Building a Malicious Code Analysis Capability
Transcript

Part 1: The Malware Arms Race

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 senior researcher at CERT, working on operational resilience and software assurance.

Today I'm pleased to welcome Jeff Gennari. Jeff is a member of CERT's Malicious Code Analysis team and today Jeff and I will be discussing techniques involved in analyzing malicious code and some considerations for building up an organizational capability to do this.

So, welcome Jeff. We're really glad to have you with us today.

Jeff Gennari: I'm glad to be here.

Julia Allen: So Jeff, what exactly is malicious code? I know in the literature it's often referred to as malware. What kinds of software are included in that category?

Jeff Gennari: So malicious code, or malware, is a piece of software that runs without the user's explicit consent, maybe without the user's knowledge. Typically it's used to conduct some illicit activities. Historically, you've seen nuisance-type things where it would just get in the way. But more recently, in the last decade or so, the focus has been on committing illicit activities, crime, identity theft, taking control of the computer to send spam or otherwise use it in that way.

So the words you typically hear to describe malware are computer virus, Trojan horse, rootkit, backdoor -- those types of things -- spyware, adware, and all that are subcategories of malware in general.

Julia Allen: Okay, and what are some of the things -- I mean, you started to mention a little bit about this -- but now that it's moved beyond nuisance value, what are some of the things that malware can actually do?

Jeff Gennari: Well, so it can harvest information. What I mean is a lot of commerce is done on computers nowadays -- networked computers. And malware is designed to steal the credentials you use to do that commerce, something like personal banking, credit card ordering online. Those are very common. There's a lot of money involved in that. And the adversary, the bad guy, will -- they acknowledge that's a big business. There's a lot of money to be made. So they write malware or create malware to take that information, to steal that information.

There's also a lot of money to be made in spam, sending out spam email. So what they'll do is they'll -- the bad guys will go out and take control of the computer, install malware on that computer, and that malware will then send spam out, those types of things. So I think a lot of motivation today is financial.

And there is a pretty -- there's other motivations too but I'd say that in terms of crime, the financial motivation is certainly big. So I think that's -- there's an entire underground economy, as we like to call it, that's built up around malicious code and selling malicious code. It's really become a commodity. And you're seeing the bad guys play in that space to make money. It is very lucrative for them.

Julia Allen: So would it be fair to say that that's probably what's really pushing the rise in malware, the sophistication of malware, why malware threats are on the rise, that just like the criminal element that we see in the physical domain and in today's normal crimes, that's moving into the Internet domain? Would that be a fair characterization?

Jeff Gennari: Yes, we talk about it really as an arms race. So there's a few different dimensions. One, yes, the money's there so that's bringing the criminal element. And too, the adversaries get better, and that cat and mouse game just continues on, just like in the physical world.

So you're seeing more sophistication because malware has to do more to evade detection and removal. You're seeing more diversity because more and more activities -- there's more opportunities to make money through the Internet and there's more opportunities to take advantage of that, for malware to take advantage of that.

For instance, the rise of online banking has provided a whole new avenue of, or whole new opportunity, to get personal information as more and more of this type of business is done by individuals online.

Julia Allen: And what about botnets as a class of malware? Can you say a little bit about how they're used and how they're being proliferated?

Jeff Gennari: Yes. Well, first off, a botnet is essentially just a collection of computers that are all running the same piece of malware. We call that “the bot.” And it's basically about coordination. So you'll have one person or a group of individuals we call the "bot-herders," who install bot software, which is malware, on any number of machines. Some of them are in the tens of thousands. Larger ones are even bigger than that.

And what they do is they coordinate activities among those machines. So they have a piece of control software that controls these bots and then they just have an army of computers that they could do whatever they want with. So they'll use that to send spam across the Internet, they make money on that. They will also use the botnet -- they'll lease that botnet for services. A popular service among them is denial of service.

So for a fee, you can lease their botnet or their service, which is the botnet, to go and attack a website or a site on the Internet to effectively take it offline by flooding it with traffic or some other denial of service technique.

Part 2: Prioritizing Malware; Analysis Techniques

Julia Allen: Boy, that's a great example. So you lay out this landscape for us and clearly the malware's on the rise. There's all kinds of different motivations for it. And so I suspect many organizations are faced with this but they can't tackle all the various pieces of malware.

So I know -- as I prepared for our conversation today -- I know in your Malware Analysis Apprenticeship course, you actually talk about criteria for selecting the highest priority malware for analysis and action. So I think it'd be helpful for our listeners if you could just walk through your thinking about how to prioritize malware for analysis.

Jeff Gennari: So our malware apprenticeship course is a five-day course that covers all the parts of malware analysis, from the beginning to end. It's meant to be a real primer for people that are interested in the space. And it came about because we work with a lot of people who are designing malware analysis capabilities and one of the questions they have, as you mentioned, is "How do I pick -- there's so much malware coming in. How do I choose what to pay attention to?"

And the best answer depends on the needs of the organization but beyond that we have a few criteria we use to decide what to pay attention to. So of course if you are tasked with analyzing malware, if your mission is analyzing malware, you'll pay attention to analyzing malware. So, for instance, the CERT Malicious Code team is tasked with analyzing malware. So the criteria for us is our mission. Our mission says we'll analyze the malware.

Some people assign numeric weights to it. We don't have an official scale but things you may pay attention to is how widespread is the infection. If you, for instance, can detect it on every machine in your network, you may pay more attention to it than something you aren't currently experiencing. So this gets into the number of reported incidents, antivirus hits, things like that.

You may look at the nature of the malware itself. So if it's something that spreads rather prolifically, something like an automatically spreading worm, you may pay more attention to that because the infection -- even if you're not currently infected, you may be infected quickly and then it may spread very quickly.

You may look at what the malware actually does in terms of its impact. So if it is something that is more of a nuisance piece of software -- maybe it pops up a window or otherwise doesn't damage the system -- that may take a less, that may be less important than something that, say, is looking for specific information to steal or otherwise damages the machine by deleting important files, like that.

Other things that may eat time or may factor into how you would prioritize what to look at would be how difficult it is to remove the malware, what kind of steps you need to do to remedy the situation. That could range from reimaging the machine or wiping the machine clean to deleting specific files. Or in some cases, even just rebooting.

And then some of it is organizational- specific things. So if you are recovering from an incident and you notice during that recovery that a piece of malware is looking at, again, sensitive data, you may spend more time analyzing that piece of malware because you want to understand the motive of the attacker, why they're looking for that data, the capabilities, what else they can do -- things like that.

Julia Allen: So do you find as you're doing malware analysis on the CERT team, do you actually have ways of tagging, categorizing, archiving -- a repository type of approach that you use as new pieces of malware come to your attention?

Jeff Gennari: Yeah. We collect lots of malware and we store it, for historical reference. So we're big advocates to understand what's going on you need to understand where you've been. So what we do is we collect lots of malware and we categorize it, put it together, and then we try to cut it in terms of malware families, so things that are very related to each other. We find this helps us expedite analysis because if we've seen it before or something that is similar to it, we can speed up the analysis process. Because then we're not rediscovering everything again; we're leveraging past insights. So storing malware, at least for historical record, is we think very important.

Julia Allen: Great. Great. Well, as I think about the stuff coming in, analyzing it, prioritizing it in some way, clearly there are a number of different analysis techniques. Again, I saw this in some of the materials you've provided.

So could you take our -- without going too deep into the technical details -- could you briefly describe the different analysis techniques that your team uses, just to bring everybody along the learning curve?

Jeff Gennari: We slice the malware analysis spectrum, if you will, roughly along three lines. We start with surface analysis, which is trying to get a sense of what you're facing. So the word I like to use is we want to characterize what we have. So this is looking at the statistics of a file.

Malware is delivered as an executable file or some other file, maybe an exploit, a malicious document of some type. Anyway, the artifact you typically see is some type of file that is the malware. You could run various processes, lightweight processes such MD5 hashing, even file size, in some cases file name.

That can give you insights, quick, easy to retrieve insights on, to answer questions like, "Have I seen this exact file before? What type of file is this? Is this a malicious document? Is it an executable? Is it a dynamic link library, or a DLL?" Those are data points that help you guide the rest of the analysis.

And again, they're very easy to attain right off the bat, especially compared to some other types of analysis. Now, they come at the cost of they don't tell you a whole lot about what the file actually does. The ideal situation is the file you receive is an exact copy of a file you already know about and then your analysis effort is pretty much zero. You know what it does if it's an exact copy.

But oftentimes you don't get that -- that's the lowest of the low- hanging fruit. Oftentimes you don't get that, so you need to progress and do a little bit more work. Typically what we call the second stage of this is runtime analysis. This is where you run the malware in some controlled environment and observe what it does to a machine. This will tell you information about network activity, the impact it could have on the system, things like that.

Now, there's some clear drawbacks to this approach. Number one is you have to be properly instrumented to gain the insights. So the runtime analysis is really tied to the quality of tools you're using to do the analysis. So if you have a good environment, something that is instrumented to monitor file system activity, if you're running on Windows, registry activity, network activity, then you'll get some decent granularity. You'll get some decent information back that'll give you insights into what the malware's actually doing or maybe indicators to detect the malware.

For instance, if it's talking to the network and you have sensors in place, you can go and look at your sensors, assuming the sensors are recording the right kind of data, and see if that malware is active on your network.
A second drawback is the quality of the runtime analysis depends on what the malware does. So if the malware detects that it's running in a controlled Environment (which some malware does -- it can detect virtual machines and things like that), it may not do anything exciting. It

may just sit there or it may just terminate or delete itself or something like that. So another major drawback is you are at the mercy of the malware executing and it may not do anything terribly interesting.

So the third and most advanced stage of malware analysis is what we call static analysis, which is more appropriately or most people, a lot of people call it reverse- engineering. This is where you take the code of the malware, the executable code, and break it down into its machine instructions in a process called disassembling. And you end up with very low-level instructions that the computer interprets to make the program operate.

So if you're skilled enough, you can read -- or if you have the right kinds of skills, you can read these instructions just like you're reading source code, just at a different level, and understand authoritatively what the software does. There's many advantages to this, which is, number one, the code doesn't lie. A computer program is a machine; the machine has instructions. If you read the instructions, you can tell what the machine does. Number two, you aren't limited by what the malware actually does when it runs. You can follow all the control paths in the code and figure out completely what every instruction does, and then that will tell you what the entire program does and reveal all the capabilities of the malware.

Julia Allen: This is pretty fascinating. It never occurred to me that when, for example, when you talked about the environment for runtime analysis that the malware can actually detect the type of environment that it's running in and behave in a way that's not very interesting. I hadn't really considered that before.

But let me ask you, this may not be a fair question, but of all the malware that comes into CERT for analysis, are you able to just ballpark how much of it is subject to all the different types of analysis that you described or how you decide which ones to subject to which levels of analysis? Is there any way to characterize that?

Jeff Gennari: So those three techniques -- surface analysis, runtime analysis and static analysis go in order from easiest to hard and most lightweight to most heavyweight. So you don't need a tremendous amount of skill to do surface analysis techniques. A lot of that can be automated. You run some process like an MD5 hash. We do that for basically every piece of malware we see. We'll do these basic similarity measurements like MD5, file size, things like that, to gather data points, to see if we've seen it before. Because, again, that'll take the analysis effort to zero. If we've seen it before, we don't need to do much more; we can just refer to the old report or the old analysis.

Runtime analysis -- we strive to get everything through some type of runtime analysis effort. I can't say that that works all the time. It is a little time-consuming and we're fairly well-instrumented to do that

Static analysis is very expensive because the static analysis basically, as I said, it requires you to read the instructions and make some -- it requires a bit of creativity, a bit of technical skill, a bit of experience with hardware and software.

And in our experience, the talent pool for that isn't very deep, so there's only a few people, at least at CERT anyway, doing this kind of work. So that's typically very expensive in terms of effort and resources and we reserve that level of analysis for the things that are very important. And we identify those things based on the criteria I mentioned before, which is: who's asking for the analysis; what are the attributes of the request; is this an active incident; is this something we consider high priority for whatever reason. In other words, who's asking for it and why are they asking.

Part 3: Why Invest in a Malware Analysis Capability; First Steps

Julia Allen: That makes good sense and I suspect that kind of criteria and prioritization is true for most organizations.

So heading in that direction, lifting us a little bit out of some of the technical descriptions, why should a business leader consider investing in either building their own internal malware analysis capability or possibly -- I don't even know if you can buy it as an outsource service. I would think that some of this stuff is pretty sensitive, so you may not want to do that. But how should a business leader think about this and whether or not to invest in this capability?

Jeff Gennari: Well, that's a great question because there's some schools of thought out there that people don't particularly care what the malware does. If it's detected and removed, that's fine. But we take the approach that -- well, and for some people that is perfectly fine. But if you really want to understand what the malware's doing on your network -- if you want to gain some insights into how do you remedy this in an intelligent way, not just wipe the machine which can be very expensive, or restore from backup which can be impractical if not expensive -- then it really helps to know what the malware does.

You can get a real -- when you're doing the damage assessment of your network, if you have very rough data, it would be very hard to accurately estimate what you've lost, report what you've lost. Whereas if you've actually analyzed the malware, you'll get a good sense of what exactly has happened on your network, on your systems.

Then you can get a better, a more accurate idea of what kinds of things you need to do to, what kind of things you have to report, what kind of things you need to do to actually clean up the network

So, for instance, you may identify an infection on one machine on a network of a few hundred machines. And if you just clean that machine, you may end up reinfected very shortly thereafter because the malware spreads in a certain way that you didn't know.

So analyzing the malware, developing a malware analysis capability, or having those skills onsite, would let you understand how the malware propagates and let you detect all the infections, or at least take you in steps to detect all the infections, and maybe even give you a strategy to prevent new infections depending on the infection vector.

Julia Allen: Well, and I would assume too that part of this analysis goes to root cause and being able to determine what caused the malware to infect your system in the first place would help you prevent its recurrence, right? So not just new malware but seeing the same malware over and over again.

Jeff Gennari: Yeah. Yeah, certainly. That gets back to if the malware propagates a certain way, can you figure that out from analyzing the malware, and defend against that particular vector?

Julia Allen: So Jeff, you've laid out a real nice approach for some of the -- what it is, some of the technical considerations, some of the reasons why an organization might want to build this capability. So if I was a listener to this podcast and I said, "Okay, this is the path I want to proceed down," do you have some recommended first steps for getting started on building such a capability?

Jeff Gennari: Yes, I do. So as I mentioned before, maintaining a list of malware, building some kind of infrastructure to store malware so you have a historical record of the threats facing your organization is a great first step. You don't need expertise in malware analysis to do that kind of thing; you just need some infrastructure.

Beyond that, I think -- in the last five years, there's been quite a bit of literature put out there on malware analysis in terms of books, things like that. So a search of the Internet for malware analysis will yield some good results.

And then as I mentioned in the stages of analysis, the first stage of surface analysis, those types of things, those are relatively easy to do. They're lightweight, and you can get a baseline take, to gauge what threats are really facing the organization, or at least start to detect them.

Julia Allen: Excellent. Well, I really appreciate your taking the time to introduce this topic and provide some insight for our listeners.

So, I know you haven't published a whole lot in this arena but do you have some places where our listeners can learn more about what we're doing?

Jeff Gennari: Well, the best thing I think we have on the public website is our Malware Analysis Apprenticeship course. It's offered three or four times a year in the Washington, D.C. area. And like I said, that's a fairly comprehensive course.

And it goes through the three stages of analysis I mentioned a little bit more, in a lot of detail, and talks about some specific tools that we use, we like to use, addresses some challenges and some techniques. It's a very hands-on, lab- oriented course. People can expect to actually analyze some malware, pull some information out of a file, things like that. So I'd say that's the best we have available.

Julia Allen: Okay, Jeff. Well again, thank you so much for your time and your expertise today. I really appreciate it.

Jeff Gennari: Thank you. No problem.