Computational Trust for Enhanced Network Security

Tom Wallis

Reactive, human-like security

Information security is an infamously complicated topic. When we design security systems, we have to be aware of all potential vectors of attack. Indeed, the job of building a security system is much harder than the job of breaking that system, because an attacker need only find one attack vector the designer did not consider to gain access.

Security systems are getting smarter, however. Machine learning for information security is a topic with many interesting implications, and the field may well hold the answer to creating autonomous guardians of information systems which are capable of adapting and reacting faster than any human-managed system.

However, there are gaps in our knowledge of machine learning systems as they pertain to information security. Particularly, human factors are often researched from the perspective of humans interacting with the machine learning system [1]. However, another research opportunity which may prove very fruitful is the building of human factors into the security system itself. Research on computational models of trust and comfort have rarely been applied to a security mindset — with only a little published work so far [2] — and future developments in the field will undoubtedly require computational concepts of the responsibilities and trustworthinesses of other intelligent security agents that a software system interacts with. Research on fields such as computational trust are already well-established.

[1] Marco Barreno, Blaine Nelson, Russell Sears, Anthony D. Joseph, and J. D. Tygar. Can machine learning be secure. In *In Proceedings of the ACM Symposium on Information, Computer, and Communication Security (ASIACCS)*, pages 16–25. ACM Press, 2006

[2] Zheng Yan, Peng Zhang, and Teemupekka Virtanen. Trust Evaluation Based Security Solution in Ad Hoc Networks. 2003

Solutions we can trust

These hypothetical security systems would need to embody traits such as trust or responsibility in a human-like way, because they would likely also interact with human agents in their day-to-day roles. Therefore, a socially-based formulation of traits like trust need to be applied to security systems with an eye to eliminating human-like flaws and retaining enough human likeness that the system can be reasoned about by a human agent. Some work in the field of sociotechnical research attempts to create these likenesses in algoithms for anthropomorphising traits like Trust[3]. Newer paradigms for trust also examine notions such as Distrust and Mistrust, and involved logics for trust [4] allow for more nuanced models of trust which still

[3] M Stephen. *Formalising Trust as a Computational Concept*. University of Stirling, 1994

[4] Seifeddine Kramdi. A modal approach to model computational trust. URL `https://tel.archives-ouvertes.fr/tel-01328169`

retain computability. Work is done in other fields, too, such as computational comfort [5].

In deciding whether to trust an external agent, a security system can be enhanced in a number of ways. Security levels can become dynamic, to greatly increase the work needed to access a system when an agent's activity is suspicious. In addition, human operators of this system need to understand why a security system with non-deterministic behaviour is acting in a certain way; using human-like traits to make decisions around security allows the human agent to better understand the computer's actions, leading to less human-centric error. Using machine learning and trust over and above existing security techniques would also allow previously unseen security threats to be mitigated quickly by a computer, which can react and learn faster than a human agent can, without the downsides of fatigue and bias that a human security analyst might experience during an attack.

My own masters research involves a computational formalism of responsibility — work I would like to use in conjunction with existing trust and comfort research to create a security system with three human-like elements. Trust, responsibility, and comfort influence each other and the actions we take, and with computational formalisms of each, this could be used to develop a security layer which interprets activity and intent for a better understanding of potential threats. A system which has a responsibility to prevent attacks and is presented with strange instructions from an agent it would usually trust would weigh up its comfort with the strange instruction against its trust of an agent[6]. A modestly trustworthy agent may be presented with additional security measures before the secure system is comfortable enough to proceed with the request. This is one example of how a security system can use anthropomorphic traits to make intelligent assessments of security threats.

My suitability for building this solution

My own interest in sociotechnical systems is backed by research in the field. In 2015 and 2016 I completed my honours project, a workflow modelling system which used my own novel approach to program mutation to simulate human error in a sociotechnical model. For this, I won the 2016 Best Software Product prize for my year.[7] I am now pursuing an MSci, for which my project is an adaptation of Trust models to create a model of computational responsibility. In doing this, I have grown an appreciation of the breadth and nuances of various models of trust, and look to employ this knowledge in PhD research.

[5] Stephen Marsh, Pamela Briggs, Khalil El-Khatib, Babak Esfandiari, and John A. Stewart. Defining and investigating device comfort. *Information and Media Technologies*, 6(3):914–935, 2011. DOI: 10.11185/imt.6.914

[6] On a similar note, agents that were designed to reason in this way might mitigate risks of developing unfriendly AI through a more human-like reasoning system — these issues are elucidated in Nick Bostrom's book, Superintelligence.

Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, Oxford, UK, 1st edition, 2014. ISBN 0199678111, 9780199678112

[7] A paper on the work is currently being written, alongside my project supervisor, Tim Storer. A working copy of the paper can be compiled from LaTeXsource at https://github.com/probablytom/fuzzimoss-paper

Having already pursued sociotechnical research in the subfield of trust and responsibility, and having developed personally the responsibility formalism required to begin this proposed work, I am uniquely suitable as a candidate to devlelop this next generation of security models. In addition, the research is poised to be valuable. It stands to provide advancements in cyber security, trust modelling, interaction between artificial agents, and possibly future development of friendly AI. I am therefore confident in the value of the proposed work — not only in the academic sphere, but also in the realm of industry and defence — and that my understanding of the required fields and enthusiasm to commit the research makes me an ideal candidate.