**Interview Transcription**

Date: February 19th, 2013

Time: 9:00 AM – 10:20 AM

Interviewee: Dr. Salamah

Interviewer: Dr. Ann Gates and Software Engineering class

The transcription below is taken from an audio-visual recording. The interviewee and interviewer were asked to review the transcription and to check for correctness. Additional information has not been added—only minor edits and clarifications. The transcriber used “…” when the audio was not clear and could not be transcribed. Brackets are used when the reviewers added clarifying words. Words or phrases that are not adding information may have been deleted.

Dr. Salamah: So, today my goal is just for five to ten minutes to go over what these mean and then I guess we’ll open up for questions.

Dr. Gates: So what we’re going to do this time is put together your questions, we’re going to go by each team and ask a question from your team that you feel has not been answered yet.

Dr. Salamah: Okay so, the idea of this pattern and scope combination is to help the scientist to come up with properties of interest. There are many properties that can be specified. So the use of pattern and scope helps them [the scientist] focus on what they want and actually triggers some questions within them to come up with the best possible property that matches the original intent.

So in the first description that I have here is well multiple diagrams of what a scope is. And a scope is just the portion of the data in this case that you care about the property holding in. So in some cases you care about the whole data set, you want that property to hold at all the data set and that’s where the global scope comes in. So you’re saying I want that property. I want the temperature level to fall below, I don’t know, zero over the whole set of data. So take that property over every single reading. That’s what the global scope is for that’s the simplest one.

Ah, the next one is the before R scope. So before a specific event happens, before something happens, that’s the part of the scope that we care about. From the beginning up until that point where that event or condition I’m concerned about holds. So if you see here this diagram has in the before R I have the blue parts is the scope of interest, the part of the data I care about that I’m gonna check my properties over, the rest of this I don’t care about. The property can hold or not hold I’m not concerned, it’s not my concern. Questions on that so far? Something important about the before R scope is that the scope does not include that reading where R holds. So the scope is only for those readings or readings within that set that receive what R holds. When R holds is not part of the scope. Questions? I am not sure you are getting this because of the way you’re looking at me. Are, or is it too easy?

Dr. Gates: So if we we’re looking at *before* *R* and we’re looking at real near time, there’s two ways we can monitor near real time and after the fact we have the data reading from last year or a week before and now you are going to go in and evaluate it. For *before R*, can you do that check in near real time?

Dr. Salamah: Probably not, it has to be over a certain data. The *before R* has to be over a certain data and is mostly about the pattern and less about the scope. Scopes come in when you’re dealing with historical data. You’re limiting what part of the data that you’re looking at. The next one we have to go over is after L.

So the next scope is your set of interest, set of data of interest is the one that happens after the condition L happens. So let’s say that the condition L is the temperature reaches one hundred, after that then look at all the data that comes in after that, that’s where my property has to hold. And that’s what we were talking about, if that’s the condition then that’s applicable to real time, once that condition happens now I have to report any time my property is violated .

Then comes the next two interesting scopes. The between L and R scope and the after L until R scope. Amm, for the first three scopes those are not repeating scopes meaning we have only one, single scope. For example the before R I only have one before R scope even if R the associated condition with R holds many times, I don’t care I care only about only the first occurrence of R.

You understand that?

Okay, so only the first occurrence of R is the one I hold up that I care about and my scope becomes every data set, every reading that happened before that R and the same thing for am after L. Of course global scope only has one, which is the whole data set. Amm this is a little bit difficult for the between L and R scope. The between the L and R scope I care about the interval between the occurrence of the condition L until the occurrence of the condition R. That’s my scope. But this scope could repeat so L followed by R holds I have a new scope. Right? So if I have L temperature is fifty, R temperature is one hundred then I build a scope. If those two sub conditions hold after that then now I have another scope after that. And with these I can have nested scopes. This is why it’s a little bit hard. Ah because the last L is merged with the first R. So if you see on the diagram that I have I have two scopes one with light blue and one with dark blue. My dark blue is actually this whole scope. It’s from the first L to the last R. Including that big scope that includes the light blue scope. But I’m checking within two scopes I’m checking the big one and I’m checking also the small one. You are rarely going to have nested what you will have the situation that I was describing before. You have an L followed by an R and after that another L followed by a R. That’s more typical than having nested scopes.

Questions so far?

Dr. Gates: Will that create a problem?

Dr. Salamah: The nested?

Dr. Gates: Yes. If we have the temperature, what we were talking about yesterday, temperature equals one hundred. So if this reading is one hundred, is it going to be expecting on the right hand side temperature reaching eighty, we might not be able to nest it right?

Dr. Salamah: And even if you nested this it’s really not coming up with the spirit of this, you’re going be just checking the smallest scope which is the last L, and again this is, this is not the typical type of scope that you will be facing. What I think that you will see more is condition L holding followed by condition R, that becomes your scope and then you start all over again at some later point. Okay? We’re good?

Dr. Gates: Did you talk about the open [interval] on the right?

Dr. Salamah: Oh! That’s a good point. So remember when we say that the scope for the before R, so remind me, the scope for the before R does it, what’s special about it?

Stops right before the R.

Good. With the after L it includes that state of the data reading where L holds. So it’s not the same, with the after L it does include that reading where L holds. Where the before R does not include the data reading where R holds. And this is the same for between L and R. It includes the reading for L but it doesn’t include the reading for R. Yeah? And if you see the definitions also the textual definition not only the diagrams that specify them. The after L until R is similar, very similar to between L and R. The only difference here is if R never holds then the scope is between the first L all the way to the end of the arrow. Or until you get an R if that’s the infinite truth. Or if you don’t you keep checking until you get an R then it’s until the very end of the readings. We’re good? Yeah? Okay.

Alright so, let’s move to patterns, for today we decided to just give you a subset of this and hopefully elicit some questions from you and we can go to more examples. So, we talked about scope is a range of interest, what I want my property to hold, well my property is specified by a pattern. A pattern is a sort of property that repeats over and over. So we are breaking these into time dependent and not time dependent and today we are giving you four examples of the not time dependent ones. So for example universality, that’s the first example that we have. Over that scope, now the scope can be whatever one of the five we described before. Over that scope, over that set of data readings I want that universality pattern to hold which means I want that property to hold over every reading in that scope. So if I have a global scope I want that reading, universality property to be true over every reading in my data set. So we talked about scope as array of edges where I want my property to hold, where my property is specified by a pattern. Pattern that you will specify over and over and over. So we are breaking this into time dependent and not-time dependent and today we will give you four examples of the not-time dependent once. So for example, universality that’s the first example that we have. Over that scope, I mean scope it can be one of the five we described before, over that scope, over that set of data, data reading, I want that universality pattern to hold, which means I want that property to hold over every reading in that scope. So if I have a global scope I want that universality property to be true over every reading in my data set. Okay? If I have universality between L and R, then I don’t care what happens before L, I don’t care what happens after R, I only care that this property holds in every reading between the L up until just before the R. That’s what universality means, that property and for example, here I said that temperature has to be smaller or equal to 35, that temperature has to be smaller or equal to 35 in every reading within the scope. So again, if I am doing global scope, the temperature has to be 35 or less on every reading over all my set of data reading or if I am doing after L, I don’t care what happens before L, but the temperature better be 35 or less after L until the end of the reading, until the end of my data set. Easy so far? Okay.

Absence is a little bit the opposite. That property should never hold in my scope. So we were saying universality is it should hold for every reading, absence should never hold in any of my readings, again within the scope you are always going to read by the scope and existence it has to happen at some point, it has to happen at least once. If it doesn’t happen at least once, then there is an anomaly or there’s a violation, and then the interesting one is response. If you notice on all the ones before this one, before the response, in all the ones we have discussed so far we are only targeting one type of reading, right? It’s either temperature or its humidity or something, but we are doing just one type of reading, with response we have two scopes because we have two data readings. I am reading about temperature, I am reading about humidity for example, and I am saying in response it’s a cost-effect, for example if temperature falls to 15, then humidity will behave in a different way. That is why we are saying that we are performing or we are checking two readings, two data sets: the data set for temperature and the data set for humidity. Now, somehow your job will be to calibrate those two data sets to make sure that that order is preserved, if the temperature happens at reading number 5, the temperature becomes 15 in reading number 5, then make sure that humidity becomes 80% at that same reading number 5 within the humidity data set or at reading number 6. It has to happen at that same reading or the reading immediately next. Questions?

Student: I think I’ll wait until you’re done.

Student: Yeah I think we can wait until the end.

Dr. Salamah: Well, I think this is my last scope.

Dr. Gates: Yeah maybe. In that handout, there are several things you’ll notice. One is that it tells you what the scope is, and it tells you what the Boolean statement is and the parameters. Let’s see what else . I want you to look at absence for instance.

Dr. Salamah: Absence is number 2.

Dr. Gates: If you look at absence, it is giving you a property right? It is giving you a Boolean relationship, but notice that absence is saying that that [the property] does not hold. I am not putting the “not” there. I am just saying this is what should not hold.

Dr. Salamah: Right, so we are not negating that property. What we are saying is that the temperature should never be smaller or equal to 15. What Dr. Gates is saying, when you are specifying the Boolean, the Boolean does not say that temperature is not smaller or equal to 15. That is something that has to be done during your implementation based on what the pattern that has been selected. Okay?

Dr. Gates: What you are trying to do is to figure out how to get this information, or how to help the scientist specify these properties, and there are a lot ways you can do it. You may use a pattern to somehow cue them, or maybe questions that can cue and these are things we want you to think about. We are at the point of saying what you need to do, the “what”, not the “how”, but you still should be thinking about what the system is able to gather. You are going to have to come up with some relationship, right? We are going to give you some other examples that are a little more complicated and it may be that you will have to do some math or some transformation with the variables that you are checking, sensor reading that you are checking. But we will--we are going to add another layer next week. We want to make sure that you understand the basics right now.

Dr. Salamah: One thing that I want to clear up and I was hoping that you will be asking this and maybe you were going to ask but I will clear it up anyway. When we say “not time dependent”, those are the type of properties that we are explaining today, then you see the example where it says “during the daytime of May 12th”, that is time right?, but that is not the time we are talking about. Time dependent we are talking about are within the property itself, so the property holds multiple times, the property holds a certain number of times. That is the time dependency we are talking about. Of course we didn’t give you any of these examples today, because we gave you the not dependent today, but time dependent doesn’t mean that there is no time associated with the scope because the scope has to be built with some sort of a time stamp over reading. So don’t let that confuse you. “Not time dependent” doesn’t mean that the property will not have any time in the textual description of it. Is that clear? Okay. All right.

Dr. Gates: We will open it up for questions. First questions from what Dr. Salamah has said, and then we will go back and you can ask questions off of your interview questions. So you had a question.

Student: From these terms that we are using just today, are we able to assume that the scientist or the users of the program are gonna know these terms or at what level of vocabulary?

Dr. Salamah: That’s a very good question. I think that is what Dr. Gates was talking about. That is something that you’ll probably need to investigate and see. The most important thing is for you to understand what this means, right? And try to interpret it the most with the language of the scientist and to what is more appropriate for the scientist.

Dr. Gates: So the scientist, as Dr. Pennington said last time, did not want to use these words. So when you study, you’ll notice and see the different levels of users. You should try to develop this where someone doesn’t have a background and then maybe other people, as they gain knowledge or expertise, then maybe they can very easily see “oh this is absence”. You won’t know that until the scientist reviews and you start prototypes. That is a really good question.

Student: For the response relationship, is that going to be over 2 or is it just going to be a two sensors relation between each other?

Dr. Salamah: At this time it is two.

Dr. Gates: Design for change, remember that it is one important principle. We are just sticking with two, but you should be assuming that it could be multiple.

Dr. Salamah: I am going to add for that point of design for change. If you notice this list of data that we have which is not final by any mean, this is different from what you got last two weeks or so. So after you develop the system this set of data is expanded, so design for change.

Dr. Gates: When we have two different sensors-- when we talk about two scopes

Dr. Salamah: That will be later when we talk about relations

Dr. Gates: Two different sensors-- we assume that they are in sync

Student: so they are parallel scopes

Dr. Salamah: Right so the Boolean statement might be in one single statement, so one part may apply to one scope and second might be applicable to the second scope

Dr. Gates: That’s a good question.

Dr. Salamah: typo for the existence. The formal definition should read “should hold” at some reading with the scope

Dr. Gates: We’ll correct it and post it on Piazza.

Dr. Gates: Team 1 questions

Student: question about the time stamps, is it just one gigantic reading from one period to the next, or is it some discrete intervals of time stamps that are read

Dr. Gates: That’s set up by the scientists. They can say were going to collect it by the hour or were going to collect it continually, or every 5 seconds, or every minute

Dr. Salamah: However it is set up, the scope is being built by distinct time stamps. If we’re reading every hour for example then we have 24 readings. Now the scope could be any one of those, and subset of those 24 hours

Student: Question for the formats. We talk about the interface having to display the information, but did we need to have us implement the need to download the information on a different format, like a text file document file.

Do you we need to provide ability to download reports on pdf or.

Dr. Gates: It depends on what you’re talking about with respect with information. If you’re talking about the data itself, you’re not concerned with downloading the data in any format--that’s the job of some other system. What you’re concerned more about is capturing the properties and then possibly graphing the data to show the relationship to the properties--doing a visualization of that data.

There should be a lot of questions on how do you visualize the data, and how you set that up. Dr. Pennington said that I could be in front of a machine, or I could be in front of this visualization wall. I may be showing temperature data. I might be showing a lot of data on this wall; if I’m doing near real time I may be viewing a continuous feed of data and how its mapping to my properties. I’m going to stop there I’m trying to prompt you to ask more questions. We’re more concerned about the properties, displaying the properties and getting a report about what the anomalies were.

Student: So instead of logging back into the system and having to request the same properties being displayed we could have a document that’s already premade, they can open that up and it already has that information

Dr. Gates: So if I hear you right, I’m a scientist in the field and I’m working at the Jornado [site]. I should already have a lot of properties. I should be able to reuse those properties, and should be able to look at those [properties], and change them-- select from that and create a subset [of properties].

Student: What I’m getting from him is that it’s ok to implement a caching system, he wants to download something and re-use it instead of submitting another request, so should we have a way to cache certain information

Dr. Gates: Caching is only so large and it [what is cached] goes away. If you keep on adding new information, that goes into your cache. The reuse for me is more of a repository. I don’t think it’s a cache. You could be monitoring over months. I mean you set it up, and you’re monitoring for months. Unless you going from site to site to site, what you set up is what you set up for that period of time. You’re not setting it up for 24 hours typically. You’re setting it up for weeks, so these files are huge. That’s a big problem for the monitoring side of it and were not going to worry too much about that. Still that’s more of a solution then …

Student: You were saying that the files of the storage might actually be of a greater size so that would effect the performance and effect the actual time advancement. So how would we go about informing the clients about that

Dr. Gates: Again separation of concerns at one level. The separation of concerns is how you deal with large files, and how you do the monitoring. It’s more of a concern when we’re starting to look at near real time. When we’re looking at near real time its streaming--you’re streaming information. Your system is not concerned about storing that information is being streamed and your just pulling it in real time. The data logger system is the one really storing that, or it may be transmitting it to some storage. So I would separate that because we’re not concerned with that side of how it’s storing it. If you start looking at historical data, then that’s a different story. Now we’re saying I’m looking at what I have in real time and I want to compare it against data that was collected last year, so that it’s stored some place. So far we haven’t heard about any requirements that deal with performance. Performance is a quality attribute. It’s important to talk about, and you want to ask the customer if they are concerned about performance. But it all depends on how you implement it. You may have this huge file, but you may just pull out this area/time of interest, which is not as large.

So again we go back to an implementation issue. It’s a good question to ask if performance is an issue; that goes back to the customer/to the scientist.

I like the kind of questions that you’re asking I think they are really probing so I want to hear from more people. Team 4?

Student: I know you touched on if the Rf does not exits, what if the Rl does not exist, do we just not do anything about that and keep moving forward every time?

Dr. Salamah: Your scope is not built, you’re saying Rl the left side of the scope. Until you get that ‘l’, until you get that first condition that builds the left side of the scope, your scope is not built, and you don’t care about the property. Once that happens then you wait for your ‘r’, or the reading corresponding to the ‘r’, and then your scope is complete. But if any one of these, if your ‘l’ is not there your scope is not built. If the ‘r’ for the between ‘l’ and ‘r’ is not there, your scope is not built, and you don’t care about the property anymore. You first have to make sure that your scope is built.

Student: So that’s always determined by ‘r’-’l’?

Dr. Salamah: ‘l’-‘r’ left-right

Dr. Salamah: L,R. Left and Right. But now the question that was asked is if we’re doing real time. You’re not going to wait until the scope is built, right? That’s the harder question.

Dr. Gates: You have to think about that.

Dr. Salamah: You have to think about how that’s going to be done. Do you understand the problem? You’re thinking existing between L and R, well L happened, now I’m looking for existence of some sort of condition. I don’t know, if I really cared about it enough, because I don’t know if the scope has been built yet because of the R. So that’s an issue that you have to resolve.

Dr. Gates: You’re going to be creating use cases out of this problem, right? One use case is near real-time. You have a display-- you’re finding potential anomalies because you reached R but you haven’t gotten to R yet. I mean go to L

Dr. Salamah: You reached L

Dr Gates: And then you don’t know if R is there, but you’re finding that this property is not holding. That’s a scenario right? What does the display look like? Just think about it. This is going more into the prototyping part because you’re going to be looking at potential solutions and you’re going to be presenting them to the user. When you prototype, to say this is what we think we should do, or you don’t do anything, right? There are a lot of decision points here, but [its] part of dealing with complex problems. Are you convinced that this is a complex system that you’re trying to build? I mean it’s not that easy-- you can’t just go out and build it. Maybe originally we thought, oh yeah we can create this, but once you start digging in, there are the different scenarios. That’s the complexity of it and trying to nail that down this semester before you start implementing it next semester.

Dr. Salamah: I’m just going to add something in terms of software engineering. What you’re doing right now is really one of the hardest things, is making sure you understand what needs to be done and whether or not you can do it. Whether it is feasible or not. So one of the things you really have to learn is to push back with the customer, and say you know what this, we just can’t do this. But now, you have to know how to do this and you cannot just say we can’t do this, you have to document why. You have to bring a convincing argument: why this can’t be done. The performance, for example, the customer requires specific performance attributes saying it has to be performed within this second, you have to do some calculations and say you know what, this is just not feasible. You have to learn that not everything the customer says has to be done , preferably yes, but you can push back, that’s part of your job, but you have to say why you are pushing back, why is this not feasible.

Dr. Gates: [Deleted discussion about Amelia Flores]. Okay so what questions do you have?

Student: Going back to that maintenance and change, I don’t know if it’s’ been asked already but, let’s say you have these sensors and they measure W, X and Y. Is it possible that a new sensor will come in and measures [everything plus] Z, is that likely at all?

Dr. Gates: Yeah that’s likely, that’s very likely. They’re adding sensors all the time. I don’t know if you’ve seen it; we’re missing our visualization person, but I can show you a [video] on this wall--you can see robotic trams that are miles long that are out in the arctic and collecting data (reflectance data). Then there's satellite imagery that’s also collecting data. There's a relationship between those two types of information. They have towers up in Jornado that are really high and then there's sensors at all levels of the towers. So they may be adding new sensors to that tower and they may want to say--okay now I want to look at the relationship between this reading with another reading or an historical reading to see if there’s something interesting that I can deduce from that. It’s likely that they could add new sensors, and so when you add new sensors, you still may want to reuse a property, maybe change the threshold. So reuse is going to be real important

Student: It was talked about last time, a little, that the sensor itself will feed data in a specific format, is it our job to turn that format into a unified format that our system can use or will that already be happening? And if a new sensor does show up, we have to somehow give the user the ability to create some sort of conversion?

Dr. Gates: Good questions, I’ll have to think about that one. I think it’s the job of the scientist. If we have a new sensor, to set it up in a way so that they can do comparisons. So I think it’s more their job to set that up, not your job, and not the system’s job. The system’s job should be thinking about cues or thinking about questions to prompt. So the question you asked I think is if you can prompt the user to say, is this calibrated with this [format] or this sensor is recording at a different level than this sensor-- providing information back. That’s my answer, that’s not the scientist’s answer. So I think you should document that question, it’s a good one, and I’ll validate it with the scientist.

Student: When adding new sensors, if the units of measurement don’t match, how are we going to display the sensors? Do we display them as two different units of measurement?

Dr. Gates: Right, we could overlay them on the same graph, that’s for sure, right? Kind of goes back to your question, are we responsible for doing translations and I don’t really think we’re going to do translations with the data. I think that, you can’t do near real time if they’re not the same [time stamps or measurements], but you can inform the user that the units are not the same and then a different part would be doing translations. That’s something that’s added, it’s not what we’re asking to do. The solution would be to do a translation of that data which is done quite often when you take a data set and then you transform it to whatever the time units [you want], because sometimes its Julian days sometimes its Roman and there's differences. If something is all the way to seconds and you only want hours, then you’re going to pull out only on the hours, that’s preprocessing-- it would be more preprocessing. So you would have to provide a warning statement.

Student: So they're going to want, not just the anomalies displayed but also just the regular data displayed?

Dr. Gates: No. I think that if you’re doing monitoring and they want to see it [the results] as data comes in, what it looks like and you're graphing that. They don’t want to see the data itself--it would be on a scaled graph. There’s two things as I understand it, and again this is a question that goes back to the scientist, that you would stream in the data and show it. Now there’s a lot of questions you're going to have to ask about how you’re going to graph it. Are you graphing it by the second, by the minute, by the hour?