Routledge
Taylor & Francis Group

Journal of Cyber Policy

ISSN: 2373-8871 (Print) 2373-8898 (Online) Journal homepage: http://www.tandfonline.com/loi/rcyb20

Cybersecurity, trustworthiness and resilient systems: guiding values for policy

Adam Henschke & Shannon Brandt Ford

JOURNAL OF CYBER POLICY, 2016
http://dx.doi.org/10.1080/23738871.2016.1243721

Routledge
Taylor & Francis Group

Cybersecurity, trustworthiness and resilient systems: guiding values for policy

Adam Henschke and Shannon Brandt Ford

National Security College, Australian National University College of Asia and the Pacific, Canberra, Australia

ABSTRACT
Cyberspace relies on information technologies to mediate relations between different people, across different communication networks and is reliant on the supporting technology. These interactions typically occur without physical proximity and those working depending on cybersystems must be able to trust the overall human–technical systems that support cyberspace. As such, detailed discussion of cybersecurity policy would be improved by including trust as a key value to help guide policy discussions. Moreover, effective cybersystems must have resilience designed into them. This paper argues that trustworthy cybersystems are a key element to resilient systems, and thus are core to cybersecurity policy. The paper highlights the importance of trustworthiness for resilient cybersystems. The importance of trustworthiness is shown through a discussion of three events where trustworthiness was the target or casualty of cyberattacks: Stuxnet, hacking of communications and the Edward Snowden revelations. The impact of losing trust is highlighted, to underpin the argument that a resilient cybersystem ought to design in trustworthiness. The paper closes off by presenting a general set of policy implications arising from recognition of the interplay between trust, trustworthiness and resilience for effective cybersecurity.

ARTICLE HISTORY
Received 15 April 2016
Accepted 13 September 2016

KEYWORDS
Trust; value sensitive design; resilience

> I very much fear that the internet, which has been the most powerful force for political, economic, and social change in my life, and maybe even all of history, will not be what my kids inherit. The internet is built on a system of trust and it is threatened like never before. (Singer 2014)

> If poor cybersecurity erodes trust and confidence in cyberspace, the economic opportunity of a connected Australian economy will suffer. (Commonwealth of Australia 2016, 14)

1. Introduction

Trust is essential for cyberspace to function. Key government, military and economic activities are now highly dependent on cyberspace to enable critical operations. The economic and structural consequences of an attack that disrupts national cyberinfrastructure has the

CONTACT Adam Henschke adam.henschke@anu.edu.au

potential to seriously impact upon trust in such systems. As a result, any discussion of cybersecurity policy must include a discussion of trust. In other words, cybersecurity policymaking should be attentive to the central role of trust for maintaining resilient cybersystems and have structures in place to repair trust. There is always a potential vulnerability, human mistake or malicious insider that can challenge even the best cyber defences. We need cybersystems that can survive attacks. This means that one of the priorities of cybersecurity policy is resilience. And we want to design systems that can survive such attacks with the minimum disruption in their operation. A key element of resilience is trust, which can be understood in a range of ways. The technical understanding of trust is largely synonymous with reliance, whereas the strategic understanding is synonymous with weighing up risk versus benefits. But there is also what Uslaner (2002) calls 'moral trust', where the focus is on the moral character and motivation of those people we are trusting. If the daily operations in cybersystems are to be effective, then the people involved must trust what they are seeing, who they are talking with and how their commands are being carried out.

The notion of 'moral trust' is the focus of this paper. First, the paper examines the importance of 'trust' to the effective functioning of cyberspace. Second, it presents three commonly discussed examples of cybersecurity incidents, frames these examples with reference to the importance of moral trust and considers ways in which trust relations can suffer. Third, the design of morally trustworthy systems is introduced as a policy priority because it generalises and formalises system resilience to cybersecurity incidents.

A focus on trust, trustworthy systems and resilience is becoming more common in discussions of cybersecurity. In 2011, the U.S. National Science and Technology Council stated that

> [a]ssuring continued growth and innovation in cyberspace requires that the public has a well-founded sense of trust in the environment. Increasingly frequent malware attacks and financial and intellectual-property thefts must be addressed in order to sustain public trust in cyberspace but address real threats to national security. (2011, ix)

Following that, in 2013, the White House advocated three strategic imperatives to help improve the resilience of critical infrastructure:

> 1) Refine and clarify functional relationships across the Federal Government to advance the national unity of effort to strengthen critical infrastructure security and resilience;
>
> 2) Enable effective information exchange by identifying baseline data and systems requirements for the Federal Government;
>
> 3) Implement an integration and analysis function to inform planning and operations decisions regarding critical infrastructure. (Office of the Press Secretary 2013)

More recently, the 2016 Australian Cyber Strategy stated that '[a]ll of us – governments, businesses and individuals – need to work together to build resilience to cybersecurity threats and to make the most of opportunities online' (Commonwealth of Australia 2016, 4). This paper adds to these cybersecurity policy discussions by looking at the relation between resilience and trust through the values designed into trustworthy systems.

Consequently, the methodology of this paper applies ethical analysis to the problem of resilience in cybersecurity. That is, it examines the conceptual and normative elements of trustworthy systems. In particular, we use value sensitive design (VSD), which is 'a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process' (Friedman et al. 2013, 56). The VSD method allows us to offer a brief conceptual analysis of different concepts of trust. We then use the concept of trustworthiness (i.e. an existing attitudinal belief about the motivations of those working in the given system) as the conceptual foundation upon which to design resilience into a computer system. There is a great deal of academic literature that looks at the technical elements of trust in cybersystems and the relations between humans and those technical systems.[1] Our focus, however, is on the ways that those technical elements play a role in the relations between people operating and using such systems. This is similar to David Clark and Marjory Blumenthal's shift from the End-To-End design principle to the Trust-To-Trust design principle, where trust is considered a cornerstone of good design (Clark and Blumenthal 2011). Hence, this paper advocates an approach to policy that recognises the trust relations that occur between different users. Such a system is more likely to be resilient when a breach or failure of cybersecurity occurs.

2. Trust beyond technology: trustworthiness and people

Three key concepts of trust are outlined for the purposes of this paper. A standard conceptualisation of trust is technically focused and interchangeable with the notion of reliance (McLeod 2015). 'Reliance' is a term typically used for the function of inanimate objects. But the limits of this concept are that it does not take into account human factors. For example, I rely on a pen to write, whereas I trust my friend to keep secret confidential information shared with her.

> Reliability is a technical concept, and relates to the correct operation of a component or system under specific circumstances. The concept of trust is a broader concept. A component may not be trustworthy even though it is technically reliable, because it is operated by an agent with interests and motivations that are not aligned with the end user. (Clark and Blumenthal 2011, 365–366)

Although the technical elements of trust are important and play a key role in the trustworthiness of humans operating the relevant systems, this paper argues that cybersecurity policy must include the human element of trust.

The human element of trust is also conceptualised in terms of what we might reasonably expect in our relations with people. This understanding of trust is sometimes called the 'strategic' model (Uslaner 2002, 20–26). For example, is it reasonable to predict my friend will keep a promise? If he has a good track record of keeping his promises to me in the past then it seems reasonable to predict that he will keep a promise he makes to me now. But if he has typically broken his promises to me in the past, then it might seem unreasonable to predict that he will keep a promise now. In the first instance, it is reasonable to conclude that my friend can be trusted, but in the second instance it might be more reasonable to judge that he cannot be trusted. According to Uslaner, '[t] rust on this account is an estimation of the probability that you will keep your promises, that you are trustworthy' (2002, 3).

But there is also a third way to conceptualise trust. This is what Uslaner calls 'moralistic trust'. He describes this approach to trust as

> a commandment to treat people *as if* they were trustworthy […] Moralistic trust is not a prediction of how others will behave. Even if other people turn out to be untrustworthy, moral values require *you* to behave *as if they could be trusted*. (Uslaner 2002, 18–19, emphases original)

From this perspective, trust is more than a form of risk analysis; it is not simply a matter of what we might reasonably expect from others given their past behaviour. I might reasonably predict that an untrustworthy friend will break his promise to me but I would still feel betrayed or wronged in some way if he did. Moralistic trust says something about the character of my friend – that he should care for me in some important way. Even if he has broken his promises in the past, I still hope that he feels positively disposed towards me and is concerned with my well-being.

We see the reliance and moralistic accounts contrasted here: 'People receive an e-mail but worry that it is spam or contains a virus. They are willing to receive it (because it is worth the risk), but they do not trust the sender' (Clark and Blumenthal 2011, 363). With the strategic concept of trust, the receiver calculates that the risk is worth the reward, but they do not judge the sender as worthy of trust – the receiver's attitudinal state lacks moralistic trust.

Having described three conceptualisations of trust, we are now better positioned to understand *trustworthiness*. We can see from the brief discussion above that moralistic trust can be the result of seeing someone *as* trustworthy. That is, *if* I believe my friend is trustworthy, *then* I trust them. If I see them as untrustworthy, *then* I might distrust them. Rather than considering that trust is the cause of seeing someone as trustworthy, this approach reverses that approach. It places the attitudinal stance that a person is trustworthy as prior to trusting. Trust understood this way is the *result* of seeing people as generally trustworthy. Part of Uslaner's point is that the background conditions of belief have an important impact on how people assess the statements and actions of others (Uslaner 2002, 76–114). This has particular relevance for cybersecurity policies insofar as they are concerned with resilience. If trust is necessary for effective operation in, and use of, cyberspace and related technologies, then the background beliefs about whether the given technologies and systems are trustworthy will impact upon how and when these systems are used or not. This in turn depends on whether the users see other users as trustworthy. The resilience of a cybersystem – here measured by reference to users continuing to use the given operations, applications or system following an attack – is dependent in part upon how trustworthy the users judge that system to be.

This focus on people and trusting each other tells policymakers that trust in cyberspace involves more than the reliability of a technical system, and more than a strategic calculation of risk versus reward. Trust in cyberspace involves people and the relationships between them. This means that the motivations of those people operating in cyberspace (in some cases deliberately disrupting them) should be considered as important to any discussion about trust, cybersecurity and resilience.

3. The implications of losing trustworthiness

Having given an overview of three different ways to conceptualise trust, we can now begin to analyse specific aspects of the connection between resilience and moralistic trust in cybersecurity. In particular, there are three overlapping layers to consider. The first layer is concerned with one-to-one relations, such as those between coworkers. This is the 'interpersonal layer'. The second layer is concerned with the way that intermediary groups, such as third parties, mediate trust relations between people. This is what we call the 'intermediary group layer'. The final layer is concerned with the special role of governments and other critical institutions that mediate trust relations between citizens and the state. This is what we call the 'policy layer'. The examples below illustrate these different layers of analysis and the importance of moralistic trust relations for the development of resilience in cybersystems.

3.1. Stuxnet

Stuxnet is a highly complex computer virus that was designed to disable Iran's nuclear program by infecting specific centrifuges. Rid describes Stuxnet as functioning in the following way, 'The malicious code caused the machine's circuit breakers to cycle on-and-off in rapid succession, causing permanent damage through vibration … ' (2013, 46). But it was not simply the Siemens centrifuges that were targeted for attack. According to Rid, the intention of Stuxnet was both to disable the centrifuges and also undermine the trust of Iran's nuclear scientists in their own skills to develop and implement a nuclear programme. Rid states,

> [*Stuxnet*] *was intended to undermine trust*, the trust of scientists in their systems *and in themselves*, and the trust of a regime in its ability to succeed in its quest for nuclear weapons. When Stuxnet started successfully damaging the Iranian centrifuges, the Iranian operators did not know what was happening for more than two years […] The rationale was that once a few machines failed, the Iranian engineers would shut down larger groups of machines […] because they distrusted their own technology and would suspect sabotage in all of them. (2013, 32, emphases ours)

Had Stuxnet not been discovered, Rid argues, its capacity to provide operators with false information could have effectively delayed Iran's nuclear programme for many more years (Rid 2013, 33). So in the Stuxnet case, a key target was trust in the nuclear programme itself. The information being communicated to the operators did not match the actual output of the system. As a result, the programme was failing. Such failures undermined the operators' capacity to rely on the system, trust in their own capacities and trust in the overall integrity of the nuclear programme. Moreover, the Stuxnet attack undermined the belief that the workers had the capacity to do their job. The Iranian workers were not intentionally harming the system. But any one operator could not *trust* that their coworkers had the technical and scientific capacity to do their job effectively. Trust relations *between the workers* were undermined, and as a result, so was the resilience of the nuclear programme itself. According to Rid, this was part of the intended outcome.

Consider now the impact of a similarly motivated attack on a nation's critical cyberinfrastructure. Should a key energy provider within a country's energy sector suffer a Stuxnet-like attack, what might be the outcome? Were the operators of the given energy provider to lose trust in critical infrastructure (and the human–technical system as a whole) it is not unreasonable to envision a long-term shutdown in order to overhaul that sector. The resultant costs (in terms of money spent, jobs and other support mechanisms) could be highly detrimental to the nation's economic stability. The basic point here is that cyberweapons (such as computer viruses) can use technical elements of a system to target *human* vulnerabilities in interpersonal relations. This can undermine the sense of trustworthiness and degrade the overall system.

3.2. Ashley Madison hacks

The second layer of analysis is concerned with the ways that cyberattacks on intermediaries or third parties can impact on the sense of trust that users have in other people using such services. These attacks undermine the belief that intermediaries can guarantee that particular conditions existing around the relations between users are maintained. To demonstrate this, we look at the hacks of the Ashley Madison dating website.

The Ashley Madison hacks targeted AshleyMadison.com, the self-proclaimed 'most famous name in infidelity', a website that specialised in setting up dates for people who were either married or in relationships. It was hacked in July 2015 and 10 gigabytes of user information were subsequently posted online. A group called the Impact Team claimed responsibility for the hack, explaining that they wanted to highlight the site's fraudulent business practices and because of their opposition to the immoral practices it encouraged and enabled (Hackett 2015).

In this case, an intermediary group was hacked: the services – particularly confidentiality and anonymity – offered by the Ashley Madison website. While these hacks targeted the technical providers, and undermined faith in those intermediaries to provide a particular service, the impacts were intended to extend to the human–human relations of users of the service. The intermediary guaranteed anonymity. The technical and strategic accounts of trust can explain failures at the technical level, and the risk versus reward of using given services. We miss an important aspect of resilience however, if we overlook Uslaner's moral concept of trust; the hacks caused a loss in trust between humans – a loss of belief in honest dealings between people, and a betrayal by one's partner.

By recognising this intermediary layer of trust, we can start to identify different trust relations that occur with one set of behaviours. For the Ashley Madison hacks, we see a range of trust relations impacted – the users could not trust the service provider to assure anonymity. Furthermore, making the lists of users public was intended to impact on the trust between partners and spouses. The point to draw out here is that there are at least two different types of trust relations impacted by these hacks.

In describing the role of intermediaries in human–human trust, Herwitz writes that

> the mechanism for establishing trust upon which online encryption is based requires users to ask themselves a more complicated question: to be effective, Internet browsers must ask, 'Do we trust the person vouching for party A?' instead of, 'Do we trust party A?' … . (2013, 1604)

What we want to suggest is that the moralistic trust concept flips this to mean that the question, 'Do we trust the person vouching for party A?' has a direct impact on 'Do we trust party A?' and failures of trust at the technical level will impact upon the sense of betrayal of trust at the personal level. That is, we need to recognise that different sorts of trust relations occur above and beyond the technical and strategic.

3.3. The Snowden leaks

The final layer of analysis concerns the trust relations between the state and its citizens, considered here as the 'policy layer'. The Snowden leaks have allowed previously hidden intelligence collection practices to be revealed in public, which has provoked widespread alarm, condemnation and embarrassment (Greenwald 2014). The cost here is in the public's willingness and capacity to trust in state mechanisms of national and international security, especially those intelligence programmes involving cyber-surveillance. The loss of trust in this area is particularly detrimental to the effective functioning of national security. This is because the state relies on a twofold presumption of trust. Firstly, that intelligence surveillance is necessary for the state to protect its inhabitants (whether it is national security, military or law enforcement). Secondly, that the state must keep the majority of its intelligence activities secret from the public (Lester 2015, 5–7). The public trusts that secret government surveillance, which sometimes infringes privacy rights, is necessary to protect them. But governments that are caught out abusing this trust, risk losing the goodwill of the public. If the public's trust in this area is lost and the government is judged as untrustworthy, then effective governing can become more difficult.

The revealing of secret surveillance programs (such as PRISM and xKeyscore) by Edward Snowden has led to a variety of concerns being aired (Greenwald 2014). Our focus here is not on the legal or moral permissibility of such state surveillance programs, rather it is the consequences this has for public trust. Large sectors of the global public have responded angrily to such surveillance programs (Harding 2014). Some argue that this proves governments cannot be trusted with secret surveillance (whether their own or others) and that the Internet is not trustworthy because any Internet activity might be monitored (Greenwald 2014, 170–209). The problem here is that short-term impacts on trust might reduce people's use and alter, or 'chill', the content of their online communications and activities (Greenwald 2014, 173), greatly reducing its benefits. This comes back to the statement from the Australian Government's cyber strategy policy that we opened with: 'If poor cybersecurity erodes trust and confidence in cyberspace, the economic opportunity of a connected Australian economy will suffer' (Commonwealth of Australia 2016). Once lost, the public's trust is very difficult to get back.

Loss of public trust is ultimately counterproductive for effective national security. One key area of national security policymaking that can be impacted in this way is surveillance. Necessary and justified surveillance might be hampered to the extent that a preventable tragedy is not stopped. Consider here a comparison to international cooperation on pandemic preparedness. Mass public surveillance is a central plank of any pandemic preparedness programme (Barnett and Sorenson 2011). Should the problematic use of surveillance cause a loss of trust that spills over into other state surveillance programs, these essential public health programmes could be negatively impacted. For instance, a polio vaccination programme was used to help gain intelligence on the occupants in Bin Laden's compound, leading to the targeted killing of Osama Bin Laden. As a result of this becoming public knowledge, citizen participation in the polio vaccine programmes in Pakistan dropped (Anonymous 2013; McNeil Jnr. 2012). Some hold that this has needlessly exposed many children to the virus and ensured that the disease persists in the population for years to come, potentially delaying 'polio eradication for 20 years, leading to 100,000 more cases that might otherwise not have occurred' (Anonymous 2013). The point to draw from this is that loss of moralistic trust – the notion that governments are trustworthy – from poorly thought out cyber-surveillance could have similarly broad and dispersed impacts.

In sum, the three layers of analysis here illustrate the significant consequences that a loss of trust has on the way people interact with each other, their service providers and governments. It can lead to decreased trust in coworkers' competence and honesty (which degrades workplace effectiveness); intermediary groups (which obstructs the provision of corporate services) and agencies of government (which undermines good governance).

4. Designing trustworthy systems

Trust is an essential element of any sustained discussion of cybersecurity. In order for people to go about their business online, and for people to feel secure in markets and governments at large, they must have some minimum level of trust in the ICT systems, other public systems, system operators and oversight. One way to develop and maintain trust is to design trust into ICT and governance systems; what we term 'trustworthy systems'. This concept of designing values into systems is emerging in literature on technology and design (Friedman, Kahn, and Borning 2002; van den Hoven 2007). Trust is not a novel value to be incorporated into the design of ICT – trusted systems are commonly discussed in ICT (Cho, Chan, and Adali 2015; Granatyr et al. 2015; Khan 2016; Wang and Yu 2015; Yan et al. 2016). These approaches, however, have typically focused on narrow aspects of trust. That is, '[t]rust management systems must offer certain guarantees to secure information, as well as processes that create, manage distribute, and govern information and services, in a reliable and efficient manner' (De Paoli et al. 2010). In contrast, the approach we advocate here incorporates a broader account of trust than merely ensuring reliability of information and communications. We are advocating the view that moral trust, and its role in developing resilient systems, should be considered a priority for any national government cybersecurity policy. But what does this mean in practice?

4.1. Acknowledge the reality of vulnerability

Policymakers should not make the mistake of assuring the public that their cybersystems are impenetrable. No cyber-based system can be made absolutely secure – whether it relies on security methods, such as constant software patches, or so-called 'security-through-obscurity' (Rid 2013, 72–74). Cybersystems are always vulnerable to intrusion and attack. Furthermore, these are human–technical systems – humans are arguably the key point of vulnerability for attack, and no system can ever effectively anticipate all potential human behaviour. As events such as the loss of top secret government information (Anonymous 2009) and the Snowden leaks illustrate, a forgetful or disgruntled employee can make the technical security features of a cybersystem much less effective. Rather than aiming at absolute protection, systems should also be designed for resilience. This is why cybersecurity policymaking should be attentive to the role of moralistic trust: we suggest that it plays a central role in the resilience of cybersystems.

Given the importance of the human–human element in cybersecurity, we need to look beyond trust as technical or strategic to include the human element. We propose a method for designing trustworthy systems by identifying the trust relations between users of the given system in order to anticipate the likely impact to the network as a whole if a given trust relation is to fail.[2] This process begins with the identification of key actors in a given network. The trust relations within the network are categorised by reference to the expectation of the role the trusted agent is expected to fulfil. Then a 'taxonomy of impact' can be produced; what would happen should that trust relation be disrupted and that actor or actors leave the system? This approach operationalises trust in the given system in order to increase the load bearing capacity to trust of important relations and to better prepare the systems for cyberattacks into the future.

Critical systems would at least require operators and supervisors, or others on site, to respond quickly to aberrant results in order to identify if an attack has occurred, and that the human elements can be trusted. As the Stuxnet incident demonstrates, it is risky to frame cyber incidents as purely technical affairs and ignore vulnerabilities in 'social engineering'. Most cybersecurity breaches involve human failure in some way (Hutchinson 2011). The systems that underpin critical infrastructure involve humans in key roles. This human element is not only important in creating system-wide vulnerabilities to intrusion, but also in designing system resilience. As operators, users and those reliant upon critical infrastructure, we need to be able to trust that those systems are resilient to cyberattack. As part of this resilience, we need to be able to trust that, should a cyber incident occur, our systems have the capacity to respond to such cyber incidents in a way that reduces impact and inoculates similar systems with similar vulnerabilities against intrusion, thereby maintaining the overall belief that the system is worth using.

4.2. Map trust relations

Policymakers should devote significant resources to the task of mapping trust relations. This is likely to be expensive and time-consuming. How do we map the trust relations in an event like the Ashley Madison hack described above, where an intermediary service provider is hacked? The first step is to identify the key actor types in such an event, such as employees, intermediaries and management/employers. Identifying specific actors would occur in planning for (and responding to) specific incidents.

The second step is then to categorise how a trust relation between each actor type operates under normal conditions. This says that the 'trustor' trusts that the 'trustee' will act in an expected way. For example, in a workplace scenario, an employee (trustor) *trusts that* the ICT network that facilitates communications (trustee) will (1) protect the employee's account from an attack; (2) know when accounts have been attacked and (3) have some way of alerting the proper account holder (employee) of the attack. Furthermore, other stakeholders (such as employers) then trust that system's weaknesses are recognised and fixed. Finally, all stakeholders trust that the relevant security agencies would rapidly identify the source of an attack and respond within the law. For their part, the relevant security agency trusts that the ICT network provides accurate information for the purposes of an effective investigation. The point here is that, like the relations between coworkers, a service user and intermediary, and citizen and government, the human–human relations involve trust relations, and should the system be resilient to attack, those different human–human relations, across the different layers, need to be recognised.

The third step is to maintain the ability to evaluate the taxonomy of impact. This is the capacity to judge (in advance) the extent of damage caused by the loss of trust relations between actors if disrupted or broken. The loss of trust between coworkers in a nuclear power plant is likely to have different impacts to the loss of trust between a citizen and their government.

4.3. Develop and protect a reputation for trustworthiness

Developing and protecting a reputation for trustworthiness should be fundamentally important for policymakers. That is, national government policymaking should emphasise the design of trustworthy ICT and governance systems.

This methodology of characterising trust relations already exists, for instance,

> reputational models are among the most successful responses to concerns about trust online. One of the early pioneers of these models was eBay, which relied on parties to leave publicly viewable feedback about their interactions on its auction website. Users could rely on buyers and sellers with established histories of satisfactory dealings as more trustworthy than those without established histories (or, worse, with less-than-satisfactory histories) […] Similarly, parties that want to interact online but do not trust their data to be handled by untrusted intermediaries have long turned to encryption to protect against those intermediaries. To use encryption, the parties need to first coordinate the use of an encryption algorithm. This is done by relying on a trusted intermediary, called a certificate authority […] Both of these mechanisms (indeed, all mechanisms that rely on actors within a system to establish trust) are built upon the fundamental assumption that parties being relied upon to establish trust are independent from the party for whom trust is being established. (Herwitz 2013, 1603)

What our analysis shows is that there are in fact two different sorts of trust relations being identified here – one of reputation and one of confidentiality. Yet, in order for certain human–human relations to be trustworthy, perhaps confidentiality is not what should be prioritised. The Ashley Madison service, for example, was targeted *because* it allowed for other human–human relations to be exploited. That is, if confidentiality affords betraying a spouse or loved one, then this favours one set of trust relations (intermediary) over another (human–human). And if we are to have resilient systems, we need to know which of those trust relations need to survive cyberattacks.

4.4. Promote oversight and good governance

A priority for effective national security policymaking are *effective* oversight mechanisms (Lester 2015). These are necessary to build public trust in the government agencies responsible for collecting sensitive information. Some use of cyberspace for surveillance is justified. Consider the use of Internet surveillance as part of investigations into child pornography rings. Although such investigations typically include traditional policing measures, most people would agree that infringement of some level of privacy is justified if it is necessary to investigate and prosecute members of child pornography rings. The danger, however, is that surveillance is expanded beyond what is justified. National security is of vital importance, but the Snowden revelations have undermined existing trust relations between the public at large and the activities of security agencies.

The U.S. government's response was to refer to FISA (Foreign Intelligence Surveillance) court oversight of the PRISM and xKeyscore programs. They claimed that the FISA system provided a sufficient amount of oversight (Clapper 2013). However, given the U.S. government's earlier statements that they were not involved in large-scale surveillance programmes, while at the same time that PRISM and xKeyscore were operating (Kaminski 2013), this places a great burden on the trust relation between the public at large and the U.S. government. Revelations that the U.S. had been spying on allies such as Angela Merkel (Greenwald 2014, 141) and on economic targets such as Petrobras (134–135) undermine the claims that bulk surveillance is being carried out to protect against terrorists. This is not to say that state surveillance is never justified. Rather, the concern is that misrepresenting what is being done and why undermines public trust in government. As a result of the public outcry, governments might cut necessary surveillance programmes, or we might see harmful leaks from disgruntled employees with access to sensitive information. As discussed with the Bin Laden/polio vaccine example, such loss of trust in one arena of government and security can have potentially dire impacts on other areas.

4.5. Develop resilient systems

Intelligence and national security services should aim to be trustworthy so the public are right to trust that the government can manage the sensitivities of national security surveillance. And they should be largely independent from interference for short-term political gain. This is the best way to ensure that government programmes rebuild trust over time and are seen as trustworthy enough to manage the sensitivities of national security. This means that there should be organisational guarantees that the relevant oversight and governance mechanisms are independent from short-term political gain. National government policymaking should therefore seek to promote accountability through the development of functional national security oversight mechanisms. And national government policymaking should seek to build public trust over time.

We can start with the notion of tailored trustworthy spaces (TTS) where the

> user is informed of the levels of trust available and chooses to accept the protections and risks of a particular tailored space. The attributes of each available trusted space must be expressible in an understandable way to support informed choice. The attributes must be made manifest and readily usable to support being customized, negotiated, adapted, and enforced. All parties to the transaction must agree on the level of trust enforced by the underlying infrastructure. (National Science and Technology Council 2011, 7)

What we suggest is a step up from this, where we ask, 'What is that the end-to-end users are being trusted to do?' In order to develop a TTS that is *resilient* to malicious attack, the human–human expectations of trust must be recognised and understood, because if you lose one trust relation between people, then resilience requires a response that repairs that particular relation (rather than a different trust relation). This requires mapping of the trust relations between key actors in the given system. Secondly, a system may have to trade in different trust types to develop a resilient system. For instance, a system designer may have to choose between anonymity (resilient because individual identities are protected) and reputation (resilient because those who display bad character are punished).

For instance, the U.S. National Strategy for Trusted Identities in Cyberspace favours identification over anonymity, preferring one element of trustworthy relations over another: 'Identity solutions will be privacy-enhancing and voluntary, Identity solutions will be secure and resilient, Identity solutions will be interoperable, Identity solutions will be cost-effective and easy to use' (White House 2011). The point here is to recognise the role played by specific trust relations in system resilience. Should trusted identities not be valued by the users, or be lost due to malicious actors, repairing a trust-resilient system is going to be different to establishing a set of trust relations that expect, and respond to, anonymity.

5. Conclusion

Cybersecurity policymaking at the national level should be attentive to the central role of trust for maintaining resilient cybersystems – and it should have plans in place to repair trust. But trust in cyberspace is not simply about the reliability of a technical system; trust in cyberspace includes people. A cyberattack can target *human* vulnerabilities by undermining trust. The relatively low technological know-how needed for some types of attacks on trust is sufficient to make it a policy priority for any national government. If it is true that trust can be subject to cyberthreats, then it follows that trust should be designed into cybersystems. Furthermore, loss of trust from poorly thought out cybersurveillance could also have impacts on issues relating to trust. So governments should aim to ensure that national security oversight mechanisms are functional and largely independent from interference for short-term political gain. Good governance requires that intelligence and national security services protect a reputation for trustworthiness.

Notes

1. See, for example, Cho, Chan, and Adali (2015), Granatyr et al. (2015), Khan (2016), Wang and Yu (2015) and Yan et al. (2016).
2. This builds from Clarke and Blumenthal's trust-to-trust, where they suggest we ought to

   > look at the range of options that each participant in the communication can take, based on their individual choices about trust, and then we look at the range of options that arise jointly, depending on the degree to which the various communicants trust each other. Trust-to-trust acknowledges that … there is more reason for one end to question the trustworthiness of another and therefore more reason to seek something beyond simple end-to-end communication. (Clark and Blumenthal 2011, 370)

   Our approach seeks to recognise the different human relations that are enabled by, and dependent upon, cybertechnologies, such that the value of moral trust between people is designed into the relevant systems.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Dr ***Adam Henschke*** is a lecturer with the National Security College at the Australian National University and secretary of the Asia Pacific Chapter of the International Society for Military Ethics. He received his Ph.D. from the Centre for Applied Philosophy and Public Ethics at Charles Sturt University. He was a visiting assistant professor in the Department of Politics and Public Administration, University of Hong Kong, and has been a visiting researcher at the Hastings Center in New York, U.S.A, the Brocher Foundation in Geneva, Switzerland and Delft University of Technology, Delft, The Netherlands.

Shannon Brandt Ford is President of the Asia Pacific Chapter of the International Society for Military Ethics. He was previously a Research Fellow at the Centre for Applied Philosophy and Public Ethics, Charles Sturt University, where he led a research project on the ethics of cybersecurity. Before that, Shannon spent 10 years as a Defence Strategist and Analyst with the Australian Department of Defence. He is completing his doctorate at the National Security College with the Australian National University. Shannon was recently awarded a contract with the Australian Army Research Scheme to write a report that examines 'Emerging Weapons Technologies: Political, Ethical and Legal Dilemmas'.

References

Anonymous. 2009. "Previous Cases of Missing Data." *BBC News*. http://news.bbc.co.uk/2/hi/uk/7449927.stm.

Anonymous. 2013. "How the CIA's Fake Vaccination Campaign Endangers Us All." *Scientific American*, May 1. http://www.scientificamerican.com/article/how-cia-fake-vaccination-campaign-endangers-us-all/

Barnett, Tony, and Corinna Sorenson. 2011. "Infectious Disease Surveillance in the United States and the United Kingdom: From Public Goods to the Challenges of New Technologies." *Journal of Health Politics, Policy and Law* 36 (1): 165–185. doi:10.1215/03616878-1191144.

Cho, Jin-Hee, Kevin Chan, and Sibel Adali. 2015. "A Survey on Trust Modeling." *ACM Computer Survey* 48 (2): 1–40. doi:10.1145/2815595.

Clapper, James. 2013. *DNI Statement on Activities Authorized under Section 702 of FISA*. Office of The Director of National Intelligence.

Clark, David D., and Marjory S. Blumenthal. 2011. "The End-To-End Argument and Application Design: The Role of Trust." *Federal Communications Law Journal* 63 (2): 357–390.

Commonwealth of Australia, Department of the Prime Minister and Cabinet. 2016. *Australia's Cyber Security Strategy*. Accessed 13 October 2016. https://cybersecuritystrategy.dpmc.gov.au

De Paoli, Stefano, G. R. Gangadharan, Aphra Kerr, and Vincenzo D'Andrea. 2010. "Toward Trust as Result: An Interdisciplinary Approach." *Proceedings of ALPIS* 10 (8).

Friedman, Batya, Peter H. Kahn, and Alan Borning. 2002. *Value Sensitive Design: Theory and Methods*. University of Washington Computer Science and Engineering Technical Report 02-12-01. Washington, DC: University of Washington.

Friedman, Batya, Peter H. Kahn, Alan Borning, and Alina Huldtgren. 2013. "Value Sensitive Design and Information Systems." In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by Neelke Doorn, Daan Schuurbiers, Ibo van de Poel, and E. Michael Gorman, 55–95. Dordrecht: Springer Netherlands.

Granatyr, Jones, Vanderson Botelho, Otto Robert Lessing, Edson Emlio Scalabrin, Jean-Paul Barthes, and Fabricio Enembreck. 2015. "Trust and Reputation Models for Multiagent Systems." *ACM Computing Surveys* 48 (2): 1–42. doi:10.1145/2816826.

Greenwald, Glen. 2014. *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State*. New York: Metropolitan Books.

Hackett, Robert. 2015. "What to know about the Ashley Madison Hack." *Fortune*. Accessed 17 June 2016. http://fortune.com/2015/08/26/ashley-madison-hack/.

Harding, Luke. 2014. *The Snowden Files: The Inside Story of the World's Most Wanted Man*. New York: Vintage Books.

Herwitz, Justin (Gus). 2013. "Trust and Online Interaction." *University of Pennsylvania Law Review* 161 (6): 1580–1622.

van den Hoven, Jeroen. 2007. "ICT and Value Sensitive Design." In *The Information Society: Innovation, Legitimacy, Ethics and Democracy in Honor of Professor Jacques Berleur s.j.*, edited by Philippe Goujon, Sylvian Lavelle, Penny Duquenoy, Kai Kimppa, and Veronique Laurent, 67–72. Boston, MA: Springer.

Hutchinson, James. 2011. "Social Engineering Remains Biggest Cyber Threat." *Computer World*. Accessed 15 April 2016. http://www.computerworld.com.au/article/380867/social_engineering_remains_biggest_cyber_threat/.

Kaminski, Margot. 2013. "PRISM's Legal Basis: How We Got Here, and What We Can Do to Get Back." *The Atlantic*. Accessed 15 April 2015. http://www.theatlantic.com/national/archive/2013/06/prisms-legal-basis-how-we-got-here-and-what-we-can-do-to-get-back/276667/.

Khan, Minhaj Ahmad. 2016. "A Survey of Security Issues for Cloud Computing." *Journal of Network and Computer Applications* 71: 11–29. doi:10.1016/j.jnca.2016.05.010.

Lester, Genevieve. 2015. *When Should State Secrets Stay Secret? Accountability, Democratic Governance, and Intelligence*. Cambridge: Cambridge University Press.

McLeod, Carolyn. 2015. "Trust." Stanford Encyclopedia of Philosophy. Accessed 15 April 2016. http://plato.stanford.edu/entries/trust/.

McNeil Jr., Donald G. 2012. "C.I.A. Vaccine Ruse May Have Harmed the War on Polio." *New York Times*, July 9. http://www.nytimes.com/2012/07/10/health/cia-vaccine-ruse-in-pakistan-may-have-harmed-polio-fight.html?pagewanted=all&_r=0.

National Science and Technology Council. 2011. *Trustworthy Cyberspace: Strategic Plan for the Federal Cybersecurity Research and Development Program*. Executive Office of the President.

Office of the Press Secretary. 2013. *Presidential Policy Directive – "Critical Infrastructure Security and Resilience*." The White House.

Rid, Thomas. 2013. *Cyber War Will Not Take Place*. London: Hurst.

Singer, Peter. 2014. *Peter Singer on Cybersecurity and Cyberwar, Part III*. Edited by Sam Roggeveen. The Interpreter: The Lowy Institute, February 6. http://www.lowyinterpreter.org/.

Uslaner, Peter. 2002. *The Moral Foundations of Trust*. Cambridge: Cambridge University Press.

Wang, Zhong, and Qian Yu. 2015. "Privacy Trust Crisis of Personal Data in China in the Era of Big Data: The Survey and Countermeasures." *Computer Law & Security Review* 31 (6): 782–792. doi:10.1016/j.clsr.2015.08.006.

White House. 2011. "National Strategy for Trusted Identities in Cyberspace." Accessed 20 July 2016. http://www.nist.gov/nstic/

Yan, Zheng, Wenxiu Ding, Valtteri Niemi, and Athanasios V. Vasilakos. 2016. "Two Schemes of Privacy-Preserving Trust Evaluation." *Future Generation Computer Systems* 62: 175–189. doi:10.1016/j.future.2015.11.006.