Carnegie Mellon University
Software Engineering Institute

SEI Podcasts

Conversations in Artificial Intelligence, Cybersecurity, and Software Engineering

The Impact of Architecture on the Safety of Cyber-Physical Systems

featuring Jerome Hugues 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. My name is Suzanne Miller, and I am a principal researcher in the SEI Software Solutions Division. Today, I am joined by Jerome Hugues, a senior architecture researcher and also in the SEI Software Solutions Division. We are here to talk about his work in assuring cyber-physical systems, especially autonomous ones. Welcome, Jerome.

Jerome Hugues: Thank you, Suzie.

Suzanne: You have done previous podcasts with us. We will link to those in our transcript. For those who didn't catch your earlier podcast, would you start just by telling us a little bit about your background, why you came to the SEI, and what's cool about the work that you do here?

Jerome: Thank you. It is actually my second podcast with the SEI. I joined the SEI four years ago to work mostly on model-based techniques. I am interested in the architecture of cyber-physical systems as a way to ensure that they are safe, correct by construction, and that we can also derive an implementation of the systems. My background is obviously in computer science, mostly middleware operating systems, formal methods, really, with this emphasis on architecting systems and demonstrating that they can hold multiple types of properties in terms of performance seals, safety, security. For this, demonstrating that they have natural semantics, explaining what the semantics is and deriving tools from this semantic is critical to my work. I joined SEI four years ago as a senior researcher, and I have run a couple of projects on this particular topic. Today we are discussing one of those on the assurance of cyber-physical systems, autonomous ones, as you mentioned, and the couple of contributions we have made in this field. Thank you for this opportunity.

Suzanne: No, this is excellent. Many of us at the SEI work on cyber-physical systems, which are systems that rely both on the hardware and the software elements to be able to perform their functions. So neither alone will actually perform the functions that we need, and those tend to be very safety-critical, mission-critical kinds of systems. Assuring that the quality attributes—security, safety, reliability—are present throughout the system is one of the biggest challenges in those systems, because we tend to build things, you know, in components and you're not always thinking about how your component affects everything else. The work that we do and that you do in the model-based arena is very critical to exposing to people and illustrating what the effects can be if we ignore those kinds of quality attributes when we are building these systems. I am thrilled to have us doing this kind of research, and I am absolutely thrilled to talk to you about what we've done today. So let's get into the topic. We like to think about having more data. More data, more data, and more data is a good thing. We outfit autonomous, which are cyber-physical systems that actually work on their own in many of the settings that they are in. We give them more and more sensors and better sensors to give us higher fidelity data and give them more data to operate in their environments. But that's not enough, right? Can you talk about the challenges that come with adding more data, more sensors to these systems as we become more technologically mature and being able to outfit them?

Jerome: Of course. When you think about it, first of all, why do we need more data actually for those systems? We need more data because we want to be more precise in evaluating the situation and making decisions. We want our systems, ultimately, to replace us for a couple of mundane tasks, the one we don't really want to do. Autonomous driving being one example out of many, many other types of autonomous systems we can think of. We want to have more data, so that we can take more precise decisions. As we are collecting this data, we have to think in terms of what is the implication of the system we are building because more data means, first of all, more computations that we will have to execute more complex ones. In terms of algorithmics, of course, in terms of mathematical domain, it has a lot of implications. There is a whole area of research on how to do reinforcement learning, how to do autonomous decisions, how to build a representational environment, for instance, for making maps, to build maps out of the environment... But this is usually the visible functional aspect of the iceberg that we are facing. Usually below this iceberg, there are all types of other situations that we have to think about. We have more sensors, so we do more computations, so we need more powerful CPUs. With more powerful CPUs come more tricky questions in terms of energy management, for instance, or in terms of reliability of the processor that we want to embark. Some of us may have already experienced the system, for instance, our cars crashing just because some tornado, some thunder strike happened nearby. Electricity, electromagnetic compatibility, are impeding reliability of the systems. We have to think about this. This is not related to the type of research we do here at the SEI. But many other FFRDCs [federally funded research and development centers] are already concerned with electromagnetic compatibility. Closer to what we do, more sensors mean, obviously, that we will have to aggregate value streams of information, value streams of data. Ultimately, the software architecture, the architecture of the cyber-physical systems, will be more complex.

That data will have to be synchronized. We have to make sure that, yes, the information we get from the left side of the car is compatible and is arriving at the same time as information coming from the right side of the car, for instance. We have to ensure that the data is correct, arriving at the right time, so without a timing issue. So the data is correct. It has not been tampered with external indication. For instance, many of us...again I am taking the car as an example. We all love autonomous cars. I am really looking for having a fully autonomous car, even level four type of thing, but it is not for today. But many of us have pressure sensor on tires. We love to check that this information is correct. This information is coming from a wireless sensor. It could be possible, for instance, if we look at it from a cybersecurity perspective, just to tamper with this sensor, send an incorrect data to the car so that you got a flat tire even though it is obvious that you don't have one, but the sensor has been jammed. Ultimately, the car will not stop because the car cannot allow you to start. There are all types of issues in terms of security that come with these additional sensors because the more sensors, the more complex architecture, the more difficult it is to ensure that it is secure from several security perspectives. The same issue will be observed from a safety perspective—more information, more sensors. If one of them is not providing the right information, not just at the right time or just even the right value, the system may consider that it is in a specific state or specific position, if you will, but it's not accurate. It may go on the left, even though there is a wall on the left, and you may have a safety situation in this particular case. More sensors are nice from a functional perspective, but it is really a nightmare or a challenge, so to speak, from the safety, security and performance perspective, which is what we are here to solve and address. I love those architectural challenges.

Suzanne: I would argue that one of the reasons that not everyone may be as enthusiastic about autonomous cars as you are is—especially if you are not really directly involved in this work—when something like that happens, when the tire sensor says you have a flat tire and you don't, it reduces the trust in those autonomous elements of the system. It's like, *Why does it keep telling me that my tire is flat? It's not flat.*

Jerome: You are right. In this particular case, trust is critical. It is a reason why we have to remember that in cyber-physical systems, there is this last world system. We have to consider it from a multidisciplinary perspective. It is a system, so, ultimately, what we are delivering is not just a piece of hardware or a piece of software. It is a system, which is made of hardware and software, eventually an assurance case, a convincing argument that can be understood by an external party, that the way we have engineered the system, the way we have tested it and the underlying theories that we use to get evidence is consistent. This was part of this activity of this project that we are about to wrap up at the end of FY23 to build this crazy story, which is not just to say we want to work on the architecture work, do software as usual, but really consider the system dimension of it. And you are right, ultimately trust is something that we want to achieve. And trust is something that you can achieve by convincing people and by providing arguments that what you have done is correct by construction.

Suzanne: All right. Let’s get into more of the details about this work. There is a learning component aspect that is fundamental to an autonomous cyber-physical system, right? Part of what makes it perform better is as it gathers data, not only about the environment, but about its own performance, it can make corrections, it can add precision based on learning. We build that kind of learning into these kinds of systems. Can you talk about the role of those learning components and the challenges that they actually pose to things like safety and security and your approach to verifying the learning components?

Jerome: Yes, let me start first with this idea of learning components. There is a lot of enthusiasm around AI topics, as we all observe. It is something great. Just the idea of dropping a couple of software elements, learn this data, and make something out of it is really exciting. The truth is, they are not really learning. If you look at it from a mathematical perspective, they are just solving some optimization problem, and CPUs are very good at it. The second answer I may give is that I am not really interested in verifying grounding components in this particular project. Many other groups are working on this topic. It is part of a very like large research portfolio, some colleagues, for instance at Collins [Aerospace], made tremendous contribution in that particular domain. What we observe, though, is that in most cases, they are interested in the correctness of the running component itself. This is fine because this gives a verified component that we may integrate and verified with some stochastic properties, such as probabilistic property. The problem we want to address, or that we are addressing in this particular program, is a question of integrating the surrounding integral component within a system. What does it mean to take a component for which we may have a characterization of its inputs and outputs, all the data streams I am collecting and the data stream I am outputting, with some performance matrix? I know that it will take up to a couple of milliseconds for me to tell you whether or not this is actually a horse, a bicycle, or just a tree that is on the road. How can I integrate this component knowing that I may be wrong into classifying this object that is on the road? By knowing this, what we want to achieve is to integrate this component and mitigate any bad decision, any bad outcome, coming from this element, so as to minimize the risk of bad events happening in your system. This is really the system perspective that I was discussing just before, which is basically how can we define the architecture of the system so that for any given learning integral component that is part of the system we can add proper fences around it so that whatever this component is receiving, we can assess that the value is correct, has not been tampered with, is arriving at the right time? And whatever the system is outputting can be controlled for some notion of correctness. *Is the acceleration of my car within the speed limits of the state of Pennsylvania*? for instance, or within the body limit if you want to have a cool way of driving. All of that is part of this fencing that we want to put on around the system so that, ultimately, we can be confident, we can trust that the system itself is correct, or, not that wrong, compared to verifying the running component because this part we know is, as of today, not possible.

Suzanne: One of the challenges you didn't mention is doing all the things you're talking about at real-time speed.

Jerome: Oh, yes.

Suzanne: For those that are not really familiar with real-time systems, one of our continuing, continuing, continuing challenges is integrating more data, integrating more function and not reducing the speed of performance. Because if we are talking about a car, an airplane, a ship, there are time constraints that if you miss them, people can die. So we miss the mark of whatever we are trying to do. That is the other thing about learning systems is learning systems that certainly I am aware of in the lab, really tend to just say *more data, more data*. We were talking about that earlier, right? Just *more data, more data, more data* to make the learning more precise, but that actually slows down the learning.

Jerome: Yes.

Suzanne: The real question is, how much data from where gives us the best balance of timing, accuracy, and precision, and then safety? The fence. How do we prevent anyone interfering with the precision accuracy of that data so that we don't end up having malicious or just incidental failures because something interfered with the data coming into the system and the decision-making that goes along with the system. You are taking on some pretty big challenges there, Jerome.

Jerome: It is, yes. In this particular domain, because we are all dealing with autonomous systems, actually NASA made a tech report on increasingly autonomous system because there is a graduation between basic autonomous system, which are nothing but controllers, to fully autonomous systems. Relating to this question of timing and decision making, before joining the SEI, I was actually teaching in France, in Toulouse, so a big place for aerospace industry. Every year we were reusing the same video of the Ariane 5 first flight. For those of you who may not know about this one, it was a complete mess. At T-zero plus 45 seconds, the rocket collapsed because of some software/system/testing error. So probably a topic of its own, but it has been addressed in many reports. But it is definitely relevant to this discussion of autonomous system, because at the time, the operator in the room was saying that the trajectory was perfect, and all parameters were all nominal. The rocket took the decision to self-destruct because by the time the human reacted to the information he was receiving that everything was on track, this onboard CPU detected an anomalous situation and this idea that the only correct course of action at this stage was self-destruction. So the timing is going in both directions, timing to react to external events, timing to engage with human to make sure that we can protect them. This is creating all types of interesting challenges to address in terms of timing, safety, security. We give the feeling that we are obsessed by those three aspects, but they are here to ensure ultimately the trust of the public that may use our systems.

Suzanne: Absolutely. In doing this work, you have been collaborating with researchers at Georgia Tech on the approaches to these challenges. Why don't we talk about how this collaboration came about? What are the different areas of expertise that the team brings to the table and your overall vision for taking on the kinds of challenges we have been talking about?

Jerome: Sure. On our end, my team at the SEI developed a couple of contributions on the architecture of cyber-physical systems. We started by taking the AADL [Architecture Analysis Design Language] that we covered in many podcasts with you in the past, Suzie. What we did was to extend the semantics of AADL so that we can make a formal semantics out of it. Basically, the semantics in such a way so that we can simulate AADL model in a very precise way, but also so we can verify some properties on AADL models. When we have done this programmatically, we can speak about or reason about the architecture of cyber-physical systems. What we decided to do, as I mentioned, was to establish a fence around those running-enabled components. But still, we need to perform full detection, just to make sure that if those inputs are incorrect or tampered with, all the outputs are invalid. This is what Georgia Tech brought to this project. It's not very well known, I would say, deterministic or statistics-based techniques for detecting faults. The easiest one we may think about is some form of voting mechanism. You attribute the system, and you do a majority vote on the decision on whether you will go left or right for your system. As we are increasing the level of or the number of sensors or the precision of those sensors, we may want something that is less binary. We may want something that is looking at more precise evolutions of the signals that we receive, for instance. This is what Georgia Tech is bringing to this project. They have developed a couple of strategies to detect faults or attacks, cyber-attacks, if you will, so that you can build fault detection mechanisms that are based themselves on reinforcement learning, so that we can look at specific patterns that are representative of either a cyber-attack or a fault happening in the system. This is where the boundary between security and safety is becoming fuzzy. Because from the perspective of the system, if I have a faulty sensor that is sending me periodically a value of zero, for instance, from the perspective of the system, it is difficult to tell whether the sensor is faulty or whether it is an attack, and someone else is tampering with my sensor. That is why we are blurring those two aspects. The idea of using reinforcement grounding for looking for a more aggressive fault or attack pattern is definitely relevant. Of course, there is the question of regressivity here because we want different learning component with another learning component that may itself be wrong. But, again, using architectural patterns, we can address those questions by expanding the fence around those components. That was basically the contribution of Professor Kyriakos Vamvoudakis to this project to help us implementing and testing those full detection, resolution, and recovery mechanism on some mission that we have implemented on UAV platforms.

Suzanne: One of the things that you began at the beginning to talk about and reinforced here is that role of architecture as being one of the determinants of… What you are speaking about now, one of the things I know is that if your architecture is not modular, does not meet certain properties, then being able to isolate fault tolerance kinds of things, you have to have a very big fence if you don't have a very good architecture, which means you are actually going to be doing a lot more processing to gain that trust and verification. Whereas if you have an architecture that you can verify is amenable to these fault tolerance techniques, then you can actually not degrade the performance by actually adding in the fault tolerance. That is fantastic trade-off if you can get that.

Jerome: Yes, it is a fantastic trade-off. One of the works that we are wrapping up for this project is basically a collection of design patterns that are known in the safety community for detecting mitigating faults to switch, for instance, using a simplex architecture to switch from one version of a component to another. You are right, there is definitely a trade-off here in, first of all, establishing the smallest perimeter. Because, as you said, if the fence is too big, we are not doing a good job in terms of resources that we are using, but also in defining what is a good trade-off for each of those patterns. For instance, patterns will differ in the number of redundant components needed. As we know, more redundancy is more costly because we need more CPU, more wires, more energy. Some of the patterns will take more time to make a decision or to switch from the nominal mode to the degraded mode, for instance. We are doing an evaluation of all those patterns, taking into consideration the cost to implement them in terms of resources, our resources mostly, and the timing aspect time, so the time to go from detecting an event to reconfiguring the system. This is this type of practical contributions that we are delivering as part of this project. Not just the format semantics of AADL, which is nice for computer scientist, but also this trade-off analysis that is made possible by providing a careful evaluation of each and every pattern that have been documented in the literature.

Suzanne: What is your ultimate vision for this search? Where are you really trying to get to over time? Because this is not a single two-year research project to get all the answers.

Jerome: No, it is just one milestone out of a very, I would say, large research roadmap that I want to develop. I am thankful for the SEI to allow me to execute it. At the completion of this project, basically, what we have developed is this idea of, first of all, this trade-off analysis on design patterns for safety, which is one aspect. The aspects that we have developed are this formal semantics of an AADL that has been expressed using the Coq therom prover, which means that it is not just a document, but it is also something that we can experiment [with] using a theorem improver, so that we can develop all type of analyzing. We already did connections with scheduling and analysis tools or fault analysis. The idea behind all of that is to address this question of what can we do with models when we do model-based software or system engineering? We have all type of techniques, and how far can we go into building a toolbox, so that designing a system can be supported by evidence step by step?

You want to build the architecture of your system which is nice. You will start building couple of boxes, wires connecting all of them. But can you make sense of this diagram? Yes, because we have this formal semantics. You have this formal semantics. You have this model. Great. What can you say about the safety or the timing of the system? *Well, because I have this formal semantics, here is a proof that the time to detect my error is X*. All of that contributes to building this trust that we are discussing at the very beginning. It is really this idea of rigorous model-based system engineering, going from requirements to a model and evidence so that we can build this trust package that can be given to an external auditor, so that you can check if it works. You may need multiple PhD in computer science to understand all of this. But, ultimately, we have a chain of trust across all those engineering artifacts that we have built. That is really this long-term vision that we are contributing to. The SEI by trade, by history, developed a lot of contributions in that domain. It is really this idea of pushing this forward, this idea of giving the right tool for making decision on asserting systems. This is one of those contributions that are making to this line of research.

Suzanne: I am involved in some programs that have a very long lifespan. Once the initial delivery is made of the cyber-physical system, that's not the end of it, right?

Jerome: It's never the end. Yes.

Suzanne: Right? One of the things that model-based systems engineering brings to that environment is the ability to analyze and play with the model before you actually make decisions about what the next evolution in the system or the next modernization is. As technology evolves, can we bring this technology in? We can model things before we make those decisions, which is much cheaper than a flight test for example. I am coming back to this again, the trust level on those models, their ability to evolve, their ability to reflect different aspects and to be able to mitigate security issues, safety issues, in a trusted way, the importance of that just keeps going up. Having these more formal ways of assuring people that, yes, we really do understand the behavior and, yes, we can predict the behavior and, yes, we can safely change the behavior—those are just critical. Don't go anywhere. We need you for quite a while here.

Jerome: I won't go anywhere. I am happy to be working at the SEI.

Suzanne: Good, good, good. All right. So you know that we like to emphasize transition. Now, we are very early in this research. I am guessing that a lot of the things you're talking about, if I were to look at them, I would roll my eyes up in my head because they are not in my area of expertise. But if someone has become interested through this podcast or other means in learning about more about assuring autonomous components in cyber-physical systems, where do you go to learn more about this besides going and looking at Jerome and having a coffee? What are the resources that are available for learning more about this area?

Jerome: First off, it is part of a project. I just realized that we did not name this project. So it is called SAFIR for safety analysis for time-intensive cyber-physical systems.

Suzanne: Every project has to have an acronym.

Jerome: Oh, yeah. And SAFIR is a small version of it. And, actually, we made already two research review videos on this project. So different things that are starting resources that are relevant to give you some insights of what we have done in the past. There are a couple of technical reports that we are finishing as well. As usual, any of the papers we have written during the project is also available in the SEI Digital Library. All of this will be linked, I suppose, when this video will be edited. All of that will be available there. As we did this project, we made something like 15 different scientific papers, different level of depths, technical depths, and scientific depths. All of that is a good starting point and some ways for them as well looking for the SEI. We are always excited to receive emails from interested parties to push the discussion further.

Suzanne: Yes. Excellent. Beyond moving forward with SAFIR, what else are you working on that we can talk about in the future? I know the answer, but I am going to let you tell me.

Jerome: Actually, it is kind of a continuation. I mean, this project I built is one stone in this big garden. In this project, we were interested in capturing the model itself, its semantics, and connecting it to other aspects. In the follow-up project, what we will look at is the modeling process itself. What can we tell about the modeling process so that instead of saying, *Oh, let's see this model, let's do A, that model*, we can tell you, *OK, this is your problem, and this is a way you may address it by applying modeling techniques this way*. The idea of this project, ultimately, is to give a map, so to speak, of modeling processes and modeling activities. Because by training, by experience, I know how to model a system so that I can perform safety analysis. But this is something that is very difficult to convey in a particular way. We will be looking at techniques, first of all, to document those approaches and to train people to do this. As we are defining those modeling processes, the other question that is relevant for our DoD colleagues is, how much does it cost in time and money? And how can I be sure that whatever I receive from my contractors is what I expected? There will be this question of defining some quality attributes for modeling processes. I am really looking forward to this new project, a little bit less formal but definitely addressing what I believe is interesting, which is how to make model-based techniques much more efficient and easier to transition in the industry.

Suzanne: Excellent. Yes. I am looking forward to that research as well since that is a particular area of interest for me. Jerome, I thank you so much for talking with us today. It has taken us a little while to be able to get together to do this, but so worth the wait. I want to remind our viewers that we will have links to SAFIR and all the resources that you mentioned from the project in France in the transcript, so people will be able to access those. A reminder to our audience that our podcasts are available everywhere you can think of, SoundCloud, Spotify, Apple, and, of course, my favorite, the SEI's YouTube channel. I hope that you will enjoy viewing this video and learn some things and be able to expand your views of what is important about assuring these autonomous cyber-physical systems as we move more and more into the autonomous world. Thank you once again, Jerome. I look forward to our next chance to have a conversation.

Jerome: Thank you, Suzie, for your time and enthusiasm helping us advocating for what we are doing. Thank you so much.

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. Thank you.