On Security Singularities

Wolter Pieters
TU Delft
Delft, Netherlands
w.pieters@tudelft.nl

ABSTRACT

In future studies involving artificial intelligence, the so-called technological singularity is a key theme. It refers to a hypothetical point in the future where technological progress becomes automated through the creation of a new form of intelligence. Under the assumption of adversarial behaviour, this could pose an existential threat to humanity. More modestly, singularities and tipping points refer to thresholds beyond which the behaviour of a system changes in a qualitative way. The nonlinearity of the behaviour causes existing control mechanisms to become obsolete, guiding the system towards a new balance, if this exists. In this paper, we ask the question to what extent the notions of singularity and tipping point can contribute to an analysis of security in 2038. Can we expect to have seen such phenomena in twenty years time, and will they have changed our perception of what security entails? Or are they useless forms of speculation diverting our attention away from the day-to-day best practices that are needed to keep our basic security up-to-date? We discuss examples of singularity-style developments, characterise them in terms of acceleration mechanisms and discontinuities, and discuss whether and how these characteristics should be used to prepare ourselves. We conclude that a broad discussion on potential security *singularities* and associated general adaptation strategies is more useful than focusing on one big *singularity*.

CCS CONCEPTS

• **Security and privacy → Social aspects of security and privacy**; *Economics of security and privacy*;

ACM Reference Format:
Wolter Pieters. 2018. On Security Singularities. In *New Security Paradigms Workshop (NSPW '18), August 28–31, 2018, Windsor, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3285002.3285008

1 INTRODUCTION

We are fascinated by big and irreversible changes. The meteor that wiped out the dinosaurs, the fall of the Roman Empire, the atomic bomb. Although some of these have "big causes", small and gradual developments may also cause a system to transition to an entirely different state during a relatively short timespan. In scientific and technological developments, these events are often referred to as revolutions. Revolutions do not only change our world; they often also change our position within it. This holds for Copernicus as well as social media.

In this context, it is unsurprising that we are already looking forward to the next big change. The domain of artificial intelligence (AI) seems to be particularly fascinating. If we would build something that is more intelligent than we are, what would happen? Again, this would fundamentally change our position in the world, and under some scenarios, the world would never be the same again, since we would be unable to control it. This so-called *technological singularity* is a big theme in futurist studies of AI, and also sparks philosophical and ethical debate in terms of (ir)responsible design [11]. At the same time, people are wondering whether this speculative scenario doesn't draw our attention away from more pressing ethical issues in AI, such as discrimination via algorithms.

Losing control to intelligent machines could also be seen as a security issue, under the assumption that these machines could have adversarial goals that would harm our own. At the same time, there may be other developments in security that could be thought of in terms of singularities, not because we lose control to machines, but because we may *lose control to other actors through machines.* The possibilities of uncontrollable cybercrime or hijacked democracies would fall under this category, and power is a central property in those scenarios. This thought is the starting point for the exercise described in this paper: to what extent is the concept of singularity useful for futurist studies on security?

Just to be clear, we are not aiming to analyse cybersecurity in the context of *the* technological singularity referred to in AI research. There are indeed papers that aim at defining cybersecurity approaches in the face of the technological singularity or more generally advancements in AI [22]. The aim of the present paper is to identify whether it makes sense to speak of singularities *within* the security domain itself.

There are several other concepts that have similar connotations to the notion of singularity, although they seem to be somewhat more modest. Without claiming to be exhaustive, it seems that the concept of *tipping point* [14] is useful to include in this exercise, because it also refers to situations in which a system quickly transforms based on relatively small developments. We will explain the concepts and their relation in more detail in the next section.

In order to evaluate the usefulness of the singularity concept, we first expand on the existing definitions and uses of key terms. Then, based on several examples, we try to derive a taxonomy of key properties of security developments that could be understood as singularities. We then evaluate the benefits and drawbacks of looking at security developments through the singularity lens. We end with implications for the governance of security, and some speculations on what might happen until 2038.

NSPW '18, August 28–31, 2018, Windsor, United Kingdom

ACM ISBN 978-1-4503-6597-0/18/08…$15.00
https://doi.org/10.1145/3285002.3285008

2 BACKGROUND AND RELATED WORK

As outlined above, the main inspiration for applying the notion of singularity to security comes from the AI domain, and the discussion on *the* technological singularity involving runaway forms of intelligence. However, despite the often seen use in the singular, different singularities (plural) can be distinguished. Eden et al. [11] provide an overview of singularity hypotheses, and suggest that there are two main scenarios. In the first scenario – the one we refer to here – the singularity is a point beyond which intelligent machines can rapidly improve new generations of themselves. In the second scenario, the key development involves human enhancement, leading to a new posthuman race.

If there is more than one singularity, an obvious question is what different singularities have in common. Based on their investigation, Eden et al. try to identify which elements are key in labelling a development as a singularity. The authors distil *acceleration* and *discontinuity* as key notions in singularity claims. Depending on the specific development, what exactly is claimed to be accelerating and what exactly is claimed to show a discontinuity may differ.

Singularity claims have sparked a lot of debate in the AI domain. According Eden et al., a key argument in favour of singularity hypotheses states that technological progress has been accelerating for a long time. In essence, the proponents show that relevant progress curves become steeper and claim that this trend can be extrapolated (J-curve). On the other hand, critics state that every such progress will level off at some point (S-curve). And even if something would change fundamentally, how would that be different from the small and large revolutions we've seen in history all the time? The latter criticism also links up with the suggestion that singularity discussions lead to a waste of time that could have been spent on more pressing moral problems related to technological developments.

If one doesn't believe that a future singularity conceived is fundamentally different from what we have already seen, one may instead look at the potential of using *singularities* in the plural, taking into account that "singularities are part of the natural scheme of things" [6]. In this context, singularity would refer to a major evolutionary milestone (or revolution), without claiming that the particular development is unique. This is compatible with the combination of acceleration and discontinuity, and the question then becomes whether we can expect to observe this combination within the security domain, and if so, how.

From this more modest angle, it seems to be worthwhile to also look at the similar notion of tipping point. Gladwell [14] states that the key idea of the notion of tipping point is to think of social changes in terms of epidemics. Again, acceleration seems to be a key ingredient. In fact, Gladwell lists 3 (or 4) properties: little changes plus contagiousness, big effects, dramatic speed. Human behaviour is at the forefront in his work, and the translation is thus from pathogenic epidemics to social epidemics, which may explain some phenomena in security as well, such as attackers and defenders going after the same types of vulnerabilities. Using epidemics analogies in cybersecurity is definitely not new (see e.g. [12]). However, what we are after here is not the spread of infections in computer networks, or immunology for computer systems, but rather how the socio-technical system, including human actors, may change quickly due to causes that are somehow related to cybersecurity.

A key aspect in security singularities or tipping points is the role of strategic players, in particular adversaries/attackers. We are not only dealing with uncertain self-reinforcing phenomena and defensive responses, such as in the climate change debate, but also with strategic behaviour of the threats themselves. In mathematical terms, we are facing a game theory problem rather than a (probabilistic) decision theory problem. In game theory, the changes we are speaking about may be thought of in terms of shifting equilibria: under slightly changed circumstances, optimal strategies for players may be radically different. To what extent people in fact exhibit optimal behaviour can be questioned; however, no matter the exact drivers or value tradeoffs, many people may change their behaviour in a short period of time. So, the acceleration of social processes may be explained from contagion effects, but these can sometimes in turn be explained as rational behaviour changes in the face of a changed game, rather than mere imitation.

Tipping points do not necessarily involve large groups or herd behaviour. Friedman [13] mentions tipping points in security in relation to online crime reaching a point where service providers may decide to stop certain services or infrastructures altogether, leading to substantially larger societal damages than due to the crime alone. While using the term tipping point, this paper doesn't refer to social contagion as the main acceleration mechanism.

So, singularity claims typically combine acceleration of and discontinuity within developments. The notion of tipping point focuses primarily on social contagion as an acceleration mechanism. Both may be relevant in a security setting, but we need to be more precise about how the properties can materialise in a security setting to benefit from such an exercise. Based on a suggestion by Goodman [16], Danaher [10] already discussed the possibility of a "singularity of crime" in terms of exponential growth in connectivity and thereby scale of crime, as well as individual risk of becoming a victim. When we lose control over those developments, "[w]e might all be permanent and potential victims of crime". In the current paper, we try to take the discussion beyond connectivity as a driver and crime as the result.

3 EXAMPLES

In the previous section, we saw that singularities have been understood in terms of acceleration and discontinuity, and tipping points in terms of little changes, contagiousness, big effects, and dramatic speed. Let's now look at some security examples and see whether it make sense to describe them in these terms. These examples are by nature controversial, and the explanations are not meant to be objectively accurate, but rather to serve as a starting point for discussing the (non)sense of considering something a security singularity.

One seemingly obvious example of a discontinuity would be the development of a quantum computer. Such a machine is claimed to make several known security mechanisms (cryptographic algorithms) obsolete [8]. The actual achievement of a quantum computer breaking RSA is clearly a discontinuity. The more small successes in the process towards such an achievement, the more effort is put into it, providing an acceleration dynamics as well. However, the cause of this discontinuity can hardly be understood as a "little change" leading to such acceleration. So this could be an example of a potential singularity (discontinuity after acceleration) that is not a tipping point (epidemic/contagion).

A second example could be sought in the increasing scale of cyberattacks. The term "Cyber Pearl Harbor" is used for an attack of such a scale as to serve as a wake-up call due to the damage done [23, 24, 38]. Like exemplary singularity claims, it also represents a clear example of a doomsday scenario [23]. However, apart from being the apotheosis of a gradual increase in the size of attacks, it could be questioned whether this example has the characteristics of contagion or acceleration through system dynamics. Instead, it could be said to be a one-off event planned by adversaries that believe they are capable of organising such an event. Although the aftermath may lead to substantial systemic changes, the gradual acceleration process towards the singularity seems to be lacking. What *could* be conceived in terms of singularities is the gradual increase in scale of attack infrastructures such as botnets; when these reach a critical mass in terms of the magnitude of the attacks made possible, thereby enabling a massively bigger attack, this could be conceived in terms of acceleration and discontinuity, especially when the attack infrastructures would be used to infect more devices in order to add them to the infrastructure (self-reinforcement).

A third example is provided by the developments in the area of the Internet-of-Things (IoT). Currently we see massive amounts of new smart devices being connected, without much attention being paid to their security and the associated maintainability. Smith and Erickson [32] argue that next to the possibility of a Cyber Pearl Harbor, we should also be worried about "segments of our cyberinfrastructure, rendered uninhabitable". Those "cyber brownfields" may have unexpected interactions with other parts of our infrastructure. The analogy here is with environmental contamination. There is an accumulation effect: low levels of contamination are relatively harmless, but beyond a certain threshold the environment may quickly transform into a no-go area. In addition, the contamination may spread to other parts of the network.

A fourth example is the current fear of democracy being hijacked by online forms of targeted persuasion. If at some time in the future the possibilities of online influencing would reach a point where critical voices could be completely overruled, and additional censorship could be put in place to keep it that way, this could be called a singularity. Formulated in terms of the AI singularity: we are already being controlled by machines/cyborgs, but more through profiling algorithms than through self-replicating machines. The acceleration here lies in the fact that the more people who can be influenced to support a particular political movement, the easier it becomes to silence opposing forces. Although depending on the definition this may not be seen as a cybersecurity issue in the strict sense, the fact that this is currently possible could indeed be evaluated as a failure to organise proper security online.

A fifth example is the servicification of cybercrime. When components of criminal modi operandi are standardized and marketed [17], these components can be produced and consumed at a much larger scale. Here the singularity corresponds to the emergence of a market. Rather than technical issues or radical technical progress, the singularity is in this case much more closely related to social processes, and therefore also to the original notion of tipping point.

In addition to examples that worsen the situation for the defender, singularities might also represent acceleration and discontinuities in defensive efforts. For example, when developments in artificial intelligence would enable much more sophisticated automation of vulnerability checking in software, this may at some point decrease the number of exploitable vulnerabilities disruptively. In such a situation, the question is whether offensive technologies can seek similar disruptive changes, keeping the arms race intact, or whether cybercriminals might leave the cybercrime domain altogether in favour of easier targets.

As we have seen, these examples illustrate different types of developments that exhibit different characteristics in terms of acceleration/contagion and (potential) discontinuity. In the next section, we'll try to approach this more systematically.

4 CHARACTERISTICS OF SECURITY SINGULARITIES

Based on the examples discussed, we can try to systematise the relevant properties of singularity-style developments, in order to support evaluating them. What are the key features that may help us characterise potential security singularities or tipping points, and how can these features be used to discuss the (ir)relevance of those developments?

Causes. We observe different types of causes in the example developments. In the quantum computing example, development of (fundamentally) new technologies may also cause qualitative changes in the security landscape. Attacks of unseen magnitude may be enabled by (gradual) increases in the size of the attack infrastructure. Cyber brownfields may emerge due to the inability to update the security of legacy devices, scalable forms of persuasion may enable the transformation of open democracies into something else, and specialisation in cybercrime may induce new markets. Finally, on the defensive side, radical improvements may have a disruptive effect on the available tools for attackers. The initiating events may thus be related to new technological developments, changes in attack infrastructures and markets, constraints imposed by existing technologies (and associated vulnerabilities), and the ability of adversaries to leverage technological infrastructures for their own purposes.

Contagion/acceleration. Somehow the initial developments need to undergo acceleration to create a singularity or tipping point (J-curve or hockey stick). This acceleration may have a physical or technological reason (as claimed in climate change), but may also be due to social contagion, which is the key feature in the concept of tipping point. One key contagion type involves attackers and defenders concentrating on the same issues. For example, when macro viruses emerged, a lot of both attacker and defender effort was put into exploiting and mitigation this type of vulnerability. This behaviour leaves other potential vulnerabilities unexplored. When these become more popular later, they have already become much more widespread than necessary. In this sense, arms races between attackers and defenders facilitate tipping points, both because they are contagious in themselves (concentrating attacker and defender effort on the same type of attack/defence), but also

Table 1: Characteristics of security singularities

Development	Causes	Contagion/acceleration	Effects/discontinuity
Quantum computing	emerging technology	small successes	computer that breaks crypto
Cyber pearl harbor	size of attack infrastructure	infection spreads	large-scale attack
Cyber brownfields	legacy technology	accumulation of no-go systems	unanticipated interactions
Hijacked democracy	persuasive technology	silencing opposition	loss of democratic control
Servicification	specialisation	criminal business case	emergence of a market
Automated defence	artificial intelligence	exponential improvement	disruptive decrease of exploitable vulnerabilities

because they draw attention away from other attack/defence areas, leaving those open for future arms races.

In the examples, we observe different forms of acceleration. In the quantum computing and automated defence cases, small successes may accelerate the development. For attack infrastructures and cyber brownfields, the infrastructures may be used to infect more machines and thereby self-reinforce the network. The existence of a criminal business case may quickly draw more bad guys into offering certain specialised services. Finally, for hijacked democracies, the more people can be (technologically) persuaded to support a particular movement, the easier it becomes to silence the remaining opposition.

Effects/discontinuity. The other key feature of a singularity or tipping point is a discontinuity, forcing the system to a qualitatively different state. The discontinuities in the examples may also take different forms. In the hijacked democracy case, the discontinuity lies at the point where we would be unable to regain democratic control. In the quantum computing case, this would be the point where existing cryptography could be broken on a large scale. For large-scale attacks, this is when an attack occurs of such a magnitude that it leads to a qualitatively different approach to Internet governance. The unanticipated interactions of cyber brownfields with other infrastructure may have similar implications when a certain size of the no-go zone is reached. The emergence of an efficient market constitutes the discontinuity for servicification. Finally, a discontinuity for automated defence would happen if methods are developed that radically reduce the number of vulnerabilities, *and* these methods are suitable for widespread adoption.

In Table 1, we list the abovementioned properties for the examples we discussed. Again, this is not meant to be an exhaustive or objectively correct characterisation. Rather, these are ideas on how the example developments can be characterised, which can and should serve as the basis for further discussion.

For each of the table entries, a discussion can take place on:

- whether the causes, acceleration mechanisms, and effects make sense as characterisations of the developments;
- whether that justifies labelling the developments as potential singularities or tipping points;
- whether the occurrence of the sketched characteristics is realistic or not; and
- what could be done about the potential developments if we find them both realistic and undesirable.

5 WHY WE SHOULD OR SHOULD NOT CARE

So we may be able to evaluate whether a particular development in security could count as a singularity or tipping point, and what the relevant characteristics are. Based on these characteristics, we may even be able to anticipate the development and prepare ourselves. The big question that remains is whether that is even needed. How serious should we take security singularities? And how likely are they to materialise? Wouldn't the problems solve themselves, as the opponents of the AI singularity also claim?

We have to admit that there is an aspect of fearmongering to discussing possible future developments that will change the world. It's also fairly easy to make such claims, and it's hard to refute them by proving that they are completely impossible. We may be able to speculate about their likelihood, but that's about it. Like many other claims in the security domain, claims on possible singularities are inherently counterfactual [18, 19], making it hard to refute such claims on scientific grounds.

One reason to refute singularity claims is that everything is eventually an S-curve rather than a J-curve. Something will stop the exponential growth eventually. In the long term, this may show up in a form similar to a step function, a series of plateaus. The question is what the world will look like in the new stable situation, and whether we find that situation acceptable (existence of the human race, level of societal disruption, collapse of technological infrastructure, etc.) Part of such a judgement is the reversibility: will we be able to recover to an acceptable extent? If not, can we find ways to trigger the stabilisation earlier, thereby increasing the manageability of the events? To this end, we may want to investigate (a) how to recognise the precursors of exponential growth, and (b) what type of effects can lead to stabilisation. In other words, can we adapt to the time scales of the events we are dealing with? Can we think of passive safety/security, such that developments will be slowed down by their own growth, or trigger defensive singularities that cancel out the offensive ones? Or can the attackers help us, because they don't favour certain extreme outcomes either (like viruses don't want to eliminate their host population)?

We may decide to look towards the past in order to get some impression of similar events. When "hard" scientific or mechanistic evidence is lacking, narrative and historical explanations can provide some guidance as to what may or may not happen in the future [3, 15]. It can be questioned whether information and cyber security have already seen major singularities, although there certainly have been cases in which attacker and defender behaviour have reinforced each other. Looking a bit broader, have there been singularity-style developments in the field of general security?

For past developments in the security space, several arguments have been made on how relatively small developments revolutionised the security arena. Existing defences may become useless in the face of new weapons: city walls were torn down worldwide after the invention of gunpowder. However, the developments need not initially involve new weapons. A case in point is the analysis by Lynn White [37] on how the adoption of the stirrup changed warfare practices (knights), and then also society in general (feudal system). Although the evidence for these claims is disputed, what we use the example for is to illustrate how relatively small changes that seem fairly remote from security could potentially change the security battlefield and the environment around it. Like in the stirrup example, the discontinuities may occur both within what is commonly seen as the security domain (knights), but also well outside this system (feudal system). Loss of democratic control is clearly an example of the latter.

Another example of a changed battlefield is the vulnerability of the Roman aqueducts when the empire became raged with war. Assante [2] argues that while the aqueducts were originally built underground, there was a move towards overground building for reasons of changed perception of risk and the ability to show off. There is an acceleration dynamics here in terms of cities wanting to show at least the same amount of architectural greatness as others, effectively covering up the need for protection. Germanic invasions were then able to target the critical infrastructure of the aqueducts as parts of their attacks. Obviously, when adversaries learn that this is a successful strategy, there is an again an acceleration dynamics in terms of using the same strategy for other target areas, creating an opportunity for major disruptions of the Empire, forcing the social system to a new state, with the ability to support only a much smaller population. Because of the dynamics of acceleration/contagion and discontinuity, this could again be conceived as a singularity or tipping point.

So, a central theme in security singularities is this: they change the battlefield (rather than just the weapons). Such changes may indeed have happened in the past. Therefore, it may be worthwhile to proactively assess possible future (cyber)security singularities. This is the rational version of the argument, but emotions seem to play a key role as well. In this context, it is worthwhile to engage in a small digression on possible responses to developments that constitute potential singularities.

Claims about singularities may evoke different types of responses, in which emotions play a key role, not in the least because there is often little hard evidence. In relation to new technological developments, Smits [33] claimed that such responses reflect mismatches of the sketched development with our existing cultural categories. That is, our conceptual framework doesn't have a "place" for these things, making it difficult to evaluate them rationally. Smits outlined four typical types of responses to such developments: exorcism, adaptation, embracement, and assimilation. Exorcism tries to ban the mismatched phenomenon from the world (e.g. prohibiting a new technology), adaptation tries to make the phenomenon fit existing categories (e.g. changing the technology to fit existing legislation), embracement tries to emphasise that the mismatch is actually good and exciting (e.g. celebrating the fundamentally different properties of a new material), and finally assimilation tries to resolve the mismatch by changing both the phenomenon and our cultural categories.

For singularities, similar responses may be evoked, and we see them in the AI space. The AI singularity may be banned, celebrated, recast as just another step in a logical development, or seen as a reason to rethink our notions of agency and responsibility. The fact that the idea of machines taking over "doesn't fit" makes it exciting to discuss these matters. The presence of different types of responses ensures that the debate can last for a long time. At the same time, we see another contagion mechanism at work here, making a large crowd focus on the same issue. Again, this may leave other relevant developments underexplored.

So, like many technological developments, potential singularities may lead to polarised discussions based on emotional responses that reflect our inability to make sense of them. Next to the rational argument, a second reason we should care about singularities is therefore that if we don't, others may waste useful resources on a polarised discussion. Even if we don't think singularities make sense, we may need to canalise the discussion on potential singularities. Therefore, one or two major singularity claims, as in AI, may not be the best way forward. Instead, can we identify a set of potential singularities that can help us discuss possible threatening future developments and prepare for them? If so, what do we need to watch out for, and how should we prepare?

6 WHAT TO WATCH OUT FOR UNTIL 2038?

Discussing security singularities may thus be found to be relevant for two reasons: because they may actually happen, and because we may want to canalise the *discussion* that may happen, even if the suggested singularities would be ridiculous. Without claiming to be exhaustive, the previous discussion highlights a number of issues that we may want to take into account if we would want to discuss a healthy ecosystem of singularity claims in security land, and possible ways of dealing with those. These issues also pose challenges to security governance in the face of possible singularities or discussions about those. We will discuss three themes below, and some initial ideas about responses.

6.1 Legacy issues

Many of the potential singularities rely on outdated, insecure systems that are out there in large numbers. These legacy systems provide opportunities for large-scale attack platforms, which can also be leveraged to expand the platform itself (acceleration/contagion). Thus, this seems to point to the idea that we should somehow prevent orphaned devices by ensuring updateability.

A related question is whether the standard discussion on updateability of devices, for example in the Internet-of-Things, is sufficient. That is, is the requirement that software can be updated sufficient if something of the magnitude of a singularity hits the security domain, such as quantum computing? Or would this require some more rigorous forms of adaptability or reconfigurability?

Path dependency is a key notion here. Whatever we create now in terms of security solutions also fixes constraints on what is or isn't possible in the future. The effects of what we design are not limited to the artefact; we need to think about design in terms of the constraints that we throw into the world rather than the products. It's about the "circulation of effects" [26]. Within this context, deployment strategies are as important as design strategies. In particular, how do we evaluate the effects when the scale is increased? We didn't get this right for the Internet; we didn't get this right for social media. How is it possible that we are now surprised that sensitive websites contain tracking code from social media services? Given the constraints and incentives that were thrown into the world, it's rather obvious that this would happen. Did we prefer inaction over action, and if so, could and should we become more cautious [29]? And what does the possibility of quantum computing mean for our current deployment strategy of cryptographic solutions?

Singularities never happen in the design stage. We need comprehensive security engineering, we need to pay attention to lifecycles, and we need more attention for deployment and maintenance. Scaling up needs to happen more consciously, as we can also learn from the Roman aqueducts case [28, 35].

6.2 Emerging technologies and actors

A key feature in at least some of the singularity examples are emerging technologies. This is especially obvious for the case of quantum computing. In this case, the emerging technology could be used as a "weapon" to break encryption schemes. In other cases, emerging technologies could be used as new targets for attacks. The Internet-of-Things shows potential for both: infected IoT devices can be used as weapons, but they may also be interesting targets in themselves. We see a potential here for battlefield changes.

If some potential technology would actually materialise, this could have large-scale effects for what is (im)possible in terms of security. There are two acceleration effects in this case. First, in the R&D stage, the more credible it becomes that the technology will actually work, the more effort is being put into its development. Second, in the deployment stage, the more people use a technology, the more people will want to use it (a spectrum from early adopters to late majority). In security, however, there is an additional effect here: the more people use a technology, the more attractive it becomes for adversaries.

So, emerging technologies can play several roles in security singularities, and there are several acceleration mechanisms. However, it is not just technologies that are emerging. In the AI singularity, a key role is reserved for new forms of intelligence. In this case, these "emerging actors" would be AI programs or machines which instead of serving their human designers would develop goals and strategies of their own. In the human enhancement variant we mentioned earlier, human-technology cyborgs would emerge as hybrids of human actorship and machine capabilities.

One could argue that also in security singularities, a key role is played by new actors created in the form of human-technology alliances [36]. Cybercriminal networks emerge around technological structures such as botnets, cryptocurrencies, and ransomware. New business models are created, and capabilities are offered as-a-service to others. Contagion takes place because successful business models are quickly imitated by others. This acceleration effect forces defenders to create new actors of their own: high-tech crime units, cybersecurity centres, etc. Emerging actors and institutions are thus both a characteristic of accelerating developments, and a possible means to achieve a new balance.

This also means that security singularities are neither technical nor social, but rather socio-technical. Technical and social infrastructures co-evolve. If more institutions focus on quantum computing, new technical solutions are more likely to emerge. If new technical solutions are close or even already available, more institutions will focus on those. Markets will be formed around the new technical possibilities, increasing availability of both offensive and defensive applications. Cybercriminals form networks with each other and with new technologies, enabling new ways of doing business, and new opportunities to take control of part of the infrastructure. New actors emerge that operate in the context provided by the new developments, again both offensive and defensive ones. The hybridity of those networks allows for different forms of acceleration and discontinuity, as we have seen in the examples. In order to understand and possibly even predict future singularities, we need to take this hybrid dynamicity seriously.

6.3 Implications for governance

The key governance question is obviously how to prepare for potential singularities. In particular, what preparation is needed to be able to respond quickly enough when developments accelerate? We have already discussed reconfigurability of technology in this context. In fact, one could point to recent discussions on resilience as a form of governance that resembles this idea. Resilience also adopts the idea of recovering by finding a new balance that is acceptably close to the old status-quo. However, it is not clear whether current resilience initiatives are capable of dealing with developments that have the acceleration and discontinuity characteristics of a singularity. With resilience, aren't we rather preparing for more of the known types of incidents rather than fundamentally new developments? If so, how could the notion of resilience be extended to account for singularity-style developments?

Apart from specific governance for addressing singularities, singularities may also have impact on other governance mechanisms for cybersecurity. Within the economics of security, insurance receives an increasing amount of attention as a possible incentive for organisations to improve their security. The idea is that if (small and medium) enterprises wish to protect themselves against bankruptcy due to a major breach, and therefore buy insurance, they may be interested in reducing the premium by meeting some minimal level of protection. However, offering cyber insurance is only interesting for insurers if they can rely on some patterns regarding the materialisation of risk, and in particular they are wary of insuring risks that are correlated [4]. When, for example, the emergence of quantum computing could spark widespread criminal activity via cracking keys, this may have implications for the willingness of insurers to cover damages. So, the existence of security singularities matters for the feasibility of widespread cyber insurance.

At the same time, when cyber insurance *would* become widespread, insurance companies have a clear incentive to "fight" singularities. If they want to avoid correlated risk, they have reasons to invest in trajectories that aim at maintaining security after potential singularities, trying to lead the system to a new stable state. Alternatively, and less inspiringly, they may simply exclude damages caused by events that could be characterised in terms of singularities.

So, if insurance companies would be disadvantaged by potential singularities, they have an incentive to do something about them. This thought provides the basis for a broader theme: how can we incentivise dealing with potential singularities? Which actors could be stimulated to help out, and how? Can we actually avoid constraints and lock-ins that force stakeholders to maintain the status quo and engage in ostrich policy? And which means could be offered to analyse the future in order to make sense of this exercise?

6.4 Identifying and evaluating singularities

The million dollar question thus seems to be whether we can make any sense of potential future singularities, and if so, how. This question carries elements of both traditional risk management and technological forecasting. On the one hand, we are facing uncertain future events that we may try to analyse in terms of likelihood and impact, even though any assessment will be very imprecise. At the same time, the uncertainty can be of a magnitude that requires reliance on narratives and comparison with historical situations rather than on numbers. No matter where we end up in this spectrum, the steps of risk management may provide some guidance for a process of dealing with singularities as well.

A necessary first step in dealing with potential future singularities is identifying them. This could be done through traditional forms of brainstorming, imagination, and similar [27]. Possibly certain modelling techniques could help here as well, but this requires the ability to show emergent behaviour that was not conceived by the designers of the models. The result of this exercise would be a list of scenarios that represent possible future security singularities. In the present paper, we have only provided a non-exhaustive list of examples we could think of, without much of a rigorous method.

As a second step, the scenarios can be analysed to assess key properties. From a risk management perspective, this would involve assessment of the likelihood and the impact of the scenarios. Of course the uncertainty margins are very large in the singularity case, and historical arguments and narrative explanations may play a key role.

As a third step, we may evaluate whether the properties of the scenarios demand any kind of response. Maybe we deem it safe to just let the developments happen; maybe we are so worried about the possible consequences that we want to do something. For example, is the disruptive potential of quantum computing big enough to invest in preparation already?

When certain scenarios demand action, a fourth step would consist of identifying possible responses. This could consist of making existing systems more adaptable in case a singularity materialises, starting to develop alternative technologies, preparing for new regulation and incentivising adaptation, etc.

Finally, we should keep an eye on actual developments to see whether our initial judgements still make sense, and if necessary adapt response strategies accordingly. If progress in building a quantum computer slows down, or alternatively speeds up, we may want to reconsider our coping strategies.

Several tools may be used to support the process outlined above. We have already discussed historical narratives as a possible tool for imagining the future. In addition, several futurist methodologies might be leveraged for timely identification of potential singularities. For example, Markley proposes a methodology for identifying what he calls "STEEP surprises". A STEEP surprise is a "plausible future event that is estimated to have low probability but high impact should it occur", with STEEP standing for Social/demographic, Technological, Economic, Ecological and Political [25]. The proposed method includes for example snowball surveys and imaginal time travel as a means to explore disruptors that may emerge from the interacting forces. Several variants of the STEEP approach exist, including STEEPV and PESTLE, with their own variants of explorative methods around the constituent factors of the acronyms.

In addition, modelling approaches could provide possibilities for investigating acceleration and possible discontinuities in terms of emergent behaviour in the complex system being modelled [1]. These modelling approaches need to take into account what we are looking for in terms of characteristics of singularities and tipping points, possibly in the form of system breakdowns [30]. For example, agent-based modelling claims to be able to explore emergent behaviour, derived from simulations of behavioural rules and interactions of individual agents. Some applications to the security domain start to appear [7, 21, 31], but they are not specifically focused on discovering acceleration and discontinuities.

Moving to a more normative or design approach, appreciative inquiry [9] and causal layered analysis [20] seek not only to identify "what might be", but also "what should be". The lesson we can learn from such approaches is that it may not always be necessary to identify problems first and then come up with solutions, but that we may also start with where we want to be in the future. In that sense, we could aim at identifying *positive* singularities (that may help the defenders) and steer towards those, rather than waiting for offensive singularities to emerge.

6.5 2038

Whether this exercise has any level of usefulness can only be judged when we reach 2038. Reconfigurability, emerging actors, or extended resilience may retrospectively be seen as key ingredients of a singularity-aware security strategy, or as outcomes of a meta-form of useless speculation *about* useless speculations. In any case, discussing security developments in terms of singularity-related concepts such as acceleration, contagion and discontinuity seems to be helpful to make sense of at least some future developments, although the examples discussed in this paper might be quickly replaced with more promising ones.

Regardless of what may happen around the bigger singularities discussed here, smaller-scale singularities will continue to make risk management in the cybersecurity domain hard. We can try to put numbers on the risk we are facing and adjust our investment and control selection accordingly, but the situation may be different tomorrow. We will still see small-scale accelerations and discontinuities in terms of major vulnerabilities found in widely deployed systems. Some developments follow cycles, in which disruptions and adaptations follow each other through time, for example spam [5] and more generally connection and disconnection of systems [34]. In this sense, the uncertainty lies mostly in the *size and impact* of the discontinuities that we may see in the next 20 years.

In terms of the bigger singularities, one example that we haven't covered is maybe the one that is closest to the AI-style singularity. What if an uncontrollable virus with "super powers" would emerge that would be able to adapt to whatever controls we come up with? This is another "AI taking over"-type scenario, but now more with a security sauce. Do these more radical scenarios make sense whatsoever? We may know in 2038. Or maybe we'll have messed up so badly that nobody will be able to evaluate except unrecognisable cyborgs.

7 CONCLUSIONS

In this paper, we investigated to what extent the notion of singularities, and the associated concept of tipping points, is useful in discussing developments in security until 2038. Keeping in mind the key features of acceleration and discontinuity, we looked at potential security singularity candidates. We identified relevant dimensions of those examples, and suggested an initial framework that helps to start a discussion on the singularity characteristics of potential developments and associated governance and design responses. Based on the framework, we highlighted some key issues that may play a role in the development and mitigation of security singularities in the next 20 years.

Although we believe the singularity lens on security developments could lead to useful discussion regarding the future of security, the examples, characteristics and issues outlined are by no means exhaustive. In fact, contrary to what's happening in AI, it seems to be beneficial to discuss loads of potential singularities first, and then converge on a set of seemingly important ones. In order to come up with general strategies, and prevent tunnel vision, having more than the two currently conceived in AI seems to be a good idea. Because of this, there may be reasons to prefer the concept of tipping point over the concept of singularity, as the latter is referred to in AI as *the* singularity. The disadvantage is that the tipping point concept seems to be constrained to social contagion as an acceleration mechanism.

In order to improve our understanding of potential security singularities, we think the following topics are worthy of future research (not exhaustive):

- Possible acceleration mechanisms and discontinuity types;
- Relevant characteristics of singularities that are currently not in the framework;
- Governance strategies for dealing with potential singularities;
- Tools (technical, psychological, futurist) that can be used in the process for coping with singularities;
- The existence of singularities that we know we cannot deal with;
- Connections between security singularities and singularities in other domains.

Based on a better understanding of security singularities, we may be able to engage in a singularity-aware form of security without falling into the trap of tunnel vision. Any *single* singularity has the potential of hijacking attention, polarising the discussion and drawing necessary resources away from other matters. In a sense, this represents the same flocking behaviour on a meta-level that we see when attackers and defenders concentrate on the same type of vulnerabilities. However, discussing *multiple* singularities and possible coping strategies could benefit adaptive strategies. The framework outlined in this paper could be a first step in developing a process for pro-active assessment of singularities and tipping points in the security space.

Acknowledgments

The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under Grant Agreement no 740920 (CYBECO). This publication reflects only the author's views and the Union is not liable for any use that may be made of the information contained herein. Thanks to the TPM "emerging actors" group for inspiration, and to Volker Roth, Jonathan M. Spring and Heather Vescent for useful comments.

REFERENCES

[1] Chris Arney et al. Using rare event modeling & networking to build scenarios and forecast the future. In *Network Science Workshop (NSW), 2013 IEEE 2nd*, pages 31–36. IEEE, 2013.

[2] Michael J. Assante. Infrastructure protection in the ancient world. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, pages 1–10. IEEE, 2009.

[3] John Beatty. Narrative possibility and narrative explanation. *Studies in History and Philosophy of Science Part A*, 62:31 – 41, 2017. SI: Narrative in Science.

[4] Rainer Böhme and Gaurav Kataria. Models and measures for correlation in cyber-insurance. In *Workshop on the Economics of Information Security (WEIS 2006)*, 2006.

[5] Finn Brunton. *Spam: a shadow history of the Internet.* MIT Press, Cambridge, MA, 2013.

[6] Eric J. Chaisson. A singular universe of many singularities: Cultural evolution in a cosmic context. In Amnon H. Eden, James H. Moor, Johnny H. Søraker, and Eric Steinhart, editors, *Singularity Hypotheses: A Scientific and Philosophical Assessment*, pages 413–440. Springer, Berlin, Heidelberg, 2012.

[7] Konstantinia Charitoudi and Andrew J. C. Blyth. An agent-based socio-technical approach to impact assessment for cyber defense. *Information Security Journal: A Global Perspective*, 23(4-6):125–136, 2014.

[8] Lily Chen et al. *Report on post-quantum cryptography.* US Department of Commerce, National Institute of Standards and Technology, 2016.

[9] David. L. Cooperrider and Diana Whitney. *Appreciative Inquiry: A Positive Revolution in Change.* Berrett-Koehler, San Francisco, 2005.

[10] John Danaher. Are we heading towards a singularity of crime? Philosophical Disquisitions, http://philosophicaldisquisitions.blogspot.com/2016/03/are-heading-towards-singularity-of-crime.html, 2016.

[11] Amnon H. Eden, Eric Steinhart, David Pearce, and James H. Moor. Singularity hypotheses: An overview. In Amnon H. Eden, James H. Moor, Johnny H. Søraker, and Eric Steinhart, editors, *Singularity Hypotheses: A Scientific and Philosophical Assessment*, pages 1–12. Springer, Berlin, Heidelberg, 2012.

[12] Stephanie Forrest, Steven A. Hofmeyr, and Anil Somayaji. Computer immunology. *Communications of the ACM*, 40(10):88–96, 1997.

[13] Allan Friedman. *Economic and policy frameworks for cybersecurity risks.* Center for Technology Innovation at Brookings, 2011.

[14] Malcolm Gladwell. *The tipping point: How little things can make a big difference.* Little, Brown, 2006.

[15] Stuart Glennan. Ephemeral mechanisms and historical explanation. *Erkenntnis*, 72(2):251–266, Mar 2010.

[16] Marc Goodman. *Future crimes: Inside the digital underground and the battle for our connected world.* Random House, 2015.

[17] Chris Grier et al. Manufacturing compromise: The emergence of exploit-as-a-service. In *ACM Conference on Computer and Communications Security*, CCS '12, pages 821–832, Raleigh, North Carolina, USA, 2012.

[18] Cormac Herley. Unfalsifiability of security claims. *Proceedings of the National Academy of Sciences*, 113(23):6415–6420, 2016.

[19] Cormac Herley and Wolter Pieters. If you were attacked, you'd be sorry: Counterfactuals as security arguments. In *Proceedings of the 2015 New Security Paradigms Workshop*, pages 112–123. ACM, 2015.

[20] Sohail Inayatullah. Causal layered analysis: Poststructuralism as method. *Futures*, 30(8):815–829, 1998.

[21] Stef Janssen and Alexei Sharpanskykh. Agent-based modelling for security risk assessment. In Yves Demazeau, Paul Davidsson, Javier Bajo, and Zita Vale, editors,

Advances in Practical Applications of Cyber-Physical Multi-Agent Systems: The PAAMS Collection, pages 132–143, Cham, 2017. Springer International Publishing.

[22] Carl E. Landwehr. Cybersecurity and artificial intelligence: From fixing the plumbing to smart water. *IEEE Security & Privacy*, 6(5):3–4, 2008.

[23] Sean T. Lawson, Sara K. Yeo, Haoran Yu, and Ethan Greene. The cyber-doom effect: The impact of fear appeals in the US cyber security debate. In *Cyber Conflict (CyCon), 2016 8th International Conference on*, pages 65–80. IEEE, 2016.

[24] Ronald P. Loui and Terrence D. Loui. How to survive a Cyber Pearl Harbor. *IEEE Computer*, 49(6):31–37, 2016.

[25] Oliver Markley. A new methodology for anticipating steep surprises. *Technological Forecasting and Social Change*, 78(6):1079–1097, 2011.

[26] Rolland Munro. Actor-network theory. *The SAGE handbook of power. London: Sage Publications Ltd*, pages 125–39, 2009.

[27] John B. Noftsinger, Kenneth F. Newbold, and Jack K. Wheeler. Future implications: Imagination, integration, and improvisation. In *Understanding Homeland Security: Policy, Perspectives, and Paradoxes*, pages 175–196. Palgrave Macmillan US, New York, 2007.

[28] Wolter Pieters, Dina Hadžiosmanović, and Francien Dechesne. Cyber security as social experiment. In *Proceedings of the 2014 New Security Paradigms Workshop*, pages 15–24. ACM, 2014.

[29] Wolter Pieters and André van Cleeff. The precautionary principle in a world of digital dependencies. *IEEE Computer*, 42(6), 2009.

[30] Chiang H. Ren. The characteristics of systems breakdown. In *How Systems Form and How Systems Break: A Beginner's Guide for Studying the World*, pages 103–175. Springer International Publishing, Cham, 2017.

[31] Marlies Rybnicek, Simon Tjoa, and Rainer Poisel. Simulation-based cyber-attack assessment of critical infrastructures. In Joseph Barjis and Robert Pergl, editors, *Enterprise and Organizational Modeling and Simulation*, pages 135–150, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[32] Sean W. Smith and John S. Erickson. Never mind Pearl Harbor – what about a Cyber Love Canal? *IEEE Security & Privacy*, 13(2):94–98, 2015.

[33] Martijntje Smits. Taming monsters: The cultural domestication of new technology. *Technology in Society*, 28(4):489–504, 2006.

[34] André van Cleeff and Roel J. Wieringa. Rethinking de-perimeterisation: Problem analysis and solutions. In *Proceedings of the IADIS International Conference Information Systems 2009*, pages 105–112. IADIS Press, 2009.

[35] Ibo Van de Poel. Society as a laboratory to experiment with new technologies. In Ibo Van de Poel, Donna Mehos, and Lotte Asveld, editors, *Embedding new technologies into society: A regulatory, ethical and societal perspective*, pages 61–87. Pan Stanford Publishing, Singapore, 2017.

[36] Wytske van der Wagen and Wolter Pieters. From cybercrime to cyborg crime: Botnets as hybrid criminal actor-networks. *The British Journal of Criminology*, 55(3):578–595, 2015.

[37] Lynn Townsend White. *Medieval technology and social change*. Oxford University Press, 1962.

[38] James J. Wirtz. The Cyber Pearl Harbor. *Intelligence and National Security*, 32(6):758–767, 2017.