Carnegie Mellon University
Software Engineering Institute

SEI Podcasts

Conversations in Software Engineering

A Penetration-Testing Findings Repository

featuring Marisa Midler and Samantha Chaves as Interviewed by Suzanne Miller

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.

Suzanne Miller: Welcome to the SEI Podcast Series, a production of the Carnegie Mellon University Software Engineering Institute. My name is Suzanne Miller, and I am a principal researcher in the SEI Software Solutions Division. Today, I am happy to be joined by Marisa Midler and Samantha Chaves, both of whom work as penetration testers in the SEI CERT Division. Today, we are here to talk about their work on a penetration-testing repository, something that I know some of the people in this field are going to be very appreciative of.

Welcome to you both.

Marisa Midler: Thank you.

Samantha Chaves: Thank you.

Suzanne: Marisa, you have done a few podcasts with us previously including one for our My Story in Computing series that we have recorded. We are going to include a link to that or go to the web and search on Marisa Midler if you have not seen it already. But Samantha has never been with us before. So Samantha, I am going to let you start to tell our audience a little bit about yourself, what brought you to this SEI team, and what is it that is so cool about the work that you do? Samantha, go ahead..

Samantha: Sure. I have a little bit of a crazy background, but I have been with the SEI for almost three years now. I started as an intern straight from Florida State. I graduated with my masters in cybersecurity. I interned first on the Monitoring and Response team and then navigated into the Applied Network Defense [AND] team. I have been full-time with AND for a year and a half now. I was fortunate enough to get put on this project with Marisa, and it is definitely, like you said, something that is going to be very useful for the cybersecurity industry. But I would say that the most interesting thing about the job is one, there are so many projects. It is not just focusing on one thing. There are so many tools between coding, assessments, and all that stuff. I get a feel for a little bit of that, which is very nice. Some people, they do not have the opportunity to be on so many projects. I would definitely say that is a cool thing. It is always something different every day.

Suzanne: I tell people that if you are bored at the SEI, it is your own fault because there are so many problems to solve, so many different ways to go about it that there is no excuse for being bored. You are definitely reinforcing that perspective.

For those that have not seen your My Story in Computing, Marisa, give us a little summary of your background and what brought you here, and what you think is cool about working here.

Marisa: Oh, sure. Yes, I am just going to give a very abridged version of it. You can see the full My Story in Computing if you are interested in learning more about me. I went to grad school in the Heinz Information Security Policy and Management graduate program. I started interning at the SEI, and including my internships, I have been here for probably about five years now. Like Sam said, I really like all the different projects that we get to work on. There is definitely a variety, and also echo what you said: If you are bored here, that is probably your own fault because there are plenty of projects to hop on if you are looking for them.

Suzanne: Yes, there are. I have been here for over 20 years, and I am still not bored. So there you go. OK, today, we are here to talk about a penetration-testing repository that you recently built. You have described this as a source for active directory, phishing, mobile technology, systems,services, web applications, and wireless-technology weaknesses that could be discovered during a penetration test. This is essentially giving, I mean, in some ways, I look at it as it is almost like a very expanded checklist for people of, *Have you seen this? Have you seen this? Because you might*. Is that a quick summary of what you have done with this repository?

Marisa: Not quite. The repository is built out for different assessments. You have your standard penetration-testing assessments where active directory, phishing, system or service weakness, and web-application weaknesses; those categories would fit well into just your standardized penetration test. But then the mobile assessments and the wireless assessments, those are not always included on penetration-testing assessments. They are typically their own type of assessments. We built out common findings that assessors will find on those types of assessments as well. It includes a little bit of everything, to be honest, and it is meant to help the assessor give standardized language overall to the assessed organization. Ultimately, we are just trying to standardize these finding names and help out the assessors, save them time on report generation by having these descriptions and standard remediations and everything else that is in the repository in there for them to use.

Suzanne: The question comes, why is it important to do that? What is the cyber landscape like right now that it makes it important to have that kind of standardized view of findings from penetration testing?

Marisa: I would say it is important because right now, everybody is naming these things whatever they are coming up with on assessments, which is not wrong. It is just that if we are wanting to try to do some data analysis on assessments overall, a vast majority of these assessments, standardizing these naming conventions is very important to identify which findings are identified on all these different types of assessments.

Suzanne: Data analysis and then probably also in terms of metrics reporting. *How many of this type? How many of that type?* If you do not have any kind of standardized nomenclature then you are not going to be able to provide any of those standardized measures. Is that one of the intentions of the repository to support that?

Marisa: It depends on how the organization uses it. The repository itself is just a tool for people to use. Right now, the format is an Excel document. And people can ingest that into a report web application like a report-generation web application and use it that way. Or, they can use it as an Excel file and just use it as a reference. I think Sam had something that she wanted to say.

Samantha: Yes. I was just going to bounce off of what Marisa said because you did ask why is it so important to have standardization. When we did our data analysis, and we were actually looking at all of the findings that we had originally, we realized that assessors on these assessments worded the findings in different ways. Some people would find a specific finding, and the name of it would be a process or it would be an attack rather than the actual vulnerability that is being found on these assessments.

When Marisa and I were looking at them, we said, *OK, maybe we can take all these 10 attacks that are in the original lists and actually map to one specific vulnerability*. So essentially creating the standardization amongst all the findings to pinpoint the exact vulnerability that is being found on the assessments rather than the processes that we found originally in the flat list.

Suzanne: You have given us a little bit about it is basically an Excel spreadsheet as the sort of foundational form of the repository. But [please] give us a little bit more of an overview of the repository, how it works, and how you actually built it?

Marisa: The findings repository is a tiered classification system. We have those six categories that you mentioned earlier, which are active directory, phishing, mobile technology, system or service weakness, web-application weakness, and wireless-technology weakness. That is layer one. Then within those layers, we have general findings, which are exactly what they sound like, a more generalized finding. Then the next layer down is specific findings, which are more specific findings that pinpoint one particular type of finding, and these layers—without seeing the actual repository, it is hard to describe it verbally—but there are 27 general findings, and then there are 111 specific findings.

This covers a lot. It is not all-encompassing, but the idea is that the community, since this is open source, they can look at this, and add to it. We will want to add to it too as there are new findings and vulnerabilities that are identified as the years progress. But ultimately, it is a tiered system. Assessors, when they are looking through this thing, if they are like, *Oh, I found something with active directory*, they are going to go into the active-directory category and then look through the general findings. They might see something that sounds like generalized what they are looking for and then look within the specific findings to see if there is something specific to their finding. If not, they can use the generalized finding.

Suzanne: [Do you have] a mechanism for people to give back to you new findings because that is one of the things about this whole area is the threats change. So, do you have a way for people to bring back, *Hey, I think I found a new finding for the repository; other people are also seeing this?*

Marisa: Right now, the findings repository is hosted on CISA's GitHub. I do not know if CISA [Cybersecurity and Infrastructure Agency] has opened up that functionality but I do believe that has been the plan. So off the top of my head, I do not know.

Suzanne: OK, that is fine. But people should look at CISA's GitHub; we will give the GitHub website in the transcript, of course. That is where they would look to see if they wanted to provide some feedback as to other things that they are finding.

Marisa: Absolutely, yes.

Suzanne: I can certainly imagine if I am a new penetration tester that this is going to be a godsend for me because it helps me to organize my thinking, it helps me to look at things. But what are some other use cases besides the beginner penetration tester that you see this repository being used for?

Samantha: I do fall into that beginner level, so it is deviating from your question a little bit. But during the time that I was actually working on this with Marisa, I went on my first pen test. As a beginner coming out of school and never actually conducting an assessment, you are kind of like, *Well, what am I supposed to be looking for? Or what are a list of common vulnerabilities that you do find on assessments?*

While I was working on that and got placed on the assessment, it was a little bit easier as far as pinpointing, *OK, what am I looking for? How can I remediate this or find that standardized language to put into a report for an assessment?* So, that helped a lot.

Another pro or another useful way to use this is just for report writing in general. When you are on these assessments, assessors spend so much time trying to figure out, *How do I word this specific finding?* Or, *What are some resources that, whether those are blog posts, tutorials, how is it going to help me* *mitigate this vulnerability?* Essentially what the repository is supposed to help with is to minimize that effort for report writing and trying to figure out spelling and grammar and all of that stuff. That is probably one of the biggest usages of the repository for people on an assessment.

Marisa: I can add to that. With the report-writing aspect, the reports—what the customer actually gets their hands on—and with those additional resources of how to remediate it, it is also helping give the assessed organization better resources to fix these findings or vulnerabilities.

Suzanne: So in the repository, you not only have the vulnerability but you have suggested resources to look at for that. I love this because so many times in the past, the SEI has been accused of, *Well, you tell us where all the problems are but you do not tell us how to fix them*. This is a counter case where you actually are helping people to understand, *How am I going to fix this?* I know that our audience will not think that it is a be-all end-all, but it is a place to start when you are looking at those kinds of vulnerabilities.

Samantha: It was just going to say [that] there are several things that are within the repository. You have the description. You have your remediation, your references and resources. In addition to that, we actually have controls from cybersecurity frameworks that maybe Marisa would like to elaborate on. There are so many resources and things that you can find from this repository, not just your vulnerabilities, because like you said, there are people like, *You just tell me what is wrong but how can I fix it or what are some resources*? Those are pretty much all the components that go into the general findings and the specific findings of the repository.

Suzanne: Marisa, did you want to say anything about the standards?

Marisa: Yes. I can give you a high-level overview of what is in there. We have the NIST Cybersecurity Framework. We mapped findings to that control framework. Version 1.1 because when we did this work, that was the most up-to-date version. We will eventually have to go back in there and update it to Version 2 that recently came out. We also have a map to the NIST Special Publication 800-53, Revision 5. We also mapped it to the Center for Internet Security, Critical Security Controls, Version 8. Each finding has mappings to those three frameworks.

Suzanne: This is a very rich resource. I think that is really one of the reasons we are talking today is to make sure that people are aware of this kind of a rich resource that is going to help assessors that are doing penetration testing find and fix things that are worthy. The question that comes up in my mind, it's a little bit tangential, but how do you decide when you are out on an assessment what is worthy of being included as a finding. Does the repository help you figure out what the is that level of abstraction that I should be looking for? I can imagine, just for myself, that would be the hardest thing for me, *Is this worth talking about? Is this too fine a detail?* How does the repository help with that, if at all?

Marisa: Yes, it does. The specific findings are all commonly identified findings that we thought were important to include to be reported on when a pen tester is out there doing an assessment. But just as a general rule, what you report to a customer, in my opinion, is…I try to put myself in the customer's shoes and if I was an organization, what would I want to know is a risk to me and my organization? If I think that this organization would want to know about that, then I would report on it. It is kind of a judgment call.

Suzanne: That is a nice rule of thumb. So, thank you. Sam, I am just curious because you said you were in that newer to pen-testing mode, is that one of the things that you found helpful with the repository, that you had a better idea of what is the right level of abstraction to be dealing with on the findings?

Samantha: Absolutely. I think too, there is a column within the repository that is a severity level. That was determined by our customer CISA. I got a feel of, *OK, which vulnerability is more informational? Which one has a severity level of low or high?* So you saw stuff deep in the weeds and then you had some general ones. But the repository has all various levels of types of vulnerabilities. I think that definitely helped a lot, and like Marisa said, there is probably still more that we can add into the findings repository. But essentially, what we did create was pretty much the most common ones that assessors would find on assessments. But the severity levels definitely helped with that.

Marisa: I just want to jump in with that severity level. That is just like a starting point. It is not the end-all-be-all level. It can be subjective based on the assessment and what type of environment an organization has set up. So it could be raised to be more severe or even lowered.

Suzanne: Got you. Where have you beta-tested this? Have you had any organizations use this and give you feedback on it? If so, we always like to know not only how does it work well, but are there any drawbacks that people should be aware of if they are going to try and use this?

Marisa: We know that it is in use, but we did not really have a beta test, per se. So it is being utilized. This started from a flat list of findings that was definitely used for years. That was the beta test where we got feedback to build out this more extensive findings repository. This particular one, we are still looking for feedback on it, but we did build it from something that was more of a flat list that was used and tested for years.

Suzanne: OK, all right. So it has a strong base in terms of the content of it is well known. Really what you are doing, that standardization piece of language and the layering is really what you are adding. Then all the resources, I am imagining the flat list did not really have all the resources that you are adding to the repository now that help people fix things. It sounds like it was a big improvement to the original product that you were using.

Marisa: Yes, and this one is open source, so we are able to share it with everyone. The original was not.

Suzanne: Ah, OK. All right, that is a good point too. This takes me actually into the topic of transition. You have an open-source product so anybody can get to it. But beyond getting to the repository and playing in the sandbox, what kinds of resources are available to people from us or from CISA to help them understand how to use it, how to do some of the things you were talking about earlier like bringing it into organizational reporting and things like that? Or do we have resources for people that support that kind of transition?

Marisa: I am not aware of any right now. We are building out blog posts to help introduce people to the findings repository. This podcast we are recording will also introduce people to the findings repository. But if there is interest, I am sure we can come up with something else for the public to use to help them take advantage of this great work product.

Suzanne: I will interpret that as a call for, *If you see this, then you think you need some help with it, contact us, we might have something already built or we may have something that we can build as part of the open-source community for this*. So you are at the beginning of the transition of this really. This is a new product, and so, you are at the beginning of figuring out what people need to be able to use this productively.

Samantha: If I can add to that. Even though we may not have resources out there, there are still ways. Let's say you have an application, because we currently use the repository and ingest it into an application for good usability for front end, and stuff like that. It is easily ingestible. If you do have this application, you can select your category. You can select your general finding, then you can select your specific finding. Even though we may not have anything out there right now, it is still good to use if you have an application that maybe you use for assessments and stuff like that.

Suzanne: It is adaptable is what it sounds like to different environments. OK. Well, that is always a good start because adaptation is something that we need to make sure our technologies can support. So, you have the first version of this done. In the interest of not being bored, what is next for you two? I know you have a blog post for this in the works. What else are you working on that we might want to bring you back to talk about in a little while?

Samantha: I am actually currently working on a risk repository. Essentially, instead of honing down into the vulnerabilities that you find on the assessments, now this risk repository can focus on all of the risks that you can find. Same thing, it is going to have a description. It is going to have a recommendation. I am not sure if it is going to have cybersecurity controls. But definitely those are being used to help write these descriptions and recommendations and stuff like that. That is in the works. I do not know when it will be publicly released or if it will be, but that is something that is being done in the backend.

Suzanne: What about you, Marisa? What are you working on?

Marisa: I am building some security tools, but none of these have any plans for being open source. So, unfortunately, I cannot really talk about them.

Suzanne: Well, not all of them can be. All right, well, I do want to thank both of you for talking with us today. As everyone knows, we will include links in the transcript to things we have talked about in the podcast. I do look forward to seeing future things from the two of you. I think you are both the type of people that are going to have your fingers in lots of different pies as time goes on. So, that will be fun.

I do finally want to make a reminder to our audience that our podcasts are available lots of places: SoundCloud, Stitcher, Apple, Google, and of course my favorite, the SEI YouTube channel. If you like what you see and hear today, you are welcome to give us a thumbs up and tell your friends. I want to thank you both again for joining us today, and I want to thank our audience as well.

Marisa: Thank you, Suz.

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 at sei.cmu.edu/podcasts and the SEI's YouTube channel. This copyrighted work is made available through the Software Engineering Institute, a federally funded research and development center sponsored by the U.S. Department of Defense. For more information about the SEI and this work, please visit www.sei.cmu.edu. As always, if you have any questions, please do not hesitate to email us at info@sei.cmu.edu.