# Transparency Enhancing Technologies to Make Security Protocols Work for Humans

Alexander Hicks and Steven J. Murdoch

University College London
{alexander.hicks, s.murdoch}@ucl.ac.uk

**Abstract.** As computer systems are increasingly relied on to make decisions that will have significant consequences, it has also become important to provide not only standard security guarantees for the computer system but also ways of explaining the output of the system in case of possible errors and disputes. This translates to new security requirements in terms of human needs rather than technical properties. For some context, we look at prior disputes regarding banking security and the ongoing litigation concerning the Post Office's Horizon system, discussing the difficulty in achieving meaningful transparency and how to better evaluate available evidence.

## 1 Introduction

The theme of this year's workshop, security protocols for humans, highlights the importance of understanding the human context in which protocols are deployed. In particular, as computer systems are increasingly used to make decisions which will have significant consequences for the people involved, it is important to understand the interplay between the meaning of security for those that execute the protocol and those that are subject to the decisions of the protocol.

One important aspect of this is how failures of a system affect different parties. The fact that failure does not always affect the party responsible for the failure is unfortunate but has already been discussed in the context of security economics [1]. Last year we also broadly discussed the related role of incentives in security protocols [2]. The takeaway from this work is that of course incentives matter, but implementing them is hard. In particular, it is important to provide a way of enabling them: evidence.

Producing evidence in the context of computer systems has already been covered to some extent in the context of banking security, specifically the EMV (EuroPay-Mastercard-Visa) protocol now used for smart card payments worldwide [8]. The idea presented in this paper is to reveal some information about the state of the system and the execution of a protocol, creating evidence can be produced that can help resolve disputes. Cryptographic tools can also ensure that the evidence is "correct".

But evaluating evidence is not as straightforward as simply making sure that it is correct in the cryptographic sense. Looking at the legal notion of evidence,

many complexities arise. While those more mathematically inclined might like to consider evidence from a Bayesian point of view, doing this in practice is not always straightforward [9].

## 2 Resolving civil disputes

A scenario in which security protocols play a central role in a legal case is disputed card transactions. Here, a customer denies responsibility for a card transaction by claiming it was unauthorised and therefore is entitled to a refund under the Payment Services Directive 2 (PSD2). The bank instead claims that the customer either indeed authorised the transaction or was grossly negligent and therefore is responsible for the transaction. A central component of the bank's evidence in such disputes is the outcome of the EMV protocol run between the card issued to the customer and the terminal operated by the merchant.

As this is a civil dispute, to resolve the dispute in their favour, a party need not show what has actually happened or even that any explanations reach a particular probability of occurring. All that is required is that a party demonstrate that, given the evidence presented, the explanations in which they are not liable are together more likely than those in which they are liable. For this purpose, the odds form of Bayes' theorem – Equation (1) – is appropriate. While a judge would be unlikely to apply this formula numerically, Jaynes has shown that Bayes' theorem naturally follows from a logical application of intuitive principles [5]. If the posterior odds are less than one, then the party is found to be not liable.

$$\underbrace{\frac{P(liable|evidence)}{P(\neg liable|evidence)}}_{\text{posterior odds}} = \underbrace{\frac{P(liable)}{P(\neg liable)}}_{\text{prior odds}} \times \underbrace{\frac{P(evidence)|liable)}{P(evidence)|\neg liable)}}_{\text{likelihood ratio}} \tag{1}$$

In cases for which the bank claims that the customer is liable their records will presumably show that the EMV protocol exchange completed successfully. The bank would then argue their records shows that the genuine card was used. Indeed, we have a high degree of confidence in the security of the underlying cryptography of EMV (RSA and 3DES), and while formal methods have failed to identify some flaws in the protocols, subsequent studies have increased confidence that within the abstraction set out, the EMV protocol does what it is supposed to.

Therefore the records of the protocol outcome can be used to reduce the likelihood of some explanations to a negligible level, such that a criminal can factor products of large primes. These explanations can then be excluded from the computation of the posterior odds, but there still are many other explanations for the evidence when the customer is not liable. For example, maybe the records are not being correctly interpreted as a successful EMV exchange. Perhaps there was a bug in a computer system which was triggered by accident or criminal behaviour. Maybe an insider issued a duplicate card or otherwise tampered with

the system. All of these explanations are consistent with the evidence despite the customer not being liable in such situations. The evidence is also consistent with the customer having authorised the transaction. Therefore the likelihood ratio becomes 1 and effectively disappears from the equation.

All we are left with is the prior odds – ignoring the evidence presented, is it more likely that fraud occurred through the customers' negligence or dishonesty, or is the result of a failure of the bank. Implicit biases likely will then play a significant part in reaching a decision, probably to the detriment of customers in conflict with well-respected institutions. The likelihood that any given customer has been negligent (by the debatable definitions used by the bank) is significant, whereas a bank would argue that their systems are secure.

This situation is exacerbated through cyclic logic. Each previous dispute that gets resolved against the customer serves to increase the prior odds that the correct action is to resolve the subsequent one against the customer. The evidence, being consistent with both customer negligence and bank error, cannot override these prior odds.

When the outcome of a case hinges mainly on the prejudices of the adjudicator and not on the evidence produced by the system explicitly designed to resolve such disputes, apparently something has gone wrong with the way we design security protocols and the dispute resolution systems around them. The focus on rigorously analysing a small part of the system – the abstract security protocol between card and terminal – then arguing each dispute separately, is destined to fail.

By relaxing the level of proof required, we can dramatically expand the parts of the system we can analyse. In the words of John Tukey, "far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise" [10]. Rather than making a formal-methods based argument on the protocol alone to change the likelihood ratio, we can make a statistical argument about the system as a whole to change the prior odds.

One approach is not to argue each case individually but to examine a collection of cases. While the likelihood that any single customer is negligent or complicit is significant, the likelihood that every customer disputing a transaction is considerably less. By dealing with multiple cases at the same time, this sort of argument becomes possible.

## 3 Post Office and the Horizon Accounting System

Banking disputes of the type discussed above rarely make it to court because the risk to the customer of having to pay a bank's costs is prohibitive. In cases where banks consider that they likely must disclose potentially sensitive information, they are quick to settle and so the broader issues we have raised do not get resolved. However, an important case in the UK High Court which has the potential to change the situation is the Group Litigation against the government-owned company – Post Office Limited.

This case concerns disputes between the Post Office and subpostmasters who operate some Post Office branches on their behalf, offering not just postal services but also savings accounts, payment facilities, identity verification, professional accreditation and lottery services. These subpostmasters have been held liable for losses that the Horizon accounting system, operated by the Post Office, reports existing. The position of the Post Office is that subpostmasters are contractually obligated to compensate the Post Office for such losses. The subpostmasters claim that these losses are not the result of errors or fraud on their behalf but instead are due to malfunction or malicious access to Horizon.

So far one of the three trials is finished, with a judgement expected in January 2019. This trial focuses on the legal relationship between the subpostmasters and Post Office Limited, and the consequences of this on the validity of the contract terms holding subpostmasters liable for purported losses. The next trial, expected to start in March 2019, will deal with the Horizon system itself, but we have already learned much from the first trial in the Group Litigation, as well as previous legal proceedings dating back to 2009. For example, the Post Office has now disclosed that there have been accounting errors in Horizon, and their staff have the ability to remotely modify accounts without the subpostmaster's authorisation, both contrary to their previous statements[1].

Although this case has not attracted much media attention, its importance to future cases of disputes relating to computer evidence should not be underestimated. The Post Office has described the case as an "existential threat", and the losses of subpostmasters have been in the tens of thousands of pounds [6,7] – leading to bankruptcy, illness and even criminal prosecution. As a High Court case, the result will also act as a binding precedent. Unlike the handful of previous individual bank disputes, this trial has seen significant investment in both legal and technical expertise (total costs exceeding £10m before the trial had begun) and so it is reasonable to expect the Post Office's claims will be subject to close scrutiny. Such expense is only possible because a Group Litigation[2] allows resources of the 500+ claimants to be pooled, who are also supported by an investment fund that presumably will pay the costs of the Post Office should the claim fail.

The Group Litigation also allows a collection of cases to be examined in parallel and so has the potential to avoid the limitations of evidence raised in Section 2 and the risk of applying a what is effectively a cyclic argument in consecutive individual cases when computer error, negligence or fraud are all fully consistent with the evidence. The case will also examine whether it is fair or not to require users of a computer system to be bound by its results when their capacity to influence the incentive-design of the system or scrutinize

[1] Further details can be found through the crowd-funded coverage by journalist Nick Wallis at http://www.postofficetrial.com/

[2] This sounds like a US Class Action, but is quite different. Claimants participating in a Group Litigation Order must opt-in, are still liable for the other party's costs if they lose, and each case is still treated individually albeit with issues that are common to all

its operation is limited. As expressed by McCormack on the subject of Seema Misra's case [7], it is striking that "a subpostmaster could be held responsible for losses they incurred as a direct result of a failing to notice an error in a sophisticated computer system over which they had no control".

## 4 Conclusion and Discussion

We don't yet fully know what evidence the Post Office will present to support their case that the purported losses are genuine, but we can use this example to discuss what would be adequate evidence resulting from the security protocols supporting Horizon. Prior work has proposed systems improving transparency surrounding the transaction in dispute, which is indeed a good approach. For the reasons outlined in Section 2 we would propose augmenting these requirements to also include transparency of transactions that are the subject of other disputes (that perhaps the institution conceded) or non-disputed transactions. Such evidence could be used to make a statistical argument whether the behaviour regarding the dispute is indeed an exception or whether this case is just one anomaly out of many.

This approach creates both legal and practical difficulties. While the rules for admissibility of evidence vary, they do usually require that evidence is relevant and a case would have to be made to justify the additional effort of extracting information for transactions that are apparently unrelated. Rules for admissibility may also include restrictions for the purposes of benefit to society, such as prohibiting evidence relating to the bad character of an individual in order to help the rehabilitation of offenders. Evidence relating to the previous behaviour of an individual in dispute could fall foul of such restrictions.

Practically, disclosing information on individuals who are not a party to the dispute could violate their privacy. Here, privacy-preserving transparency approaches like VAMS [4] could usefully be applied. The current iteration of Horizon is effectively centralised following an upgrade that took place to make the system more efficient, so this would have to be changed back to a more distributed model where subpostmasters have control over their local system.

According to Ian Henderson of Second Sight, a company charged with independently investigating the Horizon system and subsequently fired by Post Office, it had around $12,000$ communication failures every year and software defects at 76 branches as well as unreliable hardware [3]. Issues with software and hardware lessen the gain of a transparency overlay on top of the system, as that would provide integrity for information that is logged, but would only show inconsistencies if events fail to be logged properly across the system. One way of resolving this would be to design the system so that subpostmasters can get some assurance that local processes were correctly executed, for example by using trusted hardware.

Some problems are however out of control of the protocol and system designer. One is how to provide incentives to organizations commissioning systems to produce better evidence. When parties are evenly matched, this could be

included in a contract but when there is a disparity, like the Post Office vs. subpostmasters or banks vs. their customers, policy interventions are needed. Courts may play some role, but they are restricted in what they can do – as the Post Office barrister reminded the judge in his closing statement, the court must apply the law, not common-sense.

## References

1. Anderson, R.: Why information security is hard-an economic perspective. In: Proceedings of the 17th Annual Computer Security Applications Conference. pp. 358–. ACSAC '01, IEEE Computer Society, Washington, DC, USA (2001), http://dl.acm.org/citation.cfm?id=872016.872155
2. Azouvi, S., Hicks, A., Murdoch, S.J.: Incentives in security protocols. In: Matyáš, V., Švenda, P., Stajano, F., Christianson, B., Anderson, J. (eds.) Security Protocols XXVI. pp. 132–141. Springer International Publishing, Cham (2018)
3. ComputerWorldUK: Post office obstructing horizon probe, investigator claims (2018), https://www.computerworlduk.com/infrastructure/post-office-obstructing-horizon-probe-investigator-claims-3596589/
4. Hicks, A., Mavroudis, V., Al-Bassam, M., Meiklejohn, S., Murdoch, S.J.: VAMS: verifiable auditing of access to confidential data. CoRR **abs/1805.04772** (2018), http://arxiv.org/abs/1805.04772
5. Jaynes, E.T.: Probability theory: The logic of science. Cambridge university press (2003)
6. Mason, S.: Case transcript: England & Wales-Regina v Seema Misra. Digital Evidence and Electronic Signature Law Review **12**, 45–55 (2015)
7. McCormack, T.: The Post Office Horizon system and Seema Misra. Digital Evidence and Electronic Signature Law Review **13**, 133–138 (2016)
8. Murdoch, S.J., Anderson, R.: Security protocols and evidence: Where many payment systems fail. In: International Conference on Financial Cryptography and Data Security. pp. 21–32. Springer (2014)
9. Steventon, B.: Statistical evidence and the courtsrecent developments. The Journal of Criminal Law **62**(2), 176–184 (1998)
10. Tukey, J.W.: The future of data analysis. The annals of mathematical statistics **33**(1), 1–67 (1962)