes and understanding the entire ecosystem of some of the things that we are going to be talking about today. Art?

Art: I am currently the technical manager of the Vulnerability Analysis team. I have been here almost 18 years, lots of coordinated vulnerability disclosure. Part of that is always, *How bad is this vulnerability?* We're counting, like, 20,000 a year, and CVSS is sort of at least notionally designed to help answer that question. This work in CVSS is an output of all of this coordination work over the years. Anyway, that is what I focus on, and CVSS is one of our hot topics right now.

Jonathan: I started my work at the SEI in 2009. I was doing largely network forensic stuff, lot of DNS [Domain Name System]. In addition to the information security background that I have, I studied some philosophy of science stuff, and what that comes down to is essentially evaluating the quality of evidence and how to make sure that you are making good decisions based on what evidence you have. I think that to some extent I have applied that to a lot of different situations: threat analysis, network analysis, vulnerability analysis. I think that that is a clear part of talking about CVSS and trying to evaluate what kind of evidence it gives you and what sort of evidence you need to make a good decision there. Or whether you are making decisions, I suppose, is one of the things we'll be talking about. But, yes, that is how my background plays in here.

Deana: Awesome. Could you guys give me a little bit of a history of CVSS?

Art: I had to look it up to be sure, but it goes back to 2005. Version 1 came out in 2005. Shortly after that, the start of CVSS, it was transitioned into the organization called FIRST, which is the Forum of Incident Response and Security Teams. Within FIRST, there are special interest groups. CVSS is operated and developed out of a special interest group within FIRST. So, 2005, version 1 moved into FIRST. 2007 is when version 2 came out. There's a longer period of time. 2015, version 3 came out.

There have been sort of incremental changes to how CVSS works at these different revisions, but fundamentally it has always been the same construct, so to speak. Actually, very recently, 3.1 just came out in 2019 with some minor additions and changes. That's sort of the history. It's been around for a while. It's been adopted in a number of places. Notably, the NIST National Vulnerability Database uses CVSS. A lot of U.S. government and other folks rely on this score for their vulnerability prioritization purposes, which is what we are going to get into a little bit further.

Deana: Can either of you describe how CVSS works?

Jonathan: Well, I can try. CVSS asks for a number of qualitative inputs. It is trying to calculate the severity of some particular technical vulnerability to the technical system. I think depending on exactly how you think about it, the goal is either to output a number between 0 and 10 or to output a category: low, medium, high, or critical. The way that you get to those numbers—because you get to the numbers—is that you answer 10 or so questions. There are some modifications based on whether you are doing what they call an environmental score, but let's talk about the base score. It asks some questions such as, *What's the confidentiality impact on the system or the information on the system from this flaw? Can they exploit it?* You can say *low*, *medium*, or *high*. You do that several times for these different questions.

Deana: So, that is just the base score. Can you go into what the temporal and environmental metrics of CVSS might be?

Art: Sure. To Jono's point, there are these vectors or these questions that are asked. There is a small handful for the base. Temporal adds some additional questions or vectors. Temporal are more related to things that change over time, so I think something about confidence in the report might be part of temporal exploit information, maybe threat information is about temporal. There is a third section, environmental, which in version 3 is basically, *take the base questions and re-answer them for your specific environment*. There is a big change between version 2 and 3 in environmental. We still score version 2 at CERT/CC because the version 2 environmental better suits how we are trying to use CVSS.

Deana: Which organizations are using CVSS and why? Are there different communities of groups that are using CVSS differently?

Art: Well, I had mentioned we have the U.S. government use, right? So NIST NVD [National Vulnerabilities Database] is putting out CVSS scores. There are a lot of users keying off of the NVD's scores. A number of software vendors also provide scores, which sometimes agree with NIST and sometimes do not, which is a fun area of conversation as well. Other groups? I am not sure we are aware of who else might be using it.

Jonathan: The payment card folks. So, the payment card industry, PCI, they have a security standard for making sure, basically, that you're allowed to process credit card information. They attach requirements to how quickly you respond to vulnerabilities based on the CVSS score to maintain a certification to process payment card information. Different people have different amounts of requirement for those things. Deana, you probably know some of the people who are in the SIG for what sort of groups represented there.

Deana: The CVSS Special Interest Group is who modifies and updates the standard. There are several patch developers and patch appliers that join together to create what we understand as CVSS. We actually just wrote a paper—our team just wrote a paper entitled *Towards Improving CVSS*—that we are going to talk about here shortly. With that, let's talk about a couple of the drawbacks of CVSS. What are some problems that we have noted in our paper that you want to discuss? What are some real-word impacts of those problems?

Art: We've covered this already a little bit. The mandated or nearly mandated use by federal government, by the payment card industry sort of has this implication, if not a requirement, to use it. There is this message that the base score, that number—actually, it's reduced to the category, the *high*, *medium*, *low*, *critical*—that is sufficient for your prioritization or your risk or your severity decision, and you're done. It came out as a 7.7, which is a high maybe or a

medium. I'm not sure. It's high. You can stop thinking and process it according to the rules. That's great. It's very simple, but it's not realistic whatsoever. Whether it's mandated or implied that this is a good methodology, that is one of the general problems that I see. The CVSS-SIG is very careful, especially in the 3.1 new language, to say that CVSS is not risk. You are supposed to bring more input to the table before you decide to take action on a vulnerability based on a CVSS score alone. However, the decade-plus of use has resulted in this almost misapplication of it as the only score you would ever need.

Deana: Do you have anything you would like to add, Jono?

Jonathan: Well, Deana, as you know, there is a whole bunch of things that you could get into. I think that it is very important, this idea that the CVSS-SIG is very clear that they want to have this be about severity and not about vulnerability prioritization. If people do, in fact, want vulnerability prioritization, then I think that is one of the things that would need to be addressed with improving CVSS. There are some other pretty important things about just the way that the scoring is done, which we talk about a fair amount in the paper.

Without getting into too many of the details, which I think is hard to follow listening, but I would encourage people to look at the paper. Basically they suggest that they have done this scoring and that they got some experts into a room, rank-ordered some vulnerabilities by which ones were most severe, and then made some sort of regression equation that matched that. Now you plug in the numbers, and then it does that.

At the very least, we don't actually have very many details on how those people were selected, how those vulnerabilities were selected, and how sure we are that they did a good job basically. They may well have done. I think that in general they are smart people, so they probably did, but we don't have any access to that, and I think that's sort of a problem.

The second thing—which Art alluded to when he said the decisions are reduced to the categories—in the document where the CVSS-SIG says what they are doing, they say that the scores were to be accurate to within 0.5 units of when they did the equation. Which means that if you get a 7, according to the SIG's own documents, a 7 is *medium* because it is the border line between *high* and *medium*. But if it is accurate within 0.5, that means it's actually somewhere between 6.5 and 7.5, which means that it's a 50/50 chance of being *high* or *medium*, and that's not super helpful if what you want to know is the category. I think that there are some real questions about how those categories get assigned out of the numbers and that sort of thing, which seems difficult.

Deana: A little bit of my experience of being a member of the CVSS-SIG, since our team has been working within the SIG, to go over some of the things in the paper, is that CVSS does not

necessarily do a good job of scoring severity for things that are not traditional IT systems, so like, safety-critical systems, some of the things that are going on, some of the discussions that are happening within the FDA. CVSS can't be applied to those medical devices. Are there any other sectors that you see that can't use CVSS appropriately?

Art: You've covered medical devices, but the medial device sector has been voicing concerns that CVSS is not sufficient for their use. The industrial control system sector, same story. One of the arguments I hear a lot is there is this discussion of confidentiality, availability, integrity—got the wrong order—of information. But what about safety, right? The medical devices, control systems we are talking about there are physical safety concerns. Does CVSS account sufficiently for that? That is one of the questions both of those sectors have raised up. There has definitely been work in the medical device space to try to adapt CVSS for use in the medical space. MITRE has some work in that area. There has been lots of concern expressed by the ICS community, but I'm not sure of any specific work on a variant or a new standard or something else for them.

Deana: We've been talking a little bit about our paper that we wrote, but we have also been trying to be a little bit more forward-looking within our team. Can CVSS be fixed, and if so, what measures are we recommending to fix that problem?

Jonathan: Well, Deana, whether it can be fixed partly depends on what people want it to be, but can we make an adequate vulnerability prioritization scheme? I think so. I'm not sure whether it will be CVSS or not. Could be. I think that the start of that, as we have been discussing, is that if we want to model decisions, we should model decisions and not severity.

Art: We are SIG members. You mentioned this. Even when we wrote our paper somewhat critical of CVSS. We gave the SIG some advanced warning and circulated some drafts with them for comments. The question as to whether what we work on is something that happens or ends up working out within the SIG or not is open, I think, at this point. I just mentioned that they just released 3.1. There is now discussion in the SIG of the next period of work to take on, aiming at a 4.0 release, and that may be a significant amount of work. One question we will have and ask within the SIG is, *How much appetite or willingness is there to make significant changes in CVSS?* I don't think we have the answer, but we are going to investigate within the SIG, and we will be working on our decision modeling exercise within or without the SIG.

Art: What do you think, Deana?

Deana: I think it's good that the SIG is willing to understand how CVSS is being applied. We are finding that in a lot of different organizations, instead of CVSS being used only as a measure of severity, it's being used as a prioritization scheme. That comes up especially in the PCI

discussions, that if you have a particular vulnerability with a particular severity, you have to do something with it or not. It is based completely off of a score. As we know, sometimes when given just a score from 1 to 10, it sometimes makes decisions a little bit too easy for people. So, can you go into a little bit more about decision modeling and how that would help alleviate some of these problems that we have been discussing?

Jonathan: I think that one of the things is picking the right pieces of information to make the decision based on. We have a hypothesis that, for example, whether or not something is being exploited has a lot of information in it, and that is an important thing to use. I think that this is actually supported by some work at WEIS, the Workshop on the Economics of Information Security, that actually Sasha Romanosky and some of his colleagues put out doing some other modeling. I don't know whether that was in response to our paper in January or not, but it seems supported by some of that. The gist of it is, I think, that what we are going to suggest is that instead of going from categorical answers to some questions like, *Is it being exploited?* to numbers back to a category, we are going to go from category to decision with a model called a decision tree.

We are going to try to keep them small enough that everyone can understand the whole tree and why each branch and each decision is there the way that it is there with some clear, argued supporting evidence, which is what we are working on: producing, a very comprehensive, clear, transparent set of, *We think that you should decide on safety first and then decide on whether it's being exploited. Then, if those two things are really bad, then you need to fix it now, and that's all you really need to know. Or, maybe you need to learn a little bit more about the deployment posture or how bad or maybe how severe it is*.

But we can plot that decision out in a transparent way that is organized in such a way that is intelligible and explainable to a person. I think that that is part of where it really needs to come in, because if you are going to make a decision, I think you need to be able to explain why.

Deana: The decisions are different, right? If you're a patch developer or patch applier or if you are dealing with a safety-critical system.

Art: We are in the midst of an internal experiment to test out our initial thoughts on these decision trees from different points of view. Roughly, we plan on publishing our findings when we're ready, so there should be another paper in our series.

Jonathan: It is going to be a lot of work to test all of this out with the appropriate level of care. Then, of course, how the entire software community is prioritizing which vulnerabilities to develop patches for and apply patches for is a fairly big of policy. I think that we should probably have appropriate available evidence for it.

Art: Agreed.

Deana: Thank you two for joining us here today. I know that this is a passionate topic for all three of us that we like to discuss at different conferences with different folks. Hopefully, we can continue that conversation. Art, Jono, myself, along with Allen Householder and Eric Hatleback, have published a white paper, *Towards Improving CVSS,* and a blog post, which is on the CERT/CC blog. We will include all links to all resources mentioned in this podcast in our transcript. Again, thank you for joining us.

Jonathan: Thanks.

Art: Thanks, Deana. Thanks, Jono.

Deana: Yes. Thanks, Art.

Thanks for joining us. This episode is available where you download podcasts, including SoundCloud, Stitcher, TuneIn Radio, Google Podcasts, and Apple Podcasts. It is also available on the SEI website, sei.cmu.edu/podcasts, and the SEI's YouTube channel.