Integrating Privacy Practices into the Software Development Lifecycle
Transcript

Part 1: Keep Privacy at the Forefront; Collect Only Essential Information

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 operational resilience and software assurance. Today I'm very pleased to welcome Ralph Hood and Kim Howell, both with Microsoft's Trustworthy Computing organization. Ralph is a lead program manager on the Security Development Lifecycle team. Kim is a director in the Privacy group.

So today, Ralph, Kim and I will be discussing, I think, something pretty unique, which is why Microsoft has added privacy practices to their security development lifecycle; what some of these are; a few guidelines for implementing them that everybody can take advantage of.

So welcome Ralph, really glad to have you with us today.

Ralph Hood: Thank you.

Julia Allen: And welcome to you Kim. Thanks for making the time.

Kim Howell: Not a problem.

Julia Allen: Okay, so let's start with you Ralph, to get us off the ground here. So I think sometimes as a community we tend to conflate security and privacy. We link them and tend to talk about them in the same breath. And clearly they are linked when it comes to protecting information. Yet when you look at what's been going on in software development, at least I haven't seen privacy pulled out and pulling the same profile as security. So how did your team and Microsoft decide to address privacy specifically in your security development lifecycle?

Ralph Hood: Well, so certainly privacy and security have been hot topics or areas of interest for Microsoft for quite some time. But as the SDL has evolved over the last several years, the overlap and alignment with privacy became more and more clear to us. And so as time evolved, we've developed SDL requirements and guidelines around privacy to help keep those privacy considerations at the forefront of people's minds, as they're also working through the various security requirements of the SDL.

So similar to the security guidance that we have, we have topics that get development teams thinking about privacy early in their development cycles: identifying who their privacy contacts are; coming up with an initial privacy assessment; and defining their privacy bug bar, much like we do with the security bug bar for security in the SDL. And those things really lead into then, as you're working on your development and verification tasks in the SDL, you're addressing your privacy design and you're ensuring that you have valid test cases for all of your privacy scenarios when you're developing test cases for all of the other things in your application. And it's really just ensuring that the alignment of those two things, and the similarities, is seen. But also ensuring that

we're keeping those things at the front of people's minds so that it's not an afterthought from any perspective for development teams.

Julia Allen: Right. It seems to me as I look at what we're doing in software assurance and software security, as you said, keeping this top of mind. They're thinking about the functional requirements, the non-functional requirements, security, availability, performance, and now we add privacy to the mix. So keeping this on the radar is what this is all about, right?

Ralph Hood: Absolutely.

Kim Howell: And also minimizing the overhead for the developers, so they don't have to look at two separate processes, one for privacy and one for security. They can think about them both at the same time. Because the way I look at it, is they're the opposite sides of the same coin. They're both about the obligations that a company takes on because they collect and use data.

Julia Allen: Excellent point. So Kim, when you talk about user data or personal data, when we talk about what it is we're trying to protect with privacy practices and guidelines, what do you include in that mix?

Kim Howell: It's pretty broad. And some people think about personally identifiable information (PII) as being the only thing that's necessary to worry about. And personally identifiable information means information like an email address or a street address - something that helps us identify, contact, or locate an individual. But privacy really goes beyond just personally identifiable information. Information that we call pseudonymous, which is information that's collected about a user even if we don't have their personally identifiable information, is still a privacy concern. So it's pretty broad how we define personal data.

Julia Allen: Right. Pretty much anything that would allow you to either directly, or inferentially or implicitly, get to an identity, correct?

Kim Howell: Correct. Because even if you don't have personally identifiable information, the more information you collect about a specific person, the more likely it is that you'll actually be able to figure out who they are.

Julia Allen: So you mentioned in your earlier remarks this idea of integrating security, privacy, other considerations, and making it more easy for developers to treat this as part of their normal development process. So when it comes to collecting personal data, what are some of the key questions that folks who are both developing systems and administering systems should be asking?

Kim Howell: I think the first question you need to ask about every piece of data you collect is why are you collecting it? Is it really essential to collect that data in order to provide the service that you're developing? And if it's not, then you really need to question collecting it at all. Because every time you collect a piece of information you create an obligation to maintain the security of that information, to make sure you retain it only as long as you need to, and to make sure that it's only used in appropriate ways. So you really need to question every piece of information that you're collecting, and what the purpose is.

Julia Allen: Yes, because as I think back historically, where we all used to freely provide our Information -- I think in our software and systems we used to collect every possible thing that we could because we thought it might have some potential future use. That has really shifted dramatically, right?

Kim Howell: That's true. And I still hear that today where people say, "Well there might be a reason that we want to use it later on." And we really make people go back and think about it. Privacy, we don't want to tell people don't collect data. We just want to make sure that again the obligation we're creating by collecting it is balanced by the value that we get from that data.

Part 2: Minimize Collected Data; Prevent Unauthorized Access

Julia Allen: Great. Well I know that in the latest version of Microsoft's SDL, one of the nice little appendices (which is what brought your work to my attention) is this "10 Things You Must Do to Protect Privacy." And of course there's a lot of content behind that. But I thought it would be helpful for our listeners to just high spot those top 10 items, walk through them briefly. So I've broken them out into chunks. And the first chunk is how might you go about minimizing the collection of personal data? You said just the smallest possible set. So what are some of the practices around that topic?

Kim Howell: Well again it's mapping data collection to a business need. And then it's not just what data you collect but in what context do you collect it? If you do not need that data in context to personally identifiable information, because you're only going to be looking at an aggregate, you don't need to collect the personally identifiable information. Or once you collect it, you get rid of it, you discard it as soon as possible. So data minimization is not only about what data points you collect, but the associations you make amongst the data.

Julia Allen: Okay. And can you say a little bit more about the temporal nature of the data collection? I know one of your practices is collect the smallest amount of data for the shortest period of time. How does that manifest over the lifecycle?

Kim Howell: So when the data comes in, a lot of times it comes in through logs, and again, you need to look at the retention period of those logs. And what you should ideally do is look at the information that's in the logs, and why do you need it? And then you parse it out, and you take the pieces that you need and put them where you need it. And each of those pieces, depending on the business purpose, you automate a retention cycle. So it's not like you have to remember to go delete it but that that's something that's automated and built into the system that's housing the data.

Julia Allen: Okay. In the second group of practices, you talk about preventing unauthorized access to personal data and also inappropriate use and transfer. What are some of the practices in that arena?

Kim Howell: Well what constitutes an unauthorized access is really defined at the point that you collect the data, and what notice and consent you got from the customer at that point in time. So if you do not keep that information about what the appropriate uses are, associated with the data, where you store it, there's no way you can know what's unauthorized access. So you need to keep your privacy preferences associated with the data, so that as it travels through your systems, you make sure that the security that you have implemented, make sure that only people who need that data actually have access to it.

Julia Allen: Okay, so is that like a metadata that travels with the personal information?

Kim Howell: Correct. So if you need - you need to make sure you know the source because each source might have a different privacy statement that has different uses noted in it. Also if you're using information for any reason to contact the person, they may have contact preferences. So you

need to make sure those contact preferences travel with that data so that you don't inadvertently contact somebody when they've told you they don't want contact.

Julia Allen: Okay. And I notice in this group of practices you also have one called "provide administrators with a way to prevent transfers." How does that work?

Kim Howell: So there are certain products that we build for ourselves and the data is our customer data, and we need to be able to control how we manage that data. But there's a lot of products and services that are built for enabling another company or entity to handle their customers' data. So you need to provide them with the right tools so that they can define their own policies about what's appropriate use, and that they can make sure that their administrators have the tools to implement those policies.

Julia Allen: Okay. And then the last group of practices -- and you've touched on some of these -- but ways to interact with users about their personal data; about how you let them know what you have, how you give them access to the data. What are some of the practices in this area?

Kim Howell: Well you really need to look at the type of data that you're collecting and the purpose for it. And you can provide notice in a privacy statement, which is simply a link somewhere at the bottom of a web page or in an options menu in software. But if the data you're collecting, and the use of it, is more privacy invasive, you need to give notice right in the user interface. So we call that unavoidable. So depending again on why you're using the data, how privacy invasive it may be seen, your notice may need to be more obvious than just a discoverable privacy statement. So privacy statements, notice in the UI, tools where people can go change their settings; those are some of the things that we use to let users know about how we're going to use their data and control those preferences.

Part 3: Gain Parental Consent; Ensure Privacy in the SDL

Julia Allen: I also notice a unique practice in this area, which certainly is getting a lot of attention, which is gaining parental consent before collecting one of their children's personal data. How does that -- that to me seemed a unique one when thinking about it from a software development perspective. How does that show up in the development life cycle?

Kim Howell: You really need to look at the product or service that you're building, and whether it's either targeted directly towards children, or if it's something that children might like. So even if that's not your target market, if what you're building is something that they would be drawn to, then you create some obligations about collecting personally identifiable information.

In the United States there's a law, COPA (Child Online Protection Act), and there's also laws in Spain and in Korea. about collecting data from minors. And what it means is that if you knowingly collect personally identifiable information from a child, you actually have to get the parent's consent before doing that.

So if you don't know that it's a child, and it's a service that's not attractive to children, then you really don't have any obligation. But if you have a product that's attractive to children, so you can assume that some of the data you're collecting might actually be coming from children, then you do have the obligation to create a parental consent mechanism.

Julia Allen: Okay, okay. Well before I let you go and we wrap up with Ralph, I would like to ask you if you can just say a little bit about -- if you think of a typical application (although there's probably no such thing), but just think of a typical development lifecycle, the extent to which, how

you work with a development team. I'd appreciate knowing a little bit about how do these privacy practices get introduced at various lifecycle phases. And I know this is part of your standard development lifecycle but how are - what are some of the gates or some of the approval mechanisms to make sure that privacy's been appropriately considered at each lifecycle phase? Could you just say a little bit about that?

Kim Howell: So through the SDL, it helps kick off a privacy review process. And early on in the design phase you would have a consult. There are privacy managers throughout the company in each business group. And this is how we implement it, is those privacy managers are responsible for working with the development teams to look at their initial design, to help minimize the privacy issues, and help them identify the right notice and consent experiences for the collection of data. And as the product moves further down the lifecycle, they actually will have to do a final privacy review and again sign off and get approval. And before any of our products ship, they are required to have a privacy signoff from a privacy manager.

Julia Allen: Okay. So obviously it's just built in to your normal development process, just like security or any other of the functional characteristics of the product, correct?

Kim Howell: That's correct.

Julia Allen: Okay, so Ralph, you get the last shot. Do you have some places where our listeners can learn more, or any closing thoughts you'd like to offer?

Ralph Hood: Sure. We have a few places that users can get more information. The first is we always have our latest SDL guidance including both the security and privacy policies in the SDL on the SDL website. And then beyond that, we have the SDL blog where we often talk about current security topics and things like that, or things that we're considering for future revisions to the SDL. And that's a great opportunity for people to read about what we're thinking about or some of the challenges that we're dealing with as a company, from a security and privacy perspective. But also it's a good means for people to give comments on what we're thinking and give feedback to us through our blog. And we love to get that type of feedback and respond to people on that.

And then lastly we have the privacy website, which has a lot of great information around what our privacy practices here are at Microsoft, as well as other general privacy guidance and consideration and resources for people to leverage.

Julia Allen: Have you found that any of your clients or customers, just in general, are starting to pick up some of your privacy guidelines, at least in terms of their own development processes or what they're expecting of their software suppliers? Have you seen any uptake like that?

Ralph Hood: We do. As we get more and more public facing with publishing the guidance for the SDL, for security and privacy, we are seeing more and more companies leverage that guidance. And that's one of the great benefits that we have with publishing this guidance is it's really about improving the ecosystem as a whole. It's not about just like promoting what we've done at Microsoft -- but it's also identifying ways that other companies or other development companies can actually implement some of those best practices or things that they should think about just for their own benefit as well.

Julia Allen: Well I just want to thank you both for your time and expertise and sharing your experiences at Microsoft with our listening audience today. So thank you Ralph.

Ralph Hood: No problem.

Julia Allen: And Kim, thank you as well.

Kim Howell: You are welcome.