Security Behaviours

COMS30038 Lecture 5 - Welcome & Humans-in-the-Loop

Dr Ramokapane

University of BRISTOL

Bristol Cyber Security Group

Unit Structure

The term is broken down into…

Week	Topic	Lecture A	Lecture B	Lab
1	Social Engineering	Methods of influence	Pretexting & Intelligence	Crafting SE attacks
2	Attack Models	Killchains, attack trees / graphs	SecOps & Log Analysis	Incident tech analysis
3	Cybercriminology	Criminological theories	Personality & Individual Difference	CTF w/ethical hacking
4	Security Economics	Economies of cybercrime	Economic Interventions	CTF w/ethical hacking
5	Human Security	Intro / Humans-in-the-loop	Usable security	Debate preparation
		Reading Week		
7	Practice & Error	Inclusive security	Error in Practice	Debate delivery
8	Biases & Mitigation	Cognitive & physiological biases	Security Ergonomics	**Mid-term Exam**
9		Coursework drop-in	Coursework drop-in	
10		Coursework drop-in	Coursework drop-in	
11		Coursework drop-in	Coursework drop-in	
12		Revision session	Revision session	

Lectures delivered by

- Marvin
- Matt

Outcomes

Our aim for this unit is to introduce you to a range of conceptual tools for tackling cybersecurity problems. We want you to be able to understand security incidents beyond the technical level, and to be able to reason and argue about them coherently.

&

Expectations

Remember that the **lecture videos and slides are not the only essential material** -- the weekly readings are part of the unit, as are the exercises in the labs that show you how to put things into practice. Beyond that, you will get the most out of this unit if you engage with the online forum discussions, and the exercises and videos associated with them -- they're all there to give you a richer, more valuable learning experience.

and so, to the learning

humans-in-the-loop

What is meant by “humans-in-the-loop”

You might have heard of things like “supervised learning” - the process of manually labelling training data for ML. It is a classic example of modern HiL. A human (or better, a team of humans) are tasked with developing the ground truth upon which the clever maths can work its magic.

What is meant by “humans-in-the-loop”

You might have heard of things like “supervised learning” - the process of manually labelling training data for ML. It is a classic example of modern HiL. A human (or better, a team of humans) are tasked with developing the ground truth upon which the clever maths can work its magic.

And HiL occurs (I would argue) in every system, but more on that later.

What is meant by “humans-in-the-loop”

You might have heard of things like “supervised learning” - the process of manually labelling training data for ML. It is a classic example of modern HiL. A human (or better, a team of humans) are tasked with developing the ground truth upon which the clever maths can work its magic.

And HiL occurs (I would argue) in every system, but more on that later.

Traditionally HiL refers to the role that humans play within the operation of a system. So a good security example might be the modal requirement for a user to change the admin password on a new Internet router, or for a user to heed some form of contextual status or warning that they should/shouldn’t take an action - an external email containing a weblink is a good example.

What is meant by “humans-in-the-loop”

You might have heard of things like “supervised learning” - the process of manually labelling training data for ML. It is a classic example of modern HiL. A human (or better, a team of humans) are tasked with developing the ground truth upon which the clever maths can work its magic.

And HiL occurs (I would argue) in every system, but more on that later.

Traditionally HiL refers to the role that humans play within the operation of a system. So a good security example might be the modal requirement for a user to change the admin password on a new Internet router, or for a user to heed some form of contextual status or warning that they should/shouldn’t take an action - an external email containing a weblink is a good example.

Where the human is expected - in some way - to reason about the world to make a system influencing decision.

What is meant by “humans-in-the-loop”

You might have heard of things like “supervised learning” - the process of manually labelling training data for ML. It is a classic example of modern HiL. A human (or better, a team of humans) are tasked with developing the ground truth upon which the clever maths can work its magic.

And HiL occurs (I would argue) in every system, but more on that later.

Traditionally HiL refers to the role that humans play within the operation of a system. So a good security example might be the modal requirement for a user to change the admin password on a new Internet router, or for a user to heed some form of contextual status or warning that they should/shouldn’t take an action - an external email containing a weblink is a good example.

[note: we cover concepts of human error, latent failure, biases and a framework for mitigation in later lectures. This lecture focuses upon what we mean by HiL]

Where the human is expected - in some way - to reason about the world to make a system influencing decision.

What reasoning roles do humans play?

Think on any system - ideally security related in some way. Now imagine all the ways in which people interface with that system.

What reasoning roles do humans play?

Think on any system - ideally security related in some way. Now imagine all the ways in which people interface with that system.

Let’s use a consumer Internet router by way of an example. There is the person buying it, possibly someone else installing it, and a whole bunch of people using it for access. And the role of users and their ability to reason, with regards security is hardly knew knowledge.

What reasoning roles do humans play?

Think on any system - ideally security related in some way. Now imagine all the ways in which people interface with that system.

Let’s use a consumer Internet router by way of an example. There is the person buying it, possibly someone else installing it, and a whole bunch of people using it for access. And the role of users and their ability to reason, with regards security is hardly knew knowledge.

Kerckhoffs defined the human as “an irremovable and critical security component” in 1883 in his piece on military encryption!

What reasoning roles do humans play?

Think on any system - ideally security related in some way. Now imagine all the ways in which people interface with that system.

Let’s use a consumer Internet router by way of an example. There is the person buying it, possibly someone else installing it, and a whole bunch of people using it for access. And the role of users and their ability to reason, with regards security is hardly knew knowledge.

Kerckhoffs defined the human as “an irremovable and critical security component” in 1883 in his piece on military encryption!

Saltzer and Schroeder way back in 1975 acknowledged the role of the human’s ability to rationalise being based in behavioural sciences.

What reasoning roles do humans play?

Think on any system - ideally security related in some way. Now imagine all the ways in which people interface with that system.

Let’s use a consumer Internet router by way of an example. There is the person buying it, possibly someone else installing it, and a whole bunch of people using it for access. And the role of users and their ability to reason, with regards security is hardly knew knowledge.

Kerckhoffs defined the human as “an irremovable and critical security component” in 1883 in his piece on military encryption!

Saltzer and Schroeder way back in 1975 acknowledged the role of the human’s ability to rationalise being based in behavioural sciences.

Adams & Sasse, in their seminal work (Users are Not the Enemy) clearly articulated how humans struggle with password policies.

What reasoning roles do humans play?

Think on any system - ideally security related in some way. Now imagine all the ways in which people interface with that system.

Let's use a consumer Internet router by way of an example. There is the person buying it, possibly someone else installing it, and a whole bunch of people using it for access. And the role of users and their ability to reason, with regards security is hardly knew knowledge.

Kerckhoffs defined the human as "an irremovable and critical security component" in 1883 in his piece on military encryption!

Saltzer and Schroeder way back in 1975 acknowledged the role of the human's ability to rationalise being based in behavioural sciences.

Adams & Sasse, in their seminal work (Users are Not the Enemy) clearly articulated how humans struggle with password policies.

Useful Stuff

Kerckhoffs, A, 1885
La cryptographie militaire, ou, Des chiffres usités en temps de guerre: avec un nouveau procédé de déchiffrement applicable aux systèmes à double clef.

Saltzer, J. H., & Schroeder, M. D. 1975
The protection of information in computer systems.
Proceedings of the IEEE, *63*(9), 1278-1308.

A.Adams, A and Sasse, M. A. 1999
Users are not the enemy,
Communications of the ACM, vol. 42, no. 12, pp. 40–46, 1999.

Why we look at humans-in-the-loop

So, we've known about humans being integral components of a secure system for a long time. Especially when they are the users.

Useful Stuff

Waste Electrical and Electronic Equipment recycling (WEEE) Health and Safety Executive https://www.hse.gov.uk/waste/waste-electrical.htm

Why we look at humans-in-the-loop

So, we've known about humans being integral components of a secure system for a long time. Especially when they are the users.

But what about, the sales person, the marketing person? Someone had to design and manufacture it. Someone else, hopefully, will go full WEEE on it and dismantle it for useful components someday.

Useful Stuff

Waste Electrical and Electronic Equipment recycling (WEEE) Health and Safety Executive
https://www.hse.gov.uk/waste/waste-electrical.htm

Why we look at humans-in-the-loop

So, we've known about humans being integral components of a secure system for a long time. Especially when they are the users.

But what about, the sales person, the marketing person? Someone had to design and manufacture it. Someone else, hopefully, will go full WEEE on it and dismantle it for useful components someday.

All these people, in some, way have to reason about how it should work, be built, be sold, be installed, used and decommissioned.

Useful Stuff

Waste Electrical and Electronic Equipment recycling (WEEE) Health and Safety Executive
https://www.hse.gov.uk/waste/waste-electrical.htm

Why we look at humans-in-the-loop

So, we've known about humans being integral components of a secure system for a long time. Especially when they are the users.

But what about, the sales person, the marketing person? Someone had to design and manufacture it. Someone else, hopefully, will go full WEEE on it and dismantle it for useful components someday.

All these people, in some, way have to reason about how it should work, be built, be sold, be installed, used and decommissioned.

Normally we only focus on the user, and understandably so. BUT as engineers - and security focussed ones at that - we need to consider the whole system lifecycle.

Useful Stuff

Waste Electrical and Electronic Equipment recycling (WEEE) Health and Safety Executive
https://www.hse.gov.uk/waste/waste-electrical.htm

Why we look at humans-in-the-loop

So, we've known about humans being integral components of a secure system for a long time. Especially when they are the users.

But what about, the sales person, the marketing person? Someone had to design and manufacture it. Someone else, hopefully, will go full WEEE on it and dismantle it for useful components someday.

All these people, in some, way have to reason about how it should work, be built, be sold, be installed, used and decommissioned.

Normally we only focus on the user, and understandably so. BUT as engineers - and security focussed ones at that - we need to consider the whole system lifecycle.

By way of an example, let us explore that Internet router a little more…

Useful Stuff

Waste Electrical and Electronic Equipment recycling (WEEE) Health and Safety Executive
https://www.hse.gov.uk/waste/waste-electrical.htm

Human reasoning influences systems

This isn't meant to be an exhaustive list and is somewhat simplified …

1. At some point someone decides a new router is needed to meet a specification so commissions engineers to design and build it.

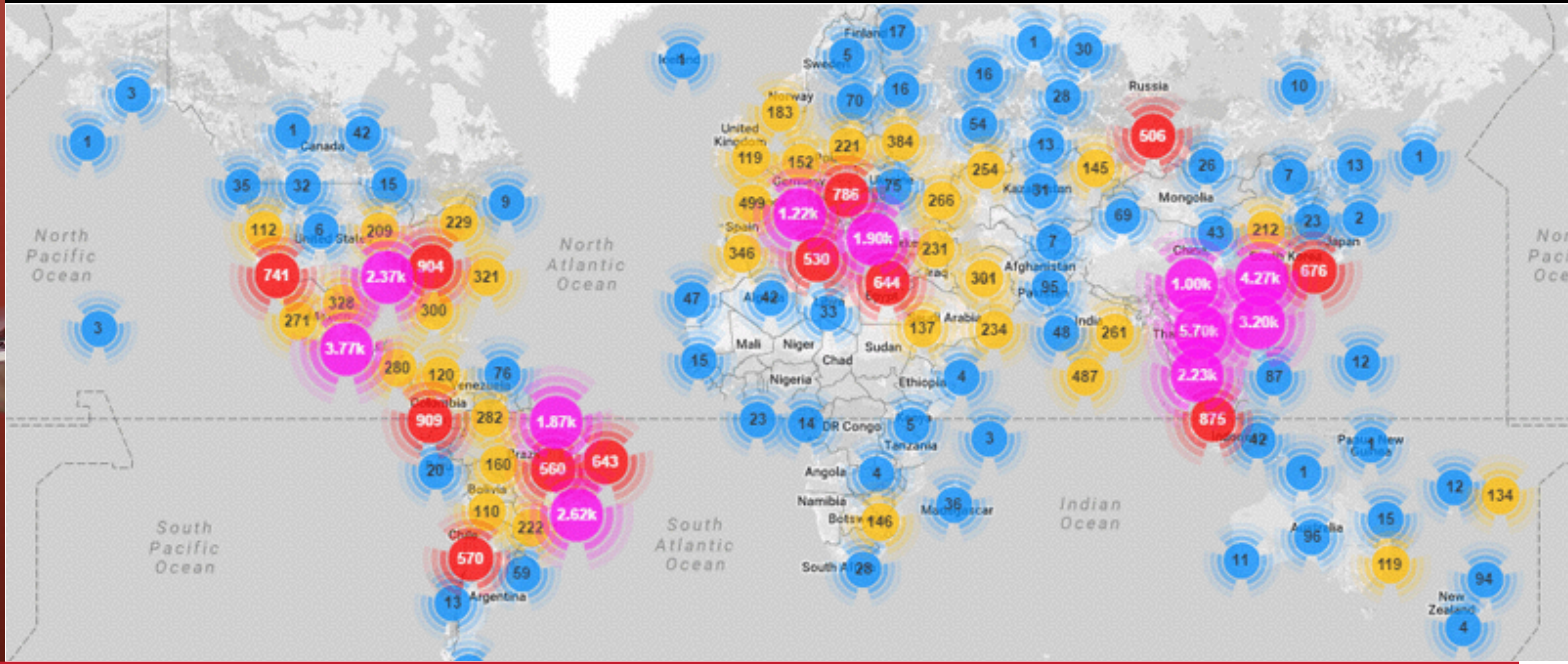
2. Being rational security aware engineers, they know i) the unit has to ship "secure" with admin credentials in place, ii) that the user needs to know what those credentials and connection details are, and iii) that some mechanism should be afforded the user to change those. They design, build and ship the units complete with a manual most likely on a CD-ROM or online.

All good common sense.

3. User A sees the router does what they need, buys, unboxes and plugs it in (likely without RT*M). The wifi SSID and password is on the box so just hooks up their laptop and away they go. It works great.

4. User B does much the same thing, only they also buy a couple of cheap security cameras online as well. The connection instructions look pretty complicated but they notice they can use the WPS button on their router to have the security handshake done semi-automagically.

But a few days later…


North Pacific Ocean
North Atlantic Ocean
South Pacific Ocean
South Atlantic Ocean
Indian Ocean
Canada
Russia
Mongolia
Afghanistan
Sudan
Chad
Niger
Mali
Nigeria
Ethiopia
Angola
Namibia
Argentina
Colombia
Tanzania

© Disney 2020 - Timmy Failure

Decisions were made…

And this is - in a condensed version - how the Mirai botnet took down many of the major web properties in 2016.

- The companies demanding new routers didn’t specify strict enough security by design requirements.
- The engineers, made assumptions about users reading manuals, taking common sense steps to change admin credentials and update firmwares.
- Users, driven by a need/want/desire for functionality, were handed working routers that fulfilled those needs, didn’t or were not compelled to take sensible pro-secure behaviours - possibly believing the engineers had already done the hard work.
- And the adversaries took advantage of a published list of just 60 sets of default IoT credentials, and know weaknesses in router firmwares.

Reasoning, however well thought out, how well meaning will often fall back to basics recognised well over 100 years ago.

[note: we cover concepts of human error, latent failure, biases and a framework for mitigation in later lectures]

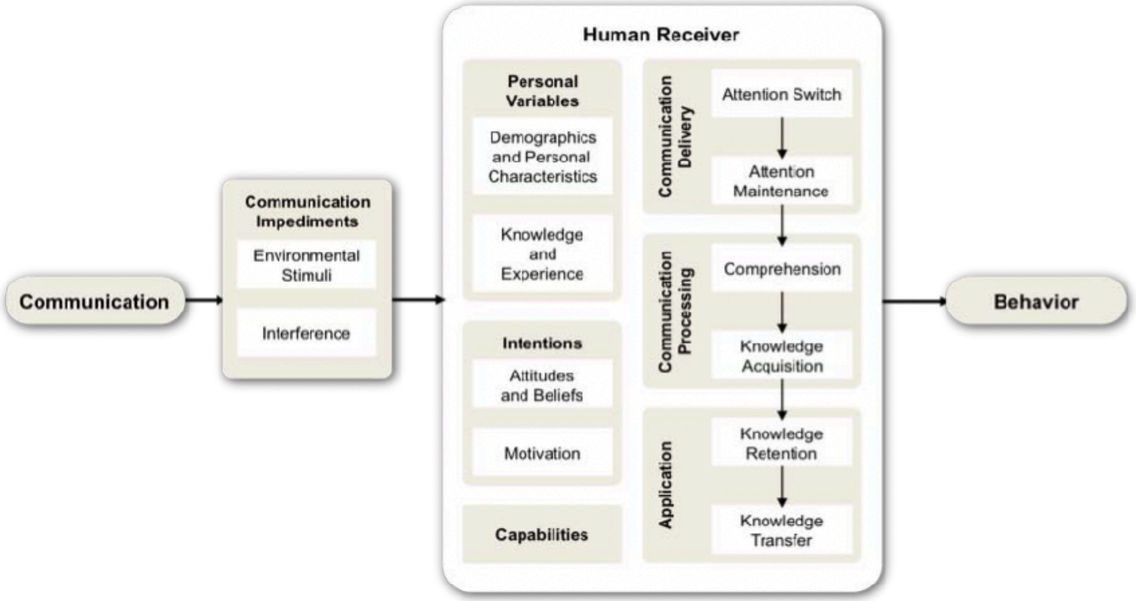
Humans-in-the-loop security framework

Cranor's 2008 *framework for reasoning about human in the loop* is designed to provide systematic approach to identifying potential causes for human failure, even looking down to possible non-malicious "human error" attributed root cause. [note: again, we will look root cause this in more detail when we cover error] Cranor tends to refer to the "user" in most cases but as we have seen other humans play a critical role, this is acknowledged by them pointing to better design practice to help with signalling but doesn't really address the developer/designer as a user in their own right.

Based on the Communication-Human Information Processing Model (C-HIP) as "*security-related actions by non-experts are generally triggered by a security-related communication*." It looks a lot to HCI practice and how those communications are transmitted, received and acted upon by the human receiver.

Broadly the framework works on the basis that security-related communications are received by a user, interpreted based on personal variables, intentions, capabilities etc.. and this results in a security-related behaviour. By applying this framework in the context of threat identification failures (human error) can be detected.

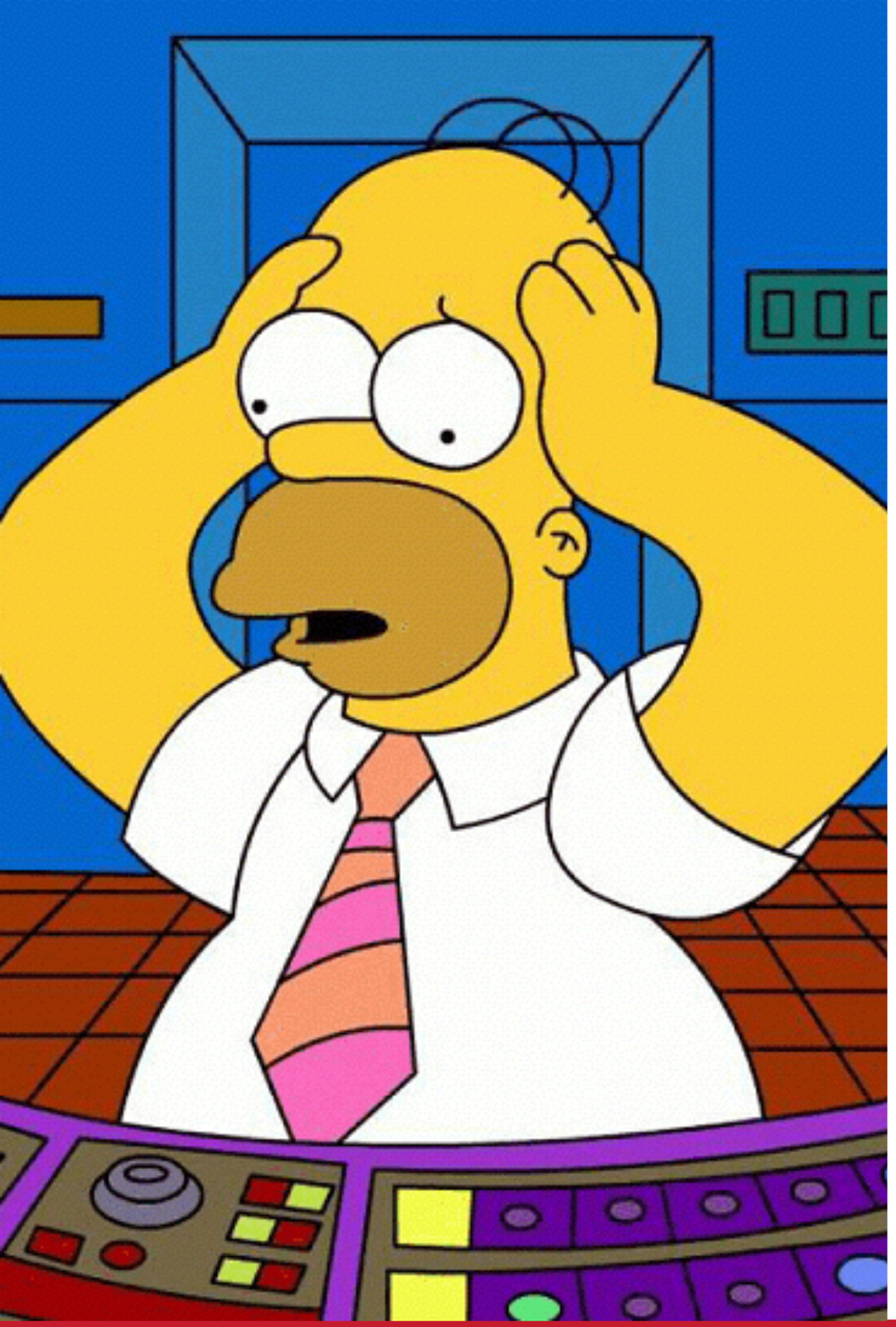
[note: we will apply this framework in a lab session on culpability]



Useful Stuff

Cranor, L. 2008
A Framework for Reasoning About the Human in the Loop
https://www.usenix.org/legacy/events/upsec08/tech/full_papers/cranor/cranor.pdf

Wogalter, M.S. 2006.
Communication-Human Information Processing (C-HIP) Model.
In Handbook of Warnings.



A trope

But the more we understand humans-in-the-loop, their limitations, behaviours etc.. the easier it is to start to conflate the inevitability of human error as meaning **humans are the weakest link**.

By studying human security behaviour we have become the proverbial “butt of our own jokes”, a trope. Many of these will sound familiar…

“…the human is the weakest link.”

Lance Spitzner, SANS

https://www.sans.org/security-awareness-training/blog/why-human-weakest-link

“…employees are making businesses vulnerable from within.”

Kaspersky

https://www.kaspersky.com/blog/the-human-factor-in-it-security/

“…humans are still security’s weakest link.”

Robert Kress, Accenture

https://www.accenture.com/us-en/blogs/security/humans-still-securitys-weakest-link

“…blockchain / ai will solve…”.

A.Non

Why the trope isn't really fair (or correct)

Many of the tropes are lazy. They are designed to grab headlines, garner clicks, or often, sell "solutions" - normally technical solutions. And technology is created by someones. Organisations are really just collections of someones.

The "humans" being referred to as weak are normally the users or operators of technology, with little introspection towards those designing or building the technology in the first place. And even less attention is normally paid to the very organisations themselves.

And simply "to err is human" (Pope, 1711). It is part of the human condition. We make mistakes. They are inevitable. So just blaming security mistakes on people is a little self-defeating. [note: we cover human error in a later lecture]

Useful Stuff

https://www.gutenberg.org/ebooks/7409

What is the real threat of HiL

Humans may well be involved in some 90% of cyber breaches (ICO, 2020) and are still often attributed as being a root or contributing cause. [note: I'm not aware of research or statistics which break that down to users/operators in the firing line vs developers vs organisations].

Humans have limited capacity. Who here actually has used same password for multiple sites just because they can't remember all of them? [note: we cover capacity in usable security]

Sometimes **humans just don't know what to do**
security related processes can be complicated, even SE struggle

Humans just are not motivated enough to do the security-related task. Security is often a secondary task to their job, and one that gets in the way or becomes habitualised (fatigue)

And.. **occasionally humans do something deliberately** - either as a workaround to limitation (single password) or an obstruction (tailgating) OR something malicious like planting a bit of malware [note: we cover deliberate & inadvertent actions in error]

So why not automate HiL out?

With so much of security's woes attributed to the human there is an attraction to security-critical systems that do away with any human interaction being necessary for their operation. Systems where for them to deliver their intended function no human is required to make a decision or press a button.

By removing the human, you supposedly remove potential for human error.

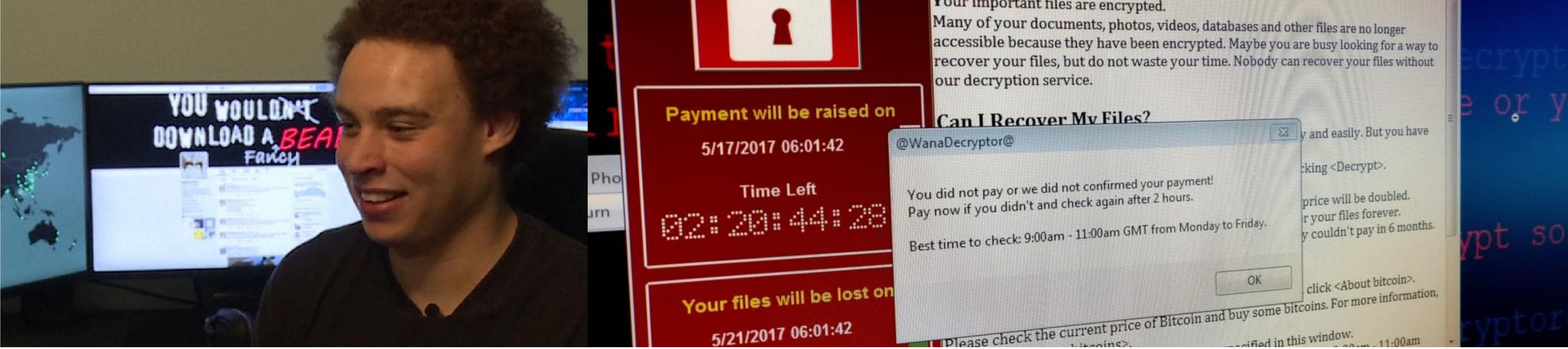
Homer always had that one critical "big red button". The one that normally says "do NOT press". Homer is the human-in-the-loop. Homer is there to do the things machines are not traditionally good at doing - reasoning. Despite advances, people are still better at a whole range of tasks than the best AI/ML. Maybe in time that will change.

Cranor gives a great example with where removing the human has worked with anti-virus tools. In the beginning they merely prompted for a human to take the decision around deletion. Today they can be all but fully autonomous. Seemingly without a human-in-the-loop.

In reality automation doesn't remove the HiL but just shifts the burden back to those developers and organisations we don't always consider to also be fallible humans.

And it will remain this way until…


SKYNET
CYBERDYNE
SYSTEMS



So are humans really the heroes-in-the-loop?

Wannacry (2017) was estimated to have cost the UK's NHS £92M to resolve. Significantly more than the US$300-600 per machine demanded in ransom. Globally well over 200,000 devices were infected within a day. Today we are still purging it from networks. What saved many, an awful lot more was the work of Marcus Hutchins.

Hutchins found a "kill switch" in the method by which the malware attempted to communicate with a hitherto unregistered domain. By registering that domain and effectively blocking that communication the malware was rendered impotent.

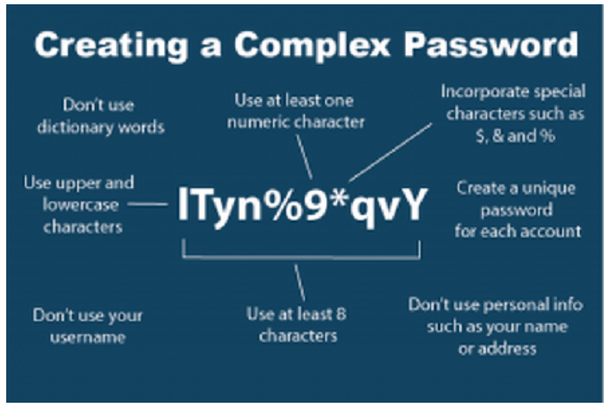
Without a human interjecting themselves into the ransomware loop the impact could have been a lot worse. And whilst Hutchins claimed to "not be a hero" his actions could be viewed as such.

Useful Stuff

Martin, G., Ghafur, S., Kinross, J., Hankin, C., and Darzi, A. 2018 Wannacry a year on. BMJ (Clinical research ed.), vol. 361, p. k2381

A lead in to usable security and error

Usable security, as defined by Sasse & Flechais, asserts **that the only effective security is that which is usable**. This hardly seems ground-breaking, but even 15 years later on it is a concept many seem to fail to implement. We forget that humans are intrinsic in the loop, and fallible.



If we want people to make sound security-related reasoning, come to secure decisions and take effective action all in the face of potentially great harm if it goes wrong (all heroic actions in my book as these are the little things that prevent the big ones) then we need not only to understand what is usable, and why - but we also need to implement it.

Useful Stuff

Sasse, M. A., and Flechais, I. 2005 Usable Security: Why Do we Need It? How do we get it? In Security and Usability: Designing secure systems that people can use. (13 - 30). O'Reilly

NCSC. Three random words. 2016 https://www.ncsc.gov.uk/blog-post/three-random-words

Images
https://www.eweek.com/imagesvr_ez/b2bezp/2019/01/Complex.password.png?alias=article_hero
https://cdn.enablis.com.au/wp-content/uploads/2019/10/passwords.png

Questions to ponder before next time

Humans have limited capacity. Can you think of a security-related situation when you have struggled to do, or even remember to do, something? Do you reuse passwords, and why? What sorts of tools might help people overcome that capacity issue?

Humans just don't know what to do. Can you think of a situation when you knew you had to do something, but just couldn't figure out what or how? How about as a software engineer - say you needed to store passwords in your code somewhere, would you know how to? Would you know where to find best practice?

Humans might just are not motivated enough. How do you prioritise security tasks in software development? What comes first functionality or security? Who makes that decision? When you are busy do you just make a placeholder note in the code and get on with the *real* task at hand? What about as a user; do you ever actually read privacy or cookie notices or do they just get in the way?

Humans do something deliberately. I'd bet you have done something deliberately that was not pro-security. How about that password reuse question above? What other decisions have you made that might not be in the best interest of security? What pressures were on you to make that decision?

next time...
Usable Security