Justifying Security Measures
—a Position Paper

Cormac Herley

Microsoft Research, Redmond, WA, USA
cormac@microsoft.com

Abstract. There is a problem with the way we reason about problems in security. The justifications that we offer for many security measures reduce to unfalsifiable claims or circular statements. This position paper argues that reliance on less-than-solid arguments is responsible for a lack of progress in several domains.

1 Introduction

A great deal of computer security involves deciding how we should protect information, resources and assets. Folk theorems and slogans often emphasize the risk in neglecting any defense; e.g., "security is only as strong as the weakest link" and "there is no such thing as partial security." Unfortunately, we can't possibly do everything. Defensive measures generally involve cost in time, money, or effort, so defending everything against all possible attacks is neither possible nor appropriate. This leaves us with hard decisions. Which measures should we choose and which should we neglect? What constitutes a compelling argument in favor of defensive action?

Consider the defense appropriate for high-value assets. The laptop of the CFO of a large company might contain unreleased information about earnings, government systems might contain citizens' tax returns and health records. In the documentary movie 'Citizen Four' Edward Snowden asks all visitors to place their phones in the fridge and places a blanket over his head before typing his password. Clearly, as the target of the national security agencies of multiple countries (and with his liberty at risk in the event of failure) extraordinary measures are appropriate for Snowden. However, for most assets and most people this level of defensive effort is obviously excessive. If the level of caution that Snowden exhibits was necessary before checking email, Twitter, or Netflix, most of us would simply close our accounts. We might enjoy these services, but the benefit we receive limits how much effort we're willing to put in.

How then should we decide? We have no difficulty acknowledging that the measures needed to protect a high-value asset is inappropriate and excessive for a low-value asset, such as an ordinary email, social networking, or even bank asset. Thus, while we may occasionally repeat slogans about absolute security, few would argue that all assets should be treated as high-assurance ones. However, this acknowledgement is not helpful unless we can say which measures we can neglect.

2 Heads I'm right, Tails you've just been lucky so far

The austrian philosopher Wittgenstein once contested that the cycle of night and day should ever have been viewed as evidence that the sun revolved around the earth: "and how would it look if instead the earth was rotating?" he asked. That is, the cycle of night and day does nothing to distinguish between these two competing theories. What looked a reasonable argument actually wasn't even evidence.

It can be hard to see the flaws in arguments, especially when the conclusions have been believed for a long time. I wish to argue that a similar phenomenon is at work in security, where we have many long-held conclusions supported by arguments that do not withstand elementary scrutiny. I'm leaning heavily on a recent paper [1]. The basic result is that claims of necessary conditions for security are unfalsifiable. To falsify the claim "you must do X to be secure" we would have to find something secure that doesn't do X. That this isn't possible is a direct consequence of the fact that we can't ever observe that something is secure.

Obvious though it is, my experience has been that this result is not embraced willingly. People who are fond of saying that "the only secure system is unplugged, encased in concrete, and buried at sea" are reluctant to think through the immediate implications of that statement. If security is out of reach then claims of necessary conditions to achieve it are unfalsifiable. This is just elementary logic; you can't have it both ways.

When confronted with this fact people often suspect sophistry; they need a lot of convincing that there's actually a problem here and not just verbal trickery. Hence, it's worth going into detail to show that the common approaches to get out of this go nowhere. For example, the idea that security is defined relative to a set of security goals or a threat model doesn't help: it merely adds a layer of indirection (i.e., one more turtle) since the necessity of achieving any of the goals is in turn unfalsifiable. The idea that security is a property to be proved rather than observed doesn't help, since proof applies to mathematical rather than empirical properties; something can be proved secure only if the term "security" is emptied of all reference to observable outcomes (e.g., Einstein: "As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality"). The idea that security is a scalar quality to be improved rather than a binary one to be achieved doesn't help, since the claim that the security of X is better than the security of $\overline{X}$ is also unfalsifiable. See [1] for an expanded treatment of these arguments.

So to summarize, the logical consequences of being unable to observe that something is secure (or more secure, or that something will not happen, or cannot happen) are that the following claims are unfalsifiable:

1. "If you don't do X you are not secure"
2. "If you don't do X a bad outcome will occur"
3. "If you don't do X a bad outcome can occur"
4. "Doing X is more secure than not doing X."

Thus, for example, we can't test the truth of the statement "if you don't use a strong password you are not secure." It rules nothing out, and is consistent with every possible observation, past and future. Equally, if I say "if you don't run anti-virus you will be hacked" I am impervious to contradiction: the only possibilities are that, heads, I'm proved right, or, tails, you've just been lucky so far.

2.1 The importance of being literal

So what should we make of this? Is computer security no better than pseudo-science? Is it on a par with homeopathy, astrology and belief in paranormal phenomena? Despite the negative connotations of "unfalsifiable" we should resist jumping to conclusions. Horoscope predictions are unfalsifiably vague because they have no basis at all in reality. In contrast, the unfalsifiable statements 1-4 above are usually used as substitutes for claims that have some real basis, and may indeed be very defensible. For example, when we talk about security being improved (e.g., #4 above):

$$\text{Security}(X) > \text{Security}(\overline{X}) \tag{1}$$

we actually generally mean, e.g.,

$$\text{Outcome}(X|ABCD) > \text{Outcome}(\overline{X}|ABCD). \tag{2}$$

That is, while the security claim is unfalsifiable it is actually meant as a (falsifiable) statement about outcomes under certain assumptions A, B, C and D. Details have been omitted in (1), but there's a huge difference between omitting details and outright pseudo-scientific claims. So is the answer then simply "don't take things so literally?" Statements 1-4 are unfalsifiable, but is it just a case of omitting details in the interest of simplicity? Unfortunately, it's more serious than that; the omission of detail does not have innocent effects.

First, it is precisely when they are intended literally that claims are most useful. A wobble in the orbit of Uranus led to the discovery of Neptune only because Newton's laws were taken literally. When taken literally, anything not explained by measurement error is a discovery. By contrast, the less literal a claim the more things it's consistent with; and with enough wiggle room it can be made consistent with anything. The history of science is largely one of finding and resolving inconsistencies [2–4]. Insofar as they make this task harder, vagueness and wiggle room in claims are barriers to progress.

Second, the errors are directional. Going from (2) to (1) isn't just a simplification, it always expands rather than contracts the claim. When $\overline{A}$ OR $\overline{B}$ OR $\overline{C}$ OR $\overline{D}$ is true, then (2) makes no claim at all about outcomes. This fact is entirely lost when we substitute (1) for (2). The restrictions implied by A, B, C and D can be severe, in which case (2) is making a very narrow claim while (1) is making a very big one (see examples in Section 3). Thus we end up claiming that X is doing far more than is actually the case.

Finally, simplified versions of claims are understandable if, when challenged, we are prepared to restate with greater precision. However, it's easy to show that this is often not the case in security. That is, (2) says that outcomes improve under certain assumptions, while (1) drops all mention of the assumptions. If we have a clear understanding of what the assumptions are, we should have no difficulty falsifying a security claim: just show that what it promised to prevent can happen anyway. For example, to falsify (2) we would just demonstrate that X makes no difference to outcomes even when conditions A, B, C and D hold. If we continue to insist that X is worthwhile when no difference in outcomes is discernible then we must acknowledge that the list of assumptions is incomplete (e.g., perhaps X improves outcomes only when E in addition to A, B, C, and D hold). By contrast (1) rules nothing out: it asks that we do X, but it offers no justification.

So, if we don't know what would falsify the justification, then we don't know exactly what the measure claims to do. If nothing falsifies our justification then either it's a tautology or were not actually claiming the measure does anything observable. Note that this is not the same as saying that it doesn't do anything.

3 Never waste a good crisis: passwords

There's been significant evolution in our thinking about passwords in the last decade or so. Users used to be advised against writing passwords down, but now most experts seem to think it acceptable or advisable. Re-using passwords was considered unacceptable, we now know it is unavoidable [5]. Mandated password expiration (e.g., every 90 days) used to be consider necessary, we now know it accomplishes little [6]. Three decades after Morris and Thompson [7] recommended composition constraints (i.e., inclusion of special characters) as a path to password strength we know that they don't have the desired effect [8]. That stronger passwords improve outcomes, in any but very narrow circumstances, is itself very questionable [5]. Even national standards organizations in the US and UK have revised long-standing guidance to reverse many recommendations.

It doesn't seem harsh to say that the history of thinking, advice and instructions on passwords appears a catalog of error. Things proclaimed with great confidence have turned out to be simply untrue. Much of the advice directed at billions of Internet users has turned out to be mis-guided or even harmful. Passwords might seem an uninteresting research area. We might imagine that they will soon be a thing of the past (although those advancing this claim have a history of being optimistic), or that password managers can eliminate many of the difficulties, etc. However, I claim that, moving on without learning from mistakes wastes a significant opportunity. The litany of errors points to profound problems in the way we reason about security measures. Unless we can be confident that the errors in reasoning that generated such a mess in the domain of passwords have not happened elsewhere it is worth carefully examining what went wrong.

3.1 What constitutes a compelling argument for a security measure?

Consider the common recommendation of using a unique password for each account. Some recommendation like this is explicitly offered by Ives et al [9] and CERT [10]. I would like to focus, not on whether we believe this measure is sensible, but on the arguments that we can make in its favor. Justification for avoiding password re-use usually is as follows:

> *If you don't use a unique password for each account, a bad guy who gets access to one can compromise your other accounts.* (3)

This is a true statement; there's no question that re-use does open an avenue to compromise. It is not, however, on its own, a convincing argument in favor of using a unique password per account. Observe that (3) is a tautology. It can be rewritten:

> *If you don't do X then a bad guy can do something that X would have blocked.* (4)

The argument (3) is simply (4) substituting X for "use a unique password for each account." However, if we're going to argue that (4) offers a compelling argument for any X we should be prepared to argue that it does so for all X. Clearly it does not. For example, the claim

> *If you don't use a Faraday cage a bad guy can get your private keys using electro-magnetic emanations.* (5)

can also be expressed as in (4). If (3) is a persuasive argument against password re-use, (5) is a persuasive argument for Faraday cages. The problem with (4) (and hence (3) and (5)) is that the argument is circular. It simply says if X blocks something, then that thing is no longer a risk if you do X. This says nothing at all about likelihood and applies equally to threats that are very real, and ones that are completely far-fetched for most of us (e.g., the necessity of placing a blanket over our head as we type passwords).

Tautologies are simply one example of unfalsifiable justification statements. Next consider the claim that choosing a strong password is better than a moderately weak one (e.g. strong enough to withstand online guessing but no more). Does the fact that many users ignore this instruction without incident falsify this claim? If not then (following Section 2.1) there are implicit assumptions unstated in the original claim. For example, there's clearly no difference in outcomes unless A) the password file leaks. There's also no difference if the password file is stored B) plaintext or C) reversibly encrypted. Even then we're far from done; the chain of assumptions actually becomes quite long [5]. We have to flush out all of the assumptions to produce a falsifiable statement like (2) from the vague starting point (1). So, it's not the case that the unfalsifiable claim is a simplified stand-in for a falsifiable one that we actually intend literally. The fact that we have to resort to reverse engineering to figure out what falsifies the claim means we just don't know under what assumptions it will improve outcomes.

3.2 What evidence would prove us wrong?

Thus, falsifying the justification forces us to be explicit and exhaustive in documenting restrictions on what a measure claims to do. Difficulty doing this reveals that implicit or vaguely-stated assumptions lurk. If we are convinced of something, but can't describe the evidence that would change our minds, our belief is not well-founded.

Unfortunately, this seems to be the rule rather than the exception with password recommendations. Consider for example the advice to:

1. Change passwords regularly
2. Avoid password re-use
3. Choose strong passwords
4. Choose passwords of a certain format.

What evidence would falsify the claim that any of these are worthwhile? If we had empirical evidence indicating that those who comply fared better than those who do not then falsification would be simple: a measurement can always be superseded by a better, more thorough measurement. However, the justification for these measures does not rest on empirical evidence. Instead, it would appear to rest on the argument that the recommended measures improve outcomes *in certain circumstances.* Since the circumstances are not stated, they are defended by an argument like (1) rather than (2).

The point is not to argue that these measures accomplish nothing, but to emphasize that uncertainty about falsifying them is possible only if our justification is muddled and we don't have a precise understanding of what is claimed.

Passwords offers a target-rich environment for those seeking tautologies and unfalsifiable claims. However, the problem is far more general. What falsifies the claim that anti-virus is necessary? That cyber-crime is large and growing? That we need something more secure than passwords? That there's a tradeoff between security and usability? That a system with a "proof of security" is better than one without? If we hold these views, but can't say what would make us abandon them then our reasons are not solid.

4 Conclusion

Falsifiability is traditionally taken as the line separating science from non-science [2, 3, 11]. While this is the almost universal practice in the natural sciences, it is not unreasonable to ask why, and whether it is equally relevant to fields such computer security?

Falsifiability is not an arbitrary demarcation criterion, and it's acceptance by other scientific communities does not rest on Popper's authority. Falsifiability represents a constraint: it restricts the kinds of statements we can make, but in return gives feedback and self-correction. Falsifiability as a criterion is simply an acknowledgement that some of the statements we make and some of the ideas we try will be wrong. Popper's description of Science doesn't say how to come

up with laws, what they should describe, or even if there should be laws at all. It simply describes the feedback mechanism that, over time, filters out the wrong statements and ideas, so that our ability to describe the world and anticipate things not-yet-observed steadily improves.

Other feedback mechanisms exist in other domains. Markets provide feedback. Good businesses flourish and bad ones fail. Business models that enjoy economies of scale push out those that don't. Engineering techniques and artifacts compete against alternative techniques and artifacts. Good ways of designing bridges, airplanes and operating systems supplant less-good ways so long as there is feedback on what is proving useful in practice. In many of these domains feedback might not be as formal as falsifiable statements, but is still strong enough to separate the good approaches from the bad.

The absence of feedback has proved a serious barrier to progress in security. The reason so many arguments about passwords go in circles is that there's nowhere else for them to go. Are lower-case pass-phrases better or worse than passwords with a mix of characters? Should passwords be written down, or changed regularly? Is defense against shoulder-surfing worthwhile? No progress is possible on these and other questions if the justifications offered for them are immune to feedback and shrink from all of the risks associated with being tested against observation.

References

1. C. Herley, "Unfalsifiability of security claims," *Proc. National Academy of Sciences*, vol. 113, no. 23, pp. 6415–6420, 2016.
2. A. F. Chalmers, *What is this thing called Science? (4th edition)*. Hackett Publishing, 2013.
3. P. Godfrey-Smith, *Theory and reality: An introduction to the philosophy of science*. University of Chicago Press, 2009.
4. C. Herley and P. van Oorschot, "SoK: Science, Security, and the Elusive Goal of Security as a Scientific Pursuit," in *IEEE Symp. on Security and Privacy (Oakland 2017)*, 2017.
5. D. Florêncio, C. Herley, and P. C. Van Oorschot, "Pushing on string: The"don't care" region of password strength," *Communications of the ACM*, vol. 59, no. 11, pp. 66–74, 2016.
6. Y. Zhang, F. Monrose, and M. K. Reiter, "The security of modern password expiration: An algorithmic framework and empirical analysis," in *Proc. ACM CCS 2010*, pp. 176–186.
7. R. Morris and K. Thompson, "Password Security: A Case History," *Comm. ACM*, 1979.
8. M. Weir, S. Aggarwal, M. Collins, and H. Stern, "Testing metrics for password creation policies by attacking large sets of revealed passwords," in *Proc. ACM CCS 2010*, pp. 162–175.
9. B. Ives and K.R. Walsh and H. Schneider, "The Domino Effect of Password Reuse," in *CACM*, 2004.
10. US-Cyber Emergency Response Readiness Team: CyberSecurity Tips, "http://www.us-cert.gov/cas/tips/."
11. K. Popper, *Conjectures and refutations: The growth of scientific knowledge*. Routledge, 1959.