Man-in-the-middle attacks evolved... but our security models didn't (Transcript of Discussion)

Hugo Jonker

Open University of the Netherlands

Hugo Jonker: Hi everyone. My name is Hugo Jonker. I'm from the Open University in the Netherlands. What I want to talk to you today about is that although we've known man-in-the-middle attacks for a very long time, we suck at them. We know man-in-the-middle attack, we think. We, security protocol people, think we understand man-in-the-middle attacks and then we find out there's all sorts of man-in-the-middle attacks.

You've probably seen Needham-Schroeder and you've seen Needham-Schroeder-Lowe. This is a typical man-in-the-middle attack. There's a man and he's in the middle. There are many examples of this in academic literature and we are aware of this attack. The original Diffie-Helmann key exchange is basically about establishing a cryptographic key such that a man-in-the-middle wouldn't know the key. In 1976, they were already basically thinking of man-in-the-middles. Lowe's attack on Needham-Schroeder's protocol is from 1995. That's really a long time ago.

There's quite a few things happening in practice as well. Just one example I put on the slide, Moxy Marlinspike. He's a guy making practical attacks. He has presentations often on BlackHat. He made a tool SSLsniff in 2002. This tool attacks Internet Explorer 5.5. We're currently past Internet Explorer 10. Internet Explorer is dead. This is how long we have known man-in-the-middle attacks. We know this. We understand that this is happening.

Let's see. Can we stop it? Well, we have distance bounding. The first theoretical approach to distance bounding is from 1993. Distance bounding means making sure that something is within a certain distance of something else. It's like when you have a smartcard to pass a door, for example, you don't want to be at the other end of the hallway. Nowadays, you have bank cards that you can pay contactlessly with. I'm still looking for a student to make sure that I can have my lunches paid for by my colleagues. This is exactly what distance bounding prevents.

Model checking. This is what Lowe did in 1995. I'm assuming everyone knows, but this is how Lowe found the attack on Needham-Schroeder. He model checked and then out pops an attack. We know how to find these attacks. Tagging, making sure that messages in protocols are used for exactly that one version of that one protocol and not somewhere else. Tool support, model checkers all over the place. There's a gazillion more.

In practice, we have certificate authorities, DNS certificate pinning. We've got this. Man-in-the-middle, ha, forget it. This ain't happening anymore, right?

Meanwhile in reality, anyone here heard of the POODLE attack? Yeah. That's a man-in-the-middle. It forces a downgrade of the protocol. That's not good. The FREAK attack, maybe also known to a few people here. It forces TSL – which is like *the* security protocol of the internet, the one thing that keeps the internet secure for most normal people – It forces it to use weak crypto. The Logjam attack, that's actually the same thing, downgrading to Diffie-Helmann export instead of to RSA export.

The DROWN attack. This is 2016. Basically, this is a bit more intricate but it's again an attack on TLS, the one thing keeping the internet secure, banking secure, anything secure on the internet for average people. Again attacked by a man-in-the-middle. All these are from the last 18 months. Four major practical attacks in the last 18 months and I didn't even look hard into it, all of them relying on a man-in-the-middle.

What happened? Seriously, we know how to do this since 1995. How can we have in 2015, 2016, how can we look back on 18 months of four major attacks that actually break, 10-20% of the servers on the internet? Where did we go wrong? How do we stop going wrong?

You might say these are all academic, in a sense. The attacks are genuine and they do break TLS. You can argue this is researched by academics. There are very clever cryptographers figuring out how to actually exploit certain cryptographic weaknesses using hundreds or thousands of cipher texts and they put it all together. This is not something your average Joe does, right? You're right. An average Joe just buys a man-in-the-middle device for cellphones.

This just blows my mind. We think we understand how a man-in-the-middle attack works. We think we can automatically detect man-in-the-middle. If you're a law enforcement agency in the US, you can buy for 20,000 euros, something that performs an automated man-in-the-middle attack on cellphones. Not only do we not really have this, we've gotten to a point where there is a commercial industry selling man-in-the-middles. We've really dropped the ball – enormously. I think this is a problem that people are making money out of the things that we're trying to make sure can't happen. I think that's a problem. It's not all that bad, I hope. Feel free to interrupt me if you completely disagree. Yes?

John Anderson: In terms of the economics of the protocol, the people who are most in control of the protocol are the telcos. I think as long as routine man-in-the-middle is only done by authorized law enforcement agencies under national remits, those national telcos have absolutely no incentive to prevent man-in-the-middle. Their customers might like it or might not like it , but it's not like the customers have any say in it. They're not going to say, "Well, I'm going to switch to somebody who is not using UMTS." When this starts to happen routinely, like it's common knowledge that this is happening routinely outside of a law enforcement context, well in that case, the telcos might want to act. This slide, I don't think the telcos care that the law enforcement can man-in-the-middle them because if law enforcement did not man-in-the-middle them here, they would just go look at things on the back end anyway, so the telcos are quite cozy with that.

Hugo Jonker: I can agree with that. But: the stuff that these things are man-in-the-middling wasn't thought up in 1995. That was later. When we had the tools. People started talking about "let's make protocols for 3G, for 4G" and then these things were invented from scratch. There were technical committees. We care about this – we should've been there. Somehow, we didn't make sure that man-in-the-middle cannot happen. I agree, now that we are in a situation where there are man-in-the-middle attacks, the business interest is going to say, "Do we care about this type of man-in-the-middle attacks? No." Unless we somehow make it a terrorist plot, businesses are going to say, "I don't care." Completely true.

We need to make sure we never get there again. We need to make sure that if a protocol is developed then it's not open for a man-in-the-middle attack. That's my view and that was my point with this slide. However, there are some things happening to mitigate part of this, fairly recently.

There's a recent paper at ACSAC14 about detecting these man-in-the-middle devices for cellphones. What they do here is they use specific non-security aspects, properties of the protocol. They're saying it's a cellphone. Cellphones have a nice feature where they say, here's a list of neighbor cellphone towers. They use all sorts of aspects like this. There is a neighbor list. You can request a neighbor list. You can request this sort of property. It has another function. Then they made a detection mechanism based on using these properties and seeing if they're consistent with what was expected.

This is one approach to do it. I like it. It gives a good answer. My problem with this is, if we were to do this then for every single protocol, we need to figure out, what are the extra properties. What in TLS are extra options you can support or not and how can we then establish a fingerprint of that and match that, every single time? There's a man-in-the-middle in between. He's in between. Why can't we figure that out? Why can't we make the protocol such that when there's a man-in-the-middle, he's in between, we see that he's in between so we're not talking to him, irrespective of the protocol used. Not relying on specific particulars of TLS or UMTS or whatever.

Another recent paper just appeared in February on the arXiv and that was looking into distance bounding for cellphone payments. The point here was they were looking into can you use a telephone sensor to actually verify at the moment of paying that the cellphone is very close to the payment terminal. Not that the cellphone is sitting there and the payment terminal is at the lunch register, paying for my lunch and not your lunch. In a nutshell the conclusion is: No. The error rates are horrendous for an individual sensor. For one individual sensor, this is no where near realistic.

This is a strong limit. The sensors such as gyroscope, light sensors, microphone, that sort of thing, if you use any of those to see if accelerometer, "Am I close? Yes or no." That's not good enough.

There's some specific examples but what I want to propose here is that we would move more towards a generic solution. There's a man, he's in the middle. He's in a place where we don't want him. That affects the protocol, not the

specific properties but that affects the communication. Can we detect how it affects the communication?

If we look at the problems I've sketched, we've seen these attacks on TLS and they were sort of all targeted at the initialization. These cellphone man-in-the-middle devices, they attack a completely different thing. They attack new properties. They attack things that we didn't think of that we wanted to keep secret but now that they're suddenly being exposed, we're looking at the protocol and saying, "we should have taken care of that." Something like call duration, contacts, location. That sort of thing.

In both cases, this underlying protocol is not claiming that they will prevent against this sort of attack. The security claims of the protocol do not cover these attacked properties. In a sense these are not attacks. They do not violate security requirements. We need to start accounting for these security requirements. I think this is where we went wrong. That there are security requirements that we did not make explicit and we did not incorporate into our protocols and therefore we were not checking for them. We can check for them.

We see two types of contexts. We have a protocol context, like initialization but also protocol meta data: signal strength, for example. We also see a different context, the context of the user. For example, location. This is something I would not like the protocol to leak, but it might. What we figured is we want to embed this type of context into formal security proofs. Based on the notion of agreement, data agreement, we have an initial formalization of something we call context agreement where two parties agree on the context. You can also do this without a trusted partner, which is a stronger requirement, harder to achieve and then you are sure about context. This is more like verification.

What is context agreement? Basically, it means, that you and your partner have the same observation about the context. This still allows a man-in-the-middle. It allows a man-in-the-middle who is strong enough to make sure that A's observation about B's context, is the same as B's observation about B's context. If someone is strong enough to fake both of that at the same time, then you can have context agreement, both parties agreed that the context looks like something, when in reality it doesn't.

Frank Stajano: I think that previously you had grievances about the fact that, "Ah we were stupid because we didn't consider all these things when we defined protocols." Now in here, when we look at the observed context, we will define what we are observing and so there will always be a case where the crook will do something which has not been in the definition of what we observed?

Hugo Jonker: Yes, true. This is the point where I'm saying, we need to account for these things. Protocol context and user context. You're right that there is a risk that we might be overlooking specifically new properties, but right now I see a whole new class of new properties, namely those relevant to a user.

Frank Stajano: What I was saying, the complaints that, "Ah, we were stupid because we didn't think about that." is too general, for anything to be done now. Whatever you put in the context, there will be things that haven't been thought of. So they could be smarter, they could exploit those next time.

Hugo Jonker: Yes. I will agree with you on this. However, I will not agree that we will get to a situation like this, where we have people, commercial institutions selling their services for man-in-the-middling. I agree that this is an arms race. You have people trying to man-in-the-middle and you have us, stalwart champions of mankind and we're trying to defend people. Here we are really losing very much and I'm not sure that we are aware.

Ross Anderson: Many of these protocols were designed 20 or 30 years ago and in fact, in 1994, I believe at the workshop and also in my thesis of the same year, I argued that robust security was about explicitness. That is about putting all the even possibly relevant context on the face of the protocol so that it would be authenticated by both parties and bound to current freshness in some way or another. My observation was that doing this eliminated essentially all the attacks that we knew about at the time. I still think that that's a reasonable principle and its violation by GSM, TLS et cetera and it just keeps on being violated.

Hugo Jonker: Yup. Essentially I can but agree and just note that apparently your warning in '94 wasn't caught well enough so I'm here repeating you.

John Anderson: If you think about the evolution of TLS. For instance if we said, the next version of TLS, we will require that you authenticate the list of protocols that you negotiated. You sent this, I sent this. Then we said let's change to this version of the protocol then we downgraded to this. Then we downgraded to that. Then we have to authenticate that record. That of course, will only exist in the next version of the protocol. By definition if you have a protocol downgrade attack, then you won't see that. Do you see this as something that, when we have to deal with the evolution of existing protocols that... Are you proposing another layer that can maybe be placed on top of this? Or are you talking about when we get an opportunity to start from scratch now on?

Hugo Jonker: No. Actually, neither. You're right that if there is a protocol downgrade attack, then you get into insecure territory. We upgraded them for a reason. If we make a new protocol, we should make sure that a new protocol does not allow a protocol downgrade attack. It allows protocol downgrade and not the attack.

Bill Roscoe: In a sense, how do you do this? If a downgrade to get to an insecure level, then the protocol is very likely to be able to manipulate the data to be able to justify the downgrade to our side. In other words, two people are going to see different data from the other one. But they will not be able to agree on the downgrade activity because the protocol was attacked, so they basically will get...

Hugo Jonker: I don't have a particular solution but I see a direction for a solution and that is that in the protocol context we should also consider all previous versions of the protocol. This is something we have to account for and we should also account for the fact whether or not someone is trying to cheat us.

Bill Roscoe: It seems to me if the protocol gets downgraded to the extent that we could reasonable expect insecurities. For example, the examples that

you gave us earlier on in your talk. Then that basically shouldn't be allowed, the user should be warned.

Hugo Jonker: Yes indeed, I agree. That's also a partial reply to John's question. There should be a point where we just say okay, this we toss out. I think browsers should be doing things like that and it is actually happening. I recently was surfing the Web and suddenly my browser said, you're not going to that website. It's insecure. You're not allowed to use https to go to that website. We're getting to that point, I think.

Context verification. The idea is that your observation of the other person's context, for example the cellphone tower, you as a cellphone user have an observation signal strength and all sorts of other variables about this cellphone tower that you're talking to. Your observation is correct. How do you define correct? That's tricky. On the other hand, I think we can do better than we are doing now, without having to rely on the fact that it is cellphones that we're talking about. So for wireless communication we have signal strength.

We have an example, just to wrap up. How does cellphone communication work? The phone and the tower share a secret key. Technically the tower doesn't but it asks the backend. What happens if you want to talk? You send your ID to the tower. The tower generates a nonce. It sends the nonce and a message authentication code using the key of that. Hash using the key of the nonce and you reply back with a message authentication code using the key of the nonce and your identity. This is the basic thing we have now. This is a gross oversimplification of the basic thing we have now and this is where man-in-the-middle attacks can happen.

Ross Anderson: You're suggesting that just as Google will now give you a Gmail warning if it thinks that the local government is trying to man-in-the-middle your email. That someone should sell you a cellular phone which will similarly give you a warning if it thinks that the local police force is trying to stingray you.

Hugo Jonker: Yup.

Ross Anderson: Well, that's an interesting idea. I wonder how many governments there are that will actually permit the unlicensed sale of such equipment to civilians.

Hugo Jonker: My point is actually, it doesn't have to be equipment. This other paper that I pointed out that already does this, this one. These people made an app. You can download it. You can install it and it gives you colors. Green, yellow, red. Green is we have information and the information is okay. Yellow is we have information, it's a bit weird. We're not sure. Red is definitely wrong and then it has gray for saying we're lacking information. We've never seen this tower before in my life. We don't know anything about it.

This is just an app. You can download it. By now, it's two years since, it's kind of hard for governments to stop this since the internet doesn't forget.

This is the oversimplification. What we are proposing to give you one example: use signal strength. The idea would be that the tower tells you the signal strength it is broadcasting at and you determine if the signal strength that is

claimed is realistic compared to what you're receiving. You have an observation of the signal strength. This is context agreement. You have an observation of the signal strength. You are receiving a certain signal and you can determine the signal strength there and you have a claim by the cellphone tower of the signal strength it is emitting at. Your cellphone then has to determine is this claim realistic given the signal strength I am receiving, given my observation of the signal strength. Therefore, you can add in a check at this point. If yes you can go on and if no, you're worried about a man-in-the-middle. That's one example of how we can actually start using this sort of ideas.

Conclusions, man-in-the-middle attacks not only exist, they've actually become commercially viable. They are preventable, I think. You pointed out the extent to which needs to be understood. As Ross said, if we just put everything that we care about in a protocol line it is genuinely everything that we care about, and we make sure the protocol secures that, then that won't leak. Your point, it'll not be genuinely everything remains standing I think.

Prevention, we should account for context. Context of the protocol, context of the user and this can be done with or without trusted partner.

Brian J. Kidney: Question about your example with signal strength. I wonder how easy it is to actually determine whether or not the context is reasonable because signal strength depends on distance which would be the easiest ... approximately how far it is from the tower so based on normal losses it should be this, but then you have buildings and different building materials and different antennas and different phones. I could be here sitting next to John and we could have totally different signal strengths and has to do with sort of the properties of the phone plus everything that's in between us and the tower.

Hugo Jonker: Well, so the properties of the phone, this is something your phone should be aware of. This is something local that can be determined. The variations between two phones for neighbors, based on the physical properties of the phone, this is something that you should be able to bake into the phone.

Brian J. Kidney: Yeah. There are other things that could come in place like different phone cases. The original antenna problem with the Iphone with the hand in the wrong way so that the signal strength drops.

Hugo Jonker: You do have a valid point. My example was not meant to suggest that just based on signal strength we will have a unique way of determining, this is exactly that one tower, there's no man-in-the-middle. Our idea here is more similar to distance bounding and distance bounding they use the speed of light, which we know is a hard upper limit, and then they say: "Okay, the speed of light would have given us this much distance, so this means this item is within 300m or 10km of the other item." Same thing here. If we correlate what we observe plus what is claimed, then we could say, "Hey. Are you standing next to the tower? Because I'm observing ten watts and the tower claims it's emitting at ten watts so you must sort of be standing next to the tower." Then you look around, and can decide "No".

It's not perfect but it is a piece of information that will limit an attackers ability somewhat. Using more of these, my hope is that we can actually limit it so that it becomes actually functional.

Frank Stajano: There's a classic problem here of the technology is giving somewhat ambiguous warning to the user and the user who has something that they want to do has to decide what to do. What is the user supposed to do, when they get the yellow warning? Well, maybe the tower is not very . . . Ignore that. I have to get on WiFi because I need to get my email. What's the response that the user should have? Should they just abort and not get their WiFi which is what they really care about? They don't care about this, you know, "Maybe man-in-the-middle. Maybe." Even the technology is not sure if there's a man-in-the-middle.

Ross Anderson: What maybe your laptop should do is turn your screen share saver into Jaws. Just be careful about what you type. The NSA might be listening.

Hugo Jonker: I think it's a good question.

Frank Stajano: If it gives a warning that doesn't really have a definite yes or no. It's just dumping all the responsibility on the users. It maybe or may not be a man-in-the-middle but now it's your fault if you get man-in-the-middle. It's not the protocol's fault anymore, it's your fault for deciding to go ahead.

Hugo Jonker: I think our first step should be to actually know that there's a man-in-the-middle.

Frank Stajano: You don't know. You're just giving a yellow warning. You're not giving a red warning.

Hugo Jonker: Right now, we're not giving a warning at all. Right now we have no idea. If we're giving a red warning, which is what the other paper is doing-

Frank Stajano: If you're confident about red warning, you should just shut it off. You should not let them proceed if you want to warn of man-in-the-middle. If you're just giving the ambiguity then the person says,"Well, if they can't figure it out, how can I figure it out?"

Ross Anderson: But Frank, ambiguity may be good. In the real world that we live in, there are many analog cues. People's body language, are they relaxed or aggressive, their tone of voice. What sort of neighborhood is it? Is anybody painting the window frames, fixing the windows around here? Or should I maybe go with three big friends and some guns in our pocket? We are used as a species to dealing with many analog signals but in the digital world we have completely cut them out. Another digital device has become evermore complex and evermore difficult to secure. Your mobile phone with, what, 50 different CPU's in it? Who knows where all of the code comes from? Perhaps, we need a studied shift towards more constructed ambiguity, so that we know when we should be slightly skeptical.

Frank Stajano: I object. I object to technology giving an ambiguous warning to the user. I'm dumping on the user the responsibility for taking a crap

decision. If even the technology can't tell whether there's an attack or not, how can the grandmother?

Speaker 3 - Bruce Christianson?: I think Ross' point is, is that in the real world, we rely on the environment to give us the cues and then our judgement to make a decision. We don't want this box to make the decision for us. We want to say, "Well, am I comfortable doing this particular transaction in this context or not? Or do I need more nuanced information about the context, rather than binary yes or no".

Ross Anderson: The problem with most digital devices is that they're built and operated by companies whose incentives are not ours. Facebook's incentive is to give you the impression that you're in a private walled garden with friends so that you'll give over as much private information as you possibly can, so that they can throw it to the spammers. That's the problem. There shouldn't be a shiny, private walled garden that should have deep shadows around the edges. That makes me very, very aware of all the lurking, menacing things out there. Of course Facebook doesn't want that to happen because then there would be less information for them to sell to the spammers.

Hugo Jonker: I completely agree with Ross and to take Ross' example: would I count my money in the street, that's my decision. If I do that somewhere, you know, in broad daylight, where I can clearly see that I'm in a secluded corner but in open view, there's no one threatening around, I might feel comfortable doing that. If I'm in a rundown, graffiti stricken, needle stricken part of town around midnight and I see all sorts of people cloaking their faces walking by and making shady deals in the corner, I might feel differently about it, but I might still do it. It's up to me, it's not up to my wallet.