Part1

Intro

hi this is Nancy made again we're going to be talking about threat modeling Now I know you've already seen a lot of material about threat modeling some from Jeff. and his overview material a little bit from me when I talked about requirements engineering, but now we'll be going into it in a little more depth. At least in terms of some other methods than the ones you've already seen and we're going to be talking about our research in the area which I think may be very relevant to you. Not necessarily while you're working on your projects in the course, but you might find what we've done to be of interest in your work when you get back to that LG.

1p

So here we go we're going to be talking about the threat modeling research that we did and, at the very end we'll touch a little bit on machine learning this of course is covering three segments of this lecture and we'll be starting out with the threat modeling.

2p

Looking at it in a little bit of detail we've got four topics spread across the three segments, the first is our threat modeling research. which we did on a few occasions, the second is a hybrid method that we developed for threat modeling and we did this after the results of our initial research project became available and you'll see why, in a few minutes. Third, I supervised some students in our master's degree in information technology and security, it was a graduate student project on machine learning. And then, finally, we looked at combining machine learning, along with our threat modeling research and seeing if we could build on that I know that I probably mentioned that I was at the software engineering Institute. And the initial research was done there so we'll start with that first.

3p

So this took place a few years back, we were funded to do a research project by our sponsor which is Department of Defense. And the research project was a one year project that you'll see why that becomes important in just a few minutes,

4p

the first thing that we did was to examine some definitions and see which ones we wanted to use for a project, as you know, i'm a big proponent of using definitions. And we picked three of them, while we actually developed our own, and then we looked at a couple of others, so our definition of a threat modeling method. Is an approach for creating an abstraction of a software system and that identifying attackers and their roles and using that abstraction to generate and catalog possible threats that the system must mitigate. So here we have some notions that you've seen in software engineering, the idea of an abstraction as a model. We have a method in the definition. And we have the notion of examining threats, specifically, so all of those things are important, and please note that the acronym that we're using for threat modeling method to TMM is one that will see throughout this lecture. The second definition is one that was started at Microsoft and then evolved at Ford motor company, I did some work with Ford and very interestingly they were using STRIDE to do threat modeling but they had the consultants from Microsoft, who had been employee an employee at Ford, and so the consultants had a very good understanding both of the software engineering work being done at Ford and the Microsoft STRIDE tool. So this definition worked for them, and I, like this one too. it's a methodology and a tool so obviously they're suggesting, you need to have a tool to come from a threat modeling. that's natural for them to suggest because they already have the tool, namely STRIDE but threat modeling doesn't necessarily require a tool, in fact, in some cases, if you're doing it very early in the process tool might not be that useful. So here we have a tool. And the method to identify and classify vulnerabilities which exploited would result in adverse business impact, so the first statement to note here a methodology and tool identify and classify vulnerabilities and look at that verse business impact. Third, again, this is pure Microsoft and repeatable process to find that address all threats to your product. So, in the case of the second two definitions its first of all, looking at business impact because these both Microsoft and Ford are for profit companies. now, when we worked with our government sponsors. Business Impact what took on a different meaning for us, but the general idea was the same. So with our own definition the first one in mind.

5p

We observed that first of all, we thought threat modeling was an activity that should be done independently of a life cycle model. Secondly, we noticed that there were a lot of threat modeling methods out there, but there was very little data indicating whether one of them was better than the other. Whether one of them was more suited for a particular type of system than another it's was pretty much everyone for themselves in terms of deciding what to use. A lot of people used STRIDE and they still do because it was well established, it was freely available. Since we did this work, some additional commercial tools have been developed, we ended up selecting three methods and you'll see why in a minute. Trying to compare them experimentally, and then analyzing the results. So we wanted to see which might be the most effective. Which one was the easiest to apply, which one got the most consistent results, which one got the most complete results, which was very important. On the one hand you don't want to identify threats that don't exist, but on the other hand, you don't want to miss important threats either. And then we wanted to allow the users to understand how good the results were. Could they be confident that they've gotten the right set of answers for their system.

6p

So we looked at three methods STRIDE which has already been discussed with you at 1st line. The second was with some colleagues at the University of Washington they had developed a method called security cards which was an aid to brainstorming that is also used fairly widely. As you'll see the cards are available for free download and you can also purchase a printed set of cards from the University bookstore, and this is kind of a fun way of thinking about threats. And then finally Persona non Grata(PnG) which was developed by colleagues at depaul university. It's been documented pretty widely in the literature, the idea is to give the modelers a particular focus so they're focusing first along the attackers. And then, on their motivations and abilities to try to come up with the threat model. With that in mind and keeping in mind that we wanted to find out if one of these was better than the other.

7p

We decided on project to look at the three of them, side by side and see what kind of results, we could produce. we had three universities participating with us, we had undergraduate and graduate students as subjects of the experiment. We had a year to do the work. and during that time period, we came up with a couple of case studies that the students could use we taught them the methods and the students learned the methods in different orders, so that we didn't have the Hawthorne effect. And they also applied them to different problems.

8p

So first we'll talk about the three approaches and then about our results i'm going to just touch really briefly on STRIDE because you've seen it a few times here's a brief history of it was invented in 1999 so it's been around for a while. Then it was implemented the Microsoft and it's very widely used typically it produces or starts with a data flow diagram. mapping is produced a threat categories and then, once the threats are determined, they can be documented and mitigation strategies can be developed STRIDE, as you know, is an acronym that stands for different types of threats and each threat. threatens the security property that's you see on the right side of this table. i've got a couple of references here at the bottom, I think the one by Sean harmon(??) is probably one that has already been assigned to you as reading. So you will be very familiar with STRIDE if you're not already so that was one of our choices with STRIDE. And we encourage the students to use the tool from their website to make their life easier.

9p-10p

The second approach was security cards, as I mentioned, this was developed at university of Washington. it's a card game that supports a brainstorming session and the way it works is the cards are used to prompt ideas from the pump participants. Each card has a front and the back. On the front there's a topic and a dimension. Dimension could be one of three. The adversaries motivations or their resources or them methods. So you've got the possibility of one of three dimensions and for each one, a topic. that's the front of the card.

11p

And then the back of the card gives more information to help you with the brainstorming, and this is done in a team like many brainstorm sessions with both these additional hence and also some examples. And, in some cases there are references as well. It also has been used fairly widely, which was in a way was surprising to us, we knew it had been used a lot of universities. But we didn't know that it had been used someone extensively in industry as well, what I would recommend is that if you would like to try it, go ahead and either download the cards order the printed once considering the distance involved I think downloading the cards and printing them yourself is probably the better strategy, I think you would have fun with it. we found in all markets, a software engineering institute that we couldn't really use it with our Department of Defense customers they didn't like the idea of mixing games in with their serious business. so it all depends on the project and the stakeholders that you're working with as to whether this would be useful.

12p-13p

The third method is Persona non Grata(PnG) that borrows on a concept that was developed in human computer interaction and the concept is that have a persona as a fictitious standing for real people. Actually, real customers of your product. And the idea is that you develop detailed descriptions of these people and their characteristics, so you can think about how they might use your system and what features are of interest to them. this is an area that's been researched very thoroughly and you can find a lot of reference material on it.

14p

Now here's one example of a persona. Thomas is 76 year old retired accountant who likes to spend time with his grandchildren. He likes to read newspapers work in his garden and stay in touch with his friends. he likes technology, but only if it doesn't get in this way. So you can use this description to think about the kinds of systems that might be appealing to this person, for example, are they interested in social media, are they interested in tracking their investments in the stock market, are they interested in having zoom meetings with their friends or using Skype. So, this would give an idea of the type of software product that Thomas is interested in and once you have that you can delve further and find out what product features you think you might be interested in. this is again a standard for real customer.

15p

Now the idea, so a persona is a user of a system, the idea of Persona non Grata, on the other hand, is a malicious user of the system so think about the diagrams that you've seen that show use cases and Miss use cases side by side, you have a legitimate user on one side, trying to use your system, and then you have a malicious user, on the other side. And of course the malicious user license to use case. So there's a structured method. and the method asks you to develop the Persona non Grata motivations, for example, do they want monetary gain do they want recognition and they just doing something nasty for their own amusement. And if you think about some of the ransomware attacks, these days, obviously, their motivation is monetary gain. Second, develop their goals. So does the Persona non Grata know what it wants to do and how it plans to do it, what skills, do they have available to choose the goal and what other assets, they have besides their own personal skills. And then finally in this structure you develop misuse cases and once you have those misuse cases, then you can proceed as before, to develop your mitigation strategies. So it's a little bit different from some of the other threat modeling methods that we've seen, but nevertheless can produce some really interesting results.

16p

We have a terrific example here, and this actually happened, this is a real case study. It took place in Australia. We have a contractor working with escape a equipment and control system. And this system controls controlled the sewage equipment for municipal authority. So he was a contractor after he left the contracting company he applied for a job, but got turned down. So he decided that he was going to get even by taking over the system and having it release raw sewage into likes and other places where you don't want raw sewage. And then he wanted to have that appear in the system as a malfunction. So that there would be a backlash against the contractor and the municipal government. So he succeeded, up to a point.

17p

He planned this out. So we have here is skills. The miss use cases and nisha the Miss use cases that actually happened. First off he stole a computer and equipment. He constructed a fake pumping station, from which to send the signals gained access to the skater system, and this will disable the alarm set the pumping stations. And then issue commands to release the raw sewage. Well, this worked, but not for very long Mike got caught. and probably went to prison is, as a consequence. However, in developing this case for the Persona non Grata we have the advantage that this is a case history. So we were able to reverse engineer this and see exactly what happened. we weren't speculating on it as a future event, it was something that had already taken place, so it was a really good validation for the idea of Persona non Grata. However, we wanted to do some more investigation of how you would use it going forward when you didn't have the Persona non Grata on hand.

18p

So we had 250 people one matt. Most of them were students in software and cybersecurity they were both graduate and undergraduate students, some of them were practitioners, who had gone back to school. Some of them were professionals, but not all of them, some of them did not have experience to depending on who they were. They all applied threat modeling methods to common test beds, we created some test cases to use for the research project, and those are freely available in our report. The first was a drone swarm case study, you can see that, on the Left unmanned autonomous vehicle and the second was an aircraft maintenance application. Each team learned and applied the approach on one test bed, and then learned a different approach and applied it on the other, test bed. That meant that the teams had to learn, only two of them. And we did mix and match to verify that a particular team didn't have an advantage when they did this. We came up with some standard templates that we asked them to use some of them did it as class exercises, some of them did it as homework assignments. it just depended on the classroom situation.

19p

Another feature of the study. was that we had some professional threat modelers is a control group. the professionals did threat modeling for a living. So it was a good control group and we looked at their results, side by side with the student results, and we were looking at in this initial set of 250 that we mentioned here(18p). We were looking at those results, collectively, but then we also drill down a little bit and look to see if we had differences from one class of students to another, for example, between undergraduates and graduates experienced people and those without experience. So the results weren't what we hoped for was that we would see one method command, is it clear winner is the best method of industry that didn't happen. What we did find out was if we found it given threat how confident, could we be that would be reported by the team. We found that there was a lot of variability with STRIDE. Security cards also found a lot of thread types. But also had a lot of variation. PnG was the most consistent. But have found fewer thread tops. And finally, no single method lead two teams reporting a majority of the valid threats and our benchmark for the valid threats was the set of threats that the professional setting found.

20p

Another question that we had was. Are we finding threats that are really important to or maybe don't even exist so. On the one hand, we want to find all of the real threats that are likely to happen. On the other hand, we don't want to identify a bunch of extraneous threats that will probably never occur if they even exist. So we looked at this in both directions. Some threats were found by the professionals, but not by the students and other cases students found threats that the professionals didn't find. So it could be that the ones that the students found them the professionals didn't, or either false positives, or maybe they were unusual.

21p

Some other impressions that we came away with, first of all. We felt that STRIDE was something that was expected to be used somewhat later in the life cycle. So you had to have some idea of the system architecture, when you were using it, you had to be able to represent the system using a data flow diagram. And then had more of a cookbook style than the other methods from our viewpoint. We like the security cards, because it encouraged thinking outside the box and the I its idea of brainstorming. However, once again we a lot of variation in the results. PnG was more narrow, perhaps because of its initial focus on the attacker, but have provided consistent results. So we thought that perhaps if we merge, Security cards and PnG we could come up with a different view.

22p

So that was the initial study results. When we get to later part of this lecture we're going to see what we did with the results and how we came up with a hybrid method, but for now, to expand your knowledge, here's a reading assignment. it's pretty readable it's a summary of available methods, this was done more recently than our threat modeling research, so I think there might be somewhere on the order of a dozen methods that are discussed here, and this is changing on a kind of a regular basis, right now, there are several commercial products for threat modeling and some additional research or other facts as well. So with that we've talked so far about our work in threat modeling method experimentation. a research, study that compared three methods so now you've seen a little bit more about Persona non Grata and security cards to add to your knowledge your STRIDE. The reading assignment will give you an overview of some of the methods that you might consider it sometime in the future. You probably won't be spending time using them during this course, but I think you could find some of them to be useful, as you go forward thanks a lot and i'll See you in a few minutes.

Part2

23p-37p

hi it's Nancy meet again now we're going to talk about our research, a little bit more will be taking a look at some further analysis that we did. With the Persona non Grata data on threat modeling and then we'll be talking about a hybrid method that we created because we weren't totally satisfied with the results of our initial research comparing the three methods so we'll get started with that right now. Initially we did some PNG analysis as part of the comparison of the three methods and since we had collected all that data, we had the opportunity. To go back and do some further analysis, using the idea of crowdsourcing to see if we could get more information out of the data that we already had. So we were looking at the results of the data from 108 students so it's a pretty good sized group. They were into introductory information security courses both undergraduate and graduate. They were a combination of novice learners returning practitioners and professionals So these are the crowd the 108 students. All of them applied Persona non Grata to the unmanned autonomous vehicle or drone swarm system scenario. As I mentioned the scenarios are available in our reports you'll have access to them in case you want to use it experimentally yourself. And it might also be interesting to take a look at So you can see exactly what scenarios, we were working with its summarized in the page so it's not a huge amount of reading. So here's a view of the threats to the Leader drone it's worth noting.

Part3

intro

we're ready to complete our visit to threat online with our third segment in this segment we're going to discuss a research project. On machine learning that at first might not look like it's related to threat modeling then we'll be looking at how it ties into threat modeling and wrapping up this whole segment with some readings and a homework assignment.

38p

Okay, so the first part of this is a discussion of a project that was undertaken by a team of master's degree students getting a master's degree in information technology and security as part of our master software engineering Program. As part of the master's degree, they were expected to work on a summer project that was a research and development project. Several of us proposed that they do something in the area of machine learning, and this was the result in 2018 so you've seen we started that in 2015 and 16 with our research comparing various threat modeling methods in 2017 we went on to develop a hybrid threat modeling method and apply it. To both a small example and a medium sized system and then in 2018 we have this opportunity to work with some of our graduate students in Pittsburgh on a machine learning project.

39p

So the goal of the project and here i'm helping myself to the materials that were developed by the students, they were very pleased to have their material included in our courses and lectures so you'll recognize that some of these may not look exactly the same as the slides i've been using up until now. But it's all part of the same set of activities that took place over that several year period, and in fact suggested some future research that we are about to start. So the scope of the project was to attempt to assess the robustness of machine learning models against adversarial examples, by which we mean threats. So there is a tie into threat modeling right away.

40p

The example that i'm showing here in was considered to be somewhat novel now there are lots of examples like this because machine learning has become such a hot research area and development area, so in this example, we have an image that should be recognizable as a Panda(1st image). If a small change is introduced to it(2nd image). It can cause the model to make a mistake and think that it's actually a gibbon a type of monkey with 99% confidence pretty remarkable that by introducing a small error, the model can make a mistake, you seen more recent examples of this, were in automated driving. A system can be modified so that a stop sign no longer appears to be a stop sign. There is a lot of potential for attacks on systems that have been developed using machine learning and we really have to take care to try to avoid those attacks from taking place at all. especially since so many machine learning models are being used in safety critical systems such as Automotive Systems and another concern that we would have is control systems such as those that are used in power engineering.

41p

The motivation for the project at that time was that there were no standardized security frameworks to prevent adversarial attacks from attacking a model. So people need to make sure their machine learning models were good enough to be safe. But there were only a free few resources available for people to use for that purpose and the student team had the idea that they could experimentally, try to develop something that was at least as good as what was already out there.

42p

They decided on an implementation of an application on a website. That would assess the robustness of the model, so this was before the model was implemented it was early in the process. And the idea was to assess the robustness of that model against adversarial examples or threats on computer vision tasks, such as the one we just looked at. The website application would then evaluate and provide feedback about the robustness of the models, using the Cleverhans library, which was a library of adversarial examples and the image net data set. which was a standard set of image. So the input was a clean image, such as the one that we saw on the left in the earlier chart and the output would be an adversarial example generated by each attack method.

43p

you'll see that we ran into some interesting questions along the way, but overall it contributed a lot to our research and to our thinking at that time. So they wanted to use attack methods that were built with different assumptions to measure the models vulnerability against different adversaries, who had different levels of capability, so it would be a single model that would be input and it would be tested against a number of attack methods. That way, they could get an idea of the vulnerability of the model and how well it was able to tolerate those attacks under different circumstances.

44p

Some of the features that were considered desirable or a scalable system, so it should be able to scale up to handle a number of inputs, not just one, at a time. A user friendly web interface, and then, comprehensive and reliable feedback to the users who were providing the models for evaluation. obviously reliability is important, because if you couldn't depend on the results you wouldn't be able to mitigate the threat successfully. One constraint of the system was that if you made modifications to a model, you had to rerun it you couldn't keep a modification in the database and do a partial rerun of the system.

45p

They came up with a list of functional requirements, as usual, when we have students teams, working with faculty there was some negotiation about which requirements were necessary. And then there were constraints that we were working with as researchers and constraints that the students who are working with. So the functional requirements that were considered to be important, or a dashboard for the user and upload form for the model. And history of submissions what the details were and then an information page. The business constraints that the students ran into. Where first of all, they had limited budget each student team was allocated a budget, but it was pretty small so they couldn't just go out and spend this much money as they wanted they had to kind of think about prioritizing how they would spend the funds that were available to them. The second constraint was we wanted to make it freely available to users, so, on the one hand we're spending some amount of money to develop it. And on the other hand, we wanted to make it freely available to users, the technical constraints were first of all, that it was based on this single library of adversarial examples, namely Cleverhans. It was deployed on AWS Amazon web services. It supports curious primary reason for selecting curious was that it's open source and hence freely available and didn't require any budget, and then it targets only computer vision models. actually these weren't really constraints that were not predictable because keep in mind, they were doing all of this work over the course of the summer, which is only a short time period.

46p

The results were that they were they completed the project, they had a demonstration of it, they made it freely available, as with many student projects it didn't have a support after the fact, and so we didn't have a way of supporting users, if they wanted to try it out and we didn't have a way of supporting users, if they wanted to install it on their own systems. So the usage of it after that was pretty limited but as a proof of concept, it was really good and that was what we were trying to do at the time. So, as I mentioned, we had these three projects our initial research project, which was a fairly large project over a one year period of time looking at the results of 250 students. assigned to student teams. And those students were using a mix of STRIDE, security cards, and Persona non Grata as threat modeling methods and we were then able to compare the results. We found that there was no single best method, which was disappointing, but a retrospect, perhaps lots of pricing. And this doesn't even consider all the methods that we didn't look at the ones that are in the paper that I referred you to at least 12 of them by now, but then this is five years later. That following year we developed the hybrid method and applied it to both a small example and the medium size system, it was useful, unfortunately. We didn't have the opportunity to look at it and say well wasn't better than what we did before, and then, finally, this student project, which was focused on machine learning and threats, but what wasn't actually. A throughout my leg method, like the other ones that we had worked with so our idea was that going forward, we would try to merge what we had done.

47p

and see whether there was a way of marrying machine learning with our research ideas for the hybrid method. This is a project or set of projects that I started discussing in 2019 and actually had some plans to work on, with a team of researchers in the UK and Ireland, but we all know what happened in 2020-2021 so that project hasn't yet started.

48p

but we looked at some potential ideas and, here are some of them one was to apply multiple threat modeling methods. To systems that employ machine learning to try to understand which ones are best. so it's an expansion on our prior research examples taking several threat modeling methods, including our hybrid method in conjunction with machine learning. It provides an opportunity for collaboration with both students, especially graduate students and practitioners, such as yourself a lot of organizations are interested in machine learning and it certainly behooves them to learn what the threats are. Secondly, we thought we might focus on one of our prior examples, the drone example and the aircraft maintenance example with the machine learning twist so we modify the example and then apply several methods to it. And then perhaps develop some new examples to be used as part of that experimentation so that's one set of ideas, applying multiple threat modeling methods in conjunction with machine learning.

49p

The other set of ideas is looking at machine learning in a different way and asking ourselves whether machine learning can be used somehow to improve threat modeling. so it's kind of the reverse, instead of using threat modeling to improve machine learning can we use machine learning to improve threat one. In order to do that we'd want to examine some existing research on machine learning and that's now a very important area, so there's research going on all the time. Maybe focus on a specific domain, or one of our prior examples, either for example power systems or perhaps the drone example when see how that would work out. And then see whether machine learning can be used in conjunction with our hybrid method to improve threat modeling. so Those are two sets of research ideas. A third research idea was to use machine learning in education and that that's kind of a different thread so i'm not going to discuss that here.

50p

So wrapping up the references for the student project are listed here. And the third one, is a video, obviously.

51p

The references for machine learning are here and there are some new ones, after we did the research. The newer ones are the SCI blog post from 2019 which was after the students had completed their project. It became an area of interest to Gary mcgraw who's a well known figure in software security. And he and a nonprofit organization of his documented the top 10 risks of machine learning security that can be found in an issue of IEEE computer. And then there is a very good podcast threats for machine learning by Mark Sherman who's at the software engineering institute it's a webinar it's available on the SCI website for free viewing. it's interesting because it is a much more mathematical approach than some of what i've seen previously and it makes use of certain textbook references. that are more toward mathematics and formal methods, if you have background in those areas I think you would find that certainly the webinar is accessible and you might be able to pick up some ideas from the webinar and the references for your future use.

52p

And finally, as promised a homework assignment and the homework assignment. reflects back on the entire they have this lecture and that the segment that we just covered. And the assignment is to use Persona non Grata to identify Persona non Grata is for your project. identify them remember we said there were four parts motivations, goals, skills, and miss use cases. so go ahead and develop the motivations, goals, and skills you don't have to develop for miss use cases, but you might want to do that if you identify something new. and ask yourself whether there was something that you missed. Using STRIDE or their new threats that came up as a result of considering the attackers or does it validate the results of STRIDE so answer that question so two questions, one is to apply the energy to your project with your team and the other is to answer the question of whether there might be new threats that you missed with STRIDE that would suggest new miss use cases, or does it validate the results that you've got before. If you are actually using Persona non Grata you would certainly need the misuse cases because that's what you need to develop mitigation strategies. but for this exercise, you don't have to do that. so with that we've completed the lecture on threat modeling we've talked about three research projects and some possible future work if any of you would like to collaborate on future work or you have suggestions for others who are interested in collaboration that's something that I want to start working on very soon, thank you and i'll see you again soon.