sipri

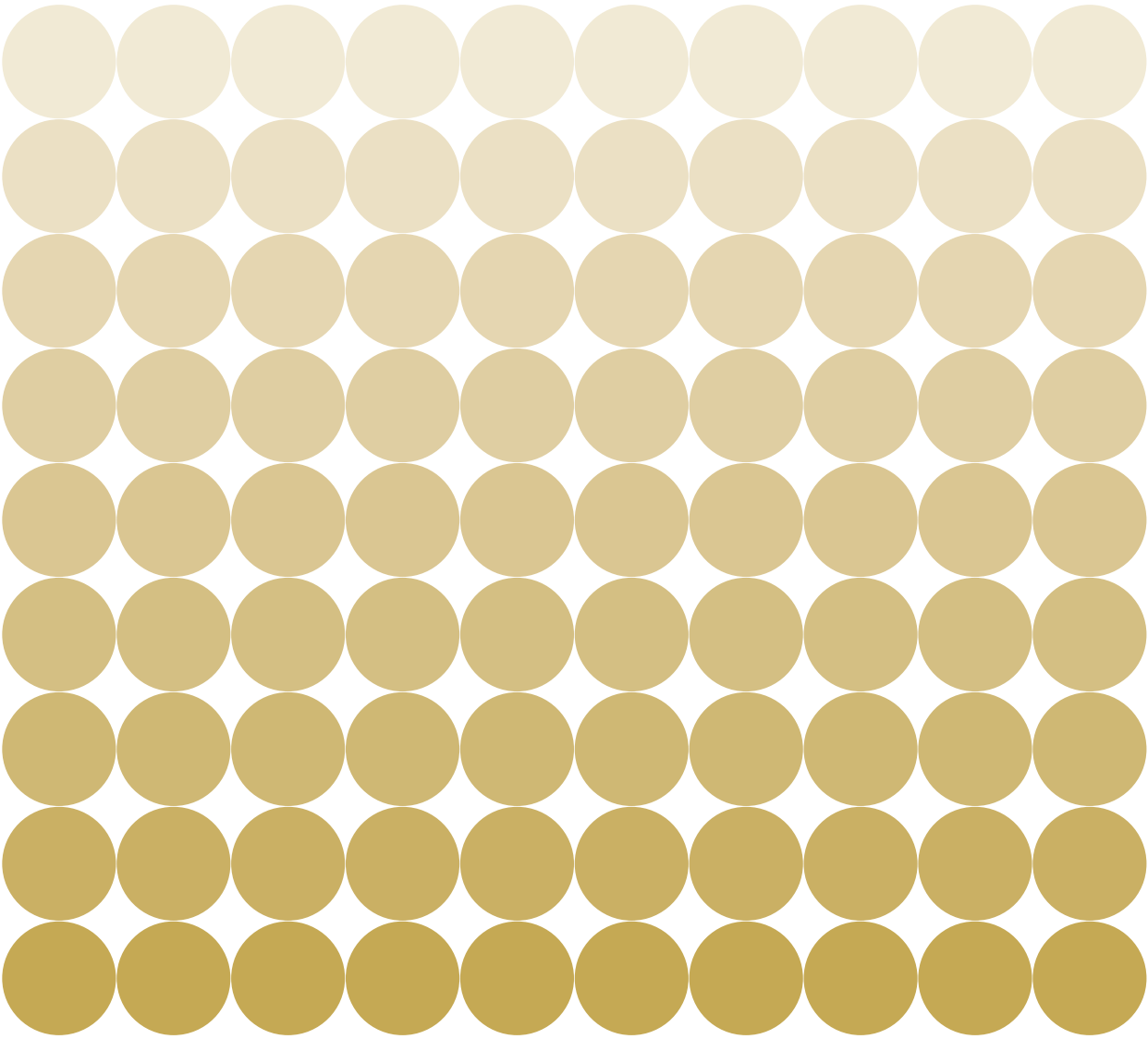
MSB Swedish Civil Contingencies Agency

SIPRI
Policy Paper
55
September 2020

CYBER-INCIDENT MANAGEMENT

Identifying and Dealing with the Risk of Escalation

JOHAN TURELL, FEI SU AND VINCENT BOULANIN



STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

SIPRI is an independent international institute dedicated to research into conflict, armaments, arms control and disarmament. Established in 1966, SIPRI provides data, analysis and recommendations, based on open sources, to policymakers, researchers, media and the interested public.

The Governing Board is not responsible for the views expressed in the publications of the Institute.

GOVERNING BOARD

DIRECTOR

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE
Signalistgatan 9
SE-169 72 Solna, Sweden
Telephone: + 46 8 655 9700
Email: sipri@sipri.org
Internet: www.sipri.org

Cyber-incident Management

Identifying and Dealing with the Risk of Escalation

SIPRI Policy Paper No. 55

JOHAN TURELL, FEI SU AND VINCENT BOULANIN

sipri
STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

MSB Swedish Civil Contingencies Agency

September 2020

Contents

Preface

Networked information and communications technologies (ICTs) have become an essential part of everyday life. The societal benefits of these technologies are beyond dispute. It is also evident, however, that these technologies generate new kinds of societal vulnerabilities.

Increased reliance on computer and networked technologies in most state and business operations, as well as the personal sphere, has created a situation where a single cyber incident affecting the normal functioning of one ICT system can cause major societal disruptions, which in turn can have major economic, societal and political consequences. The reason cyber incidents can take a political turn is that the cause of a cyber incident is not always clear. It may take some time to determine if a cyber incident is the result of cyberattack, human error, system malfunction or natural phenomenon. Uncertainty surrounding the cause of a cyber incident may then in turn lead to misperceptions and actions that may be escalatory in nature and lay the ground for conflict. The belief that a cyber incident might be the result of a cyberattack, along with pressure from the public and the media to respond promptly, might lead politicians to attribute responsibility for the incident too hastily. They may point the finger at another state or group, despite the fact that they only have incomplete information. This in turn may trigger an actual political crisis between states.

The question of how this type of escalation can be prevented has been the focus of a nine-month research project conducted by the Stockholm International Peace Research Institute (SIPRI) in partnership with the Swedish Civil Contingencies Agency (MSB). The objective of this project is to help national crisis management authorities not only to improve their strategies for preventing, detecting and handling cyber incidents, but also to equip them for managing the societal and potentially political aftermath. As part of this project, an expert workshop on 'De-escalation of cyber incidents' was held in Stockholm on 24 May 2019. This report builds on the outcomes of the project workshop and on desk research to provide insights and recommendations on preventing and de-escalating such crises.

This report is unique in its emphasis on the importance of employing de-escalatory strategies and actions for managing the consequences of cyber incidents. Although the recommendations were developed for the Swedish context, they are relevant to the larger international community of policymakers and practitioners who work on cybersecurity and crisis management. SIPRI also commends this report to researchers in politics and international relations, as well as to members of the general public who are interested in understanding the particulars of conflict escalation in the cyber context.

Dan Smith
Director, SIPRI
Stockholm, September 2020

Acknowledgements

The authors would like to express their sincere gratitude to the Swedish Civil Contingencies Agency (Myndigheten för samhällsskydd och beredskap, MSB) for supporting this project through generous research funding and participation in its 'De-escalation of cyber incidents' workshop on 24 May 2019.

The authors would also like to thank all the speakers and participants who shared their knowledge and experience at the workshop: Michael Chertoff, former Head, US Homeland Security Department; Dr Serge Droz, Vice-president, CERT, Open Systems; Mihoko Matsubara, Chief Cybersecurity Strategist, Nippon Telegraph and Telephone Corporation; Dr Gary Brown, Professor of Cybersecurity, Marine Corps University; Sebastian Bay, Senior National Representative for Sweden, NATO Strategic Communications Centre of Excellence in Riga; Dr Fredrik Blix, Associate Professor in Information Systems Security and Head of Information Security and Governance Services, Cybercom Group in Sweden; Dan Eliasson, Director General, MSB; Dr Åke Holmgren, Head of Department, Cybersecurity and Secure Communication, MSB; Fredrik Löjdquist, Swedish Ambassador on Hybrid Conflict; Sara Myrdal, Director of International Affairs, MSB; Merle Pajula, Estonian Ambassador to Sweden; Sigrún Rawet, Deputy Director, SIPRI; Samuel Taub, Manager, Cyber Strategic Analysis Programme, MSB; Erik Ryd, Analyst, Cybersecurity Strategic Analysis Programme, MSB; Dan Smith, Director, SIPRI; Luc van de Goor, Director of Studies, Conflict, Peace and Security, SIPRI; and Erik Windmar, Cybersecurity Coordinator, Swedish Government Offices.

Further, special appreciation goes to Dr Eneken Tikk, Head of Normative, Power and Influence Studies, Cyber Policy Institute, and Dr Mika Kerttunen, Programme Director, Cyber Policy Institute, for their deep insights and generous contributions at various stages of this research. Special thanks also go to Dan Smith, Dr Ian Anthony, Kolja Brockmann and Mark Bromley at SIPRI and to the MSB staff for their feedback on this paper.

Finally, the authors would like to acknowledge the fine work of the SIPRI Editorial and Publications Department, and to thank Ekaterina Klimenko, Alexandra Manolache and Stephanie Blenckner for providing logistical and communications support.

Responsibility for any errors and views expressed in this report lies with the authors. The recommendations presented here are those of the authors and are not meant to reflect the policy positions of the Swedish Government Offices, MSB or other Swedish governmental agencies.

The authors
Stockholm, September 2020

Summary

The ever increasing dependence on information and communication technologies (ICTs) in all aspects of society raises many challenges for national crisis management agencies. These agencies need to prepare not only for new cyberthreats and cyber vulnerabilities, but also for the fact that the aftermath of a cyber incident affecting critical infrastructure has its own challenges. On the one hand, the practical disruptions caused by an isolated incident can be hard to predict and control, and, on the other hand, the consequences and perceptions of an incident whose cause is not yet determined can be equally hard to manage. Uncertainty surrounding the cause of the incident and the remedial actions being taken often lead to public speculation and political pressure to respond in ways that could create political tensions, and possibly conflict, between countries.

This policy paper is the result of a nine-month research project that was jointly conducted by SIPRI and the Swedish Civil Contingencies Agency (MSB) on cyber-incident management. It explores what national crisis management authorities can do to improve their cyber-incident prevention, detection and response strategies and also how they can better deal with the larger societal and potentially political aftermath. It investigates why and how cyber incidents may lead to escalatory scenarios and how these scenarios can be avoided and contained using various de-escalatory approaches. It comprises an introduction providing background and the inspiration of this report (chapter 1); four chapters that explore the dynamics of escalation and de-escalation from conceptual (chapter 2), analytical (chapters 3–4) and empirical (chapter 5) standpoints; and two chapters that present the main findings and recommendations (chapters 6–7), which can be summarized as follows.

Major cyber incidents are prone to escalate. There are numerous reasons for this. To begin with, the security landscape in cyberspace is deteriorating. In that context, when an incident takes place, people tend to presume that it is antagonistic in nature and was caused by a cyberattack. Society's ever increasing dependency on networked ICTs has also made possible situations where a simple incident on a single computer can cause widespread disruptions, sometimes even beyond national borders. Communicating about a cyber incident in such a situation can be a complex endeavour, particularly when the causes of the incident are not known. An aggravating factor is that other actors may join the conversation and shape the narrative in a way that is escalatory. Such actors may be found in the media, private cybersecurity companies, other states' intelligence agencies and multilateral organizations. The organization that was first affected by the incident may also have a vested interest in withholding information about the incident, which in turn can fuel public speculation and misperception of the incident. On top of this comes a number of persistent vulnerabilities. The general lack of knowledge and critical perspective on cybersecurity issues outside the cybersecurity industry make it possible for farfetched claims about cyber incidents to stand largely unopposed. Most states are also still in the process of

developing their technical capabilities to detect and investigate incidents. They may also lack means of efficient information sharing and coordination between the different organizations that might be involved in the response to the cyber incident.

Certainly, the best way to handle the risk of escalation in the aftermath of a cyber incident is by making sure that cyber incidents do not happen in the first place and then, if they do happen, by making sure they do not have any long-lasting or wide-ranging effects. This report therefore recommends states and national crisis management agencies opt for an approach to cyber-incident management that focuses on broad robustness and resilience across the cyber, cognitive and physical domains. To achieve this, they could seek to:

1. Focus on building frameworks for transferring incident management responsibility during escalating cyber incidents. Such frameworks should develop a deterministic chain starting in private organizations, moving to supporting organizations, continuing to the national government level and, finally, connecting to the international political level.

2. Align interests to facilitate external information sharing and develop internal cohesiveness.

3. Build and maintain trust for the incident management system in general, and incident management organizations in particular.

4. Enhance existing structures for the management of cyber-incident escalation risks, rather than build extraneous structures.

5. Establish prior readiness and prepare for the unforeseen by implementing and maintaining a framework for systematic and risk-based cybersecurity work, a standard operating procedure for communication and coordination, clear reporting standards, continuous staff training, and appropriate legislation and policies to enable and support these measures.

6. Enable coherent response coordination by maintaining a response framework that allows inter-agency, public–private and cross-border cooperation.

7. Conduct proactive communication strategies by providing swift and fact-based information, focusing on de-escalatory and non-escalatory messaging, preparing communications as often as new information comes to light, and keeping a clearly defined division of roles in terms of who communicates about what.

8. Enable sound decision making by dedicating a function that maintains and develops an overall situational awareness, analyses and assesses trade-offs among incident response policy options, and always proposes a varied set of evaluated policy options.

9. Optimize national institutional arrangements to context-specific needs by: balancing the number of agencies involved, the number of conflicting objectives involved and the number of sectors to which cybersecurity support is given; balancing the involvement of the military, intelligence and law enforcement communities; integrating cyber, cognitive and physical-incident response capabilities; and balancing cybersecurity roles between the state and the private sector.

Abbreviations

CERT	Computer Emergency Response Team
CSIRT	Computer Security Incident Response Team
CVE	Common Vulnerabilities and Exposure
DNC	Democratic National Committee (USA)
GDP	Gross domestic product
ICS	Industrial control system
ICT	Information and communications technology
IOT	Internet of things
IP	Internet protocol
JPS	Japan Pension Service
IR	International relations
LFV	Luftfartsverket (Swedish Civil Aviation Agency)
MSB	Myndigheten för samhällsskydd och beredskap (Swedish Civil Contingencies Agency)
NHS	National Health Service (UK)
NIS Directive	Directive on the security of network and information systems (EU)
SIPRI	Stockholm International Peace Research Institute
TSE	Tokyo Stock Exchange

1. Introduction

On the evening of 15 March 2016, a Sunday, staff at the Teracom Network Operations Center in Stockholm noticed warning signals coming from network switches near the Dalsjöfors radio mast, just outside of Häglared in the vicinity of the city of Borås, Sweden. Moments later they also noticed that no television or radio signals were being transmitted from the mast, affecting, as it would later emerge, 190 000 people, or about 2 per cent of Sweden's total population.[1] This incident—now known as the Häglared incident—received much attention in Sweden for several reasons.

First, the Häglared incident was at the time, and remains at the time of writing, one of the most publicly known cyber incidents in Sweden—'cyber incident' being defined as an event having an actual adverse effect on the security of a network or information system. In this case, the incident was, in fact, the result not of an event originating in the cyber domain but of a physical act of sabotage: the police investigation eventually found that the bolts at three of the mast's wire-holds had been manually unscrewed, causing the top 200 metres of the 322 metre-long mast to fall to the ground.[2] Although the sabotage affected only television and radio signals at the time, the mast is also a relay component of Sweden's public alarm system for broadcasting warning signals to the general public in an emergency. In other words, the mast has national security importance.

Second, the Häglared incident rapidly took on an international dimension, independently from the intention of the governmental authorities in charge of its handling. When the police investigation could not determine who was directly responsible for the sabotage, Swedish media and other actors began to speculate about the possibility that a foreign power—assumed to be Russia—could be behind it. The Russian embassy in Sweden eventually issued a public statement denying any Russian responsibility.[3]

Third, the Häglared incident was followed by a number of unrelated incidents in other critical sectors, notably in the telecommunications and transportation sectors, all within a few days of each other.

On 17 May 2016, the communications service of the Swedish emergency call system, SOS Alarm, went down, severing its connection to rescue and ambulance services for 15 minutes. Swedish news media reported that the police had been asked to be on the lookout for people dressed in Teracom's work gear and clothes, after Teracom had a break-in at one of their sites in which communications gear and staff uniforms were stolen.[4] Later that day, a mysterious object was found

[1] Swedish Civil Contingencies Agency (MSB), *Mediebranschen 2016—hot, risker och sårbarheter* [Media industry 2016: threats, risks and vulnerabilities] (MSB: Stockholm, Mar. 2017), p. 49.

[2] MSB (note 1), pp. 49–51.

[3] Granlund, J., 'Internationellt spår utreds efter sabotage mot tv-mast' [International angle investigated following sabotage of TV mast], *Aftonbladet*, 16 May 2016.

[4] 'Sabotage mot master—detta har hänt' [Sabotage of the mast—this has happened], *SVT*, 18 May 2016.

outside the telecommunications mast in Skutskär, outside of Gävle. The police bomb squad went to the scene and removed the object.[5]

The following day, on 18 May, the media reported that another mast in Tranemo, in the same region, had been sabotaged in early May.[6] Rescue services had noticed that someone had cut a cable to the mast.

On 19 May, two more incidents happened: first, the booking system of the Swedish national transport operator, SJ, malfunctioned for 10 hours; and second, the communications and radar systems of the Civil Aviation Administration (Luftfartsverket, LFV) stopped working, forcing the LFV to close the airspace over Stockholm for commercial flights (cancelling a total of 122 flights).[7]

Uncertainty at the time around the causes of these incidents again led the media to speculate about the possibility of cyber sabotage. The media speculation forced the authorities to consider whether the Häglared incident was part of a larger sabotage campaign against Swedish critical assets and whether all these incidents had to be handled in a coordinated way. As each incident occurred, Swedish agencies had to consider whether they were connected and what, if anything, that meant. From a practical standpoint, it meant they had to investigate not only each individual incident, but also whether any links could be established between these incidents.

The Swedish agencies eventually found that, unlike the Häglared incident, these five incidents were the result of human mistakes, system errors and natural phenomena; they were able to issue public statements denying a connection between the incidents and pointing to the non-antagonistic nature of what had taken place. The affected services returned to normal and worries about a possible 'grey zone situation'—unconventional military attacks such as cyber sabotage—gave way to reflections on the vulnerability of a digitalized society. What if some of the incidents had actually been caused by attacks? What if there had been a demonstrable connection between those attacks? What if that connection had become commonly known? What if an authoritative or influential party had spread the message that, contrary to the claims of Swedish agencies and other organizations, there were clear signs that the incidents were the result of attacks perpetrated by a hostile foreign power? What if such a foreign power, guilty or not, had taken that opportunity to start a diplomatic row?

The Häglared incident is a clear example of how the *aftermath* of a cyber incident can develop into a difficult situation in its own right, especially when other incidents occur in the same country and near in time. Amid the uncertainty that develops before the cause of an incident has been determined, speculation turns into rumours, giving rise to unease and even panic about the cause being a possible attack on the affected operation. Suspicions that an incident or series of

[5] 'Misstänkt föremål vid telemasten i Skutskär' [Suspicious object at the telemast in Skutskär], *SVT*, 17 May 2016.

[6] 'Fler sabotage av radiokommunikationen' [More sabotage of radiocommunications], *SVT*, 18 May 2016.

[7] 'Kommunikationsproblem stoppade flygtrafiken' [Communications problems stopped air traffic], *SVT*, 19 May 2016; and 'Svenska flygstoppet utreds—krismöte i dag' [Swedish flight stoppage investigated—crisis meeting today], *HBL*, 20 May 2016.

incidents is the result of a targeted attack may spread, and specific individuals and entities may be identified and publicly named as being responsible for carrying out or instigating the attack. The situation may escalate into controversies that are both expensive and difficult to handle for the affected individuals and entities as well as the responding agencies. It may also turn out that the incident was the result of an attack—even so, there will be instances where those publicly named as possible instigators or as being directly responsible are in fact innocent. In view of the potential for such crisis escalation following a cyber incident, the ability of authorities to de-escalate the situation is crucial to control the aftermath and to ensure that decision makers always have options, and the time and resources to make rational choices.

This policy paper explores the dynamics of escalation resulting from cyber incidents, and how de-escalatory strategies and actions may be employed to manage their consequences. The key research question that this paper aims to address is why and how cyber incidents can lead to escalating situations, and what factors states and their relevant agencies should consider to strengthen their national capabilities in this field.[8]

The remainder of this paper comprises six chapters. Chapter 2 describes a conceptual and analytical framework for understanding escalation and de-escalation. It presents and discusses the definitions of the key terms used, and the actors referred to, in this paper. Chapters 3 and 4 then map out the various threats and vulnerabilities, respectively, as factors that can fuel escalation of cyber incidents. Chapter 5 presents the lessons learned from a series of case studies on past cyber incidents in other countries and the ways in which those incidents were handled. The last two chapters summarize the key findings (chapter 6) and present the recommendations to the target audience of this paper (chapter 7).

[8] The scope of this paper is limited to civilian activities. Active operations, as well as other military options, in response to an attack are not considered here. The rationale for this limited scope is that the focus is on countering unwanted or uncontrolled escalation. Even if decision makers wish to consider active operations in response to attacks, they will presumably want to have de-escalatory options available if they need more time or if they are uncertain about what to do. Even if they decide on taking retaliatory action, decision makers will presumably not want such action to cause further escalation if other, more peaceful, alternatives exist as means of conflict resolution.

2. Analytic framework: The concepts of escalation and de-escalation and the actors involved

This chapter provides a conceptual and analytical framework for understanding why and how cyber incidents escalate and also how they might be de-escalated. It starts with a brief review of the literature on the concepts of escalation and de-escalation and then introduces the concepts and definitions, as used in this paper, that together make the analytical framework for reasoning about and understanding escalation and de-escalation. Finally, it presents the types of actor that would typically be involved in scenarios where escalation could happen; some concrete examples of how escalation might look in relation to different types of actor; and ways in which the interrelations between actors might complicate matters.

Escalation and de-escalation: A brief review of the literature

In everyday language, *escalation* and *de-escalation* are straightforward concepts that refer to an increase and decrease, respectively, in intensity or seriousness.[9] In academic circles, however, the terms have much more specific meanings that also differ depending on the discipline. There are, for instance, discipline-specific meanings of the terms in international relations (IR), economics and computer science. For the purposes of this paper, the academic field of IR provides the primary backdrop.

In the context of IR, 'escalation' is generally defined as successive, visible and significant increases over time in the vertical, horizontal and political dimensions of an antagonistic dispute between two or more states.[10] Examples of escalation that can be readily observed include the additional commitment of offensive military resources to an ever more intense conflict; the geographic or temporal expansion of a conflict; increases in belligerent rhetoric and threats; greater use of economic statecraft such as trade sanctions; cyber capabilities used as a means of power projection; and the severing of diplomatic relations.[11] Misperceptions and misunderstandings between states can lead to political escalation by increasing tensions between established adversaries or creating friction between states that have insufficient mutual guarantees of certainty and confidence.

The IR literature on escalation and de-escalation is particularly relevant for this paper, not only because the field has generated the most work on these

[9] Lexico, Dictionary.com and Oxford University Press, 'escalation'.

[10] Morgan, F. E. et al., *Dangerous Thresholds: Managing Escalation in the 21st Century* (Rand Corporation: Santa Monica, 2008), pp. 7–8, 18–19. For an analysis of the effect of the balance of power on the likelihood of a conflict's escalation, see Siverson, R. M. and Tennefoss, M. R, 'Power, alliance, and the escalation of international conflicts', *American Political Science Review*, vol. 78, no. 4 (1984), pp. 1057–69.

[11] Becker, M., 'Incremental escalation as a cost-avoidance instrument in international conflicts', *Central European Journal of International and Security Studies*, vol. 9, no. 2 (2015), p. 21.

concepts, but also because it provides a number of useful associated notions—such as 'accidental', 'unintended', 'intentional' or 'deliberate' escalation—that can be useful to describe and analyse escalation and de-escalation dynamics.[12] However, these notions, as they are typically understood by IR scholars, have their limitations when applied to the cyber context. One such limitation is their origin in the context of the cold war, where the focus was on managing the relations between nuclear powers to mitigate the risk of conflict escalating to use of nuclear weapons.[13] The objectives, variables, uncertainty factors and measures that this literature discusses are therefore highly specific to nuclear conflicts and not directly applicable to conflicts playing out in cyberspace. In the conflict scenarios explored during the cold war, it was generally clear who the parties involved in a potential conflict would be. As the rest of this paper will show, the present context is much more complex because cyberspace opens the possibility for a vast set of actors to conduct destabilizing actions more or less anonymously. States do not have a monopoly on 'violence' in cyberspace, and organized non-state actors or even individuals can also be parties to a conflict. The uncertainty of the counterpart in the cyber domain makes it difficult to identify the responsible party, which creates an *attribution* issue.

The IR literature on escalation and de-escalation has evolved since the end of the cold war. Scholars generally accept that escalation of conflicts relates to more than simply nation states at international level; escalation also relates to individual behaviours and organizational performance.[14] However, little has been written to date on how the concepts of escalation and de-escalation apply to cyber incidents.[15] Moreover, the few publications that exist on the topic tend to focus on the case of cyberattacks and how they are, or might be, dealt with in the context of foreign policy and the military. These papers pay little to no attention to escalatory scenarios that may emerge from cyber incidents that are not generated by cyberattacks. They thus bypass the question of how such escalation might be dealt with by crisis management agencies. This is the knowledge gap that this paper aims to help fill.

Definitions of escalation and de-escalation in this paper

In order to structure the thinking on escalation dynamics in the cyber context, this paper formulates a number of definitions that combine into a framework for the analysis of escalation dynamics. Use of the concept of 'conflict' in this framework is wider than its use in IR.

[12] Pruitt, D. G., Kim, S. H. and Rubin, J., *Social Conflict: Escalation, Stalemate, and Settlement* (McGraw-Hill: Boston, 2003), pp. 87–91; Mitchell, C., *The Nature of Intractable Conflict: Conflict Resolution in the Twenty-first Century* (Palgrave Macmillan: Basingstoke, 2014), pp. 71–75; Kahn, H., *On Escalation: Metaphors and Scenarios* (Praeger: New York, 1965); and Morgan et al. (note 10).

[13] Morgan et al. (note 10).

[14] Bösch, R., 'Conflict escalation', *Oxford Research Encyclopedia of International Studies* (Nov. 2017).

[15] Lin, H., 'Escalation dynamics and conflict termination in cyberspace', *Strategic Studies Quarterly*, vol. 6 no. 3 (Cyber special edition, Fall 2012).

Escalation occurs when an actor either starts performing aggressive or provocative actions directed towards some other actor (thus creating a conflict or joining a pre-existing conflict), or increases the number or severity of such actions towards other actors in a pre-existing conflict.

De-escalation occurs when a majority of the actors in a conflict either cease aggressive or provocative actions directed towards the other actors in the conflict, or they decrease the number or severity of such actions.

Non-escalatory post-incident actions are peaceful actions conducted by at least one actor during or in the aftermath of an incident with the observable intent of ensuring that tensions are decreased or not raised, and that no conflict arises because of the incident.

De-escalatory actions are peaceful actions conducted by at least one actor that are observably intended to reduce tensions among the actors in a conflict or to reduce the number or severity of aggressive or provocative actions taken by the actors in a conflict.

So, non-escalatory actions concern the handling of uncertainty relating to cyber incidents. These are actions that may be conducted *before* a conflict arises with a certain actor. Examples include actions taken to calm nerves, to counter unfounded claims or to show that the incident is being handled. De-escalatory actions, in contrast, are actions that are taken once a conflict has arisen.

Some implications for understanding escalation and de-escalation

There are a number of noteworthy implications following from these definitions:

1. Aggressive reactions to inaction when action is expected may be escalatory, but the inaction is not escalatory in itself. That is, the chain of escalation cannot start through the inaction of some actor—but it can start by aggressive or provocative reactions to inaction.

2. Actors who are not parties to a particular conflict can perform de-escalatory actions. Mediators, for instance, can perform de-escalatory actions.

3. While de-escalatory actions can be performed unilaterally, de-escalation cannot be achieved by unilateral action alone. Other actors, especially the parties to the conflict, must acquiesce.

4. De-escalation can be achieved by escalatory actions. For instance, a major power may threaten two parties in a conflict with aggressive measures unless they cease hostilities.

5. Inaction might be de-escalatory, if several parties refrain from action. However, taking action that is less escalatory than expected is not in itself de-escalatory.

6. While the primary focus in this paper is on situations where a cyber incident is the 'spark' for escalation, nothing in the above framework requires de-escalatory actions to take place in cyberspace.

Non-escalatory and de-escalatory actions are conditioned on being 'observably intended' to do something. Communication therefore plays a critical role in conveying the correct intention. All communications are constituted by three fundamental components: (*a*) the sender, (*b*) the message and (*c*) the receiver.[16] In this context, communication is sender-oriented when releasing information, then becomes an interactive two-way process between senders and receivers. Ensuring that the sender's non-escalatory or de-escalatory intent is conveyed to receivers through the message, which comprises both its semantic content and contextual factors, is the key.

Whether actions are perceived as non-escalatory or de-escalatory hinges on not only what those actions are and how well their underlying intentions are communicated, but also on historic and contextual factors such as expectations and rights relating to both the acting party (sender) and the other actors involved in the incident (receivers). Such expectations may in turn be based on the reputations and public images of the respective actors, as well as other factors. For instance, if an actor who is known for aggressive behaviour attempts to take de-escalatory actions, other actors might be less inclined to reciprocate because of suspicions about the aggressive actor's real intentions.

Understanding the escalatory potential of incidents

Key factors in terms of the escalatory potential of incidents are the *visibility of* and *attention given to* the incident. To have an escalatory potential, an incident or its consequences must be both observable and observed. That is, the incident or its consequences must somehow attract attention in the first place. This may happen either during the course of the incident or at some later point when the consequences of the incident are noticed. Escalatory potential is especially heightened if there are indications that the incident itself, or some controversial aspect connected to the incident, has been intentionally hidden from the public. The alleged state-sponsored hack and theft of information from the Democratic National Committee (DNC) and the subsequent use of the stolen information to interfere in the run-up to the United States presidential election in 2016, is an example of how controversial aspects connected to the incident can be exploited by antagonists and result in an escalation. In that particular case, emails sent between members of the DNC and other members of the Democratic Party were hacked and then leaked to the public. The content of these emails then led some to believe that some people were covertly trying to cheat in order to make sure that Hillary Clinton would win the nomination.[17] These developments turned the incident from a breach (where trust in the ability of the DNC to maintain IT security and the confidentiality of communications was damaged) into a major

[16] Shannon, C. E., 'A mathematical theory of communication', *The Bell System Technical Journal*, vol. 27 (July and Oct. 1948), pp. 379–423, 623–56; and Ellis, R. and McClintock, A., *If You Take My Meaning: Theory into Practice in Human Communication* (Bloomsbury Academic: London, 1994).

[17] Mueller, R. S., 'Russian hacking and dumping operations', *Report on the Investigation into Russian Interference in the 2016 Presidential Election*, vol. 1 (US Department of Justice: Washington, DC, Mar. 2019); and Huetteman, E., 'Obama White House knew of Russian election hacking, but delayed telling', *New York Times*, 21 June 2017.

incident where trust in the professionalism and commitment to the intra-party democratic process was badly hurt in some quarters.

Once an incident has been observed by someone, a key determinant of its escalatory potential is the wider attention it receives. This will depend on a range of factors, such as how long the incident goes on without being resolved; the severity of its consequences; whether it has any particularly interesting aspects (e.g. sordid, political or antagonistic aspects); the public image of the actors involved in the incident; whether the incident was caused by an attack and, if so, whether this has been demonstrated; and whether there are conflicting narratives on what happened. The more attention an incident receives, and the longer it stays in the public consciousness, the greater its escalatory potential.

Types of actor involved in cyber incidents

The actors involved in a cyber incident can be classified into 10 main types:

1. *The incident organization* is the organization in whose information systems or networks the incident first occurs. In some cases, the affected information system or networks may uphold a service on which the systems or networks of other organizations depend.

2. *Affected parties* are the individuals and organizations affected by the incident by way of a dependency on a service provided by the incident organization, or because of obligations they have to the incident organization. (Organizations in this category may, in turn, suffer incidents of their own as a consequence of how they are affected by the first incident.)

3. *Supporting organizations* are organizations that provide incident or crisis management support to the incident organization or to affected parties.

4. *Supervising organizations* are organizations that supervise the incident organization or the affected parties. This category mainly comprises supervisory governmental or local agencies, but it may also comprise organizations that have a supervising role as part of some private agreement or contract.

5. *Law enforcement agencies* are the various crime prevention, investigation and security agencies that have a mandate to investigate criminal aspects of cyber incidents.

6. *Governments* are the governments in countries where any of the above kinds of organizations operate or are headquartered.

7. *Attributors and commentators* are the individuals, organizations and states who describe incidents as 'attacks' and attribute responsibility for the attack to a named or indicated actor. Examples of attributors

and commentators include cybersecurity companies, states (such as governments or their agencies) and the media (broadly defined below).

8. *Assailants (purported or actual)* are the individuals, organizations and states that are accused of committing, or that actually commit, attacks that cause cyber incidents.

9. *The media* includes media organizations of both the traditional and newer kinds (such as citizen journalism and investigative networks) that report on cyber incidents.

10. *The public* refers to the people who live in the country in which a cyber incident occurs.

Interactions between the incident organization and other actors

When an incident occurs, the incident organization will have various interactions with many of these types of actor. The way it handles its relationship with a particular type of actor may affect its interactions with other types of actor, with the potential for ending up in an (uncontrolled) escalating situation with one or more of them.

The incident organization (and its supporting organizations, if any) may face various kinds of escalation in its interactions with other actors. For example, affected parties may make complaints and demands for compensation to the incident organization. Supervising organizations may place the incident organization under review or impose sanctions. Law enforcement agencies may undertake criminal investigations into the incident organization. The media and the public may subject the incident organization to scrutiny and criticism. The incident organization can be made an unwitting participant in various influence or foreign interference operations, or be drawn into various conflicts because of attribution claims by third parties. It may suffer, for instance, a cyberattack, because of the negative attention it gets, and it may even come to the attention of governments which then take various punitive actions against it, such as restricting market access if it is a company, or forbidding it from working on certain issues if it is a non-governmental organization.

This set of relations makes for a complex game theoretic situation where the handling of unwanted escalation may be frustrated in various ways. For instance, if the cyber incident is an intrusion in which sensitive personal information is stolen from the incident organization, then the organization may be loath to admit an intrusion has taken place or that anything was stolen, and may be slow or fail to inform the authorities about it for fear of regulatory action and sanctions. If that stolen information is then used in ways that escalate the situation by intensifying the conflict or adding new parties to the conflict (e.g. by using the information in a way that adversely affects the individuals whose personal information it is), the incident organization and its various supporting organizations will not be as prepared and coordinated in their response as they could have been. Similar

complexities apply to public relations. When an organization suffers an incident that is intrinsically not visible to outsiders, it is faced with a choice between admitting what has happened or keeping it secret. Making an admission risks negative reactions but possibly allows the organization to influence the extent and kind of reactions that occur. Keeping the incident secret risks information about the incident leaking at some future time, possibly with a more severely negative reaction, especially if it appears the organization has attempted to hide something considered to be of public interest.

Role of actors in taking de-escalatory action

Since a cyber incident's visibility and the attention it receives are key factors in its escalatory potential, acting to mitigate and recover from the impact of the incident, in a transparent and interactive manner, will be the primary de-escalatory strategy for the incident organization. Beyond that, the de-escalatory actions an incident organization should undertake in the face of ongoing escalation or high tensions will depend on the type or types of other actors involved. For instance, informing affected parties of the incident and offering compensation may serve as a de-escalatory strategy. In relation to supervisory organizations, proactively demonstrating that the relevant regulations have been followed and that the incident was caused by factors beyond the control of the incident organization may serve to forestall supervisory action or sanctions.

It may also be the case that some de-escalatory actions may not or should not be undertaken by the incident organization itself. For example, if a cyber incident occurs in a private company, and the situation escalates to a point where a foreign government acts against that company in a public way, then it may be better or even required that the government in the home country of the company steps in and performs some of the de-escalatory actions needed to resolve the situation.

3. Escalation threats in the aftermath of a cyber incident

This chapter describes a number of threats that may encourage or accelerate escalation during or in the aftermath of a cyber incident. These threats are mostly global in character. The analysis in this chapter uses the same definition of the concept of a threat that the Swedish Civil Contingencies Agency, MSB, uses in its implementation of the European Union NIS Directive.[18] That is, a threat is the presence of something that causes or contributes to causing either an incident or unwanted consequences in the event of an incident. Four broad threat areas play into cybersecurity-related escalation dynamics: (*a*) a deteriorated security environment; (*b*) bias and the presumption of antagonism; (*c*) cyber-physical and cyber-social interdependencies; and (*d*) the complexity of the public communications arena. In the discussion that follows, specific examples are drawn from past cyber incidents.

A deteriorated security environment

The security environment globally, as well as in Sweden's immediate region, has deteriorated over the past decade. For certain parties, this may lower the bar for conducting certain actions and therefore might imply a higher risk for cyberattacks and other actions that do not quite meet the threshold, on the basis of international law, to count as armed conflict. Such a situation could also generate aggressive responses to real or perceived cyberattacks. Amid high security alert levels, there is a higher risk that an incident will be thought of as a cyberattack and also that any incident, especially one caused by a cyberattack, is interpreted as being related to the overall security environment, whether or not such is the case.

At least eight phenomena related to cyberspace, and the way in which actions conducted in cyberspace are interpreted and understood as part of the overall security environment, contribute to the continuing deterioration of the security environment: (*a*) assertive postures of major actors; (*b*) an increasing number of entities developing offensive cyber capabilities; (*c*) increased availability of tools that enable offensive cyber operations; (*d*) an increase in attacks on information and communication technology (ICT) supply chains; (*e*) use of 'deterrence' measures in cyberspace to enable counter-attacks; (*f*) obfuscatory and imitative practices; (*g*) securitization of ICTs; and (*h*) the escalatory potential of actions according to perceptions of the actor. Each is discussed in turn below, followed by consideration of their escalatory potential when different actors perceive these developments and behaviours in different ways.

[18] 'Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union', adopted 6 July 2016, entered into force Aug. 2016.

Assertive postures of major actors

Actions conducted in cyberspace may increasingly come to be interpreted and analysed in view of the 'postures' taken by major actors. Since some of these postures have aggressive elements, the propensity to interpret border-crossing incidents or cyberattacks as being related to such postures, may increase.[19] However, if interpreting certain kinds of cyber incidents in the context of aggressive postures becomes the default, several risks arise: too much may be read into an incident; conclusions about *why* an incident happened may go too far; expectations may be shaped around what certain actors will do in response to certain events; and further expectations may be created as to when and how often escalation will occur. A further risk is that, to maintain the credibility of such postures, some major actors may have to step up the frequency and severity of their offensive actions in response to attacks they and others detect.

Increasing number of entities developing offensive cyber capabilities

Given the low costs associated with developing rudimentary offensive capabilities for cyber operations (compared to other military or intelligence-related capabilities), as well as the comparatively limited risk of getting caught for those who conduct such operations, the development and use of offensive capabilities seems to have become an attractive option in many places.[20] It becomes particularly attractive to assailants whose strategic targets (whether competitors or adversaries) are well advanced in their digitalization processes; or whose targets' wealth or security is heavily reliant on a few organizations that are strongly dependent on ICT—especially if they know or assume their strategic targets are developing similar capabilities, creating a strong incentive to respond in kind.

Increased availability of tools that enable offensive cyber operations

Demand from intelligence agencies, law enforcement agencies[21] and cyber criminals[22] has created a market for the development of tools that make various forms of hacking easier (referred to by some as 'cyber weapons'), such as software solutions or code (colloquially known as 'exploits') that target ICT vulnerabilities (especially those previously unknown, colloquially termed 'zero day').[23] The

[19] White House, *National Cyber Strategy of the United States of America* (White House: Washington, DC, 2018), pp. 20–21; British Government, *National Cyber Security Strategy 2016–2021* (British Government: London, 2016), pp. 47, 51; '中华人民共和国网络安全法 [China's cyber security law]', *Xinhua*, 7 Nov. 2016; and '习近平"4·19讲话"┌含国家网络治理的"五观"[Xi Jinping's 'April 19 Speech'—five views on national cyber governance]', *The People*, 17 Apr. 2017.

[20] Clapper, R., Lettre, M. and Rogers, M. S., Joint Statement for the Record to Senate Armed Services Committee [on] Foreign Cyber Threats to the United States, 5 Jan. 2017.

[21] Larson, J. and Tigas, M., 'Leaked docs show spyware used to snoop on US computers', *Ars Technica*, 9 Aug. 2014.

[22] Insikt Group, *Bestsellers in the Underground Economy: Measuring Malware Popularity by Forum* (Recorded Future: Somerville, MA, July 2019).

[23] An example is a 'zero-day vulnerability' in software, which is a flaw known to the software vendor but for which no patch (fix) is in place. Hackers can target the vulnerability with specific code called a 'zero-day exploit'; the term 'zero day' refers to the fact that once the vulnerability is discovered, the vendor has zero days to fix the flaw before hackers can make use of it—in some cases it is already too late.

availability of such tools is augmented by various 'dumps'—releases, usually unauthorized, of data or code onto an internet server—of tools allegedly developed by intelligence agencies, which may be used for the same purposes.

The global WannaCry and NotPetya attacks famously exploited the EternalBlue vulnerability, the knowledge of which was provided by one such dump.[24] There are also dumps of various security-related credentials, such as names, usernames and passwords. Since people often reuse (or only slightly change) their credentials across many services, finding out who works in an organization and then looking them up in such dumps often results in a number of hits that an attacker may then use to try to gain access to the information systems and networks where that person works.

Another resource for offensive cyber operations is the Common Vulnerabilities and Exposure (CVE) system that security researchers use to spread knowledge about vulnerabilities and to push for fixes, which may be used by those attempting to hack systems they suspect have not been patched properly.

Compounding the risks of cyber-offensive tools being widely available is the fact that it is easier to create a tool that does indiscriminate damage than one that does damage in a restricted and targeted fashion. The Stuxnet malware is a famous exception to the norm that was very clearly optimized to destroy only centrifuges arrayed in a way that exactly matched how they were arrayed in the Iranian nuclear programme; few such finely targeted tools have been seen since.[25] However, the indiscriminate destruction caused by the NotPetya virus may have demonstrated the need to control and restrict the workings of malicious code beyond the targeted actors.

Increase in attacks on ICT supply chains

ICT supply chains—the individual components that make up an ICT system or network—are increasingly coming under attack.[26] In an ICT supply chain attack, an essential component is manipulated to expose a vulnerability or to embed a threat that the attacker may then use against targets further down the line in the supply chain. ICT supply chain attacks have a lot in common with the problems associated with insiders in organizations, in that the necessary element of trust is used against the supply chain or organization. In the case of insiders, certain individuals are trusted to access sensitive information or systems in order to do their jobs and serve the organization. In the case of supply chains, certain components need to be trusted for the system or network to be able to function as required. That trust can be exploited.

In the supply chain context, components such as hardware, firmware or software are often developed by many different suppliers in many different

[24] Computer Emergency Response Term for the European Union (CERT-EU), 'WannaCry ransomware campaign exploiting SMB vulnerability', CERT-EU Security Advisory 2017-012, 22 May 2017.

[25] Milevski, L., 'Stuxnet and strategy: a special operation in cyberspace?', *Joint Force Quarterly*, vol. 63, no. 4 (2011), p. 64.

[26] European Commission, NIS Cooperation Group, *EU Coordinated Risk Assessment of the Cybersecurity of 5G Networks*, Report, 9 Oct. 2019, p. 11.

countries, and may be refined and combined in many iterations by being sent back and forth through the global interlinkages that make up the global ICT supply chain. Components may originate in countries where it is difficult to review and monitor the development process—a task often made even more difficult by the sheer number of suppliers—or in countries that are adversaries of the component's destination country. These factors create vulnerabilities in the ICT supply chain. The risk is that a component that is compromised in the early stages of a supply chain has the potential to propagate far and wide, and end up in (potentially) millions of devices, software programs and systems. The effect is a general reduction of trust in ICTs through a process where suspicions about the intents and actions of state, or state-sponsored, actors are carried over into suppliers of ICT components. Another effect is the generation of strong suspicions about what ICTs might actually be doing without users' knowledge or control.

Use of escalatory measures to enable deterrence

The way that some states and organizations think about cyber defence contains escalatory potential. A prevalent theme is that of 'offensive deterrence'—the idea that a party will be deterred from attacking another party if it knows that its opponent will be able to immediately respond with a counterstrike of similar or higher impact. Beyond the escalation risks this poses in terms of detection and attribution in the cyber context, this strategy also often requires the defending party to establish a *forward presence*—a cyber version of a foothold or bridgehead—in the networks of the party it believes might attack. Such a forward presence may take a long time to establish, and may therefore need to be prepared for *before* the attack an actor plans to counterstrike against even begins. To establish such a forward presence, the defending party needs to pierce the networks of the party who is about to attack and then move, without detection, to the parts of the network where it needs to be to inflict damage, and then find a way to deliver a payload in that part of the network, again without being detected or blocked.[27] When an actor has established a foothold from where a payload may be delivered, that actor must then maintain that foothold until the actor needs it to counterattack. However, to use the foothold to *counterattack*, the actor must (*a*) know it has been attacked and (*b*) know that it has been attacked by the organization in whose networks it has established a foothold, or that the attacking organization is somehow linked to the organization where it has that foothold. If the actor misattributes the attack it has suffered and strikes at an actor it thinks is responsible, but which is in fact innocent, the actor may start another conflict. So, in essence, defence through cyber deterrence involves forestalling a would-be aggressor by acting aggressively first, while running the risk of starting new conflicts along the way.

[27] MITRE, 'Enterprise tactics', [n.d.]; and Infosec Institute, 'Red team assessment phases: establishing foothold and maintaining presence', 13 Dec. 2018.

Obfuscatory and imitative practices

As the ecology of cyberspace has developed, certain competitive practices have become more common in the performance of offensive operations. Examples include using resources and infrastructures (such as command and control—C2—systems) for offensive operations, in some cases deploying these in foreign countries so as to obfuscate attribution for the operations; and imitating the behaviours of other actors, also to avoid attribution or even detection. Two examples of these developments are the cyberattack on the ICT infrastructure of the 2018 Winter Olympics in Seoul, where substantial efforts were made to shift the blame to other parties,[28] and the cyberespionage campaign by the (allegedly Russian) Turla group on (mostly) Middle Eastern targets through the hijacking of attack infrastructure set up by (allegedly Iranian) hacking groups.[29]

These obfuscatory and imitative practices in particular make it harder to gain an overview of the 'theatre' and to keep track of how many parties there are in a conflict, let alone *who* they are. The uncertainty introduced in this way may hamper decision makers in initializing their official stance, and open up time and space for other actors to push agendas that suit their own interests.

Securitization of ICTs

In response to the threats described above and other developments, ICTs are becoming increasingly securitized[30]—as evidenced by recent security policy discussions on, among other things, 5G telecommunications,[31] the semiconductor and computer chip industry,[32] encryption,[33] jurisdiction over information stored in cloud services,[34] the 'internet of things' (IOT) and cyberphysical systems,[35]

[28] Greenberg, A., 'The untold story of the 2018 Olympics cyberattack, the most deceptive hack in history', *Wired Magazine*, 17 Oct. 2019.

[29] National Cyber Security Centre (UK) and National Security Agency (USA), 'Advisory: Turla group exploits Iran APT to expand coverage of victims', 21 Oct. 2019.

[30] 'Securitization' is a concept coined by Ole Wæver, Professor of International Relations at the Department of Political Science, University of Copenhagen. It describes a process where something is transformed by an actor into a matter of security in order to allow for the use of extraordinary measures. See Wæver, O., 'Standard securitization and desecuritization', ed. D. Lipschutz, *On Security* (Columbia University Press: New York, 1995); and Muller, L., 'How to govern cyber security? The limits of the multi-stakeholder approach and the need to rethink public–private cooperation', ed. K. Friis and J. Ringsmose, *Conflict in Cyber Space: Theoretical, Strategic and Legal Perspectives* (Routledge: Abingdon, 2016), p. 117.

[31] See e.g. European Commission, NIS Cooperation Group (note 26).

[32] See e.g. Swanson, A., 'Trump blocks China-backed bid to buy US chip maker', *New York Times*, 13 Sep. 2017.

[33] See e.g. Lewis, J. A., Zheng, D. E. and Carter, W. A., *The Effect of Encryption on Lawful Access to Communications and Data*, Center for Strategic and International Studies (CISS) Report (CISS: Washington, DC, Feb. 2017).

[34] See e.g. Grigsby, A., 'The intelligence collection implications of the CLOUD Act', Council on Foreign Relations Blog, 12 Feb. 2018.

[35] See e.g. Bracy, J., 'Why securing IOT is a national-security imperative', International Association of Privacy Professionals, 24 Oct. 2016.

satellite infrastructure,[36] segmentation of the internet,[37] and standard setting.[38] The securitization of ICTs in turn shapes the lens through which developments and events relating to cybersecurity are interpreted, over time raising the stakes as these technologies and others become more prevalent.[39] Future developments, such as the growth of the IOT as 5G telecommunications networks are established, will probably add to the securitization process as the number of attack surfaces increases and low security standards in IOT devices will allow hijacking of most of the things we have around us every day.[40]

Escalatory potential of actions according to perception of the actor

The analytical framework for escalation and de-escalation described in chapter 2 is only one way of thinking about such matters. For instance, another view is that an aggressive action taken in response to an incident or attack may be considered de-escalatory if it is not *as* aggressive as would be expected or justified given who the responding party is and what that party has been subjected to. This view raises several questions: who gets to decide what response is reasonably expected or justified (the 'bar', as it were); how to know whether the actors involved agree on what constitutes the expected or justified response; and is the actual response really less aggressive than what would have been expected or justified? When perceptions differ as to where the bar is and where a particular action sits in relation to the bar, there is a risk that actors misunderstand actions intended to be de-escalatory (or at least not escalatory) or disagree that certain actions actually are de-escalatory. In this situation, an action taken by one actor with the intention of de-escalating the situation may not be perceived by other actors as de-escalatory at all, and instead results in further escalation or continued high tensions. This is a general problem that applies across most domains. Amid a deteriorated security environment, however, and given the other threat aspects described above, especially the obfuscatory and imitative practices, a misinterpreted action may have particularly escalatory consequences.

Bias and the presumption of antagonism

The common idea that cyberspace is a constant battlefield is strengthened by the security perspectives presented above, as well as other related aspects. While the security outlook as regards cyberspace certainly is bleak in some senses, human

[36] See e.g. Hitchens, T., 'NSC makes cyber security for space industry "top priority"', *Breaking Defense*, 23 Oct. 2019.

[37] See e.g. Doffman, Z., 'Putin now has Russia's internet kill switch to stop US cyberattacks', *Forbes Magazine*, 28 Oct. 2019.

[38] See e.g. Greenbaum, E., '5G, standard-setting, and national security', *Harvard Law School National Security Journal*, 3 July 2018.

[39] Nissenbaum, H., 'How computer security meets national security', *Ethics and Information Technology*, vol. 7, no. 2 (June 2005), p. 63; Hansen, L. and Nissenbaum, H., 'Digital disaster, cyber security and the Copenhagen School', *International Studies Quarterly*, vol. 53, no. 4 (Dec. 2009), pp. 1155–75; and Dunn Cavelty, M., *Cybersecurity and Threat Politics: US Efforts to Secure the Information Age* (Routledge: London, 2008), p. 2

[40] European Commission, NIS Cooperation Group (note 26), p. 29.

bias may make it appear even bleaker than is actually the case. At least four factors shape this bias: (*a*) cybersecurity's historical ties with the military and intelligence spheres; (*b*) misleading statistics and claims; (*c*) human psychology; and (*d*) dubious epistemological premises.

Cybersecurity's historical ties with the military and intelligence spheres

The field of cybersecurity has historically, and to a large extent, grown from the military and intelligence spheres.[41] These connections have shaped the field to be predominantly about the protection of information systems and networks, and the information assets contained within them, against *antagonistic* threats. Protection against other threats, such as natural disasters, is not always considered a proper part of the field. This shapes the world view of cyber experts, on the technical as well as the policy levels, who think of cyberspace as a place where more or less clandestine operations are constantly being carried out.

This world view is further bolstered by the range of products offered to regular users to secure their systems and information. Anti-virus software, firewalls, phishing detection and other products are exclusively or primarily focused on protecting against antagonistic threats, supporting the tacit assumption that these kinds of threats are the only or primary threats that exist.

Misleading claims

The spread of (sometimes) misleading claims about the number of attacks happening over time reinforces the view that cyberspace is a constant battleground. Such claims may for instance relate to the number or severity of attacks being carried out against actors in a certain sector, while cyber incidents caused by other factors are ignored or downplayed. It is not always clear whether claims are based on actual statistics (which may be checked) or are mere estimates. Nor is it always clear what definition of 'attack' is being used; that is, whether there is any differentiation between intentional and unintentional behaviour (so that mistakes are not counted as attacks), or whether incidents not severe enough to count as aggressive are being excluded. For instance, when Statistics Sweden (the Swedish statistics agency) studied ICT usage in enterprises nationwide in Sweden in 2019, it found that, in the previous year, of enterprises with 10 or more employees, only 2 per cent had an incident (not necessarily caused by an attack) where confidential data had been disclosed; 8 per cent had an incident in which data was destroyed or corrupted; and 33 per cent experienced disruption to ICT services at least once.[42] While there will certainly be *some* underreporting of incidents, it is hard to square these results with claims to the effect that 'a business will fall victim to a ransomware attack every 14 seconds by 2019, and every 11 seconds by 2021' and that cybercrime globally will cost organizations and businesses $6 trillion.[43] To put these numbers in perspective, the Central Intelligence Agency's World

[41] For a brief overview, see Murphey, D., 'A history of information security', IFSEC Global, 27 June 2019.

[42] Statistics Sweden, 'Digitalisation and security in Swedish enterprises', 20 Nov. 2019.

[43] Morgan, S., *2019 Official Annual Cybercrime Report* (Cybersecurity Ventures and Herjavec Group: Toronto, 2018).

Factbook estimates that the total world gross domestic product (GDP) in 2017, measured in purchasing power parity, was $127.8 trillion.[44] Furthermore, studies of actual costs due to cyber incidents have concluded estimates that are difficult to reconcile with such claims. For instance, the US Council of Economic Advisers estimated that 'malicious cyber activity' cost the US economy $57–109 billion in 2016.[45] A systematic review of 12 000 cyber incidents that were recorded between 2004 and 2015 estimated that the annual cost to the US economy resulting from 'data breaches, security incidents, privacy violations and phishing/skimming' amounted to an average of only $8.5 billion.[46]

Human psychology

Humans tend to focus on the dramatic and the sensational, and to look for antagonistic actors behind events. It is not surprising then, that cyberattacks tend to receive a lot of attention when they come into public view, or that they may be remembered for a long time. The same goes for cyber incidents where something has happened that allows antagonistic actors to somehow gain an advantage or access to some sensitive system or piece of information.

Two recent examples in Sweden are the major cyber incidents at the Swedish Transport Agency (SJ)[47] and Svenska Kraftnät (the Swedish national electricity grid operator).[48] In the case of SJ, access to the national road traffic database, including some sensitive military and intelligence-related information, was given to a company whose staff did not have the proper clearances. No evidence was ever publicly produced of misuse or improper accessing of the information by that company. In the case of Svenska Kraftnät, access to a sensitive control system was given to a company whose staff did not have the proper clearances either. In both cases, a major theme in the resulting debate and investigations had to do with the *potential for antagonistic activities* that had arisen, rather than what happened in and of itself.

The same phenomenon applies to discussions around ICT supply chain security. The idea is often floated that some makers of components, or their governments, would *intentionally* place vulnerabilities ('back doors') in the components in order to exploit these vulnerabilities for geopolitical, technopolitical or other ends. It is true that ICT supply chain attacks happen, and that they are increasing (see previous subsection). However, there is a difference between adding a threat (something that causes an incident) into some component or software, and 'adding' a vulnerability (a lack of protection against some threat, so, essentially, failing to add some kind of protection). While there tend to be a lot of vulnerabilities in

[44] Central Intelligence Agency (CIA), *The World Factbook 2019* (CIA: Washington, DC, 2019).

[45] White House, Council of Economic Advisers (CEA), *The Cost of Malicious Cyber Activity to the US Economy*, CEA Report (White House: Washington, DC, Feb. 2018).

[46] Romanosky, S., 'Examining the costs and causes of cyber incidents', *Journal of Cybersecurity*, vol. 2, no. 2 (2016), pp. 121–35.

[47] 'Transportstyrelsens IT-upphandling' [The Swedish Transport Agency IT procurement], Wikipedia, 13 Jan. 2020.

[48] For a collection of articles on the many aspects of the incident, see 'Svenska kraftnät' [Swedish power grid], *Dagens Nyheter*.

ICT components,[49] it is often nearly impossible to know whether they were put there intentionally; it is, generally speaking, equally probable, if not more so, that any given vulnerability is the result of poor design or inadequate testing.[50] Even if some vulnerabilities have been intentionally placed into some component, the vast majority of vulnerabilities found in ICTs are there because of mistakes, lack of security focus or an inability to predict ways in which a particular set-up may be exploited.[51] But such nuances do not seem to generate the same amount of attention.

Dubious epistemological premises

With its historical ties to the intelligence community, hacking is sometimes thought of in ways that resemble romanticized ideas about spycraft, especially in the sense that there is an expectation that there are certain actors who are able to conduct operations and then disappear without a trace.[52] The idea that hacking organizations with such capabilities exist (which, to a limited extent, they do),[53] and are very active, may lead to the mistake of thinking that a lack of evidence of an attack (given the supposition that the organization that lacks this evidence has proper active detection and logging capabilities) may be an indication of the skill the attacker, rather than the obvious explanation that there was no attack.[54] This idea is commonly expressed through a dictum that system operators have to assume that the outer boundaries of their networks have been breached—even if they cannot see any signs of a breach.[55]

The presumption of antagonism

The above biases, individually or in combination, may lead to a *presumption of antagonism* when a cyber incident occurs.

This presumption of antagonism may have come into play in the aftermath of the Häglared incident discussed in chapter 1. Since it was established early on that the mast had been sabotaged, it became all the easier to think of the subsequent incidents as probable attacks. The character of the incidents themselves possibly strengthened this presumption. Many of the incidents were visible and relatable, and involved disruptions in systems that would be essential during a major crisis or even war: emergency call services, air traffic control and rail transport. It was

[49] See e.g., Stubbs, J. and Bryan-Low, C., 'Britain rebukes Huawei over security failings, discloses more flaws', Reuters, 28 Mar. 2018.

[50] Thomas, S. L. and Francillon, A., 'Backdoors: definition, deniability and detection', eds M. Bailey et al., *Research in Attacks, Intrusions, and Defenses: 21st International Symposium, RAID 2018, Heraklion, Crete, Greece, September 10–12, 2018, Proceedings*, Security and Cryptology series, vol. 11050. (Springer: Cham, 2018); and Decloedt, H. E. and van Heerden, R., 'Rootkits, Trojans, backdoors and new developments', *Proceedings of the Workshop on ICT Uses in Warfare and the Safeguarding of Peace 2010 Forever Resort, Bela Bela 11 October 2010*, pp. 4–11.

[51] British National Cyber-Security Center, 'Understanding vulnerabilities', 14 Oct. 2015.

[52] See e.g. Netsurion, 'The Assume Breach paradigm', EventTracker.com, 20 Jan. 2016.

[53] See e.g. Poulsen, K., 'Russian cyber unit that went dark after hacking DNC is still spying', *Daily Beast*, 17 Oct. 2019.

[54] See e.g. 'The absence of evidence in breaches', Rapid7 Blog, 20 Aug. 2015; and '5 ways hackers can breach your company undetected', InfoSec, 22 Feb. 2018.

[55] See e.g. Pompon, R., 'Living in an Assume Breach world', Help Net Security, 24 Aug. 2017.

only natural to assume that a hostile foreign power would attempt to disrupt such services in the run-up to a conflict.

However, the evidence on the causes of cyber incidents does not justify such a presumption in terms of service disruptions. For instance, in the Council on Foreign Relations Cyber Operations Tracker (a database on state-sponsored cyberattacks), the vast majority of the incidents listed are espionage operations.[56] About 20 per cent of the ICT incidents reported to the MSB are categorized by the reporting party as an attack; the other 80 per cent have other causes.[57] When the MSB analysed those reports that were categorized as attacks, the agency found that the vast majority were profit-driven (such as ransomware attacks, attempts at gaining access to bank accounts, or coercing staff to transfer money). For the most part, those attacks failed in the sense that affected parties refused to comply, even though they did cause fairly limited disruptions to some services or workflows.[58] Among the rest of the reports, about 50 per cent were caused by human mistakes or system errors. Further, the incidents caused by human mistakes and system errors frequently had more severe consequences.[59]

Not only does the presumption of antagonism lead to a tendency to quickly hypothesize that an attack is the cause of a cyber incident, it may lead to a reluctance to accept other hypotheses. The biases behind this presumption function as a cognitive vulnerability that may result in the spread of false narratives, some of which may be resilient to attempts to debunk them. This threat has a clear escalatory potential.

Cyber-physical or cyber-social interdependencies

As societies digitalize, the potential for cyber incidents to have physical effects and cause physical incidents increases. For example, a cyber incident that affects digitalized industries may trigger industrial incidents in high-impact processes, which in turn may cause further incidents. At least four related factors contribute to the escalatory potential of interdependent systems: (*a*) ICT dependence on electricity, cooling systems and data flows; (*b*) dependence of essential services and industrial processes on ICTs; (*c*) the complexity of interdependent relationships; and (*d*) siloed approaches to incident management.

ICT dependence on electricity, cooling systems and data flows

To function, information systems and networks need electricity and cooling. Electricity used to be the sole inflow on which ICTs depended. The higher levels of cooling that many ICTs now need means that many organizations rent cooling services (often delivered as a coolant directly into data centres), making ICTs

[56] Council on Foreign Relations, 'Cyber operations tracker', [n.d.].

[57] Holmgren, A., Head of Department for Cybersecurity and Secure Communications, MSB, 'Hoten mot den digitala förvaltningen' [The threats to digital management], Speech at eFörvaltningsdagarna conference, Stockholm, 9–10 Oct. 2019.

[58] Holmgren (note 57).

[59] Holmgren (note 57).

dependent on a constant inflow in that regard as well. Most ICTs also depend on data flows—whether input from human operators or from other information systems or networks—to function as intended. Every service that depends on ICTs (the vast majority) will experience disruption and sometimes damage if there is a cyber incident affecting the inflow of electricity, coolant or data. Since many ICTs depend on the same sources of electricity, coolant and networks, and in some cases the same data, a single cyber incident that damages or disrupts one or more of these inflows may have cascading and far-reaching consequences.

Dependence of essential services and industrial processes on ICTs

Modern essential services and industrial processes depend on a number of ICT systems. A cyber incident that disrupts an ICT system on which an essential service or industrial process depends may lead to one or more physical incidents.

For example, many municipal water supply systems rely on an industrial control system (ICS) to control the filtering of drinking water in a water plant. A cyber incident that disrupts or damages the ICS—whether caused by bad original coding, malware or a bad software patch—may cause the ICS to allow contaminated water to flow into the municipal water supply system. If the problem is not discovered and handled quickly, people may drink contaminated water from their taps and become sick or even die. As a further consequence, businesses and services that require clean water, such as restaurants, hospitals and care homes, may have to close or evacuate.

Complexity of interdependent relationships

The above dependencies are rarely simple one-to-one relationships, but complex, interdependent sets of relationships. It is a significant challenge to understand these interdependencies and map out how changes in one system in one organization effect changes in the organization's other systems and also in external systems, which in turn cause even further changes in other systems. While risk analysis within an organization may enable the organization to identify, monitor and address the effects of a cyber incident within its own systems, performing these tasks for downstream effects in external systems is beyond the capabilities of most organizations. It may be possible for the organization that suffers the cyber incident in the above example to identify how the cyber incident may cause some of the effects it brings about beforehand (in a risk analysis, for instance). But it may not necessarily be as easy for organizations downstream to understand how cyber incidents upstream may disrupt their services. Even if they understand the potential for upstream disruption to their services, there are likely to be many different kinds of events that could cause the same kind of disruption, which means identifying the source of a disruption will still be difficult.

To continue the above example, the municipal water supply system could have been sabotaged, or there could have been a leak somewhere that allowed poisonous materials to flow into the system. Thus, if the first incident to be noticed is that people become sick, then medical services will need to analyse people's medical conditions, formulate the hypothesis that contaminated water may be the

cause and communicate this to the utility which runs the municipal water supply system. That utility will then need to confirm that the water is contaminated, and formulate a number of hypotheses on what the possible cause may be (including that the filtering systems are not working), and then the utility will need to investigate these hypotheses and find that, indeed, the filtering systems have failed. Clearly, this may be a time-consuming process.

Siloed approaches to incident management

Managing concurrent developments in complex, interdependent systems requires sophisticated coordination across incident management teams in several organizations. If incident management teams are siloed—that is, not communicating and coordinating with each other—the cascade effects of one cyber incident have much greater potential for escalation.

Again using the water supply example, how the water contamination should be handled will depend on how long it may take to fix the cyber incident in the utility's ICS, and whether hospitals and care homes need to be evacuated will depend on the time it takes to decontaminate the water supply once the ICS is functioning again. If siloed cyber-incident management teams and physical-incident management teams fail to cooperate and coordinate with each other, or fail to communicate with a broader set of organizations (in the example, restaurants, shops and other businesses that depend on readily available clean water), the risk of escalation increases. By failing to mitigate the respective incidents as quickly as possible, the incident organization, affected organizations and supporting organizations may invite criticism for lack of preparedness, incompetence and allowing avoidable damages to be incurred.

Escalatory potential of complex, interdependent systems

As the water supply example shows, the complex ICT interdependencies that exist in digitalized societies mean that the potential for cyber incidents to cause real harm has reached an unprecedented level. The many challenges involved in monitoring systems for cyber incidents, and investigating and coordinating responses to a cyber incident, may mean that even if a cyber incident is resolved quickly, its effects might take longer to resolve, and escalation continues or increases.

The public's, as well as other actors', patience can be limited when faced with long disruptions to essential and other services, justifiably so when the response to a cyber incident is less than stellar. One recent event that illustrates this kind of escalation is the dysfunction in the ordering systems at the Swedish pharmacy and medical goods supplier Apotekstjänst in 2019, which caused severe materials shortages in vital medical services in five Swedish regions. The ICT system used for ordering medical goods proved badly designed from the start and when the problems grew, the company 'violated' its own procedures to fix the problem—but

made it worse instead.[60] The shortages resulted in operations being cancelled, a barrage of social media complaints, and the media describing Apotekstjänst as a threat to national security.[61]

Another example of interdependencies causing serious physical harm is the series of failures that led to the two Boeing 737 Max aircraft crashes in late 2018 and early 2019, killing all people on board. It was only after the second fatal crash happened in similar circumstances that all 737 Max planes were grounded pending investigation, despite the cause of the first crash still being determined and early signs indicating faulty sensor readings were to blame. The fallout from these incidents is ongoing.[62]

The potential for these interdependencies to result in physical harm means that, in the event of a cyberattack that actually managed to damage an information system or a network which maintains (for instance) an essential service, the attack could legitimately meet the threshold for starting an armed conflict according to international humanitarian law. The probability that such attacks will occur is higher than before—and increasing the more that societies digitalize.

The complexity of the public communications arena

Following a major cyber incident, information about the incident must be communicated to affected actors and the wider public in a way that does not escalate the situation. The incident organization and supporting organizations will need to communicate that there has been an incident, how they are handling the investigation and any consequences from the incident, and the cause of the incident (once known). In many cases, other actors—supervising organizations, law enforcement agencies, governments—will also need to communicate with the public on the actions that they are taking (or not taking). And depending on the kind of cyber incident that has occurred, the media and a number of attributors and commentators will also be communicating with the public about the incident, while assailants and other antagonists may also make public statements. With so many voices shaping the discourse, there will be conflicting messages and it will be hard for any one actor to control the narrative or for actors to ensure there is a coordinated narrative, or for actors to know which messages to trust.

At least four factors drive escalation dynamics in this area: (*a*) high expectations of, but low trust in, public institutions; (*b*) the difficulty of getting communication right; (*c*) the number of actors joining the conversation; and (*d*) dynamics between different types of actor.

[60] Pettersson, M. G., 'Mejlen avslöjar: Apotekstjänst förvärrade läget genom att "våldföra" sig på it-system' [Email reveals: Pharmacy service exacerbated the situation by 'forcing' itself on IT systems], *SVT Nyheter*, 22 Oct. 2019.

[61] Ericson, P., 'Apotekstjänst är ett säkerhetshot' [Pharmacy service is a security threat], *Aftonbladet*, 23 Oct. 2019.

[62] 'System failure: the Boeing crashes', Al Jazeera, 16 Oct. 2019; 'Boeing 737 Max Lion Air crash caused by series of failures', BBC, 25 Oct. 2019; and Liebermann, O., 'Investigators spread blame in Lion Air crash, but mostly fault Boeing and FAA', CNN, 26 Oct. 2019.

High expectations of, but low trust in, public institutions

In today's high-speed communications environment, the public expects to be informed quickly about any issue of public interest generally and about issues that affect them directly, by both the media and the public institution responsible for the issue. Service users also expect swift notification from service providers, including those that provide public services. At the same time, trust in public institutions is lower in some groups.[63] Overall, the generally high levels of trust Swedes have in the state and in public institutions[64] hinge to a large extent on expectations of competence, fairness, transparency and efficiency.[65] However, when institutions and organizations fail to live up to those expectations, trust may quickly be eroded and escalation may ensue.[66]

The difficulty of getting communication right

Actors trying to manage the situation following a cyber incident will need to make decisions about when to communicate, what to include in that communication and what to hold back. Misjudging the timing or the content of the communication may lead to accusations of incompetence or of an attempted cover-up, especially if it appears the message has been delayed or that information is being withheld. But the same accusations apply if communications are made too early or contain too much information, especially if it later turns out that what was said was incorrect or premature. Even when messages about the incident have been communicated well, mistrust of the communicating actor—whether the incident organization or the authorities—may mean that its messages are met with disbelief, allowing any narrative that runs counter to the message to gain more credence.

The number of actors joining the conversation

In addition to the actors who must communicate about a cyber incident, a number of other actors will inevitably join the conversation and shape the narrative. Examples of possible attributors and commentators include private security companies, states and multilateral organizations expressing concerns or calling for attribution, as well as other organizations and individuals speculating about the incident, its causes and consequences, and who is responsible.[67] The media will be reporting on the incident itself but may also report on and give airtime to some of these attributors and commentators. Those publicly named as possible

[63] Edelman, *2019 Edelman Trust Barometer: Global Report* (Edelman: New York, Feb. 2019); and Ortiz, E. and Roser, M., 'Trust', Our World in Data, 2019.

[64] Chipperfield, N., 'Trust in democracy and institutions strong despite fears: Report', Radio Sweden, 27 June 2018.

[65] Statista, 'Share of people with very/rather large trust in the government in Sweden from 2008 to 2018', Statista survey report, July 2019.

[66] See e.g. Annergård, M., 'Invånarnas inställning till digital service i välfärden' [Residents' attitudes to digital service in welfare], Sveriges Kommuner och Regioner, 25 Oct. 2019; and Palm, L. and Falkheimer, J., *Företroende Kriser: Kommunitionsstrategier Före, Under och Efter* [Confidence Crisis: Communication Strategies Before, During and After] (MSB: Karlstad, 2005).

[67] For an interesting case, see Starks, T., 'Criticisms, questions surface over cybersecurity firm's report on Burisma hack', Politico Pro, 14 Jan 2020.

assailants will also want to spread counter narratives. As a result, the discourse will become shaped by speculation, rumour, false claims and influence operations. These narratives may restrict the set of available options for decision makers, as well as reduce the time they have for deliberation, and sometimes, possibly, even force their hands.[68] Poor decisions made in these circumstances may result in new incidents and further escalation.

Dynamics between different types of actor

An incident organization may have various incentives to withhold information about a cyber incident from other actors, such as fear of reputational damage or unwillingness to incur regulatory sanctions or attention from law enforcement agencies. This dynamic may lead to further, avoidable consequences for other actors. For instance, an ICT service provider may know of a severe vulnerability in its own systems that has resulted in exposure to unauthorized parties of sensitive data belonging to its customers, which may be private organizations or public service agencies. This sensitive data may in turn have been collected from the organization's customers or the agency's users (i.e. members of the public). If it is then revealed that the data has been exposed to and possibly accessed by unauthorized parties, then the service provider's customers may face strong public reactions and accusations for not having done their due diligence—and those customers may be entirely, and unnecessarily, unprepared.

[68] Huetteman (note 17).

4. Escalation vulnerabilities in the aftermath of a cyber incident

This chapter describes four broad vulnerabilities that indicate a lack of capacity to prevent or counter escalation. They are: (*a*) a lack of cognitive robustness in relation to claims about cyber incidents; (*b*) a lack of interactive response capabilities; (*c*) a lack of formal, complementary crisis response structures; and (*d*) a lack of security focus in digitalization and security maintenance in critical systems. In this context 'a lack' does not mean a complete absence of capacity but an insufficient level of capacity that could, and should, be developed. Some of the vulnerabilities listed below have not been well studied in the literature to establish definite conclusions; in those cases, statements about those vulnerabilities are necessarily qualified.

Lack of cognitive robustness in relation to claims about cyber incidents

Cybersecurity is often perceived to be a technical and secretive field, and thus one that is difficult to understand.[69] Most people therefore rely on experts, purported or actual, to understand cybersecurity issues. As society increasingly depends on ICTs, this dependence on experts also increases. Scrutiny of claims made about cybersecurity in general and a cyber incident in particular may therefore lack cognitive robustness, and farfetched claims about the threat posed by various cyber-related phenomena may stand unopposed. This lack of cognitive robustness is therefore closely tied to the threat areas of bias and presumption of antagonism and the complex arena for public communications (see chapter 3). This vulnerability can be exploited by gaps in four areas: (*a*) common language using well understood and defined concepts in the field of cybersecurity; (*b*) statistics and research on the magnitude of the problems in the field; (*c*) knowledge about cybersecurity issues; and (*d*) informed and critical perspectives.

Lack of a common language

The cybersecurity field uses a large number of concepts and terms, but few of those have commonly accepted and categorical definitions (there are, however, a few influential ones).[70] For instance, important terms such as 'threat', 'vulnerability', 'incident' and 'attack' all have different and varying definitions and thus apply to different concepts. The potential problem arising from lack of a common language is that ordinary words (such as 'attack') used by cybersecurity experts to refer

[69] Smith, A., 'What the public knows about cybersecurity', Pew Research Center, 22 Mar. 2017.

[70] Examples of influential definitions include the US National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC) definition of 'security incident': NIST CSRC, 'Security incident', [n.d.]; and the European Union NIS Directive definition of 'incident': NIS Directive (note 18), Art. 4(7). On the definitional debate around cyber-security, see Dunn Cavelty, M. 'Cyber security', ed. P. Burgess, *The Routledge Handbook of New Security Studies* (Routledge: New York, 2010), pp. 154–55.

to something fairly specific may be understood by non-experts in terms of their everyday meaning.

For instance, if cybersecurity experts use 'attack' to encompass everything from scanning ports in networks to see if there is any way into the network (something that in and of itself has no potential to cause an incident and probably happens frequently), to making changes within the network or exfiltrating information (something that either is or causes an incident and probably happens much less frequently), then it would be understandable if experts claim that attacks happen often. But if non-experts understand 'attack' to mean some sort of action that has a definite and intentional harmful impact, they will misunderstand what the experts are saying and be unnecessarily worried. This in turn may mean that people over time come to think of cyberspace as a battleground, and expect that incidents are caused by attacks.

If people have better knowledge of cybersecurity terminology and concepts, they can gain a more detailed understanding of how cybersecurity experts think about attacks, which means they will have a greater capacity to be critical when they come across expert commentary and speculation in relation to a cyber incident.

Lack of statistics and research

Ideally, organizations would develop their cybersecurity strategies on evidence-based and rational foundations. In this ideal setting, organizations would be able to access statistics and research relevant to their sector, size and workforce that tells them what factors lead to the most frequent and most severe incidents, which problems are easy to manage and which are hard, and so on. Correspondingly, governments would have the necessary data to see where state-led efforts could add the most value. However, while there is a lot of research being done on specific cybersecurity issues,[71] there is an overall lack of data to review on the prevalence and scope of the problems.[72] The lack of commonly accepted and categorical definitions (among other things) means that it is hard to combine research results into aggregated sets that show the overall picture of what problems there are, and how big or small those problems are compared to one another. It also means that there are few major knowledge bases that may be used to question certain statistics used for marketing purposes in the field. Furthermore, since the field is sometimes considered to only be about protection against antagonism, attempts at providing holistic overviews within the field tend to leave out issues that do not have anything to do with antagonism.

[71] See e.g. Khandelwal, S., 'New attack lets Android apps capture loudspeaker data without any permission', *Hacker News*, 17 July 2019; Bushwick, S., 'New encryption system protects data from quantum computers', *Scientific American*, 8 Oct. 2019; and Romanosky, S., 'Examining The costs and causes of cyber incidents', *Journal of Cybersecurity*, vol. 2, no. 2 (2016).

[72] White House, Council of Economic Advisers (note 45); and Romanosky (note 71).

Lack of knowledge about cybersecurity issues

The lack of a commonly accepted and categorical language, combined with few resources on statistics or research on the challenges posed by cyber issues, as well as other aspects, means that it is hard to educate the public, decision makers and others on the issues according to their respective needs. When overall knowledge about cybersecurity issues probably remains fairly low in most populations, it is hard to counter threats such as bias and presumption of antagonism, especially in the complex arena of public communications (see chapter 3).[73]

Lack of informed and critical perspectives

The three gaps described above together result in a lack of the foundations needed to develop critical perspectives on cybersecurity issues. In more mature fields such as epidemiology, climate science and financial risk management, which have groundwork that allows for strongly evidence-based policies and risk management, unfounded claims about the prevalence of some problem or the societal impact of some phenomenon can be evaluated on the basis of previous studies of those problems or phenomena, as well as already established facts and large data sets. Different practitioners in the same field will be able to evaluate the same claim and come to (roughly) the same conclusion about it. But the lack of this kind of groundwork in cybersecurity means that, in the context of a cyber incident, the resources needed to maintain a critical mindset in the face of misleading statistics and unfounded or false claims are lacking. Also lacking is the means to, at scale, overcome and counter the presumption of and bias for antagonism in cybersecurity.

Lack of interactive response capabilities

There is a lack of detection capabilities and interactive information sharing among organizations and the cybersecurity, intelligence and crisis management communities that could help organizations to determine when they may be at higher risk of cyberattack and various escalatory activities; understand why they have become a target for certain kinds of cyberattack and what to expect in terms of escalatory activities; and prepare for and counter such escalatory activities.[74] This vulnerability is closely tied to the threat areas of a deteriorated security environment and the complexity of the public communications arena. Some gaps relating to this vulnerability are: (*a*) lack of detection capabilities; (*b*) lack of proactive information sharing between organizations and the cybersecurity community; and (*c*) lack of synergistic uses of information flows.

[73] Smith (note 69).

[74] See e.g. Wåhlberg, K. and Limmergård, R., 'Industrin behöver bättre skydd mot cyberattacker' [Industry needs better protections against cyberattacks], Teknikföretagen, 11 Feb. 2019.

Lack of detection capabilities and awareness of target potential

An incident organization needs to be prepared to respond to a cyber incident in a way that does not escalate the situation. That means first that it must be able to detect a cyber incident. Cyber incidents range from those that are very easy to detect to those that are near impossible to detect. Even when an incident is detected, it may be hard to quickly determine whether or not it was caused by an intentional, antagonistic act (an attack). Even if it is quickly determined that an attack was the cause, the organization may find it hard to determine the impact, if any, that the attack will have on some ongoing process or development.

In addition, organizations may lack the awareness to understand why they would be made into targets and it may be very hard to determine why and how they may be drawn into escalating situations, especially beforehand. For instance, if a company is targeted because it is seen as a symbol of something that is deemed by many to be controversial, then the company might expect to be targeted more than once. Organizations may also lack awareness of how they may prepare for and respond to escalatory actions to which they may be subjected.

Lack of proactive information sharing

Correspondingly, incident organizations may fail to share information, or share insufficient information, with the cybersecurity and intelligence communities about cyber incidents they have suffered (beyond those incidents required by law to be reported to regulatory bodies and law enforcement agencies). One possible reason for this lack of information sharing is that an incident organization may deem the information vulnerable to exploitation by various parties for escalatory purposes. If information sharing about cyber incidents was more widespread, other, possibly affected organizations could receive advance warning and take appropriate action.

Lack of synergistic uses of information flows

There is also potential to make better use of already existing incident reporting schemes to share the information among the cybersecurity, intelligence and crisis management communities and deliver value back to reporting parties. Currently, incident reports are sent to various agencies, but sharing that information between agencies may be difficult or even prohibited. However, if various already existing information flows (such as incident reporting schemes) were shared between agencies and other organizations, this information could be used to detect, predict, avert and mitigate cyberattacks. For instance, instead of an incident report being viewed in isolation, it could be viewed in the context of other incident reports and other kinds of information. From these various sources, patterns can be detected, such as a particular kind of cyber campaign being in the works, and the kinds of organizations being targeted. In this way, cyber campaigns may be hampered enough to become unsuccessful, which in turn may make it more difficult for any escalatory purpose to succeed.

Lack of formal, complementary crisis response structures

The geographical orientation of Sweden's emergency and crisis response system means it is optimized to handle 'local' disruptions in essential services and various incidents in the physical domain (such as fires, floods and storms).[75] Unlike physical incidents, which tend to be localized to a particular region, cyber incidents are not limited to a physical area. Cyber incidents tend to occur across sectors or among actors that use a particular kind of ICT service, or in a particular kind of ICT system, or in systems that employ a particular kind of software or that depend on a particular kind of ICT supply chain. Consequently, incidents caused by the spread of malware or a dysfunctional patch may appear near simultaneously across several regions in the country. For example, during the 2017 WannaCry attack, several medical facilities across the UK were hit almost simultaneously by ransomware (discussed further in chapter 5).

Currently, there is a lack of formal, complementary and cross-cutting structures that allow for coordinated cyber-incident management all over the country.[76] As a consequence, malicious code can spread to more devices, and incidents can last longer or lead to more damage, than would have been the case if such structures existed. This vulnerability is closely connected to the threat area of cyber-physical or cyber-social interdependencies.

Lack of security focus in digitalization and security maintenance in critical systems

Cyber incident-related escalation will, for the most part, only happen after a cyber incident has occurred. Unfortunately, many organizations do not make sure that their cybersecurity efforts are in lockstep with their digitalization efforts, and so the risk that a cyber incident will occur is higher than necessary. This vulnerability is closely connected to the threat area of cyber-physical or cyber-social interdependencies. Some gaps related to this vulnerability are: (*a*) lack of cybersecurity focus in digitalization efforts; (*b*) lack of cybersecurity focus in the maintenance of existing infrastructures; and (*c*) lack of training in handling serious cyber incidents. The following discussion applies to the Swedish case, but may well be generalizable to other states.

Lack of cybersecurity focus in digitalization efforts

Organizations tend to focus their digitalization efforts on creating digital solutions for new and existing services. Digitalization projects are driven by a combination of factors: to simplify services, to make services more efficient or accessible, and to save money. Cybersecurity is rarely prioritized from the outset, and services are

[75] For a comprehensive description of how the Swedish emergency and crisis response system is set up, see MSB, 'Gemensamma grunder för ledning och samverkan vid samhällsstörningar' [Common foundations for management and cooperation in the event of social disturbances], 9 May 2019.

[76] Sweden's emergency and crisis response system has a number of compensatory arrangements to handle problems such as those described here, but the geographical orientation and the lines of communication it facilitates are still at the core of how the system is meant to work.

rarely developed with a focus on making their functions secure.[77] The same may be said about IOT devices, which rarely come with ambitious built-in security features or security updates.[78]

Lack of cybersecurity focus in the maintenance of existing infrastructures

Organizations often lag behind on upgrading existing infrastructure to the extent needed for those organizations, as well as those parts of society that depend on the infrastructure in question, to reach a proper level of security and continuity.[79] This leaves systems unprotected against system errors and exploits of vulnerabilities that had not been discovered or fixed at the time of the release of the hardware and software versions on which the systems run. The broad failure to update, upgrade or remove unprotected systems was realized when the WannaCry ransomware attack spread globally in 2016, using the EternalBlue exploit that targeted certain vulnerabilities in outdated versions of Windows XP.[80]

Lack of training in handling serious cyber incidents

Organizations need not only to maintain their critical ICT systems but also to practise shutting down and restarting systems in the event of a serious cyber incident. If an organization's ICT systems support services that are critical, or the organization will be subject to regulatory sanctions and potentially fines when those services are not functioning, the organization may be tempted to not perform full shut-down exercises and other drills for major incidents, for fear of not getting the services back up, in time or at all.[81] As a result, organizations might not sufficiently train their incident response and mitigation teams to handle the most severe cyber incidents.

This means that organizations often are ill-equipped to quickly recover from serious incidents, and that, as a consequence, major incidents may last longer or have more severe consequences than they would otherwise have had, with possible further incidents and consequences.

[77] See Spiezia, V. et al., 'Digital security policy', *OECD Reviews of Digital Transformation: Going Digital in Sweden* (OECD Publishing: Paris, 2018), p. 122.

[78] Maples, C., 'Security and privacy in the internet of things', *Journal of Cyber Policy*, vol. 2, no. 2 (2017).

[79] See e.g. Swedish National Audit Office (SNAO), *Föråldrade it-system: hinder för en effektiv digitalisering* [Obsolescent IT systems: an obstacle to effective digitialization], Report RIR 2019:28 (SNAO: Stockholm, 1 Oct. 2019), pp. 20–28.

[80] Newman, L. H., 'The leaked NSA spy tool that hacked the world', *Wired*, 3 July 2018.

[81] See e.g. Swedish National Audit Office (note 79), pp. 20–28.

5. Lessons from past cyber incidents and country studies

As a way to explore possible elements of best practices for cyber-incident management and non-escalation as well as *de*-escalation, this chapter presents nine case studies of past cyber incidents that occurred in countries which share some similarities or connections with Sweden in terms of geography, demographics or state of digitalization, namely Estonia, Finland, Japan, Singapore, South Korea and the United Kingdom. Eight of the nine cases investigated were cyber incidents caused by malicious activities and one case was an incident caused by system malfunction (see box 1). This chapter also examines how these countries addressed the issue of escalation and the goal of de-escalation in their national policies and practices. While each incident was unique in the way it unfolded and was managed, and despite the large diversity in national cyber-incident management practices, five lessons can be learned from these case studies.[82]

Lesson 1. The importance of prior readiness

Prevention and preparation are crucial in limiting the impact of a cyber incident and in improving the efficiency of incident response. Specifically, it is critical to have appropriate cybersecurity awareness and a predefined response plan that covers multidisciplinary engagement and indicates what to do when additional resources are required.

Cybersecurity awareness and cyber hygiene

Many of the cyber incidents could have been stopped earlier, or even avoided entirely, if users and system administrators had behaved differently. In the WannaCry case, damage would have been largely avoided if security patches had been applied in time.[83] In the Japan Pension Service (JPS) case, an employee from the JPS opened an email containing a virus which infected the employee's computer.[84] Staff training in cyber hygiene would have prevented the cyber incidents from occurring or reduced incident duration.

Moreover, it is important to improve the kind of cybersecurity awareness that helps in detecting the signs of a cyberattack, understanding the potential security risks and reporting the incident in a timely fashion. In the Singapore case, the team at SingHealth responsible for both reporting and initial response did not realize the significance of the security issues they were facing. Their indecisiveness about

[82] The focus in this chapter on presenting the five lessons identified from analysis of these case studies does not imply that these lessons are entirely new for Sweden, nor that they are *complete*; that is, it will always be possible to improve the lessons by expanding on or adding to them.

[83] British House of Commons, Committee of Public Accounts, *Cyberattack on the NHS*, Report no. 32 of Session 2017–19 (British House of Commons: London, 18 Apr. 2018), p. 6.

[84] Nippon Telegraph and Telephone Corporation (NTT), *2016 Annual Cyber Security Report* (NTT: Tokyo, 2016), pp. 59–60.

whether to report the incident to higher authorities delayed the incident response and led to a larger data breach.[85]

Organizational readiness

In each case analysed, the cyber incident had implications that went beyond the immediately affected parties. Many of the incidents were handled in concert with multiple actors from different sectors. As a result, the handling of the incident came to be, at least in part, elevated from the organizational level to the national or international level. In several of these, a striking observation is the ad hoc nature in which this process of elevation was handled. The lack of organizational readiness and dependency on pure contingencies led to incoherent signalling and responses.

During the WannaCry ransomware attack in 2017, the impact on the hospital systems of the British National Health Service (NHS) was so huge that incident response was both raised to the ministerial level for coordination and broadened to involve police departments and outsourced ICT support.[86] Furthermore, these authorities followed the protocol for handling major national incidents, despite the differences in dealing with cyber incidents compared to other major incidents, especially in the phases of threat reporting and incident response.[87] As a cross-border cyberattack, the response involved not only national agencies, but also warranted international cooperation. It provided the first test for information sharing and coordination on cyber incidents among the EU member states, through the Computer Security Incident Response Team (CSIRT) network under the NIS Directive.[88]

In the case of SingHealth, there were also several problems with the incident reporting process. Although there were guidelines on how to conduct incident reports, SingHealth staff were not aware of the existence of these documents. The timing for raising the incident with higher authority was another issue. While investigative results showed that elevating the incident to higher authority was necessary, staff opinions varied as to whether that was the case. There was no clear guidance on how or when to report to higher authorities.[89]

[85] Singaporean Government, Ministry of Communications and Information, *Public Report of the Committee of Inquiry into the Cyber Attack on Singapore Health Services Private Limited's Patient Database on or around 27 June 2018*, 10 Jan. 2019, pp. 96, 101 and 137.

[86] British Government, Department of Health, *Investigation: WannaCry Cyber Attack and the NHS*, Report (National Audit Office: London, 25 Apr. 2018), p. 9.

[87] Smart, W., *Lessons Learned Review of the WannaCry Ransomware Cyber Attack* (British Government, Department of Health and Social Care: London, 1 Feb. 2018), p. 31.

[88] Council of the European Union, Note on cybersecurity from the General Secretariat, no. 9621/17, 31 May 2017.

[89] Singaporean Government (note 85), p. 316.

Box 1. Cases of past cyber incidents

2007 Estonian cyberattack
Several Estonian government and private websites were disabled by a series of distributed denial-of-service (DDOS) attacks, amid political tensions between Estonia and Russia. This event is considered to be a landmark for the country in handling cyber incidents and has led the trend of securitization of ICTs at the international level.[a]

2013 Finnish spyware incident
The internal network of the Finnish Ministry of Foreign Affairs was infiltrated. Although no valuable information was compromised, the cause was established to be an espionage activity with political motivations.[b]

2013 South Korean cyberattack
The websites of South Korean broadcasters and banks were attacked, causing economic loss. A similar event happened three months later, in which several South Korean government and media websites were attacked. The later event happened before the anniversary of the start of the Korean War, and it was suspected that political motives were behind the attack.[c]

2014 Seoul metro system hacking
The computer workstations and servers of lines 1 and 4 of the Seoul metro were compromised for five months in 2014. No report was made until one year later and very limited information about the incident is available.[d]

2015 Japan Pension Service data breach
A cyberattack resulted in 1.25 million cases of personal data being leaked outside the Japan Pension Service. The data breach had significant implications for the organization's national cybersecurity strategy and induced internal reforms.[e]

2017 WannaCry attack
WannaCry was a worldwide ransomware cyberattack that not only required organizational and state-level emergency responses, but also information sharing and communication at European Union and international levels. This case study focuses on the WannaCry attack on the National Health Service (NHS) of the United Kingdom.[f]

2017 security flaws in Estonian ID cards
A vulnerability in digital ID cards was overlooked by the government and posed potential risks to the 2017 election in Estonia. The later replacement of the ID cards led to a loss of public trust and damage to the national reputation.[g]

2018 SingHealth data breach
The patient database of Singapore's largest group of private healthcare institutions, SingHealth, was breached in a cyberattack. Among the patient records released online as a result of the breach, it was the Singaporean prime minister's personal and medical data that was specifically targeted.[h]

2018 system malfunction at the Tokyo Stock Exchange
This case presents, for reference, an example of how a cyber incident due to system errors (rather than cyberattack) was handled efficiently at an organizational level.[i]

[a] Ottis, R., 'Analysis of the 2007 cyber attacks against Estonia from the information warfare perspective', Paper presented at the Seventh European Conference on Information Warfare and Security, Plymouth, UK, 30 June–1 July 2008.

[b] Haapala, T., 'MTV3: Suomen ulkoministeriö laajan verkkovakoilun kohteena vuosia' [The Ministry for Foreign Affairs of Finland has been subject to extensive cyber-espionage for many years], MTV3, 31 Oct. 2013; and 'Tuomioja: Cyber espionage embarrassing for Finland and MFA', YLE, 31 Oct. 2013.

[c] Branigan, T., 'South Korea on alert for cyberattacks after major network goes down', *The Guardian*, 20 Mar. 2013; 'Cyber attack hits South Korea websites', BBC, 25 June 2013.

[d] Yi, W., 'NK denies allegations on Seoul subway hacking', *Korea Times*, 8 Oct. 2015; AFP, 'North Korea suspected of hacking Seoul's subway operator last year', *Straits Times*, 5 Oct. 2015; and Hyun-jeong, L., 'Seoul subway server allegedly hacked by NK', *Korea Herald*, 5 Oct. 2015.

[e] '1.25 million affected by Japan Pension Service hack', *Japan Times*, 1 June 2015; and Otake, T., 'Japan pension administrator's lax response worsens data theft', Nikkei Asian Review, 2 June 2015.

[f] British House of Commons, Committee of Public Accounts, *Cyberattack on the NHS*, Report no. 32 of Session 2017–19 (House of Commons: London, 18 Apr. 2018); and British Government, Department of Health, *Investigation: WannaCry Cyber Attack and the NHS*, Report (National Audit Office: London, 25 Apr. 2018).

[g] Luis, P., 'ID-kaardi kiibis peitub teoreetiline turvarisk' [There is a theoretical security risk in the ID card chip], ERR, 5 Sep. 2017; 'Security flaw in Estonian national ID card', Schneier Blog, 5 Sep. 2017; and e-Estonia Briefing Centre, 'What we learned from the eID card security risk?', May 2018.

[h] Singaporean Government, Ministry of Communications and Information, *Public Report of the Committee of Inquiry into the Cyber Attack on Singapore Health Services Private Limited's Patient Database on or around 27 June 2018*, 10 Jan. 2019.

[i] Tokyo Stock Exchange, 'Regarding the equity trading system glitch occurring on October 9, 2018', Press release, 24 Oct. 2018.

To proactively prevent a cyber incident and to coordinate cyber-incident responses from different actors, the following measures need to be in place:

1. *An implemented framework for systematic and risk-based cybersecurity work*. Organizations need to use established standards and best practices to avoid, prevent, mitigate and recover from cyber incidents. To handle escalation scenarios, they particularly need to be trained in how to detect, report on and engage incident response functions to handle incidents in their early stages.

2. *A standard operating procedure for communication and coordination*. Organizations need to know how, when, where and to whom they should disseminate information, as well as what information to disseminate, and how, when, where and from whom they should expect to receive information.

3. *Clear reporting standards*. Organizations need clear, written guidelines that dictate how and when they need, or indeed are required, to report a cyber incident to higher authorities

4. *Capacity training*. Organizations need to ensure staff receive adequate education, training and practice in complying with existing provisions and protocols and to prepare for the unforeseen.

5. *Legislation and policy*. Governments need to support the above four measures with appropriate regulations, as well as provide support in terms of infrastructure and aid to enable organizations to conduct the training and practice that they need.

Lesson 2. The importance of coherent response coordination

Coherent response coordination among the involved parties is key to improving incident handling and to showing the public that the situation is under control.

In many cases, the parties involved lacked experience of major cyber incidents. The lack of preparedness for handling such incidents often led to incoherence in coordinating the responses among the supporting and affected parties. The result was that a number of more or less ad hoc incident management schemes ran concurrently, with little or dysfunctional coordination happening between them. This in turn led to confusion, as well as inconsistent and incoherent communication about what was going on and what had happened. It is not unlikely that this prolonged the overall incident handling process, as well as leading to an avoidable loss of trust by the public.

In the case of the British NHS WannaCry attack, it was not established who had the lead on coordinating the response until fairly late into the process. Moreover, the internal communications between the NHS and local authorities faced challenges that complicated collection of information, which reduced situational awareness. Meanwhile, affected hospitals faced an unnecessary burden when they needed to report similar information to different authorities.[90]

The contrast is significant compared to the reference example of the system malfunction at the Tokyo Stock Exchange (TSE) in 2018. Here, the incident organization was clear about whom it should inform, what it should check, what the solutions were and when to make information public. The immediate action taken by the TSE meant that the system failure was resolved quickly.

Achieving more coherent response coordination requires *a cooperation framework* that covers inter-agency, public–private and cross-border cooperation. The framework needs to be clear, easy to use and efficient. It needs to make clear who communicates about what, and what needs to be checked before organizations communicate. It needs to make sure that unnecessary redundancies, such as reporting the same thing multiple times, are eliminated or limited. It needs to delineate which inter-organizational or co-organizational activities should happen 'automatically', which ones require active decision making, and how that decision making should happen—including who has the authority to make each decision.

Lesson 3. The importance of communication and public perceptions

In several of the analysed incidents, communications played a major role in determining public perceptions of the incident, and how those perceptions would develop over time. It is critical to demonstrate to the public that the relevant parties are facilitating recovery efforts. During the data breach of the JPS, timely updates were made during the entire recovery process to provide affected people with assistance.

Moreover, if the incident organization does not share information regarding an incident that has no publicly visible impact, then information about the incident will often reach the public through a suboptimal channel. In the case of the Seoul Metro incident in 2014, while the investigation was being carried out by South Korea's National Intelligence Service, the incident became public in statements

[90] Smart (note 87), p. 32.

made by Saenuri Party representative Ha Tae-keung in October 2015. Another similar case is the 2013 Finnish spyware incident, where the infiltration of the internal network of the Finnish Ministry of Foreign Affairs was not reported until the news was leaked to the media. Not only did this mean that the incident organization, affected parties and supporting organizations missed their chance to present the incidents in non-escalatory ways, it also meant that those organizations were caught unprepared when the stories broke.

Leaving aside the issue of whether or when incidents that are not immediately visible to a wider audience should be publicized, there are also a number of lessons relating to how communications should be conducted when an incident *has* come into public view. When that has happened, it is important that communications are swift, signal ongoing action, only tell of things that are known with certainty, and are made often, but only as often as necessary to say something new each time. The information does not need to be comprehensive, but must be soundly based on established facts. Inconsistencies in communications may lead to considerable damage to reputation and trust. In the Estonian ID card case, the security flaws were identified in July 2017. A month and a half after the discovery, the Estonian Government announced the vulnerability, but stated that the ID cards were 'completely secure'. The Estonian Government subsequently backed away from this position and suspended 750 000 affected ID cards in November 2017.[91]

This can be contrasted with the successful handling of communications in the SingHealth case. In this incident response, the public announcement was coordinated with all involved parties, including the timing of making the announcement. Prior to the announcement, to avoid public panic, the information remained classified until the authorities understood what had happened and until the situation was contained.[92] It is, however, worth noting that the involved authorities took a fairly big risk when they chose to withhold the information until they were ready, since a leak could have caused serious damage to their approach and to public confidence.

Key factors that should shape post-incident communications are:

1. *Swift and fact-based communications*. From the moment the incident becomes known to an outside audience, communication should be conducted continuously, using active verbs (to signal that work is being done continuously) and presenting verified facts as they are found. The contents of communications should focus on what the target groups care about, and should be expressed in unambiguous and understandable terms.

2. *De-escalatory or non-escalatory messaging.* Messaging should be crafted to avoid escalation and address public concerns. If possible, it should pre-empt attempts by third parties to shape the narrative about what has happened.

[91] Luis, P., 'ID-kaardi kiibis peitub teoreetiline turvarisk' [There is a theoretical security risk in the ID card chip], *EER*, 5 Sep. 2017.

[92] Singaporean Government (note 85), p. 196.

3. *Preparedness*. Major incidents will be met with public reactions and media attention, as well as, possibly, attempts by outside influencers to shape the narrative, for instance by attributing the incident as an attack perpetrated by a named party. There need to be readymade answers to media questions, to counter false or unfounded claims.

4. *Division of labour*. There need to be clearly defined roles in terms of who communicates about what, so that different organizations do not undermine each other's efforts or cause confusion.

Lesson 4. The importance of considering the consequences of actions and making balanced choices

When handling major cyber incidents, it is important to consider societal, economic and political consequences before actions are decided on at the policy level. In many cases, actions undertaken during the course of incident response can have unwanted and unavoidable or unforeseen side effects. Typically, incident response involves decision making under severe time constraints, with a lot of outside pressure of various kinds. Acting before enough is known may yield good results, but can lead also to unforeseen consequences, whereas waiting for more information may lead to pushback because of perceived inaction. In general, decision making during times of crisis, especially at the policy level, may run a high risk of being either rash or overly careful. Inability to maintain awareness of the entire situation, and overly focusing on certain aspects, may yield short-term gains in some respects, and long-term losses in other respects.

In the British NHS case, many hospitals shut down their systems as a precaution. This might have prevented the further spread of the attack, but it also affected the overall operations of those hospitals. Meanwhile, alternative solutions for limiting the impact of the attack were found. Internal communications were recovered to some extent via alternative channels, such as the encrypted WhatsApp application.[93]

Among the analysed incidents, several turned out to be caused by attacks. This raised the question of whether and how to attribute those attacks, as well as whether efforts should be made to explain the probable motives underlying the attacks. On the one hand, a lack of attribution allows speculation as to possible sources and motivation of the incident. Moreover, abstaining from public attribution may be interpreted as acceptance or even endorsement of malicious and hostile behaviour. In the case of the JPS, no immediate attribution was made after the data breach in 2015. Instead, an investigation that followed three years later closed without any named suspects. Meanwhile, some reports pointed out footprints from Chinese actors.[94] On the other hand, hasty attribution can create tensions and cause escalation. The technical limitations to determining

[93] British Government (note 86), pp. 9, 24.

[94] Harold, S. W. et al., *US–Japan Alliance Conference: Strengthening Strategic Cooperation* (RAND Corporation: Santa Monica, CA, 2016); and Nippon Telegraph and Telephone Corporation (note 84).

attribution further amplify this effect. An internet protocol (IP) address was traced to China after the cyberattacks on South Korea's broadcasters and banks in 2013.[95] Speculation about Chinese involvement came early in the investigation. Shortly after these allegations were shared with the public, the South Korean Government attributed the attack to North Korea.[96] Later, South Korea stated that the IP address was not from China, but from a virtual IP address used internally at the bank which coincidentally matched an address registered in China.[97]

Considering the consequences of the available policy options in terms of the trade-offs between preferred outcomes and risk, in the short as well as the long term, is key to navigating the way out of the kinds of crisis that may arise as a consequence of a cyber incident. In particular, decision makers need to consider:

1. *The overall outlook*. In times of crisis, it is very hard to focus on anything beyond the immediate matter at hand. Yet doing so makes for unbalanced choices that would not have been chosen had a more comprehensive view been taken.

2. *Trade-offs*. Most incident response policy options will come with trade-offs. Those trade-offs will vary on a case-by-case basis and may involve unforeseen or unwanted consequences in terms of security, political, societal or economic issues.

3. *Alternative solutions*. In stressed situations, even finding just one policy option may be difficult. Yet having access to several options allows for stronger adaptability as the situation develops, and may also help in evaluating the pros and cons of individual policy options.

Lesson 5. The importance of optimizing institutional arrangements

An overall key lesson that emerged in the course of analysing the case studies is the need to balance centralized and de-centralized handling of a cyber incident, as well as balancing the legal, political and technical mandates of various authorities and other stakeholders.

In Finland, the Finnish Government, the cross-governmental Security Committee, the Ministry of Finance, the Ministry of Transport and Communications, and the National Emergency Supply Agency are the key stakeholders in the Finnish de-centralized cyber-incident management system. The Finnish regional authorities do not have any explicit cybersecurity role. Meanwhile, in Singapore, national cybersecurity coordination and measures are centralized. The small geographical size of the nation and the tradition of strong state control make such an approach appealing.

[95] 'China IP address link to South Korea cyber-attack', BBC News, 21 Mar. 2013.

[96] 'North Korea "behind cyber attack" on South websites', BBC News, 16 July 2013.

[97] '[3.20 사이버테러] 사설IP 구분 못하는 정부… 보안정책 실종' [(3.20 cyber terrorism) private IP missing security policy], *ChosunBiz*, 26 Mar. 2013.

Given the almost simultaneous appearance of harmful cyber effects in interconnected and interdependent systems, it is clear that some level of national centralization is needed to be able to pool scarce competencies. However, it is also crucial to keep a certain level of flexibility, where decisions can be made deliberately at the appropriate management level.

There is also a question of how closely cyber-incident response capabilities ought to be connected to the military, intelligence and law enforcement communities. There may be great benefits in having close connections when dealing with individual cyberattacks or targeted campaigns, but some organizations may be more reluctant to share information directly or indirectly with the military or with intelligence or law enforcement agencies. Any of these actors may also wish to keep certain capabilities and operations secret, which may hinder cyber-incident response in certain situations.

An important aspect to consider when trying to balance centralization and de-centralization of management approaches is the issue of *conflicting objectives*. Overly de-centralizing cybersecurity capabilities among different agencies with different objectives may result in the agencies impeding each other's work. For instance, tying cyber-incident response capabilities to the military or intelligence spheres results in a greater need for secrecy, which may hamper flexibility and reduce transparency in certain situations. Similarly, if cyber capabilities such as vulnerability analysis are managed by the intelligence community, difficult trade-offs may arise between the objective of disclosing vulnerabilities to strengthen defences and the objective of keeping such information secret for espionage purposes. Even when cybersecurity capabilities are centralized in a single organization, conflicting objectives within that organization can lead to inertia or hinder operations.

Beyond the cybersecurity realm, it is important to keep in mind that major cyber incidents often have effects in both the physical and cognitive domains. However, incidents in the cyber, cognitive and physical domains are commonly treated independently. With increasing digitalization, much critical infrastructure is connected to computer networks; and a cyber incident that has effects in the physical realm will be observed and cause reactions in the cognitive realm. Thus, there is a need for coordinated response capabilities across the cyber, cognitive and physical domains. Interdisciplinary or strongly coordinated teams performing dispatch and assessment functions will assist in achieving this goal.

To complicate matters further, major incidents with physical effects will generally be related to disruptions in essential services such as water supply, energy, transport and telecommunications. Understanding cyber incidents in such services, which use highly specialized ICTs and cyber-physical systems, will require specialized knowledge of equipment, procedures, skills, threats and vulnerabilities specific to that sector.

Finally, since most essential services are privately owned and operated, and there is a market for private cybersecurity solutions that cater to such essential services, there is a question of how to balance the role of the state and the role of private enterprise in cyber-incident response as well as in other cyber-related

respects. Whatever the division of roles, strong coordination between the two will be highly beneficial.

So, to build effective capacity for integrated deployment of resources and capabilities, it is important to perform a number of balancing acts.

Balancing centralization and de-centralization of cybersecurity capabilities

This involves balancing:

1. *The number of responsible agencies involved*. The fewer agencies there are, the less time will be spent on inter-organizational deliberations. And the more agencies there are, the higher the risk that no one agency, or only a few, manages to maintain enough competent staff to make a difference. However, the more agencies there are, the more streamlined and the more clear-cut their missions can be. The more streamlined and clear-cut their missions are, the less intra-organizational inertia those agencies are likely to suffer.

2. *The number of competing objectives involved*. The more conflicting objectives given to a single agency, the more organizational inertia there will be. However, the more streamlined different agencies are in terms of their objectives, the harder they might find it to cooperate when they have to do so.

3. *The number of sectors to which cybersecurity support is given*. Since cyber incidents in cyber-physical systems in essential services tend to lie at the heart of the kinds of incidents the state will want to be able to handle, and since understanding how such cyber-physical systems work and what role they play in the sectors where they are used requires deep sectoral understanding, the more sectors that are to be supported, the stronger the case for de-centralization to avoid intra-organizational inertia.

Balancing the involvement of the military, intelligence and law enforcement communities

Involving these communities can lead to significant boosts in the capability to handle antagonistic incidents, but may come with the disadvantage that certain organizations will be less keen to share information, and that a stronger need for secrecy may hamper efficiency in communications and transparency.

Coordinating responses between cyber, cognitive and physical domains

Major cyber incidents will have repercussions beyond the cyber domain, and unless silos are removed or strong coordination is established between the three, incident management may be impeded.

Balancing cybersecurity functions between the state and the private sector

The private sector can play a strong role in strengthening cybersecurity, and often does so regardless of the views of the state. However, some functions should

exclusively be held by the state, and in some areas there is room for public–private partnerships. The state should find ways of dividing roles between itself and private enterprise, coordinating with such organizations where appropriate and harnessing the market forces in this area to maximize the advancement of the field.

6. General conclusions and recommendations

This chapter presents the overall conclusions of the report and provides general recommendations for managing the risk of cyber-incident escalation. The final chapter (chapter 7) presents those recommendations in a form specifically tailored to Sweden and its particular institutional arrangements.

Key findings

To summarize, this project has found that escalation dynamics relating to cyberspace are characterized by four broad threat areas. The first is an overall deteriorating security environment in cyberspace. The second is a presumption of antagonism as the primary explanation for events that occur in cyberspace, as well as a bias for overly focusing on incidents caused by cyberattacks. The third is the ever increasing dependency on ICTs and the growing complexity of mapping and managing such dependency relations. The fourth is the complexity of the public communications arena, which is characterized by a combination of independent actors, high expectations, communication challenges and fragile trust. These threats mean that even cyber incidents that have not been caused by antagonistic acts may escalate into various kinds of conflict.

The project has also found four broad areas of vulnerability, three of which are (probably) of general relevance, and one that is more specific to Sweden. These vulnerabilities mean that Sweden, and probably many other states, are not as prepared to face the above threats as they could be. The first vulnerability is a lack of cognitive robustness in relation to claims about cyber incidents. The lack of a commonly used cybersecurity language with well-defined and understood terms, as well statistics and research on cybersecurity issues and, through them, knowledge, means that most organizations lack a solid foundation on which to devise cybersecurity strategies, while many people lack the resources to take a critical perspective on cyber issues and not accept misleading or provocative claims about cyber incidents out of hand. The second vulnerability is that organizations generally lack interactive response capabilities that would foster better preparation for and management of unwanted and uncontrolled escalation in the aftermath of cyber incidents. The third vulnerability is Sweden's lack of formal nationwide cyber-incident response structures that would complement the geographically oriented structures of its current crisis and emergency response system. The fourth, and perhaps most crucial, vulnerability is the lack of a sufficient focus on cybersecurity in digitalization efforts and maintenance of critical ICT systems.

The main conclusion of this paper is that managing the risk of unwanted and uncontrolled escalation in the aftermath of cyber incidents should be a proactive strategy that focuses on *broad robustness and resilience*. This strategy should apply across the cyber, cognitive and physical domains, and also be the main focus because it is inherently non-escalatory and de-escalatory. This approach

also comes with the added benefit of *solving a greater set of problems*. In the cyber domain, network segmentation, backups and redundancy systems are all examples of measures that may mitigate *both* cyberattacks *and* bad patching, for instance. In the cognitive domain, raising critical thinking and crisis communication skills means stronger protection against both accepting disinformation and spreading misinformation. In the physical domain, fire-resistant materials protect against both arson and accidental causes of fires, for example. Overall, the best way to handle escalation scenarios is by making sure that they do not happen in the first place, and by making sure that if they do happen, the effects are of short duration and limited range. Therefore, robustness and resilience should be the overall objective in any escalation risk management strategy.

Three factors are key to achieving robustness and resilience on the national level:

1. *A framework for transferring escalation management responsibility during cyber incidents*. Since escalation scenarios may start through cyber incidents that occur in private organizations, then come to involve various supporting organizations and thereafter require the participation of government, it may be necessary to spell out when escalation management at the incident organization should be supplanted by management at supporting organizations or the government, and if there are cases where escalation management should be transferred from the incident organization to some other organization. These lines of responsibility are often fairly well determined in national emergency and crisis response systems (at least when public, rather than private, organizations suffer incidents), but it is less clear how the interfaces work or should work when there are foreign actors taking part in an escalation scenario. National emergency and crisis response systems should address management responsibility for cyber incidents specifically, including when foreign actors may be involved.

2. *Alignment of interests*. To build interactive response capabilities and be better prepared *before* controversies arise, there is a need to find ways of aligning the interests of the actors involved. Incident organizations may face strong incentives to keep information about cyber incidents they suffer to themselves, such as fear of regulatory sanctions for poor cybersecurity practices. This prevents other organizations from being warned and having the opportunity to prepare or learn from others' mistakes. Such incentives may prove to be an efficient factor to exploit for actors who would try to shape perceptions, reactions and narratives, whether by spreading rumours or making public statements, or through obfuscatory or imitative practices used in some cyberattacks. To achieve alignment of interests, the probable gains of collaboration must outweigh the risks. This means that organizations that fall within the purview of various regulations need to be offered an incentive, and not be

unduly penalized, for sharing information about cyber incidents they experience. In other words, supervisory practices and sanctions schemes should be shaped to reward organizations that try to do the right thing.

3. *Building and maintaining trust*. Central to efforts aimed at pre-empting and countering unwanted and uncontrolled escalation is the goal of *building and maintaining trust*, especially in terms of the trust accorded to supporting organizations. Supporting organizations play a key role in escalation management since they may add an authoritative voice to conversations where trust between the other participants is low. This is important when cyber incidents occur and the incident organization comes under severe pressure because of perceptions about its competence, its willingness to do what is required or to take requisite responsibility, secrecy practices, or the consequences for affected parties. It is especially important when outside actors use the opportunity to spread disinformation or attribute the incident to an attack perpetrated by some party.

Using this paper's framework of escalation dynamics, the contextual factors which raise the probability that an actor's attempts at achieving de-escalation through de-escalatory actions and non-escalatory post-incident actions will be successful—in the sense that other actors observe, interpret and reciprocate with de-escalatory and non-escalatory actions—are being trusted, having a good reputation, and being considered predictable and understandable.

General recommendations for cyber-incident management

The overall strategic direction, when developing cyber-incident escalation risk management capabilities, should be towards increased national robustness and resilience in the cyber, cognitive and physical domains. To achieve this, states should:

1. Focus on building frameworks for transferring incident management responsibility during escalating cyber incidents. Such frameworks should develop a deterministic chain starting in private organizations, moving to supporting organizations, continuing to the national government level and, finally, connecting to the international political level.

2. Align interests to facilitate external information sharing and develop internal cohesiveness.

3. Build and maintain trust for the incident management system in general, and incident management organizations in particular.

4. Enhance existing structures for the management of cyber-incident escalation risks, rather than build extraneous structures.

5. Establish prior readiness and prepare for the unforeseen by implementing and maintaining a framework for systematic and risk-based cybersecurity work, a standard operating procedure for communication and coordination, clear reporting standards, continuous staff training, and appropriate legislation and policies to enable and support these measures.

6. Enable coherent response coordination by maintaining a response framework that allows inter-agency, public–private and cross-border cooperation.

7. Conduct proactive communication strategies by providing swift and fact-based information, focusing on de-escalatory and non-escalatory messaging, preparing communications as often as new information comes to light, and keeping a clearly defined division of roles in terms of who communicates about what.

8. Enable sound decision making by dedicating a function that maintains and develops an overall situational awareness, analyses and assesses trade-offs among incident response policy options, and always proposes a varied set of evaluated policy options.

9. Optimize national institutional arrangements to context-specific needs by: balancing the number of agencies involved, the number of conflicting objectives involved and the number of sectors to which cybersecurity support is given; balancing the involvement of the military, intelligence and law enforcement communities; integrating cyber, cognitive and physical-incident response capabilities; and balancing the cybersecurity roles between the state and the private sector.

7. Targeted recommendations for cyber-incident management in Sweden

This chapter translates the overall recommendations set out in chapter 6 on how to develop national cyber-incident escalation risk management capabilities into particular recommendations for the Government Offices of Sweden and for the MSB and its partner agencies.

Recommendations for the Government Offices of Sweden

The relevant sections of the Government Offices should review their current response structures to see if there is any need for changes in relation to managing cybersecurity incidents. In particular, such a review could include: (*a*) whether there is an adequate framework for handling quickly escalating situations with hybrid components, such as impacts in cyber, cognitive and physical domains; (*b*) how the interface with the new national cybersecurity centre[98] should look during escalation; and (*c*) how international aspects relating to escalation scenarios should be handled. The review should take into consideration the lessons from past cyber incidents, outlined in chapter 5, in determining how the relevant agencies should be steered as they develop new capabilities to counter hybrid threats.

Recommendations for the MSB and its partner agencies

Currently, there is an ongoing process of centralization and integration of cybersecurity capabilities in Sweden through the formation of a national cybersecurity centre (the Centre). The Centre will be a joint venture between the MSB, the Swedish Armed Forces, the National Defence Radio Establishment and the Swedish Security Service. The development of the Centre is also being carried out in close coordination with the Swedish Police Authority, the Swedish Defence Materiel Administration and the Swedish Post and Telecom Authority. The Centre will, at least initially, focus on countering antagonistic cyber threats on the technical level.[99]

While it is building the Centre with its partners, the MSB also has the overall responsibility for coordinating and, in some respects, leading national cybersecurity efforts on the basis of the NIS Directive and the related Swedish legislation. This means that the MSB will be able to serve as a conduit between the work being conducted within the Centre and the work being conducted nationally in the sectors designated in the NIS Directive, and their respective supervisory agencies. To strengthen this institutional arrangement and increase its overall

[98] Löfven, S., 'Statement of government policy, 10 September 2019', Speech by the Prime Minister of Sweden, Stockholm, 10 Sep. 2019.

[99] MSB, 'Fördjupad samverkan för ökad cybersäkerhet' [In-depth collaboration for increased cyber security], 16 Dec. 2019.

capacity, following are three sets of recommendations for the Centre, the NIS Directive agencies and the MSB.[100]

1. Recommendations for the national cybersecurity centre

The agencies involved in the Centre should continue to develop joint capabilities that include escalation prevention and de-escalation management support to the Government Offices, particularly in terms of:

1. Monitoring the development of (*a*) assertive postures, (*b*) offensive operations, (*c*) tools that enable offensive operations, (*d*) obfuscatory and imitative practices used in such operations, and (*e*) the securitization of ICTs, and how these and other matters shape policy.

2. Helping to detect and counter offensive operations, in terms of identifying when forward presence has been established, more far-reaching cyberattacks and ICT supply chain attacks against Swedish networks and information systems, and obfuscatory and imitative practices used in these operations.

3. Facilitating information sharing, pooling of detection capabilities and capacity building in terms of escalation risk analysis, between the public and the private sectors.

4. Maintaining continuous situational awareness to enable detection of deviations in terms of increased or decreased offensive cyber activity.

5. Being a 'one-stop shop' in terms of providing and communicating swift and fact-based information about threats and vulnerabilities as they are detected, as well as about specific cyber incidents and what is being done about them.

6. Encouraging and supporting organizations across society to increase their cybersecurity awareness, maintain and update their existing infrastructures, and organize regular staff training in handling cyber incidents and escalatory scenarios.

2. Recommendations for NIS Directive agencies

The MSB and the NIS Directive supervisory agencies continue to raise cybersecurity standards in critical sectors, so that the cyber robustness and resilience of essential services is continuously strengthened. In this respect, the MSB should use its position as a conduit to facilitate interaction between the Centre and the NIS Directive supervisory agencies as well as the NIS Directive sectors.[101] This interaction will:

[100] Disclaimer: none of the authors of this paper is involved in the development of the Centre, and the recommendations presented here are not necessarily in line with the policy positions of any of the above mentioned agencies, or the Government Offices of Sweden.

[101] NIS Directive (note 18), Annex II.

1. Foster a continuous conversation about sectoral and general cybersecurity issues between the Centre and the NIS Directive agencies and their sectoral actors.

2. Efficiently spread information about current threats and vulnerabilities to sectoral experts who may analyse, assess and forward such information to relevant parties in their respective sectors.

3. Strengthen the Centre's ability to monitor and analyse the sectoral impact of threats and vulnerabilities it detects.

4. Help operators of digitalized essential services secure the external flows on which they depend, such as electricity, cooling and data flows.

5. Collaborate with operators of digitalized essential services to map out chains of dependencies and find ways of preventing and mitigating cascading effects resulting from cyber incidents, including proactively securing coordination between relevant incident response functions.

6. Enhance the security focus in digitalization efforts as well as in the maintenance of existing information infrastructures, including facilitating staff training in the kinds of major cyber incidents that could result in the absence of a security focus.

3. Recommendations for the MSB

The MSB, possibly in conjunction with its partners in the Centre, should continue to build forums for information sharing and joint capacity building to complement the current Swedish crisis and emergency response system in terms of handling cross-sectoral or other cross-cutting cyber threats that are hard to counter in a geographically focused response system. The MSB should bring these perspectives with it as: (*a*) new operative coordination and incident response capabilities are discussed and developed among EU member states within the NIS Directive framework; and (*b*) the Swedish emergency and crisis response system undergoes reviews and updates in the coming years.

The MSB should work to counter the escalatory potential posed by the presumption of and bias towards antagonism and by the lack of cognitive robustness in relation to cyber incidents by:

1. Working with relevant stakeholders to develop a commonly accepted and used language about cybersecurity issues, and to provide statistics and support research on the magnitude of the problems in the field.

2. Continuously and widely providing information about (*a*) the prevalence and impact of cyber incidents, (*b*) the importance of

demanding a strong security focus in digitalization efforts, (*c*) what can be done to prevent and mitigate such incidents when they occur, and (*d*) what the public can expect when such events occur.

3. Raising general knowledge and fostering the development of informed and critical perspectives about cybersecurity issues, including the importance of being vigilant about the risks and not accepting dubious claims at face value.

4. Further developing the ability to detect, analyse and help organizations counter misinformation and disinformation relating to cyber incidents.

5. Further developing crisis communication capabilities that allow for informative and accessible communications about cybersecurity issues and specific cyber incidents.

CYBER-INCIDENT MANAGEMENT

The ever increasing dependence on information and communication technologies in all aspects of society raises many challenges for national crisis management agencies. These agencies need to prepare not only for new cyberthreats and cyber vulnerabilities, but also for the fact that the aftermath of a cyber incident affecting critical infrastructure has its own challenges. On the one hand, the practical disruptions caused by an isolated incident can be hard to predict and control and, on the other hand, the consequences and perceptions of an incident whose cause is not yet determined can be equally hard to manage. Uncertainty surrounding the cause of the incident and the remedial actions being taken often lead to public speculation and political pressure to respond in ways that could create political tensions, and possibly conflict, between countries.

This policy paper is the result of a nine-month research project that was jointly conducted by SIPRI and the Swedish Civil Contingencies Agency (MSB) on cyber-incident management. It explores what national crisis management authorities can do to improve their cyber-incident prevention, detection and response strategies and also improve how they deal with the larger societal and potentially political aftermath. It investigates why and how cyber incidents may lead to escalatory scenarios and how these scenarios can be avoided and contained using various de-escalatory approaches. Its main chapters describe a conceptual and analytical framework for understanding escalation and de-escalation; map out the various threats and vulnerabilities as factors that can fuel escalation of cyber incidents; and present the lessons learned from a series of case studies on past cyber incidents in other countries. Finally, it provides recommendations to national crisis management authorities for managing the risks of cyber-incident escalation.

Johan Turell (Sweden) is a Senior Analyst and Research Coordinator at the Swedish Civil Contingencies Agency (MSB) in Stockholm.

Fei Su (China) is a Researcher in SIPRI's China and Asia Security Programme.

Dr Vincent Boulanin (France/Sweden) is a Senior Researcher on emerging technologies at SIPRI.