Cyber Security Norms: Trust and Cooperation

Allison Wylde
Cardiff University, UK
wyldea@cardiff.ac.uk

Abstract: As cyber crime becomes ever more sophisticated and a significant asymmetric threat, the need for effective cyber security is of vital importance. One important cyber security response is through cyber norms. At the same time, calls for multi-sector and multi-domain trust and cooperation are widespread. Yet research on the nature of trust and cooperation in cyber security norms appears to be underdeveloped. Key questions remain concerning the emergence and nature of trust and cooperation in norms. In addressing this gap, the article first considers how we can understand trust and cooperation in cyber norms through leveraging well-established theory from management research on trust building. Next, the paper examines the SolarWinds breach, as an example, to evaluate norms, trust and cooperation. The paper then applies principles from prominent trust-building theory to examine the antecedents, processes of outputs involved in building trust and cooperation. The contribution of this work presents a foundational conceptual framework, to allow the dynamics of norms, trust, and cooperation in managing cyber crime incidents to be studied. In doing so, the literature on examining trust and cooperation in norms is extended. Other researchers' interest is encouraged as is an agenda for further research on norms, trust, and cooperation to support cyber security management. Implications may help the cyber security community as they construct and manage norms, trust, and cooperation.

Keywords: cyber security, norms, trust, cooperation, dynamics

1. Introduction

The SolarWinds breach of 2021 highlights an instance of an ongoing and large-scale extraction of sensitive material from carefully targeted organizations across government departments, financial institutions and public health and education institutions. The full impact may never be known due to the sensitivities involved. Government responses were rapid. Although the implementation of norms in cyber security is assumed, in practice, explicit understanding of the central and underpinning elements of trust and cooperation is limited. The puzzle at the heart of this paper concerns an important question; if norms for cyber security rely on implicit trust and cooperation, how is this understood by the various international actors involved? Indeed, is there agreement on a shared understanding? The SolarWinds breach is dawn on to provide an example through which to explore trust and cooperation. The author explored the SolarWinds breach during as part of the UN IGF workstream 2. Several other breaches were also examined along with policy (UN IGF BPF, 2021a; b).

The development of norms has progressed. As a start, the United Nations Government experts 2015 (UN, 2015) and later the Global Commission on Cyberstability (GCSC, 2019) set out important norms for cybersecurity. However, arguably, the continued proliferation of norms without an underpinning framework for evaluation poses a challenge for a common understanding.

To answer the questions raised, this paper next revisits the central literature on norms associated with cyber security to explore the roles of trust and cooperation. This is followed by a review of trust building theory drawn from the management literature. The Integrative Trust Model (IT) (Mayer et al. 1995), together with foundational research in conflict management Deutsch (1958; 2006) are selected due their ability to disentangle the various elements and processes involved in trust building. Given that trust is central in cooperation, these models are reexamined with a focus on extracting the key elements that can shed light on processes of cooperation. As the central terms trust and cooperation are multiplex, definitions in the context of this paper are next presented. The study is operationalised through an exploration of the SolarWinds breach with a focus on the norms as implemented together with the underlying trust and cooperation. In the final section, the conclusion, the work as executed is discussed together with limitations, future directions, and implications.

2. Key literature

2.1 Norms for cyber security

Although norms are widespread and generally understood to relate to shared beliefs and actions founded on what is correct or proper, some confusion remains. To add clarity, what follows is a brief overview of norms theory and a focus on the development of norms for cyber security.
Social scientists view social norms as constructs of three main types, descriptive norms, simply representing what other people do and injunctive norms, what people should do (Smith and Lewis, 2008). A third category is subjective norms. These norms concern an individual’s perception of another actor’s approval (or not) of their own actions and the motivating factors that are involved (Cialdini et al, 1991). In subjective norms, an individual’s perception of approval (from a referent) may prompt that individual to behave in a way that may be encouraged by the referent (Cialdini et al, 1991). This thinking, extended in the Technology Acceptance Model (TAM) (Ajzen, 1991), suggests that individuals may be motivated to copy or mimic the behaviour of a referent group out of a belief that in doing so their own status will be enhanced (Venkatesh and Davis, 2000). This is often seen in the release of new technology - people want the latest gadget, ‘everyone has it’ and individuals may feel better, and indeed enhanced, once they possess the item. Indeed, a similar way, some individuals may act as enforcers and implement sanctions if a particular norm is not followed (Ellickson, 1999) while other studies have found that the negative action may be performed if an individual believes a referent group approves of those actions, for example, illegally downloading streaming services (Wired, 2021).

Summing up, norms are viewed as beliefs and motivators - in some cases the right thing. Norms are dynamic. Change may arise through change agents, from self-motivated leaders to norm entrepreneurs and finally, opinion makers (Ellickson, 1999). In provide a working definition, this paper views norms as based on the key considerations; a norm itself is agreed, norms act as informal rules (Smith and Lewis, 2008), norms act as drivers and can prompt referents to behave in a particular way (Ajzen, 1991; Cialdini et al, 1991), and importantly, that norms are subject to change by change agents (Ellickson, 1999).

For cyber security, norms are seen as agreed methods and shared beliefs of how to behave and operate in the cyber domain and referents are motivated to follow the guidelines (Smith and Lewis, 2008; Ajzen, 1991; Cialdini et al, 1991). Active norm creation has occurred through the establishment of agencies such as UN the Internet Governance Forum (UN IGF) (UN, 2015) which aimed to help support the development and practices of norms for cyber security. Numerous individual states and organizations developed their own norms along with, in 2015 the UN’s Framework for Responsible State Behaviour (UN, 2015; GCSC, 2019). The foundational cyber security norms include (1) the non-interference of the public core of the internet (2) the protection of electoral services (3) the avoidance of tampering (4) agreement not to commandeer ICT devices into botnets and (6) a reduction and mitigation of significant vulnerabilities (GCSC, 2019). In 2021 the UN IGF BPF undertook research to identify and map the frequency of policy actions for each norm, the highest frequency was cooperation, adherence to human rights, reporting vulnerabilities and providing remedies (UN IGF BPF, 2021a; 2021b).

As highlighted norms are dynamic, norms also have arisen organically as best practice among practitioners, notably the adoption of zero trust approaches among cyber security practitioners (Wylde, 2021). A driver to establish a norm may arise internally through practitioners’ actions as norms entrepreneurs (Ellickson, 1999). As a gap is identified practitioners cooperate to provide a solution, as in the example of zero trust. Alternatively, policy makers may implement a new norm (NCSC, 2021a). This gives rise to questions concerning underpinning assumptions in cyber security norms. Are all norms trusted? Is trust and cooperation necessary for the formation and subsequent operation of norms, are all norms based on trust and cooperation, which would in turn promote trust, so forming a positive feedback loop? Figure 1, below, presents a first conception of norms formation and the role of trust and cooperation. The question here, concerns how to understand the processes trust and cooperation involved in norms.

### 2.2 Norms: Trust and cooperation

Although recent findings point to the underpinning and implicit role of trust and cooperation in norms (UN IGF BPF, 2021a) there remains important gaps concerning what this may mean in practice. In the extensive literatures on both trust and cooperation, questions concerning definitions and dimensions remain. Important studies are considered next with a focus on teasing out and drawing together the key strands that will form the conceptual basis for this paper.

#### 2.2.1 Trust

In an extensive literature, trust is viewed as related to uncertainty and risk (Mayer et al, 1995) in an interactive process (Dietz, 2011), as an enabler of cooperation and an alternative to formal governance (Vanneste, 2016) and indeed control (Mayer et al, 1995). The definition, context and conceptual model for this paper, cyber security, is discussed next.
A definition for trust as viewed in this paper, is provided by the integrative trust building model of Mayer Davis and Schoorman (1995) and the foundational research by Deutsch (1958; 2006). These authors define trust as “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other party (Mayer et al, 1995, p. 712). The element of willingness to be vulnerable is further clarified as “a psychological state comprising the intention to accept vulnerability based on positive expectations of the intentions or behaviour of another” (Rousseau et al, 1998, p. 395). Drawing these themes together, a psychological state and intention (Rousseau et al, 1998) and willingness to accept vulnerability; irrespective of the ability to monitor or control (Mayer et al, 1995). From the conflict resolution literature from which norms have arguably evolved (Wylde, 2021) trust is viewed as founded on an individual's ability to trust along with their experience of trust (Deutsch, 1958), cooperation or a lack of cooperation (Deutsch, 2006). Important also are considerations of norms in society and alignment with the individual trustor’s beliefs (Deutsch, 1958; 2006). This approach allows researchers to evaluate key antecedents, processes, and outcomes in these multi-dimensional and understudied concepts.

Important research has also argued for clarity regarding the scale and referent in trust relations. Fulmer and Gelfand's (2012) work set out the different levels of trust relations, trust in individuals or teams or organizations and at the level of institutions. Non-person-based trust has also been studied, including the examining nature of trust in technology (Mckight, 2011). For this paper, the scope of the trust relations examined are confined to trust in institutions and in policy. Importantly as Vanneste (2016) clarifies, it is the people in the organizations who trust and prompt others to trust.

For the context of cyber security, important beliefs about trust are seen as based on confident positive expectations of the other’s trustworthiness based on an assessment of ability, benevolence and integrity moderated by an antecedent actors’ propensity to trust (Mayer et al, 1995) and psychological intention and willingness to accept vulnerability; irrespective of the ability to monitor or control (Mayer et al, 1995; Rousseau et al, 1998). Together with a trustor’s experiences and beliefs (Deutsch, 1958; 2006).

Yet, this foundational view of trust may be at odds with current practices in cyber security, which rely on zero trust, in other words, non-presumptive trust (Wylde, 2021). In a similar sense as trust, in zero trust, the central thinking is founded on a psychological state based on experience and societal norms. Importantly in implementing zero trust vulnerability or risk are not accepted, rather, there is continuous monitoring, assessment, and authentication (NCSC, 2021a). In Table 1, below, the dynamics of trust and cooperation building: antecedents, processes, and goals (DT&CB) model, sets out key elements from the trust and collaboration models. Cooperation discussed next, together with trust form the basis of the conceptual frame in this paper, presented in Table 1 below.

2.2.2 Cooperation

Cooperation has been studied across different disciplines, in particular management and public administration. Like trust it is subject to numerous competing definitions. An attempt to add clarity is provided next.

Clear conceptual understanding is essential to allow the distinct elements involved in cooperation and collaboration and coordination to be discerned. Currently the terms are entangled (Castañer and Oliveira, 2020). Indeed, cooperation and collaboration are used interchangeably in many studies of policy documents on norms, (UN IGF BPF, 2021a). As Dietz (2011) highlights, even trust and cooperation may be conflated. Common conceptualisations on relationships based on cooperation include perceived risk, risk taking and that at least two-parties are involved (Dietz, 2011). This position is further confused as some researchers see cooperation as an umbrella term, encompassing both collaboration and coordination (Gazley, 2017). This is based on a relationship involving mutual goals, and the management of activity to achieve jointly agreed outcomes (Gazley, 2017) while others suggest that coordination is a deliberate process among partners based on order to achieve goals (Gulati et al, 2012) This thinking is expanded by public administration researchers, who see collaboration as involving cooperation and comprising discrete dimensions, an event horizon comprising, antecedents, processes, and outcomes (Thompson, 2006).

Coordination and cooperation are also considered in strategic alliances, with the authors’ arguing that cooperation suggests the pursuit of private goals at the expense of the collective (Kretschmer and Vanneste, 2017)
2017). Given space constraints, this discussion is not taken further here, what is important to note is the importance of disentangling these terms, as indicated, for this paper, the emphasis is on cooperation.

For clarity, in this paper, further discussion is limited to cooperation as defined through the meta study undertaken by (Castañer and Oliveira, 2020) which suggests that individuals’ voluntarily helping others (Ring and Van de Van, 1994). The authors point to the goal as the unit of analysis, in the context of interorganizational relations (IOR). This is whether a common goal or a private goal. Since the focus of this paper concerns common and/or collective goals, for parsimony, private goals are considered out of scope.

**Table 1.** Dynamics of trust and cooperation building: antecedents, processes and goals, DT&CB model (extending Deutsch, 1958; 2006; Mayer et al, 1995, and, for zero trust, Wylde, 2021)

<table>
<thead>
<tr>
<th>Antecedents</th>
<th>Processes</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beliefs and willingness to act</td>
<td>Trust ‘fit’ with personality (zero trust)</td>
<td>Trust and cooperation (zero trust)</td>
</tr>
<tr>
<td>Ability to trust</td>
<td>Trust mirrors prevailing societal rules and norms (zero trust)</td>
<td>Trust and cooperation (zero trust)</td>
</tr>
<tr>
<td>Experience of trust (zero trust)</td>
<td>Ability</td>
<td>Trust (zero trust)</td>
</tr>
<tr>
<td>Confident positive expectations of trust (negative expectations)</td>
<td>Benevolence</td>
<td>Trust (zero trust)</td>
</tr>
<tr>
<td>Propensity to trust (zero propensity to trust)</td>
<td>Integrity</td>
<td>Trust (zero trust)</td>
</tr>
<tr>
<td>Acceptance of vulnerability (non acceptance)</td>
<td>Trust and cooperation (zero trust)</td>
<td></td>
</tr>
<tr>
<td>Risk taking behaviour (no risk taking)</td>
<td>Trust and cooperation (zero trust)</td>
<td></td>
</tr>
</tbody>
</table>

Examples, the pursuit of common and or collective goals can be seen in the actions of state and inter-state organizations. As an example, the Council of Europe was set up in 2001 as a formal initiative to encourage international cooperation (EU, 2001). This was followed by the formation of the NATO cooperative defense center in Estonia (Schmitt, 2013). Subsequent bodies have followed and created policy for cooperation (UN IGF BPF, 2021a). In sum, cooperation is seen as reliant on common and collective goals.

### 2.2.3 Trust and cooperation

Taken from the discussions above and drawing from Levine’s (2019) view of the integrative social contract theory, together with Donaldson and Dunfee (1994) who suggest that individuals’ follow norms; a conceptual frame is proposed to elaborate on our thinking on norms-building in the context of cyber security.

The underlying assumptions are as follows: organizations follow norms, in this context, norms for cyber security, and in doing so promote trust, trustworthiness and cooperation (Levine, 2019). The mechanisms involved here are based on understanding the processes and elements involved in the formation of trust and, as is suggested here, by implication, in the formation of cooperation. As highlighted, other studies on norms in cyber security are founded on assumed trust and cooperation.

In addressing this gap, Figure 1 below, presents norms viewed as a continuous process. The start point arises from a driver for formation such as an incident or event, this is followed by the emergence of trust and cooperation as a necessary condition to push forward the development of the norm and then its implementation. As the norm is implemented trust and cooperation form, in doing so these processes drive the formation of additional trust and cooperation, seen here as a positive feedback loop. The figure illustrates the role of positive feedback in driving the whole process forward such that new norms are formed. The NTAC+ model, Figure 1, below, forms part of the contribution of this paper through extending our thinking on how we can examine the dynamics of norms, trust, and cooperation (notably, Vanneste, 2016).
Figure 1: Simplified framework: The dynamics of norms, trust and cooperation, and positive feedback (NTAC+).

As highlighted earlier, in this conception, NTAC+ (Figure 1, above), control and monitoring are absent (Mayer et al., 1995). Vanneste supports this view, arguing that trust enables cooperation as an alternative to formal governance (Vanneste, 2016).

3. SolarWinds

During 2020 a breach took place on the cyber security firm SolarWinds with more than 1,800 of their client organizations, healthcare, education, the military and governments and prominent listed companies worldwide, affected. The breach occurred as routine updates were released. A second wave followed, targeting organizations selected in the first wave for further exfiltration of sensitive material (NCSC, 2021b).

Responses by the US Government followed, measures implemented included attribution, financial sanctions, and the expulsion of diplomats (NCSC, 2021b). Some argued the punitive response meant that US President Biden broke the norms of US foreign policy (NCSC, 2021b). As one official explained the hack was ‘beyond the boundaries’ due to the level and severity of the breach (Volz, 2021). Subsequent reactions included the creation of the US agency responsible for cybersecurity and infrastructure security (CISA, 2021).

The breach of SolarWinds is considered to have acted as a driver for the implementation of norms (BPF 2). In this paper the subsequent responses by governments and agencies are examined to evaluate the presence of trust and cooperation in cyber norms (UN IGF BPF, 2021b).

4. Evaluation of norms, trust, and cooperation

In the immediate outcomes from the SolarWinds breach, several governments and agencies implemented punitive measures, created policy and later, established agencies to drive implementation of the norm, zero trust (NCSN, 2021b). The central question of this paper concerns understanding the roles of trust and cooperation in norms. The conceptual model as set out in Table 1 (DT&CB model), above, is drawn on next, to disentangle and explore the separate, conditions, processes and outcomes concerning the lead-up to, and aftermath of the SolarWinds breach.

If we start by examining the antecedent conditions, the first elements in the model concern the ability to trust, and experiences of trust and cooperation (drawn from Deutsch, 1958; 2006). Looking back, prior to the SolarWinds breach, we can see a history of norms for cyber security under development, reliant on trust and cooperation by participants (EU, 2001; Schmitt, 2013; UN IGF 2015; GCSC, 2019). At the same time, norms were in development organically, notably through the actions of individual practitioners, later followed by cyber security organizations, such as FireEye, Microsoft and the NCSC (NCSC, 2021b). These examples are taken to demonstrate the presence of the ability to trust. Agencies, organizations, and practitioners were indeed working together and cooperating (NCSC, 2021b); providing evidence of experience of trust and cooperation.

In engaging in norms development, cooperation can be considered to represent an expectation, and indeed what can be termed, as a positive expectation of trust. Or indeed if we consider zero trust, no presumptive expectation of trust (NCSC, 20211). If we dig deeper, the element propensity to trust, may be viewed as tempered by need and desire (Dietz, 2011). In the case of zero trust, which emerged organically from
practitioners, we see the approach being adopted and refined arguably as an indication of trusted practices as assessed through the next stage the processes (Table 1, the DT&CB model).

In the next phase, the processes implemented in the aftermath are examined. The processes include and are demonstrated by beliefs and a willingness to act. In the SolarWinds breach, beliefs were voiced, attributions made, and reactions were quick (Voltz, 2021). The processes include assessing if the measures or responses fit with the actor and the prevailing norms of society (Deutsch, 1998; 2006).

Trust itself is assessed through the three dimensions of ability, benevolence, and integrity (Mayer et al, 1995). If we consider a norm, say zero trust, the ability dimension asks, is zero trust able to do the job? In addressing a breach, a change to zero trust, though arguably not so simple (NCSC, 2021a). This is followed by the question of benevolence, will applying zero trust bring benefits that fulfill the responsibilities of an agency? Lastly the question of integrity, can the referent be sure that the provider is working with the best interests of the referent in mind? What example can demonstrate this?

Next, what evidence can be identified regarding accepting vulnerability and taking a risk? Actions in the response to an incident occur in a domain of the unknown. Respondents look for solutions that may involve accepting vulnerability and or risk taking (NCSC, 2021a). Arguably issuing a public consultation or encouraging referents to review material published by private organizations demonstrating such an example.

Turning to the final outcomes, these include trust, and cooperation, both are evident throughout the timeframe, from antecedent through processes and beyond into the future. Summing up, the model as enabled an evaluation of the complex and separate elements of the dynamics, and in doing so, extended the literature (notably, Vanneste, 2016).

5. Conclusion

The principle of norms for cyber security are to provide a trusted guide and an agreed set of rules (GCSC, 2019). Yet, as highlighted, the underpinning foundations rely on the understudied role of trust and cooperation. Through leveraging well-established theory from management and conflict resolution the contribution of this paper has addressed important gaps in our understanding to date and identified a solution.

Returning to the early research in conflict resolution in the aftermath of World War II and the use of atomic bombs in Japan literature is timely. This work highlighted the key role of mutual trust and cooperation in overcoming suspicion and creating norms of cooperation (notably, Deutsch, 1958; 2006).

In addressing our limited understanding in this domain, this paper has sought to provide a first foundation to evaluate the role of the central elements of trust and cooperation in norms. The contribution of this paper extends the literature (Vanneste, 2016) through providing a conceptual model that makes explicit the dynamics involved in trust and cooperation building, the DT&CB model (Table 1). The NTAC+ framework (Figure 1) allows an understanding of the dynamics of norms, trust, and cooperation. Application of the DT&CB model together with the underpinning conceptual framework (NTAC+) has proved capable in disentangling the multifaceted elements involved in the SolarWinds breach. The evaluation has highlighted the various actions by organizations both public and private in developing norms and in driving their implementation. The development of zero trust as a norm for cyber security was also discussed (NCSC, 2021a), in the context of its emergence as an organic norm (Wylde, 2021).

In this paper discussions are inevitably limited due to space constraints. It is important to add that further work is necessary to allow elaboration and to address any shortcomings. Several promising avenues for further research arise, notably the opportunity for empirical studies to examine in detail, the specific contexts of trust and cooperation activities as norms are practiced. For example, organizational structure, strategy or culture. Further work could usefully explore the application of machine learning to assist analysis through automation. Although these conditions could be incorporated here, their development is beyond the scope of this paper.

The significance of the contribution of this paper is reflected in the recent Vienna talks (Joint Comprehensive Plan of Action) (TASS, 22). Speaking about how “joint statement would enhance mutual trust”, one foreign minister stated that “replacing competition among the great powers would help with cooperation...and the
building of major-country relations” (TASS, 2022). Finally, it is hoped that this work will help provide colleagues, policy makers and practitioners with a starting point to help disentangle the important constructs of trust and cooperation in norms for cyber security

References


TASS (2022) “Nuclear power’s arms control efforts to boost mutual trust - China’s foreign ministry”, [online], TASS, https://tass.com/world/1383711


Wired. (2021) “Netflix’s password sharing crackdown has a silver lining”, [online], WIRED, https://www.wired.com/story/netflix-password-sharing-crackdown