Collaborating as normal: detecting systemic anomalies in your partner (Transcript of Discussion)

Simon N. Foley

University College Cork

I'd like start with an analogy of the problem that we've been thinking about recently. Consider a bank ATM. The provider is the bank who provides this service. Within the bank they use various security controls; the simplest control is your ATM card and your PIN, and maybe there's a chip there as well. The bank also has terms and conditions about how you're allowed to use the ATM to withdraw cash. To go with this is a security infrastructure that the bank has put in place in an effort to ensure your ATM transaction is secure. Our view is that for such a complex system, one will never be able to articulate, or describe, all of the necessary security controls, or security mechanisms, that make the system secure. Therefore, in practice there's a lot of other things that the users of the system are doing that contribute to the overall security of the system.

Consider how the bank ATM is used in this picture[1]. You can see that there's a social norm, which is that when someone is standing at the ATM machine, other people stand back at a distance, and this gives some sense of safety, or security, to the person using the machine. You might say, that's just security theatre, but if that's the case then, would you use this ATM[2]? Consider this second picture; there's no queue, its seems to be a free for all. We have here what we think of as an anomalous norm. The social norm of orderly queuing is broken and as a consequence we're somewhat concerned about using this particular system; it does not feel secure.

Frank Stajano: But usually, when you look at the first picture, you would imagine that the person being one metre back is so that they cannot see your PIN. Now in the case of your second picture it's like they are they going to snatch the money when it comes out of the machine.

Reply: Yes, while the bank will install some security mechanisms, there's additional things that people do that help to make the overall system secure. You can think of the bank's security controls as being the regulation, those known things that we're trying to defend against, but then there's all of these social norms, which are all these other things that are, if you like, unspoken, or unspecified, that people do to make the system secure.

Jeunese Payne: Is it not the case that people just don't like to sit next to each other, or stand next to each other, like on the bus everybody will sit in a seat. I'm aware that I'm doing it right now, but is it not just the case that they

[1] Points to a picture of an orderly queue of people at an ATM.

[2] A disorderly ATM queue http://commons.wikimedia.org/wiki/File:ATM_Masalli.jpg

don't care about the security, they just don't want to be seen as imposing on somebody else's personal space?

Reply: So would you be happy to withdraw money from this second ATM machine?

Jeunese Payne: No, I certainly wouldn't; not necessarily because of the money thing, but because everybody is around me: I'd think, oh what are they doing, this is my personal space. So do you think maybe it would be, is it really to do with, how do you know that it's to do with idea of security.

Frank Stajano: But then how would it apply to the other people in the queue who are just next to each other before the last one.

Peter Ryan: If you go to the first picture then the spacing between people, there's a very poignant space between them.

Jeunese Payne: Yes, there is a poignant space that's a good point.

Reply: It's a fair point, but if you're in a movie theatre on your own and somebody comes in and sits two spaces from you, then that's strange.

Jeunese Payne: It's a bit creepy.

Reply: Our position is that security is provided by both the the bank's security mechanisms and the social norms. In this case, because of the breakdown in the social norms and even though they're not violating the bank's security policies or mechanisms, we'd still be somewhat concerned about using this ATM.

Peter Ryan: It looks like that one's spewing out money, doesn't it.

Reply: Yes, thanks Peter. Here's another picture of an ATM machine where somebody had accidentally put in, I think, £20 notes into the £10 note slot, and everybody was getting twice the amount that they withdrew. In this case, its not that there was a breakdown in the social norm, the breakdown was in the actual security mechanism itself. And of course what's interesting here as an aside is that, in some senses, because the mechanism has failed there's quite a different social norm going on here, there's people looking over their shoulder saying, you know, how much are you getting, and people seemed quite relaxed with that.

Our paper is about a systems equivalent to social norms. The outline of the talk is as follows. Our view is that when you consider security then there are the security controls that are prescribed and put in place to defend against those known threats, but then there are also these other patterns of normal behaviour—in society we think of them as social norms—which are a sort of a compliment to regulations and laws. In computer-based systems we call them behavioural norms, and I'll show you some examples of those later. What we're interested in, is when a consumer is using a service, the consumer is interested in detecting anomalies. Not just the anomalies from failures in the known controls that are are put in place by the provider, or even by the consumer. For example, if I'm browsing the web and visiting websites, then I'm going to install various security mechanisms into my browser to help stop, for example, cross site scripting attacks. Thus there's a number of things that I should try to do myself as a consumer, but then again, I don't know about everything that might happen. What I'd like to do is to make sure that when I am using a service that there's a norm in the way I'm I'm using it. The way that I'm using it is similar to how I've used it in the past.

We're going to look at a very simple example about photograph hosting and printing services, and how people might use them. The main idea is that we've got security mechanisms, but the security of the system is more than just these controls, it's a combination of controls plus these patterns of normal behaviour.

The setup is as follows. We have a consumer who is interacting with various provider services. The consumer has access to a service log from the provider, which is a log of the actions, or API calls, a record of what has happened on the provider on the consumer's behalf. What the consumer wants to try to do is to build a collection of what we call behavioural norms that describe their patterns of normal behaviour with the provider. From the provider's point of view we know that they have some security controls in place, but they are possibly incomplete, perhaps because they've misunderstood them. Also perhaps, there's an internal attacker within the provider, or the provider itself could be the attacker. This is not just a simple internal attacker, it's what we call a systemic attacker. The provider might also get the consumer to sign up to a service agreement to avoid problems. From the consumers point of view they judge security as being normal based on their past interaction.

Consider the photograph hosting service. Frank, the consumer, is interacting with a photograph hosting service, that allows him to upload, share, comment, and view, on photographs. Its a very, very simple system. The photographs can be either public or private, so that when I upload a public photograph it means everybody can look at it, or if I upload a private photograph then it means that I have to explicitly share it with other people. And then there's the service default setting, which is private sharing. We'll assume that the hosting service has the usual terms and conditions that it won't misuse your photograph, and so on. Frank likes free services, he wants to share photographs with his friends, but he also likes his privacy, and he's going to rely on the default privacy setting. A very, very simple scenario.

Here's an example[3] of the log that might be available to Frank. What he'd like to do is to ask, what's my normal interaction with this particular service, or, what is this (what we call this behavioural norm) repeating pattern of behaviour? Frank might look at just the actions in the log and ask if there is any particular repeating pattern of behaviour here. Maybe he'll do something like an n-gram analysis and try to build a model from that. However, there's nothing terribly interesting going on when we consider n-grams built from just the action name. If we also consider the context of the action, what we have in this line of the log is that Frank, as himself, uploads photograph 23, and then as himself he shares that photograph 23 with Lucy, and then Lucy, who's a friend, views that photograph, and then Lucy, who's a friend, comments on that photograph. You can see that this is the pattern of behaviour for image 23, and is also the same pattern for image 24. Olgierd's been working on a scheme whereby you data-mine your logs to extract these behavioural norms, these transaction-like

[3] Points to the log in Figure 1 of the paper

behaviours, these patterns of behaviours in the logs. In this slide we've identified one example of a behavioural norm in the log. And of course we might identify other behavioural norms if we had a sufficiently large log. For example, here in the log a friend has a similar norm, which is: he uploads photos, he shares them, and then people can view them, or he can view it. Or here's another friend who uploads photographs and in this case Frank can directly view the photograph and comment on it, so it doesn't require an explicit share. In this case, this friend of Frank must have the public setting for his photograph sharing options.

Suppose that the previous log represented past normal behavior for Frank. The log in this next slide reflects a slightly different arrangement. The service provider decides to change the upload default setting from private to public. Perhaps their motivation is that if they can get more visitors looking at photographs on their website, then they get more advertising revenue. If everybody is sharing photographs privately then that's not enough public photographs, and not enough visits to the website. The service provider decides that they're going to change the upload default setting, and of course they send out an email to all their subscribers, and, of course, nobody reads these new terms and conditions. As a result the service log for Frank changes, you can see that: he uploads image 24, and then immediately Bob, a friend, can view it. Frank doesn't have to explicitly share. As a result we have a pattern of behaviour that's not in Frank's norms. Frank could use an anomaly detection system to alert him to these different behaviors. For Frank the social norm has changed.

Frank Stajano: Isn't this rather similar to what you were saying in the context of Dieter's talk that at the beginning you don't know what's normal, and then you sort of rely on something becoming normal, and it's a bit like the identity, you don't really know where to anchor that.

Reply: Right, absolutely. Maybe I've never used this photograph service before and I'm somewhat concerned about using it, and about using it properly. However, I see myself as being someone like you and therefore feel it would be safe if I adopted your norms as my norms. I'm happy enough following your normal patterns of behavior. But of course, the question is how do we bootstrap this.

Michael Roe: Isn't this what the intrusion detection plan is usually talking about, where they're mining logs of accesses that are permitted by policy to ensure that something that is happening is normal.

Reply: Yes, I think there are two differences. One is that intrusion detection is usually about intrusions or extrusions, where you're looking for anomalies in your system, or anomalies that your system is causing to its external environment. One slight difference in our case is that it's a consumer who's using a service who's looking for anomalies in the way that the consumer is using that particular service. That's not a big difference. The other difference is more technical. Consider anomaly detection schemes that are based on checking normal behavior, such as the original work by Stephanie Forrest on models of self. Built using n-grams, these Markov style models are less expressive as they were not used to distinguish transactional behavior in the system. Our position is that it's not just a case of discovering simple correlations between sequences of events in the log. There are separate sequences, which in our model correspond to the transaction like behavioural norms.

In addition to this, many of the existing anomaly detection systems require an a priori identification of the attributes in the log that are considered relevant, prior to mining the log. For example, in building a model of normal behavior, one might decide to mine Frank's log based on just the action attribute while not realizing that it is only by also considering the photo-id attribute that transactional like behavior is discovered. Olgierd has been investigating how, by analyzing the log, one can also discover what combination of log attributes generate the most precise behavioral norms that describe the system.

Olgierd Pieczul: One point is that some of the log attributes may not be obvious to an administrator or to someone configuring the anomaly detection system. Certain attributes may, in some unexpected way, contribute to the system's normal behaviour. Also one should bear in mind that these behavioral norms may identify completely unexpected behavior patterns in the system that works according to particular parties. Thus its valuable to consider all of the attributes and look for those that can result in repeating patters of behavior.

Alastair Beresford: What's the incentive for the provider of this free service to provide the log, why would do that?

Reply: That's a fair question. A malicious provider could decide to provide an incomplete log and while Frank can collect some of the information based on his interaction, there are other interactions he cannot easily discover for him self. For example, he does rely on the provider telling him in the log that Lucy looked at his photo.

Frank Stajano: Insofar as it's an adversarial game between the provider wanting to open up the settings, regardless of what the users want, then it does make sense to ask Alistair's question.

Reply: Absolutely, yes.

Olgierd Pieczul: The idea here was to have it like a wall with some information that the provider would show to the consumer, and in this host model it could be something like a basic blog, so its in the provider's interest to release the information. Also, there might be regulations requiring the provide to show that information.

Alastair Beresford: And there might be other, as you say, Facebook's motivation, might be to get more users as well as to see more sharing and more ads going out because they're all interlinked because in that case the log, they have some interest in doing the log.

Rebecca Wright: Certainly I think there are privacy and policy implications on the log itself. And you had the one where, I think you had Frank inferring that Bob made his photos public by default, and that may or may not be something that you would want shared, and that that policy itself of sharing that in the log could change.

Reply: Yes.

Alastair Beresford: Photo's gone private here.

Reply: Yes. There's another example related to Alistair's question. It's an example that Roger Needham gave, which was that if a bank wants to improve its internal security then it should include details of who looked at your account on your account statement. In our case have to trust that the provider is going to provide all this information.

Frank Stajano: And also you're putting a non-trivial burden on Frank to check the log in a meaningful way because, you know, the way you look at bank statements, OK, another one, that's fine, you don't even read in there unless you are worried about stuff that might happen. And so here there has to be quite a bit of analysis from Frank in order to detect those things.

Reply: Yes, and there's two related observations. The first is that one would hope that maybe there's some automated support that will do this for Frank. The second, which I think the more interesting part of your question, is, how does Frank recognise that this is an issue? Perhaps the system has identified 200 different norms based on Franks past behavior. How does Frank decide how to interpret a previously unknown behavior that doesn't match these previous norms? Should Frank be concerned? You can imagine that the more norms we have then the harder it will be for a user to determine whether or not he has a problem.

Alastair Beresford: Another thing that's just popped into my head, they did some trials I think in the US where they put fake entries into the records, of people who were in hospital that had famous names, and went to see who looked at them. I don't know quite what the analogy would be here, but whether you could inject fake items into the log, or would be a useful thing.

Bruce Christianson: Following up on Frank's point, a lot of fraud detection mechanisms are simply trying to detect a typical transaction, or transactions where it is worth bothering the user by popping up a box, not endless boxes saying, allow or deny, but boxes saying, how do you feel about this transaction. And I guess you almost want, three answers, definitely yes, definitely no, or, yes I feel ambivalent about this one, because that says you're drawing the line in the right place.

Reply: If we know what valid transactions should look like beforehand then its relatively easy to look at the log and flag potentially anomalous behavior. However, what about the situation where you start to see repeating patterns of behaviour in your logs that you never expected and you don't necessarily know what they mean, mostly they are OK, except that there's this peculiar transactional like behaviour to the others. Maybe your looking at a cloud provider and when you start going down at the lower, layers, you start seeing some very strange behaviour down at the file system level due to caching, and it's this repeating pattern, and you've never put a name onto it, but perhaps it's important for the security.

Bruce Christianson: But I guess what you wanted to flag there is, when it changes.

Reply: Yes, that's it, so you know something has gone wrong. So it's like the social norm, and like the case of the social norm, it requires a analogous study to identify and recognise true norms.

Frank Stajano: There's still a shift of the burden towards the user who might not understand and says, you know, the system is very good at flagging anomalies and it says, well I'm just a system flagging anomalies, I don't understand what is security relevant really, I just see that it's different, so you user, what do you think about this, do you feel ambivalent, do you feel it is dangerous. And the user says, I don't understand this stuff you said, file system, caching, and I don't have a computer science degree, I just want to share my photos, and I will just go to another service that doesn't ask me these difficult questions because it's just too scary for me, you're putting all the responsibility on me to ask, and, you know, decide which one I should do of things I don't understand. So I think we should protect the user from having to take the decision themselves on the ultimate top-level part, which one is dangerous and which one isn't. How are they supposed to know.

Reply: Yes, Olgierd discussed a related issue this morning. Ignorance is bliss when it comes to security, and in some sense it's perhaps a case of you're better off not knowing that your norms are changing because if we start trying to tell you about all your norms then it's an overload, and you can't cope.

Nikita Borisov: But I was trying to understand, norms have two different meanings. Here you seem to talk about usual behaviours, what you usually do, but when I think of norms I often think about what is the expected behaviour, so to see the difference, for example, I always cross my hands like this, this is my usual behaviour, but nobody expects me, nobody would say, oh you're doing something wrong, if I did it like this, right. So are we really talking about social norms in terms of the expected behaviour from a social perspective, or are we talking about things that just have some behaviour.

Reply: These were social norms. For example, safety in society is made up of not just laws, but also the social norms that people follow. We can't rely just on regulation for people to feel safe in society, there's all those other things that people do which contribute to somebody feeling safe in society.

Nikita Borisov: So then I think that there's actually maybe a distinction between what typically we think about anomaly detection and these norms. For example, if I normally log into my work account at 9 am and I started doing it at 8.30 the next week, that's nothing to do with social norms, but might show up as a big anomaly.

Reply: Yes, the social norm is the analogy. We think of society, which is made safe by a combination of regulations and social norm, and then for a computer system or enterprise, we think of the security of it as being a combination of security mechanism, which is regulations, things we know about, plus these other behavioural norms, these repeated actions that perhaps don't apparently have any bearing on security, but have this repeated transactional like behaviours which are occurring in the system.

James Malcolm: I think when somebody breaks their norms you don't just have to say, stop. Uca Moriyama, a couple of years ago, suggested that all you need to do is change the screen colour a little bit to say, you may be in a slightly dangerous area here, you're doing something unusual, just be careful.

Reply: Yes.

Olgierd Pieczul: So like with the ATM example, if you really need to get money from the ATM then you would maybe still use it if there is a guy behind you.

James Malcolm: Exactly yes.

Dongting Yu: So earlier Alistair mentioned why the provider might have to provide logs. Suppose there is such a social log and the provider is forced to provide the logs, then what's preventing the provider from basically decreasing the signal to noise ratio in the log? One of the techniques might be just to insert records that you can't trace.

Reply: Agreed. In the next example Frank uses a hardcopy print service to get physical copies of the photographs, and this print service interacts with the hosting service to get access to the photographs. You can imagine the scenario as an OAuth style protocol whereby the *intention* is that Frank gives the print service temporary access to the hosting service so that it can access Frank's photographs. Frank gets to select those photographs and then prints them out. Again, we've got service logs that are being collected from both of these services.

Kosta Beznosov did a study on how OAuth was being used in practice and he found that it was not unusual for services to issue much longer delegation credentials than were required. In our example, its the print service getting a credential that it can use over to assess Frank's photos over and over again even when Frank hasn't necessarily initiated the request.

We have two simple logs[4] where Frank visits the print service and initiates a new printing order. Following OAuth, the print service then logs in as Frank to the hosting service and gets a list of the photographs. Frank selects the images that he wants to print, and then at the print service, the print service then, at the hosting service as Frank, gets the full sized images, prints them out, Frank submits the order.

If we were to look at the log of just the hosting service, you can see that there's a norm like behaviour where the print service, as Frank, lists the photos, and then gets thumbnail and full-size of an image. There's a very simple kind of norm pattern of behaviour here. If the print service was to subsequently, at a later date, access the hosting service without being initiated by Frank, then this would still be a valid norm. There's nothing here that involves Frank's participation when the print service accesses the hosting service as Frank. But if we considered the two logs together, and build a norm out of the two, you can see here that we have a norm, whereby Frank looks for a new order, lists photos, gets thumbnails, selects the photo, gets full-size and submits. And then, if the print service was to try to login as Frank at a later date with an offline

[4] In Figure 3 of the paper.

permission, which it had gotten from a previous print transaction, then you can see that it would break the norm.

These are two very simple examples. When we think of the security of a consumer using a provider then then it's a combination of whatever security controls might be in place, plus these behavioural norms. The view from a consumer's perspective is that if there's a change in norms then that perhaps points to some kind of anomaly. This is neither intrusion detection nor extrusion detection from the consumers point of view. We are also considering is mimicry attacks whereby an attacker attempts to generate a behavior that fits within the known norms, yet at the same time does something malicious.

Dongting Yu: How do you propose to differentiate between evolving norms and broken norms?

Reply: At the moment we haven't considered that. We've focussed on how to data-mine norms from system logs. You could look at a log, and if there's a small change in norms over time then perhaps that's acceptable. However, you wouldn't be entirely sure since some of the norms could be critical, and other ones less so. That's future work.

Shishir Nagaraja: I can emulate good behaviour and then I can be malicious. How do you address that?

Reply: I think that's the ideal scenario. If the provider is a well-behaved provider and does everything exactly as one would expect then from that I can build up the set of norms that represent good behaviour. If the provider then begins to be malicious, as in the example of the photograph scenario where the provider changes the default settings, then that's something that I flag because it's different from the previous behaviour.

Partha Das Chowdhury: I presume you can also spot other users being malicious, so if you had a system where you could share pictures and your friends had never done it before, and suddenly they start sharing on your pictures then that might be something.

Reply: Yes, if it's in the logs. However, I don't think Facebook will tell me whether my photograph is being used by somebody else.

Bruce Christianson: So that's part of their privacy policy.

Reply: Yes.

Henry Tan: I'm still curious why you don't look at it from the other point of view where the provider is looking at his access logs since he has all of them, and making sure that all of his users are using it innocently.

Reply: Yes, that's a good point, we could also do that. We just focused on the consumer who is interacting with a provider, and is not entirely sure whether he can trust the interaction with that provider. Of course the flipside is a provider who wants to check whether his consumers are behaving properly.

Henry Tan: Yes, and then they don't have to provide logs or anything because they have the logs.

Reply: Yes, they have all the logs. Olgierd has been looking at discovering norms from very large logs of transactions from a large-scale enterprise systems. What he found is that one discovers the patterns of behaviour from the high-level application level calls, as expected. However, as you increase the granularity of the event log and consider the lower level/system calls, you start seeing new and interesting behaviours that suggest other patterns that were not a priori anticipated. One might ask, if there was a change in these low-level patterns then does that mean that there's something wrong with the system? We're looking into this.