Trusting the IoT: There Is More to Trust Than Trustworthiness

Piotr Cofta

Trusting the IoT: There is more to Trust than Trustworthiness

Piotr Cofta [0000-0002-4269-6590]

University of Science and Technology (UTP), Bydgoszcz, Poland, piotr.cofta@utp.edu.pl

Abstract. The emergence of the IoT as an everyday fact raises the question how the IoT can be trusted. Considering that the IoT is pervasive to the level of being secretive, practically monopolistic and that human participation is often involuntary, it hardly satisfies the assumptions that associate the often researched types of trust relationships. In order to study trust in the IoT, alternative views on trust may be needed. This paper analyses the IoT as a representation of imperfect systems, i.e. systems that by architectural choices are unable to guarantee predictably repeatable operations. This property may invalidate the metaphor of trusting technology that is constructed out of replicating human-to-human trust. This paper examines alternative views on trust that may better fit the specificity of the IoT and generally imperfect systems, adopted from psychology, sociology or ergonomics. While no definitive approach is indicated, this paper serves as an overview of possible directions in trust research.

Keywords: Internet of Things, IoT, trust models, trust metaphor, imperfect systems

1 Introduction

Internet of Things (IoT) is one of the current defining trends of the Internet, eventually linking billions of devices with cloud and fog computing into an infrastructure that will sense and control our environment. The vision of IoT assumes that it will be pervasive, invisible and monopolistic while our participation will be involuntary [4].

As such, the question of trusting the IoT is of paramount importance. This paper argues that we are currently ill-prepared to answer this question. This is because the problem introduced by the rise of IoT correlates with the paradigm shift in computing, towards imperfect computing. That, in turn, should shift a discussion from assuring the trustworthiness of the system to discussing ways of trusting systems that are currently not considered fully trustworthy, in situations that are devoid of real choices.

While the IoT is a prominent example of imperfect systems, it is not the only one. This paper starts with the introduction to imperfect computing, providing a brief summary of its common characteristics. Subsequently it analyses common assumptions about trust, mostly invalidated by the IoT. Next, it investigates some of the theories that may be applicable to explain situations of trust in the IoT. The paper concludes with some comments regarding suggested directions in research.

2 IoT as imperfect computing

Imperfect computing, as the name suggests, happens when computing systems occasionally provide incorrect or incomplete answer, not because of their fault or because they are intentionally made to do so, but because of the inherent architectural properties of those systems. Thus, imperfect computing cannot be made more perfect without significant changes to its architecture.

In a way, computing is experiencing what physics did in the last century: the transition from the Newtonian paradigm of deterministic repeatability and predictability to statistical approach that now permeates all fields of physics. In a similar manner, imperfect computing indicates the transition from the paradigm of deterministic outcome out of deterministic data to the uncertainty of it, which is not always statistical in its nature.

While the notion of imperfect computing covers several techniques, following is a brief overview of select ones.

Eventual consistency is a technical property shared by such diverse technologies as blockchain [27], NoSQL databases [14] and stream processing [16]. The state of the system is by design inconsistent, on assumption that eventually, under some conditions, it will converge to a consistent one. As the body of knowledge available to the system changes in time and is inconsistent between locations of the elements of the system, the response depends on both: time and place.

Learning algorithms are a form of imperfection associated with machine learning classification systems [11]. It is a property of the learning algorithm that its internal state changes as the result of learning, in a way that is not explainable. Thus, the response to to the same query may change in time, reflecting the learning process.

Approximate computing [13,30] is a form of imperfect computing where deterministic algorithm produces imprecise results as precise output is not required while being computationally expensive or energy inefficient. Both hardware and software solutions are available, e.g. from the area of lossy compression where perceptual limitations set the limit to required perfection.

The IoT is more than just an embodiment of some of those techniques. It provides a computational layer on top of the physical phenomena. As such, it faces the duality of imperfection: the one that comes from physics (mostly from the uncertainty of measurements) and the one that comes from the selection of imperfect information technologies [33]. If 'normal' imperfect computing deals with certain data in an uncertain way, the IoT deals with uncertain data in an uncertain way.

It does not invalidate the usefulness of the IoT. After all, an imperfect answer is quite often better than no answer at all. However, for the user conditioned to trust computers unconditionally in expectation for certain perfection, dealing with such dual imperfection may lead to doubt, and distrust.

3 Trustworthiness, trust and the IoT

The current approach to trust and trustworthiness, as discussed throughout literature (see e.g. [8] for an overview) tends to focus on the relationship between an enlightened trustor (the one who potentially trusts) and a trustee (the one that is hopefully trusted). The current approach can be characterised by the following assumptions, immediately contrasted with the properties of the IoT:

- The trustor can identify the trustee. That is, trustees are somehow distinguishable from the environment, e.g. in a form of persons, corporations or web sites. In contrast, pervasive IoT is almost indistinguishable from the environment while various subsystems and operators are also not distinguishable from each other.
- The trustor can exercise his free will in choosing to trust one of the trustors, or not trusting anyone at all, with no or little discomfort to itself. For the pervasive IoT, the trustor has no choice but to trust or not to be able to proceed.
- The trustor trusts willingly and cannot be coerced into trusting, as he subjects himself to being vulnerable to and dependent on the trustee without the ability to control them. Again, for the IoT the trustor is effectively coerced into trusting under the threat of discontinuation of vital services.
- For each trustee it is possible to satisfactory determine the extent of its internal quality of trustworthiness. Trustworthiness of the IoT, being new, technically complex and not directly observable by users, is hardly a subject of easy determination of its trustworthiness.
- The extent of trust that the rational trustor grants the trustee approximates the level of trustworthiness of this trustee. That, for the IoT, is not relevant, as it is neither distinguishable nor its trustworthiness can be determined.

Being an imperfect system, the IoT has one more hurdle to overcome: algorithm aversion [10]. This phenomenon describes the aversion to the use of algorithms that are known to be imperfect and favouring human decision-making even if algorithms lead to consistently better results.

4 How to trust the IoT

As the IoT is an imperfect system, it cannot become trustworthy (hence it cannot be started) using a current understanding of this construct, i.e. by applying the current thinking of the human-to-human relationship.

However, the fact that current considerations are not applicable to the IoT does not preclude it from being trusted, and does not absolve the research from studying trust between people and imperfect systems.

It requires, however, an alternative view on what it means to be trustworthy and what it means to trust imperfect systems. To this end, the remainder of this section is devoted to the discussion about alternative approaches that can help explain and facilitate research in trust in the IoT, and in other imperfect systems.

4.1 Trusting machines

There is a question whether the notion of trust between people can be applied to situations between people and machines, ie. whether the situation between people and the IoT can be discussed in terms of trust at all. In other words, whether humans use the same mechanism of trust for their human trustees as for the machines, specifically considering that humans have intentions while machines are devoid of them.

There is ample evidence that this is indeed the case: people behave as if they trust machines, whether this is considered real trust or a cheap substitute of it. This section highlights only some of many research streams that touched upon this problem.

Dennett [9] introduced a concept of three stances that the user can assume while exploring an artefact. The physical stance requires the artefact to be explorable on the physical level; the design stance require the artefact to be explainable through some simplifying mental models. For more complex systems, the intentional stance applies.

The intentional stance relies on treating the computing system as a human, with all the implications of it. Thus, the user may assume that the system is intentional - has intentions, moods, desires, dislikes etc., and will try to develop the relationship with the system as if the human would have developed the relationship with another human.

The intentional stance makes the perception of an imperfect system bearable for the human, but it makes trusting those systems more human too. Lessons learned from interpersonal trust can be applicable to the relationship between the human and the imperfect system, while lessons learned from studying trust in computers may be less applicable.

Ergonomics approached the problem of trusting machines from a more pragmatic angle: to what extent it is advisable to support the development of such trust in machines, knowing that they are in fact not intentional, as the 'as if intentional' behaviour can be easily traced to clever programming. The notion of an appropriate level of trust has been developed [20] (see also [8] for a similar consideration), where the main objective is not to encourage more trust than is due.

Following these lines of thoughts, trust can be considered as an explanation to some of the human behaviours while dealing with machines [32]. The emergence of animate software agents reinforced this role of trust [7]. Still, the problem of lack of actual intentionality in machines remained. Some authors (e.g. [29]), stated that it is the intentionality of the designer that modern technology exhibits, so that trusting machines is not a metaphor but an actual act of trust towards people and organisations, only conveyed through technical means.

One of the differences that emerged here is the fact that trust between people is supposed to be reciprocated [21,22]. That is, trust is not a one-way relationship between a trustor and a trustee, but a two-way trust-building exercise in mutual dependency and vulnerability. Currently, machines do not reciprocate, as IoT is not dependent on us while we are increasingly dependent on it. An interesting concept of device comfort [23] may alter this situation so that devices and systems may become partners in two-way trust relationships.

4.2 Disconnect between trust and trustworthiness

There is a trend in research as well as in the industry practices that focuses on trustworthiness of information systems, in assumption that such trustworthiness warrants trust. For reference, in that context, trustworthiness is usually defined as an objectified, collective statement regarding qualities of a trustee that may lead to trust. Such a statement is maintained by social interactions, e.g. in a form of a reputation.

Trustworthiness of information systems has been the subject of several research projects (see e.g. http://www.optet.eu/, http://www.inter-trust.eu, http://www.trescca.eu) . Despite the multiplicity of definitions [2], the key element is relatively simple: a trustworthy system does what it is supposed to do. That is, a trustworthy system delivers predictable and stable functionality.

This definition worked rather well for deterministic systems. Technical procedures such as trusted computing [28] made sure that computers never strayed from the set path, while security and reliability [3] made sure that the functionality of the system is resistant to both malicious and unintentional changes in its environment,.

By the same definition, for imperfect systems the user neither knows nor can verify that the system does what it is supposed to do, thus rendering them untrustworthy. Actually, the user cannot even distinguish between an imperfect algorithm and a sinister attack [15].

This calls for the relaxation of a popular assumption of "trust out of trustworthiness" being the only explanation for trusting. Fortunately, the relationship between trust and trustworthiness (often represented by reputation) is not straightforward, as exemplified by those two statements [17] being equally plausible: “I trust you because of your good reputation” and “I trust you despite your bad reputation”.

While imperfect systems are not trustworthy in the traditional meaning of this word, the observation that there is no direct implication of trust out of trustworthiness can be beneficial. Indeed, as trust may be extended towards untrustworthy entities as well as towards trustworthy ones, there is more to trust than just attuning to the level of trustworthiness.

4.3 Imperfect signalling systems

Trust in imperfect machines (known here as imperfect signalling systems) has been a research subject for some time in ergonomics [6]. The primary interest came from the area of work automation, where the operator should be able to trust their monitoring / advisory systems despite knowing that their advice or monitoring is imperfect.

Current study (see e.g. [18] for an overview) focuses on the impact that two categories of events (false alarms and misses) have on trust. While dealing with imperfections, trust may affect operator's strategies [1]. Experiments conducted in laboratory settings focus on operators' allocation of attention in high-load work environment where operators must split their attention between multiple concurrent tasks (e.g. [5]).

Findings from those experiments vary, and to the author's knowledge, no established model to explain the relationship between trust and the workload or the level of reliability emerged as yet. The theoretical model of trust in human-automation uses three

informational bases: performance, process and purpose, where performance refers to what automation is doing, process reflects how automation operates, and purpose describes why automation is developed [19].

For as long as the overall demand for operators' attention is bearable, operators tend to correctly calibrate their trust, i.e. they trust those systems that are more trustworthy (i.e. provide more consistently accurate information) [20]. Once the demand for attention (i.e. the workload) becomes excessive, trust is not calibrated correctly, resulting in the overall lower trust towards imperfect technology.

However, some as yet unpublished works suggest that an increase in workload can actually increase trust, specifically if not trusting is the riskier strategy. This means that operators can exhibit trusting behaviour that does not reflect the reliability of the signalling system, but rather the precarious situation of the operator.

There is an apparent similarity between imperfect signalling systems and imperfect systems in general, and IoT in particular. The ability to fail from time to time, in a way that is not easily explainable to the user is a defining characteristic for all those systems. The main difference lies in the fact that for imperfect signalling systems, the user is able to eventually determine when the system is not operating properly while for the IoT or any imperfect system it is not always attainable.

4.4 Ontological security and basic trust

While the construct of trust is inherited from human-to-human relationships, it is not the only relationship that can be called 'trust'. It may be therefore worth exploring different trusts that are definitely reported by human trustors yet directed towards the non-human trustee. Amongst them, there is a concept of ontological security that leads to basic trust, known also as ontological trust or ontological security [12].

Ontological trust is a stable mental state derived from a sense of continuity of one's experience. It is - in a nutshell - an expectation that the world is predictable. It can be disturbed by the perception of chaos, uncertainty and unpredictability. If supported, ontological security allows for person's basic trust (i.e. disposition to trust) to develop.

The interest in ontological security and basic trust is specific to psychological studies in early childhood (where it contributed to the sense of self-identity and building the disposition to trust in general [34]), learning contexts, but also to studies in international relationships [25], family stability and other areas.

Giddens [12] states that ontological security allows for the attitude where a person accepts what cannot be controlled, within the limits of some variability of its behaviour, on the basis that it is a stable, anticipated behaviour. For example an unexpected summer rain does not undermine ontological security, as it is expected that such a rain may come, even if it is not known when. In contrast, an earthquake in a geologically quiet area shatters not only the buildings but also related basic trust.

As IoT increasingly becomes an indistinguishable part of our environment, it would be worth considering whether IoT systems can be trusted 'as weather' rather than 'as devices'. That is, whether trust in the IoT would be better explained by basic trust out of ontological security than about the human-to-human trust.

To the author's knowledge there is no research in forms of ontological trust in technology. However, several trust models (e.g. [24, 31]) used in this area contain a component that is similar to the basic trust: the propensity (or a disposition) to trust.

4.5 Trust in abstract systems

Abstract systems are a concept introduced by Giddens [12]. Abstract systems use visible symbols or tokens (prescription, credit card) to represent the outcome of work of an otherwise opaque system (medicine, financial system). The average person does not know how those opaque systems work, but they know how to deal with those tokens. Hence, the person is in a position where they have to trust expert systems that they do not understand, on the basis of tokens alone. Tokens, are often, in fact, symbols of trust or evidence on which trust is assumed.

Note the precarious position of a user of such systems. Abstract systems are unavoidable and pervasive. They are usually monopolistic or near-monopolistic in nature. They are complex and their operations are hard to grasp. They are not directly controllable and yet they are trusted, with occasional complaints.

The parallel between the IoT and abstract systems is clearly visible, as the average person has no skills to comprehend their operation and is only exposed to some symbols of its operation (e.g. displays, end user devices, information).

Abstract systems do a lot to stabilise our lives and for that reason they are usually trusted. It is not because of their inherent trustworthiness, nor for the choice that the user has, but because of their usefulness. They empower people to do things that could have been otherwise impossible, whether it is a new treatment or a payment in a shop.

One may ask whether it is a genuine trust, but then how can one tell a difference between a person trusting e.g. banks 'genuinely' and trusting banks 'out of convenience or necessity'. The visible outcome of such trust will be approximately the same, while any decision may be post-rationalised by a person.

4.6 Social systems and their theory

One of the key features of the imperfect system is the radical departure from the notion of a single truth (or a single meaning). In it, there is a striking parallel between the way imperfect systems work and the way the social systems theory models the operation of a society. Considering that there is a lot of computational concepts that took inspiration from social behaviours, than this parallel is worth exploring.

Social systems theory ([22], see also [26]) assumes that society is structured into systems that consist of communications. Systems can be very abstract (such as the legal system) or more specific (such as an organisation or even a particular single interaction). Systems continuously grow by acquiring communications and by evolving their meanings.

As a result, meanings that systems hold not only alter over time, but they can also be local to various interactions - i.e. the reaction of a system may differ depending on time and place. For example, a legal system may come to different conclusions now than it did several years ago, and the conclusions may differ between countries.

Positioning imperfect system (specifically the IoT) as a technical analogue to the social system allows to define trust in the IoT in the same way as one social system can trust another one, that is to reduce its complexity.

The challenge every social system face is not to be overwhelmed with the complexity it deals with. One of the possible solutions is to rely on other systems by trusting them. That is, the 'trustor' is exporting some of its complexity to the other system (the 'trustee'), thus making itself dependent on its vagaries. Drawing from this, an analogy would be for people to export some of the decision-making complexity to the IoT and become dependent on imperfections of the technology and its decisions.

Such trust does not require the trustee to have any particular properties of trustworthiness, nor the trustor to have a choice of trustees. The trustor often has to trust someone, picking the best option it has, even if it is the only one.

However, this analogy has limitations. Trust between social systems develops as a mutual one: both systems export some of their complexity to their counterparty and both become vulnerable. This situation does not translate easily into the relationship with the IoT, unless the (already mentioned) concept of device comfort [23] will be taken into consideration.

5 Conclusions

The IoT does not fit easily into the established way of thinking of trust in technology, that essentially mimics relationships between empowered humans. This leaves several questions open. Specifically, around the monopolistic position of the imperfect IoT that asks for revisiting the concept of trust.

The author does not have a definitive solution how to approach human trust in the IoT. However, the author believes that there is more to trust than studying humans trusting humans, and that some of those alternative views better resonate with the position the IoT will take in the society.

Therefore, instead of a solution, the overview of possible approaches is presented , derived from various research domains. This paper discussed the following approaches to trust that are applicable to IoT in particular and to trust in imperfect systems in general.

- Trust by replicating human-human trust (intentionality)
- Trust out of trustworthiness (trustworthy information systems)
- Trust because of reliability (imperfect signalling systems)
- Trust in stability and predictability (ontological security and basic trust)
- Trust out of necessity and usefulness (trust in abstract systems)
- Trust by replicating social trust (between social systems)

There is no single theory that can explain the whole relationship between people and the IoT, but each one will explain some of its elements. Collectively, those approaches provide sufficient substrate to draw from in order to develop a more relevant theory of trust.

After all, there's more to trust than trustworthiness.

References

1. Bailey, N.R., Scerbo, M.W.: Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. Theor. Issues Ergon. Sci. 8 (4), 321–348. (2007)
2. Becker, S., Hasselbring, W., Paul, A., Boskovic, M. Koziolek, H., Ploski, J., Dhama, A., Lipskoch, H., Rohr, M., Winteler, D., & Giesecke, S., Meyer, R., Swaminathan, M., Happe, J., Muhle, M., Warns, T.: Trustworthy software systems: a discussion of basic concepts and terminology. ACM SIGSOFT Software Engineering Notes. 31. 1-18. 10.1145/1218776.1218781. (2006)
3. Bishop M.: Introduction to Computer Security. Addison-Wesley. ISBN 0-321-24744-2. (2005)
4. Blanter, A., Holman M.: Internet of Things 2020: A Glimpse into the Future. Available: http://aradinfocenter.com/wp-content/uploads/2017/07/A.T.%20Kearney_Internet%20of%20Things%202020%20Presentation_Online.pdf. (2017)
5. Bliss, J.P., Dunn, M.C.: Behavioral implications of alarm mistrust as a function of task workload. Ergonomics 43 (90), 1283–1300. (2000)
6. Bliss, J. P., Gilson, R. D., & Deaton, J. E.: Human probability matching behaviour in response to alarms of varying reliability. Ergonomics, 38, 2300-3212. (1995)
7. Castelfranchi, C.: Modelling social action for AI agents. Artificial Intelligence, 103, 157–182. (1998)
8. Cofta, P.: Trust, Complexity and Control: Confidence in a Convergent World. John Wiley & Sons, Ltd. ISBN:9780470061305. DOI:10.1002/9780470517857. (2007)
9. Dennett, D.C.: Intentional Stance. MIT University Press Group Ltd. ISBN: 9780262540537. (1989)
10. Dietvorst, B. J., Simmons, J. P., & Massey, C.: Overcoming Algorithm Aversion: People will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. Management Science, 64 (3), 1155-1170. DOI: dx.doi.org/10.1287/mnsc.2016.2643. (2016)
11. Flasiński, M.: Wstęp do sztucznej inteligencji (in Polish). WN PWN. ISBN 978-83-01-16663-3. (2011)
12. Giddens, A.: Modernity and Self-Identity. Self and Society in the Late Modern Age. Polity Press. (1991)
13. Han, J., and Orshansky, M.: Approximate computing: An emerging paradigm for energy-efficient design. 2013 18th IEEE European Test Symposium (ETS). IEEE, (2013)
14. Hewitt, E.: Cassandra: The Definitive Guide. O'Reilly Media. ISBN: 978-1449390419. (2010)
15. Huang, L., Joseph, A.D., Nelson, B., Rubinstein, B.I.P., Tygar, J.D.: Adversarial machine learning. In: Proceedings of the 4th ACM workshop on Security and artificial intelligence, Chicago, Illinois, USA. ISBN: 978-1-4503-1003-1. (2011)
16. Jain, A.: Mastering Apache Storm: Real-time big data streaming using Kafka, Hbase and Redis. Packt Publishing. ISBN-13: 978-1787125636. (2017)
17. Josang, A.: Trust and Reputation Systems. In: A. Aldini and R. Gorrieri (Eds.), Foundations of Security Analysis and Design IV, FOSAD 2006/2007 Tutorial Lectures. Springer LNCS 4677. ISBN 978-3-540-74809-0. Bertinoro, Italy, September 2007. (2007)

18. Karpinsky, N.D., Chancey, E.T., Palmer, D.B., Yamani, Y.: Automation trust and attention allocation in multitasking workspace. Applied Ergonomics 70. DOI: 10.1016/j.apergo.2018.03.008 (2018) 194–201
19. Lee, J. D., & Moray, N.: Trust, control strategies and allocation of function in human-machine systems. Ergonomics, 35, (1992) 1243–1270.
20. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors 46 (1), (2004) 50–80.
21. Lewicki, R. J., & Bunker, B. B.: Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.), Trust in organizations: Frontiers of theory and research (pp. 114-139). Thousand Oaks, CA: Sage. (1996)
22. Luhmann, N. Social Systems. Stanford University Press. (1995)
23. Marsh, S., et al.: Defining and investigating device comfort. Information and Media Technologies 6.3 (2011) 914-935
24. Mayer RC, Davis JH, Schoorman FD.: An integrative model of organizational trust. Academy of Management Review. 20 (3): 709–734. DOI: 10.5465/amr.1995.9508080335. (1995)
25. Mitzen, J.: Ontological Security in World Politics: State Identity and the Security Dilemma. European Journal of International Relations (12(3)): (2006) 341–70
26. Moeller H-G: Luhmann Explained: From Souls to Systems. Open Court. ISBN-13: 978-0812695984. (2006)
27. Nakamoto S.: Bitcoin: A Peer-to-Peer Electronic Cash System. Available: https://bitcoin.org/bitcoin.pdf. (2008)
28. Pearson S., et al.: Trusted Computing Platforms: TCPA Technology In Context. Prentice-Hall. (2002)
29. Rasmussen, J., Pejterson, A. M., & Goodstein, L. P.: Cognitive systems engineering. New York: Wiley (1994)
30. Ritschel, T., Grosch, T., Kim, M. H., Seidel, H.-P., Dachsbacher, C., Kautz J.: Imperfect shadow maps for efficient computation of indirect illumination. In ACM SIGGRAPH Asia 2008 papers (SIGGRAPH Asia '08), John C. Hart (Ed.). ACM, New York, NY, USA, Article 129, 8 pages. DOI: https://doi.org/10.1145/1457515.1409082. (2008)
31. Tan, Y. and Thoen, W.: Toward a Generic Model of Trust for Electronic Commerce. International Journal of Electronic Commerce, 5, 61-74. (2001)
32. Tenney, Y. J., Rogers, W.H., & Pew, R.W.: Pilot opinions on cockpit automation issues. International Journal of Aviation Psychology, 8, (1998) 103–120.
33. Varga, E., Draškovic, D., Mijic, D. Scalable Architecture for the Internet of Things. ISBN: 9781492024132. (2018)
34. Winnicott, D.W. The Maturation Process and the Facilitating Environment. Studies in the Theory of Emotional Development. The International Psycho-Analytical Library, 64:1-276. London: The Hogarth Press and the Institute of Psycho-Analysis. (1965)