TU Delft

Ecosystem-level simulation of cybersecurity incidents for critical infrastructures

An integrated agent-based modelling study based on an aggregation of ecosystem-defining concepts

Xander de Ronde

Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Article history:
Concept: 24th of June, 2018
Final: 20th of August, 2018

Keywords:
Critical infrastructures,
Cybersecurity
Ecosystem analysis
Complex adaptive systems
Agent-based modelling

ABSTRACT

The increased decentralisation and heterogeneity of critical infrastructure systems pose a threat to the safe and secure operation of critical infrastructures by complicating cybersecurity procedures. The increased frequency and impact of cyberthreats have led to the desire to develop coherent security policies. In order to explore the effects of such policies, an ecosystem-level framework for cyberincidents in critical infrastructure systems has been developed. This framework can be used to explore the effects of conceptual defensive strategy designs. The aggregation of these concepts was operationalised around a central degree of critical infrastructure operability. A practical application of the framework is provided in the form of an agent-based modelling study that is capable of simulating the effects of coherent defensive strategies. The framework by itself is limited to exploratory modelling for the sake of generalisability and requires further specification of concepts based on representative infrastructure scenarios for more advanced analysis and design of tangible security policies.

1. Introduction

Over the past decades, the role and characteristics of critical infrastructures have changed significantly. Critical infrastructures are defined as infrastructures for which the unhindered functional operation is vital to the functioning of crucial elements of society, such as electricity grids and drinking water facilities [1, 2]. Traditional critical infrastructure systems were developed as centralised and closed monolithic entities that control system operation [3-5]. Recent technological developments and the emergence of smart city thinking have changed critical infrastructures into widely interconnected networks of smaller distributed and heterogeneous systems [3, 6, 7]. These changes enable more efficient operation of critical infrastructure systems, but also increased their susceptibility to attacks from both cyber and physical domains [8-10]. Recent prominent examples of large-scale cyber-physical attacks on critical infrastructures include Stuxnet [10, 11] and the Ukrainian power grid blackout [12, 13].

The increased vulnerability from cyberattacks is the result of increased reliance on information availability and hindered agility from differences in security standards and protocols [6, 14, 15]. Additionally, integrating IT infrastructure with traditional physical infrastructure systems complicates security practices, as legacy components were never designed to incorporate cybersecurity elements [3, 5, 11]. Increasing the resilience of critical infrastructures against cyber-physical threats requires identification of threats with these issues accounted for [11, 16]. This involves establishing which issues arise from interconnected and interdependent networks of system components and how cyberthreats are dealt with across these networks. The perspective required to analyse the effectiveness of coherent security standards encompasses an ecosystem-level view on cybersecurity incidents that take place [5].

The importance of ecosystem-level analysis emerges from the lack of research the effects of coherent cybersecurity policies and control mechanisms that are capable of resisting the effects of increasingly sophisticated coordinated attacks [3]. The increased frequency and impact of cyberattacks on key assets for critical infrastructures calls for effective defensive strategies, which are currently lacking in applicability [16]. Security strategies should incorporate elements that effectively thwart intrusions by design and optimise the awareness of current threats to system operation within the ecosystem [11, 16]. There is, however, currently no major framework that incorporates and operationalises elements that should be included in ecosystem-level analysis

for critical infrastructure cybersecurity incidents. This study seeks to address the academic knowledge gap by enhancing knowledge required to conduct such ecosystem-level analysis.

The main aim of this study is to identify concepts that define an ecosystem of critical infrastructures, aggregating elements that together are capable of describing interaction within this ecosystem. The research objectives are defined as follows:

1 Identify and specify concepts that describe cybersecurity concerns in an ecosystem of critical infrastructures.
2 Formalise the set of conceptual elements into an operationalised framework
3 Demonstrate proof of concept of this framework through an agent-based modelling application

These objectives seek to address the main lack of knowledge and formulate a framework that can be used to assess the fitness and effectiveness of different types of defensive strategies. Since the creation of such a framework encompasses design science, the framework should comply with design science practices. Peffers, Tuunanen [17] lay out six activities: identifying the problem, defining objectives for a solution, designing the artefact, demonstrating the usefulness of the artefact, evaluating the artefact and communicating the artefact. As such, this study pays particular attention to these activities. The first two steps were already addressed in this section, the four remaining steps will be conducted throughout this article. First, related work is addressed in section 2. Section 3 discusses the elements incorporated in the framework, in line with the first research objective. Section 4 details the steps taken to operationalise these concepts coherently. Section 5 provides insight into a practical application of this framework in the form of an agent-based simulation model. Section 6 discusses the main insights and limitations to using the framework in an attempt to communicate its benefits.

2. Related work

Related work to this study is divided into three different categories: work proposing an ecosystem-level approach to learn more about the effects of security policies, work applying an ecosystem modelling approach and work using agent-based modelling to simulate cybersecurity issues.

Types of security policies in the form of control mechanisms are widely discussed across academia. The majority of research is performed around the creation of specific control mechanisms applicable to single systems. Cárdenas, Amin [18] describe the effects of an anomaly-based intrusion detection mechanism that adapts to a wide variety of attack types, highlighting performance in terms of detection rate and false alarms. Similarly, Ntalampiras [19] proposes an anomaly-based intrusion detection method in which an ensemble of mechanisms is used to account for multiple contributing factors, such as time or system load. Patel, Taghavi [20] provide an overview of intrusion prevention and detection systems that account for sensitivity and specificity, while discussing their connection to the resulting defensive decisions made. Specifically, this helps address the expected consequences from implementing each type of control in isolation. What these authors do not address however, is the implication of operating such mechanisms in an environment where other interconnected critical infrastructure systems are present.

The need for ecosystem-level analysis of defensive strategies requires a frame of reference for specifying the aggregation of concepts that make up a system. Since this is not done yet for the topic of this study, these articles relate to similar yet different problems. Rutkowski, Kadobayashi [21] describe a process by which decomposed cybersecurity processes are discussed in the light of their relevance to the overall process of exchanging information in a secure fashion. The authors identified four different domains in which different types of entities process information while applying different concepts. This architectural approach helps identify the complete set of relevant concepts. Singh [22] discusses a similar approach by which a cybersecurity ecosystem for multi-agent systems are specified. To this end, the ecosystem is perceived as an isolated microsociety. All elements included in the model should be closed off by their definition and relate to other concepts in at least one way and architecturally juxtapose each concept in such a way that all concepts relate to a central concept.

Lastly, agent-based modelling, which will be used to demonstrate the practical application of the framework, is an often-discussed method for cybersecurity purposes. Charitoudi and Blyth [23] establish a model simulating the impact of cascading failures within an interconnected and heterogeneous network. They found that agent-based models perform well at observing the impact of dependencies for critical systems. Janssen and Sharpanskykh [24] devised a simulation model for security checkpoint intrusion, specifically focusing on coherent impact assessment in an agent-based model. These two articles indicate that agent-based modelling can work under the desired circumstances of an ecosystem-level model for critical infrastructures, but this has not been done yet given the lack of a framework for ecosystems of critical infrastructures.

3. Integrated ecosystem-level framework

In order to understand the effects of cyberattacks beyond the intrusion of a single system, an ecosystem-level framework for the aggregation of concepts is required. This corresponds with the third step for design science, the actual design of a framework. The aggregation is depicted in Fig. 1. The aggregation distinguishes concepts between four separate entities: attackers, infrastructure nodes, infrastructure operators and users. This is done to cover real-world facets of interaction while establishing a central concept that connects all entities in the form of infrastructure nodes. The elements will be individually discussed to shed light on defining elements. A step towards operationalising the framework is mapping interaction that takes place among these entities. This resulted in the ecosystem interaction model shown in Fig. 2. Depicted are the three types of actors, interacting with the central entity in the form of infrastructure nodes.

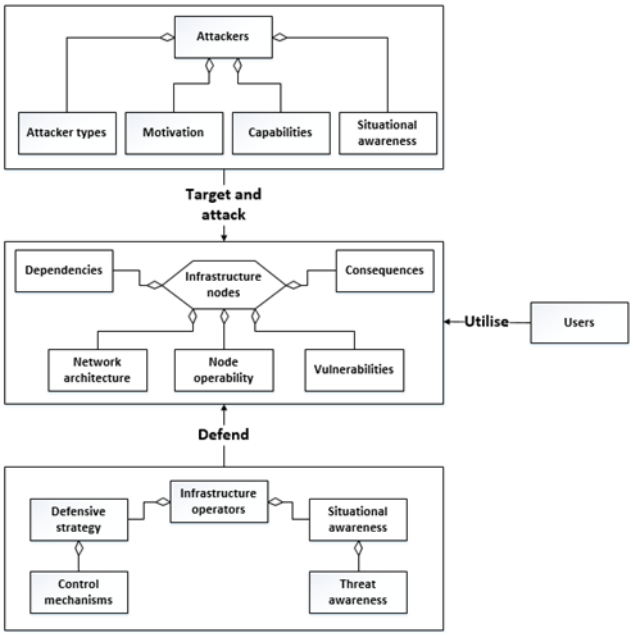

Fig. 1: Ecosystem aggregation model

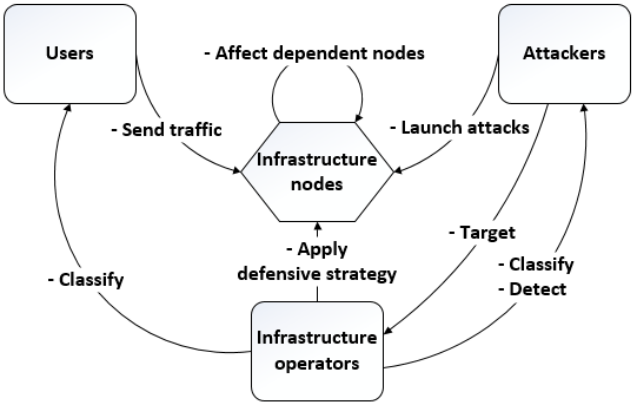

Fig. 2: Ecosystem interaction model

Infrastructure nodes

Infrastructure nodes are central in interaction within the ecosystem, as is depicted in Fig. 2. An infrastructure node is a system within a critical infrastructure network and represents any system that is to some extent linked to infrastructure operation or is capable of causing societal consequences. Infrastructure nodes are defined by five generalisable concepts.

The first and most prominent concept is infrastructure node operability. The central degree of operability represents the direct performance or productiveness of an infrastructure node [25, 26]. Disruptions caused by cyberattacks or erroneous defensive decisions directly tie in to operability of infrastructure nodes [27]. A conscious decision was made to define operability within this light, operability as a formalism enables integration of other elements, as the effects can be related to a central degree of operability and inoperability, which ensures further conceptualisation is possible [28]. Following this definition of operability, the framework maintains extensibility into more specific critical infrastructure domains, where additional factors can be defined to determine operability.

The second concept is the presence of functional dependencies between infrastructure nodes. Dependencies are functional directional relationships between two infrastructure nodes, where the dependent node relies on robust operation of the other node, and are considered a crucial element for security purposes of critical infrastructures [26, 29]. Interdependencies are then the bidirectional equivalent, or the combination of two separate dependencies [26]. Dependencies are generalised regardless of whether these are cyber, physical, geographic or logical dependencies, as the influence coefficient assigned to a dependency directly determines perturbation of node operability [25]. This generalisation of dependencies following an influence coefficient enables universal application of the effects dependencies exert within an ecosystem [30]. The importance of these effects are crucial, as they are not fully understood yet central to the challenge of improving resilience and robustness of current and next generations of critical infrastructures [5].

The third concept is the complex network architecture by which infrastructure nodes are connected. The cyber-architecture applied to a critical infrastructure network determines how nodes are organised, connected and how dependencies are structured [4]. The network topology for a critical infrastructure system determines how control is effectuated across the network, or how losses are suffered due to node inoperability [6, 31, 32]. By extent, the topology affects security practices structurally and should be included to maintain agility of the framework.

The fourth concept revolves around vulnerabilities. The presence and overall impact of vulnerabilities is rooted in the nature of a critical infrastructure system and impacts the susceptibility of a node to cyberattacks [3, 4]. Vulnerabilities should be included in a framework for ecosystem-level analysis, but could be modelled as simply as a factor for cyberattack success rates. Other possibilities include game theoretic principles that encompass a specific, in-depth representation of optimisation processes for both attackers and defenders [9].

The fifth concept is the severity of consequences for node inoperability. Consequences from successful cyberattacks and inoperability in general are significant and difficult to establish accurately [3]. Losses can be categorised as physical, economic and social losses [33]. The main implication from the severity of consequences is that a degree of tolerated risk is not representative for critical infrastructures. Any consequences sustained are undesired by definition, and cybersecurity is purely related to attack-defence scenarios as opposed to security investments [9].

Attackers

Attackers are malicious actors seeking to disrupt infrastructure node operation. This is achieved through cyberattacks that inflict damage in the form of inoperability if they manage to successfully exploit node vulnerabilities. By separating attackers from cyberattacks, the model gains another layer of interaction. In many cases, security practices depend on actions taken to secure assets as well as actions taken by cyberattackers. Attackers interact with

entities the ecosystem by targeting infrastructure operators based on their motivations and rationality. They initiate attacks on infrastructure nodes, which prompts infrastructure operators to make defensive decisions.

The first concept attributed to attackers are different attacker types. Understanding how threats against the system emerge requires mapping the background of attackers that would act against the system. Attackers capable and willing to inflict damage to critical infrastructures are typically advanced persistent threats (APTs) [34], such as foreign adversaries, cyberterrorists or cybercriminals [35]. The type of attacker determines their motivation and capabilities for conducting attacks. Establishing attacker types as an element of the framework allows for different decision-making processes to be included and forms an interface for possible inclusion of game-theoretic principles. As such, attacker types can be used to establish their capabilities, target selection procedures and other characteristics.

The second concept, attacker motivation, determines the types of losses attackers are willing to inflict and subsequently determine the types of attacks they are willing to use. Cyberterrorists are typically out to inflict physical losses and bodily harm, and will therefore select other targets and methods than attackers seeking to inflict financial harm [3].

The third concept relates to the capabilities of attackers in terms of possible attack vectors. Sophisticated attacks, such as Stuxnet, are only feasible for highly capable attackers [10, 11]. Assigning different capabilities to different attacker profiles increases the agility of the framework in dealing with typical threat landscapes [36]. Both the capabilities and motivations of attackers are directly tied into the first concept of attacker types, as the attacker type can partially or fully determine the set of capabilities, motivations and more. By separating these three concepts, the framework can be used to represent both simple and sophisticated processes at many levels of granularity by which interaction is modelled.

The fourth concept is situational awareness, the degree to which actors are capable of making rational decisions [37]. Attackers make decisions based on the information and knowledge available to them in assessing whether they should and which target is selected [24]. This is largely established based on the perceived damage they would inflict as well as their loss type preference, which together forms a degree of perceived utility [9]. Situational awareness should be included as a facet of any model for securing critical infrastructures, as the concept goes hand in hand with the desired level of sophistication for decision-making.

Infrastructure operators

Infrastructure operators, also referred to as defenders, are the entities tasked with securing resilient and robust operability of infrastructure nodes. They interact with the ecosystem by asserting control over a set of infrastructure nodes through making defensive decisions. This process involves discerning between threats and user traffic.

The first concept addressed is the usage of control mechanisms to address cyber-risk within the ecosystem. Deviating from the traditional definition of cyber-risk as an enumerable factor of risk over time, critical infrastructures instead deal with immediate decisions as opposed to investment decisions made over time. Control mechanisms are used to thwart cyberattacks and ensure node operability. Three types of controls are identified: prevention mechanisms that filter out traffic as either an attack or user traffic, detection mechanisms that attempt to classify and assess any unprevented intrusions in terms of impact and response mechanisms that provide means to remove active attacks [18, 38]. Further specification of mechanisms is possible, and in many cases required, but should follow the same high-level definition provided by this framework. This ensures that specific instances of research can be related to one another, paving the way for design of coherent technological artefacts or security policies.

The second concept is in itself an aggregation of control mechanisms, as a defensive strategy or security policy. A defensive strategy is defined as a configuration of rules and practices for control mechanisms. This defines how threats are dealt with and could be used for analysis of comparative performance for defensive strategies. The inclusion of defensive strategies is required for operationalisation of high-level interaction, while defensive strategies are more contextual for more specific analysis.

The third concept is threat awareness for defensive decisions. Threat awareness represents the knowledge about active threats in the ecosystem available to defenders. During the 2015 Ukrainian power grid attack, the presence of threats within individual subsystems was considered marginal, whereas the overall ecosystem-level threat landscape would have prompted more serious measures [12]. Awareness of threats determines defensive decisions made, and this link can be operationalised on both high and low levels of granularity.

The fourth concept, extending the degree of threat awareness is the situational awareness for defenders. Similarly to situational awareness for attackers, this determines the availability of information for decision-making. Situational awareness extends threat awareness by awareness of deviations in system operability [38] and results from defender capabilities to detect intrusions without raising false alarms [39]. The situational awareness feeds into the defensive decisions made in response to events that affect the threat landscape.

Users

Users are incorporated in the framework as an entity that ensure the functioning of an infrastructure node. They represent traffic that is crucial for functional operation for nodes. Failing to facilitate access for user traffic directly affects infrastructure node operability. Users were identified to not require further conceptual specification on an ecosystem level. However, depending on the context of desired analysis, users can play a minimal role for which this definition suffices as well as a crucial role that requires further elaboration.

4. Operationalising the ecosystem as a complex adaptive system

Without operationalisation of concepts defined as part of the framework, the ecosystem aggregation model is nothing more than a collection of generalised concepts. In order to establish a useful model for analysis of ecosystem-level effects of defensive strategies, these concepts should be operationalised into clearly defined and constrained interaction. To this end, complex adaptive systems (CAS) thinking is applied, as it provides several applicable definitions and core concepts that correspond with the devised framework. First, the CAS thinking paradigm is detailed. This is followed by defining the operationalised model concept as a collection of entities and actions.

4.1 Defining complex adaptive systems

Complex adaptive systems thinking is a school of thought that involves systems that are both complex and adaptive. Complexity in this sense means the degree by which top-down system behaviour can be understood and adaptiveness represents the degree by which parts of the system organise themselves to improve performance [41]. A key definition of CAS is given in Waldrop [42], as systems are perceived as "a dynamic network of many agents acting in parallel, constantly acting and reacting to what other agents are doing.", in which agents represent single entities that operate based on behavioural rules, actions and states [41].

The main elements of complex adaptive systems are the way in which bottom-up interaction emerges into coherent system behaviour [42, 43], chaotic behaviour from model elements [41] and the non-optimality of system elements [43]. The involvement of aggregate system behaviour being substantially different than the collection of actions taken by parts of the system highlights how such an approach could contribute to cybersecurity analysis for critical infrastructures [43]. Such frameworks should incorporate a CAS perspective, as the behaviour for coherent security policies is currently not understood despite the prevalence of analysis of single systems [16].

The applicability of complex adaptive systems to this ecosystem requires further elaboration. CAS perceives systems as dynamic networks of heterogeneous agents, which corresponds with the definition of critical infrastructure systems as vast networks of heterogeneous system components [6, 7]. Decisions made related to each infrastructure node or infrastructure operator are based on the situational awareness at the moment of decision-making. Higher-level strategic decisions are not typically made within the frame of reference for analysis, as the development of such strategies depends on uncertain future developments [44]. Chaotic elements are also identified in cybersecurity decisions for critical infrastructures, as the set of behavioural rules is consistent, whereas activities surrounding cyberattacks are inconsistent. The aggregation model could therefore be operationalised using a set of formalisms that define the states, actions and interactions in the system.

4.2 Concept operationalisation

Operationalising the framework within the light of CAS requires specification of the main actions and states used by agents. In the light of the design science practices defined by Peffers, Tuunanen [17], this step further specifies the design. Eventual practical application would require full specification of all elements involved in implementing a simulation model, but this is sensitive to requirements and assumptions applied. The list of main states and actions following descriptions given in section 3 are listed in Table 1.

Table 1: Main agent states and actions

	Infrastructure nodes	Attackers	Infrastructure node operators	Users
States	- Operability - Associated losses - Vulnerability	- Loss preferences - Knowledge - Attack capabilities	- Perceived operability - Defensive strategy - Type of control mechanisms	- Criticality of traffic - Frequency of traffic
Actions	- Affect dependent nodes - Sustain losses for inoperability	- Select target - Launch attack	- Apply prevention mechanisms - Apply detection mechanisms - Assess impact of detected threats - Apply response mechanism	- Generate traffic to infrastructure node

These states and actions involve the main interaction that takes place in the ecosystem as depicted in Fig. 2. However, these states and actions are still rather vague. For the framework to have any use, methods to operationalise the main elements should be provided. The operationalisation should not only support the original research objective, but should also be in line with the knowledge gap this study seeks to address. Factors should be operationalised in such a way that high-level interaction can be quantified, albeit to in an abstract manner, while specific, low-level interaction is still coherent with other facets of the framework.

Operability

The degree of operability can be defined as a continuous scale [25, 28]. If this scale is taken as a value between 0 and 1, an approximation of its effects can then be computed as the inverse of current impact. Being a central entity in the framework, it is important that operability is operationalised in a simple manner while facilitating interfaces for more specific driving forces. For ecosystem-level analysis, operability O can be computed from both external and internal impact components $\rho_{external}$ and $\rho_{internal}$. Multiplication between both components ensures a constraining linear relationship between impact components, as external disturbances in infrastructures have proven to disrupt operability of other infrastructure nodes completely [10]. Operability O is computed as follows:

$$O = (1 - \rho_{internal}) \times (1 - \rho_{external}) \quad (1)$$

The specification or level of granularity applied to each of these components can then be varied. Within the scope of this study, the decision was made to specify the process by which attack and defence scenarios take place as a probabilistic process. This choice was made based on the manageability of the model, as game theoretic processes require further specification of additional factors and assumptions. There are multiple ways to model such scenarios, but there is no optimal modelling approach between game theoretic processes or probability-based processes [9]. As such, whenever user traffic or attacks are created, there is a chance for an attack to be prevented correctly based on the true positive rate associated with the applied prevention mechanism and a chance for user traffic to be prevented incorrectly based on the false positive rate associated with the applied prevention mechanism. Similarly, unprevented

attacks can be detected based on the true positive rate of detection mechanisms and false alarms can be raised based on the false positive rate of detection mechanisms. This is the most straightforward way of implementing attack and defence models without establishing any factors that might not be generalisable.

Internal impact is represented by the disruption of inoperability caused by affairs within a single node, such as a successful cyberattack or erroneously blocked user traffic. Given a set of attacks A_i from the total set of active attacks *A*, a set of users U_j from the total set of currently erroneously blocked users *U*, the associated power of an attack *I* and criticality of user traffic *C*, equation 2 denotes the internal impact component.

$$\rho_{internal} = \begin{cases} \sum_{A_i \in A} I(A_i) + \sum_{U_j \in U} C(U_j), & if \sum_{A_i \in A} I(A_i) + \sum_{U_j \in U} C(U_j) \leq 1 \\ 1, & if \sum_{A_i \in A} I(A_i) + \sum_{U_j \in U} C(U_j) > 1 \end{cases} \quad (2)$$

Whereas internal impact is derived from the summated impact of all attacks and erroneously blocked user traffic, external impact for a node originates from inoperability in nodes with outgoing dependencies to this node. Dependencies are assigned a dependency weighting that directly affects the degree of operability [26]. For a node with a set of dependencies D_k from all inbound dependencies D, dependency weighting w_{D_k} for a dependency D_k and O_k the level of operability at node k, the following equation denotes the operationalisation of external impact:

$$\rho_{external} = \begin{cases} \sum_{D_k \in D} w_{D_k} \times (1 - O_k), & if \sum_{D_k \in D} w_{D_k} \times (1 - O_k) \leq 1 \\ 1, & if \sum_{D_k \in D} w_{D_k} \times (1 - O_k) > 1 \end{cases} \quad (3)$$

Equations 2 and 3 assume that the effectiveness of attacks and the weighted effects of dependencies do not increase or decrease depending on target node operability and by extent is linearly dependent on attack types and attack activity. Because operability O was previously defined as a linear product of both internal and external impact components, the overall perturbation of operability follows linear relationships. If this is not the case for a specific case that is being modelled following this framework, a logarithmic or exponential element can be added. Similarly, time-based delays could be added to these effects as well to represent successful yet inactive attacks or delayed dependent relationships. These modifications would have no direct effect to equations 2 or 3.

Impact assessment

As stated in paragraph 3, the notion of operability provides agility to involve multiple other concepts. Situational awareness is related to the notion of operability, as limitations in information availability cause infrastructure operators to make decisions on their impact assessment [23, 27]. The perceived operability $p_{operability}$ of a node then determines which defensive decisions are made. $p_{operability}$ can be modelled as the result of perceived values for $\rho_{internal}$ and $\rho_{external}$ in the form of $p_{internal}$ and $p_{external}$, respectively. Impact assessment differs from true impact, as not every active attack might be detected, and false positives from intrusion detection might lead to incorrect decisions [20, 35]. This decision was made to comply with the involvement of situational awareness as described in section 3.

Given attacks A_i from the set of detected attacks $A_{detected}$, false positives FP_j from the total set of false positives FP and *I* the impact associated with a type of attack, perceived internal inoperability $p_{internal}$ can be computed as follows:

$$p_{internal} = \begin{cases} \sum_{A_i \in A_{detected}} I(A_i) + \sum_{FP_j \in FP} I(FP_j), & if \sum_{A_i \in A_{detected}} I(A_i) + \sum_{FP_j \in FP} I(FP_j) \leq 1 \\ 1, & if \sum_{A_i \in A_{detected}} I(A_i) + \sum_{FP_j \in FP} I(FP_j) > 1 \end{cases} \quad (4)$$

Losses incurred

Losses sustained L resulting from node inoperability can be linked to this same degree of operability O. A basic operationalisation of how inoperability contributes to losses is to scale the possible damage given by respective loss factors linearly with the extent of inoperability. If desirable, the equation can be changed to contain exponential, logarithmic or other additional components. The assumed linearity enables a relatively simple model construct that ensured manageability of the simulation model discussed in section 5. Given a set of loss factors for each type of loss F_i, the overall losses are computed as follows:

$$L = \sum_{F_i \in F} (1 - O_{overall}) \times F_i \quad (5)$$

Attacker utility

As stated previously, target selection is based on situational awareness and loss preference for attackers. The target T is selected out of the set possible targets, given each assessed target i, the loss factor F_k^i for type of loss k at node i and loss preference P_k for type of loss k, following equation 6. The process by which an attacker maximises utility is heavily based on game theoretic principles and is inherently based on the presence of multiple types of losses.

$$T = \max_{T_i}\left(\sum_{F_k \in F}(P_k \times F_k^i)\right) \tag{6}$$

5. Practical application through agent-based modelling

Given the possibilities for operationalisation of the framework prevented throughout the previous paragraph, a practical application can be demonstrated for the framework in a simulation environment. This follows the 5th activity for design science: demonstration [17]. To implement this, the framework was implemented in the form of an agent-based model. Agent-based modelling is an extension of complex adaptive systems and simulates system interaction as the collection of agent-level observations and interactions [41]. First, the model narrative and formalisation is discussed. Secondly, the experimental design is detailed in the light of desired insights. Thirdly, the results from the aforementioned experiments are presented.

5.1 Model narrative

The model was created by implementing the framework with slight deviations for model manageability purposes. Infrastructure operators and infrastructure nodes were grouped together as one entity, as their tasks are directly aligned, and separate agents would result in computational issues. User agents were grouped together as a randomly occurring event, as opposed to being agents that only conduct one action based on one state. As such, the simulation model consists of two types of agents directly interacting, attackers and defenders. The full overview of their interaction is shown in Appendix A. The model was used to analyse the behaviour of entities in the framework in terms of decision-making and incurred losses based on different defensive strategies. These simulations used a discrete time step of one day for a total of 1825 steps, or 5 years. The model was implemented in NetLogo [45], which provides an easy-to-understand software environment for modelling complex systems [46].

Attackers act first, as they decide whether to launch an attack or not. If they do attack, they will first select the target that yields the highest perceived utility. The mechanism to select a target is based on equation (6) and grows more sophisticated with higher attacker knowledge, as expected consequences from dependencies can also be included. Having selected their target and attack vector, attackers will initiate an attack. Attacks are modelled as single-target disruptive attacks and worm-based attacks capable of spreading to connected nodes. If attackers already had an active worm attack, there is a chance this attack spreads further across the network.

Defenders respond to the initiation of attacks or user traffic by first applying their associated intrusion prevention mechanism. Both intrusion prevention and intrusion detection mechanisms are assigned sensitivity and specificity values that determine the rate at which false negatives and false positives occur. This circumvents making assumptions about values for vulnerabilities and threat capabilities, as these effects are instead derived from probability-based events. The intrusion prevention mechanism either blocks or allows traffic. Blocked traffic is removed completely, representing either a true positive or false positive. Blocked user traffic reduces internal operability as described in equation (2). Besides intrusion prevention, defenders go on to conduct intrusion detection. Should an attack be detected, the perceived inoperability for the defender increases, prompting them to make a defensive decision. Based on thresholds for different response mechanisms, they either make a correct decision or an incorrect decision. Incorrect decisions are either insufficient response when a response is warranted (overestimating operability) or a response when there is insufficient inoperability (underestimating operability). Two responses were implemented, alleviation and retention, respectively a light and slow intervention and a heavy and fast intervention. Responses remove existing attacks after a predetermined duration. After these procedures are conducted, losses are sustained following equation (5). The translation of the operationalised model into modelling constructs involved a balancing act between creating a manageable yet insightful model. Since the simulation model serves primarily as a proof of concept for simulating critical infrastructures following the designed framework, creating a manageable model was prioritised.

5.2 Experimental design

Since the model was designed for ecosystem-level analysis, the assumptions applied during model design involved the data points and parameter values used to assess behaviour across the ecosystem. Several factors that influence decision-making processes were also necessarily left out of the model, as implementations for these processes for ecosystem-level analysis would lose all representativeness. As such, the simulation model was not used to assess tangible defensive strategies as found in readily available sources, since these strategies are not defined that way. Instead, the model was used in an exploratory manner, following the Exploratory Modelling & Analysis (EMA) framework [47, 48]. This involves using models based on uncertain or generalised parameters to control for robustness across uncertain scenarios. Parameters used in the model were varied across to yield 250 unique scenarios. These scenarios were combined with four different defensive strategies, resulting in 1000 unique experiments. Each experiment was repeated 25 times to ensure extremely chaotic behaviour was suppressed, despite variability testing pointing out that the model showed robust variability and near-symmetrical behaviour across multiple runs. As mentioned previously, four different defensive strategies were established, in relation to threshold values for different responses. These are shown in Table 2. Prevention and detection mechanisms are each assigned values for sensitivity and specificity. For anomaly-based mechanisms, sensitivity values of 0.95 and specificity values of 0.8 were used, whereas signature-based mechanisms respectively attained sensitivity and specificity values of 0.8 and 0.95, based on the taxonomy by Patel, Taghavi [20].

Operability as a central modelling construct can be used to simulate defensive decision-making processes. In this case, perceived operability is used for deciding which response mechanism is required. Since the concept is in itself intangible, the threshold used is a simplification of a complex real-world process. Values used for these design parameters account for differences in sensitivity and specificity for each defensive strategy. These values were established through a process of trial and error, where a brief run of baseline experiments was used to establish which set of thresholds performed best in a static environment for each strategy.

Table 2: Defensive strategies in experimentation

Mechanisms	**Strategy 1**	**Strategy 2**	**Strategy 3**	**Strategy 4**
Prevention	Anomaly-based	Signature-based	Anomaly-based	Signature-based
Detection	Anomaly-based	Signature-based	Signature-based	Anomaly-based
Alleviation threshold	70%	80%	75%	70%
Retention threshold	30%	20%	25%	20%

5.3 Results

The results of the experimental design were assessed for sensitivity to certain scenario parameters, as is practice in using EMA [48]. It was found that no particular sensitivities led to extreme values or unexpected clusters of values for output parameters over time. After establishing whether sensitivities caused significant deviations, the impact of defensive strategies was analysed, in line with the main objectives for using the framework. Following EMA practices, the results were interpreted as the comparison of robustness of defensive strategy designs [49].

Results were analysed in terms of total losses incurred, cyberattack effectiveness and the quality of defensive decisions. Each graph depicts kernel density plots for all defensive strategies, indicating the distribution of performance across entire simulations. Since the parameter values were assumptious in nature, the numerical representation in outcome parameters does not say much alone. However, these outcome parameters show relative robustness among defensive strategies.

As depicted in Fig. 3, strategies 2, 3 and 4 yield comparable losses across the set of experiments, whereas strategy 1 results in substantially higher losses. This suggests that strategy 1, which employs both anomaly-based intrusion detection and prevention mechanisms, on average performs worse at the overall task of securing critical infrastructures. Since this metric indicates the average degree of inoperability under each defensive strategy, this does not help explain how interaction or decision-making affected by defensive strategies. By extent, any finding based solely on this metric is obfuscated by the assumptions under which the model was created.

Fig. 4, Fig. 5 and Fig. 6 visualise the density of the correctness of defensive decisions made across simulation runs for all defensive strategies. Interestingly, it is not the first strategy that results in the least correct decisions, as this is caused by strategy 4 instead. The pattern as to whether incorrect decisions are caused by primarily overestimated or underestimated operability varies between these two strategies. Nearly all incorrect decisions under strategy 4 result from underestimated operability, whereas strategy 1 results in substantial errors due to both underestimated and overestimated assessments. Strategies 2 and 3 are both characterised by high accuracy in decision-making, as the former results in few incorrect decisions in general and the latter only leads to some errors based on overestimation.

Fig. 7 depicts how node operational states were affected across model runs, shedding further light on differences among defensive strategies. Strategies 2 and 3 again show similar behaviour, with few recorded cases of inoperability and a majority of nodes in normal operation. Strategy 1 deviates from other strategies as it is the only strategy under which inoperability is a common occurrence. The deviation in losses observed in Fig. 3 is attributable to a significant number of inoperable nodes, whereas other strategies result in losses due to stressed operability. Strategy 4 also deviates from strategies 2 and 3, as the majority of nodes are stressed instead of unhindered. This was found to be attributed to the high frequency at which strategy 4 leads to defensive decisions based on underestimated operability, where minor action is consistently taken whereas no action should be warranted. These tendencies are consistent with the average deviation in impact assessment depicted in Fig. 8, where higher values and lower values imply a tendency to respectively overestimate or underestimate operability and values around 0 imply accurate impact assessment.

Further exploration of system behaviour involved assessing how effectively cyberattacks were dealt with for each defensive strategy. Fig. 9 shows the density of the average number of cyberattacks active within the model across entire simulation runs. This plot highlights how strategies 2 and 4 affect the ecosystem in a different manner than strategy 1 and 3. The average number of active cyberattacks is significantly higher for strategies 2 and 4, as attacks are thwarted less effectively due to signature-based prevention mechanisms. Due to the low and static frequency of attempted attacks (1 per time step on average) in the ecosystem, the way this translates over to inoperability is beneficial for strategy 2, for which underestimation of operability is not problematic for decision-making. As such, relying solely on one metric for performance is not insightful and the failure to prevent attacks efficiently should form a major caveat for any findings. In reality, the threat landscape could evolve to grow more capable or more frequent in case prevention does not work as envisioned. Likewise, Fig. 9 tracks all active attacks, which are all assumed to exert pressure on node operability constantly. Real-world attacks can be multi-facetted, surreptitiously infecting systems before such pressure is exerted. Findings for modelling studies are always a product of the assumptions by which the model is created.

The implications of these findings are not that one defensive strategy performs better than others. Instead, the results of this practical application are rooted in the operationalisation of an ecosystem-level model that enables new insights to be gathered through exploration [41, 48, 49]. While strategy 2 shows better performance in terms of decision-making and incurred losses, the operationalisation of these parameters relies heavily on ecosystem-level aggregation. Between strategies 1, 3 and 4, strategy 3 showed more robust performance overall, as the presence of cyberattacks is not substantially higher whereas the correctness of decisions leads to substantially lower inoperability in general.

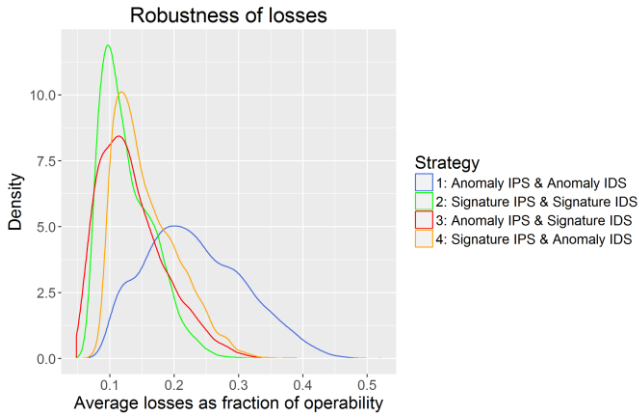

Fig. 3: Average losses

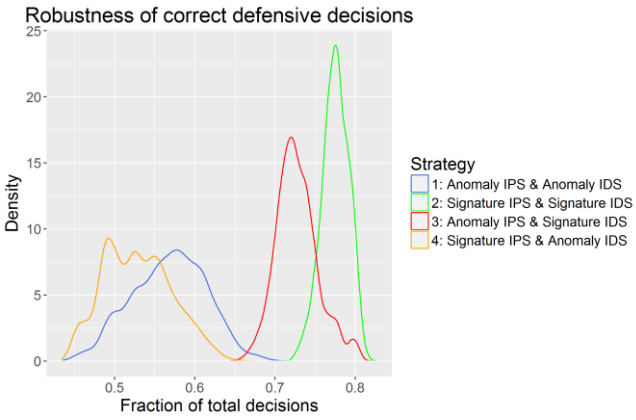

Fig. 4: Fraction of correct defensive decisions

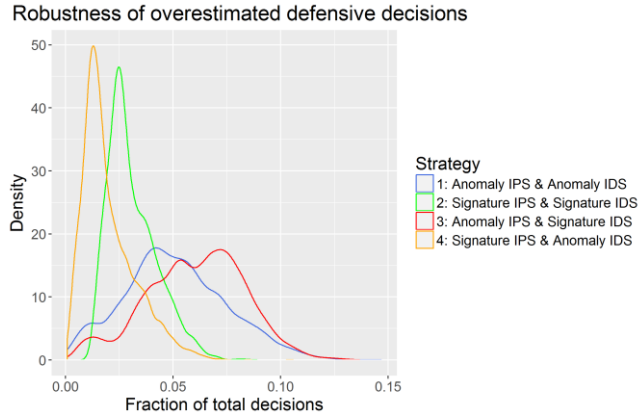

Fig. 5: Fraction of overestimated defensive decisions

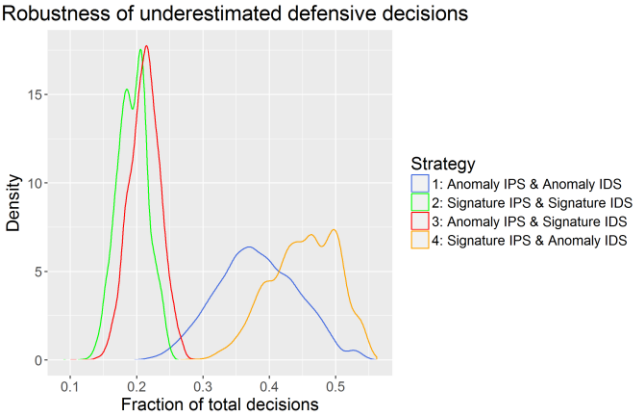

Fig. 6: Fraction of underestimated defensive decisions

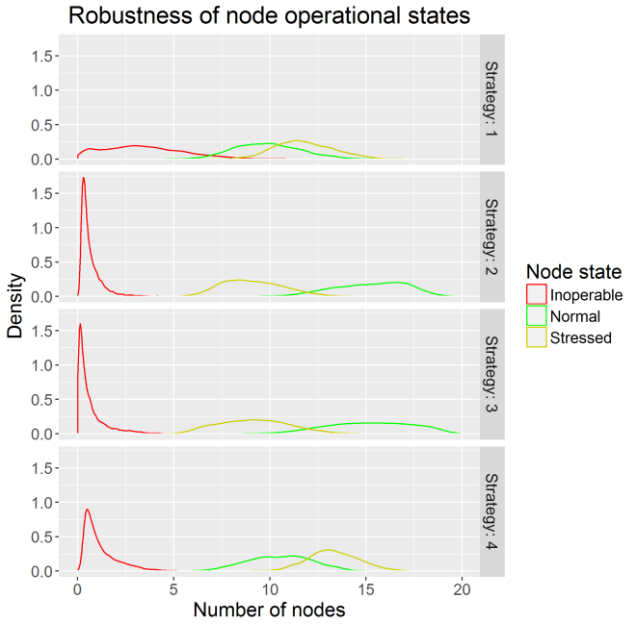

Fig. 7: Node operational states

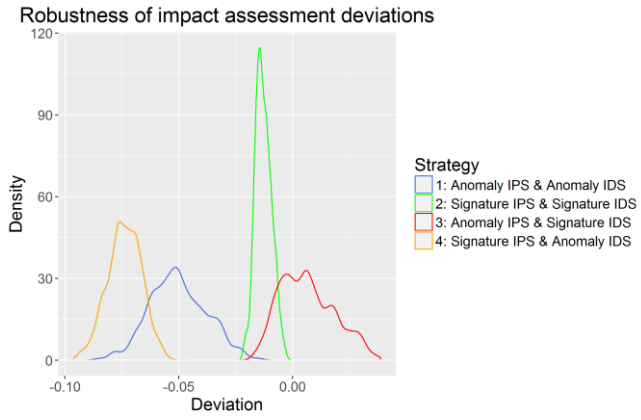

Fig. 8: Impact assessment deviation

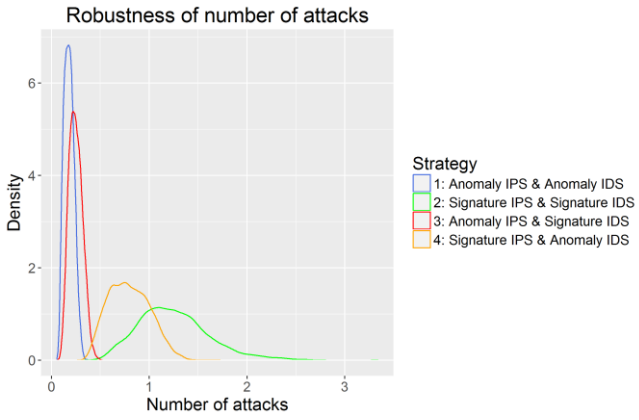

Fig. 9: Number of active attacks

6. Conclusion and discussion

It was established in section 1 that in order to cope with the evolving threat landscape looming around critical infrastructures, more knowledge on the ecosystem-wide effects of defensive strategies was required. However, there was no single framework that coherently encapsulated ecosystem-level properties for cybersecurity of critical infrastructures. To address this gap of knowledge, this study proposed a novel framework to conceptualise and operationalise a model for ecosystem-level analysis of cyberincidents involving critical infrastructures. The framework incorporates generalised elements that define most critical infrastructures to allow for analysis of coherent security policies. The operationalisation of the framework quantifies elements used for interaction modelling, based around a central degree of infrastructure node operability. This was demonstrated in the form of a practical application by creating an agent-based model capable of simulating the effects of cyberattacks and subsequent defensive decisions.

The main limitations to this framework are nested in the abstraction required to integrate elements on an ecosystem-level. Critical infrastructures are complex entities that incorporate multiple drivers for decisions made on an operational level. Boiling those decisions down to a central notion of operability enables exploratory modelling, but specific analysis and design of new control mechanisms for real-world application requires further specification of elements in the framework. This would then lead to a lower degree of generalisability of the framework. The main uses for ecosystem-level analysis are rooted in exploration of behavioural tendencies as opposed to quantification of performance indicators.

Operationalising the framework into an exploratory agent-based model proved the framework can be used to generate insight into the effects of different defensive strategies. The observed behaviour showed how different configurations for defensive strategies resulted in tendencies to either overestimate the threat landscape and incur damage through unnecessary defensive decisions, to underestimate the threat landscape and take no subsequent action. Crucially, the model assumed all attacks to be surmountable, which real-world interpretations consider problematic [9]. Given the assumptious nature of parameter values, these results do not provide comparisons of direct benefits for a certain design, but instead helps understand the effects, regardless of parameter values used. To this end, the modelling application proved that the operational framework could be used to generate consistent behavioural patterns, albeit with arbitrary values and simplifications. Further usage of the framework and creation of more specific simulation models could therefore result in tangible research benefits.

It is advisable to conduct further research into the effects of coherent, shared defensive strategies and security policies. For exploratory purposes, this study provides a framework that can be used and extended to incorporate elements that could shape the performance of such strategies. Use and extension of the framework is recommended for more advanced and specific research into preliminary designs of real-world defensive strategies. As such, the benefits of using the designed artefact can be communicated to those who can make use of it [17]. Another option is to devise similar simulation models as presented in paragraph 5. The framework presents a set of entities that should be accounted for, as well as the set of concepts that should be included in conceptual models.

REFERENCES

1. Moteff, J. and P. Parfomak. *Critical infrastructure and key assets: definition and identification*. 2004. LIBRARY OF CONGRESS WASHINGTON DC CONGRESSIONAL RESEARCH SERVICE.
2. Van der Lei, T.E., G. Bekebrede, and I. Nikolic, *Critical infrastructures: a review from a complex adaptive systems perspective.* International Journal of Critical Infrastructures, 2010. **6**(4): p. 380-401.
3. Department of Homeland Security, *The Future of Smart Cities: Cyber-Physical Infrastructure Risk*. 2015, Department of Homeland Security Office of Cyber and Infrastructure Analysis (DHS/OCIA).

4. Ericsson, G.N., *Cyber security and power system communication—essential parts of a smart grid infrastructure.* IEEE Transactions on Power Delivery, 2010. **25**(3): p. 1501-1507.
5. Fairley, P., *Cybersecurity at US utilities due for an upgrade: Tech to detect intrusions into industrial control systems will be mandatory [News].* IEEE Spectrum, 2016. **53**(5): p. 11-13.
6. Amin, S., et al., *Cyber security of water SCADA systems—Part I: Analysis and experimentation of stealthy deception attacks.* IEEE Transactions on Control Systems Technology, 2013. **21**(5): p. 1963-1970.
7. Li, X., et al., *Securing smart grid: cyber attacks, countermeasures, and challenges.* IEEE Communications Magazine, 2012. **50**(8): p. 38-45.
8. Brezhnev, E., et al., *Critical energy infrastructure safety assurance strategies considering emergent interaction risk*, in *Advances in Intelligent Systems and Computing*. 2018. p. 67-78.
9. Brown, G., et al., *Defending critical infrastructure.* Interfaces, 2006. **36**(6): p. 530-544.
10. Farwell, J.P. and R. Rohozinski, *Stuxnet and the future of cyber war*. Survival, 2011. **53**(1): p. 23-40.
11. Karnouskos, S. *Stuxnet worm impact on industrial cyber-physical system security*. in *IECON Proceedings (Industrial Electronics Conference)*. 2011.
12. Lee, R.M., M.J. Assante, and T. Conway, *Analysis of the cyber attack on the Ukrainian power grid.* SANS Industrial Control Systems, 2016.
13. Liang, G., et al., *The 2015 ukraine blackout: Implications for false data injection attacks.* IEEE Transactions on Power Systems, 2017. **32**(4): p. 3317-3318.
14. Baiardi, F., et al., *Assessing the risk of an information infrastructure through security dependencies.* Critical Information Infrastructures Security, 2006: p. 42-54.
15. Sandberg, H., S. Amin, and K.H. Johansson, *Cyberphysical security in networked control systems: An introduction to the issue.* IEEE Control Systems, 2015. **35**(1): p. 20-23.
16. Neuman, C. *Challenges in security for cyber-physical systems*. in *DHS Workshop on Future Directions in Cyber-Physical Systems Security*. 2009.
17. Peffers, K., et al., *A design science research methodology for information systems research.* Journal of management information systems, 2007. **24**(3): p. 45-77.
18. Cárdenas, A.A., et al. *Attacks against process control systems: Risk assessment, detection, and response*. in *Proceedings of the 6th International Symposium on Information, Computer and Communications Security, ASIACCS 2011*. 2011.
19. Ntalampiras, S., *Detection of integrity attacks in cyber-physical critical infrastructures using ensemble modeling.* IEEE Transactions on Industrial Informatics, 2015. **11**(1): p. 104-111.
20. Patel, A., et al., *An intrusion detection and prevention system in cloud computing: A systematic review.* Journal of network and computer applications, 2013. **36**(1): p. 25-41.
21. Rutkowski, A., et al., *Cybex: The cybersecurity information exchange framework (x. 1500).* ACM SIGCOMM Computer Communication Review, 2010. **40**(5): p. 59-64.
22. Singh, M.P. *Cybersecurity as an application domain for multiagent systems*. in *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. 2015. International Foundation for Autonomous Agents and Multiagent Systems.
23. Charitoudi, K. and A.J.C. Blyth, *An Agent-Based Socio-Technical Approach to Impact Assessment for Cyber Defense.* Information Security Journal, 2014. **23**: p. 125-136.
24. Janssen, S. and A. Sharpanskykh, *Agent-based modelling for security risk assessment*, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2017. p. 132-143.
25. Rinaldi, S.M. *Modeling and simulating critical infrastructures and their interdependencies*. in *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. 2004.
26. Setola, R. and M. Theocharidou, *Modelling Dependencies Between Critical Infrastructures*, in *Managing the Complexity of Critical Infrastructures*, R. Setola, et al., Editors. 2016. p. 19-42.
27. Sridhar, S., A. Hahn, and M. Govindarasu, *Cyber–physical system security for the electric power grid.* Proceedings of the IEEE, 2012. **100**(1): p. 210-224.
28. Puig, V., *Diagnosis and fault-tolerant control of critical infrastructures*, in *Advances in Intelligent Systems and Computing*. 2018. p. 3-16.
29. Pederson, P., et al., *Critical infrastructure interdependency modeling: a survey of US and international research.* Idaho National Laboratory, 2006. **25**: p. 27.
30. Eid, M. and V. Rosato, *Critical Infrastructure Disruption Scenarios Analyses via Simulation*, in *Managing the Complexity of Critical Infrastructures*, R. Setola, et al., Editors. 2016. p. 43-61.
31. Buttyan, L., et al., *Application of wireless sensor networks in critical infrastructure protection: challenges and design options [Security and Privacy in Emerging Wireless Networks].* IEEE Wireless Communications, 2010. **17**(5): p. 44-49.
32. Khurana, H., et al., *Smart-grid security issues.* IEEE Security & Privacy, 2010. **8**(1).
33. Stamp, J., et al., *Common vulnerabilities in critical infrastructure control systems.* SAND2003-1772C. Sandia National Laboratories, 2003.
34. Pawlick, J. and Q. Zhu, *Strategic Trust in Cloud-Enabled Cyber-Physical Systems with an Application to Glucose Control.* IEEE Transactions on Information Forensics and Security, 2017. **12**(12): p. 2906-2919.

35. Clark, R.M., et al., *Protecting drinking water utilities from cyberthreats.* Journal - American Water Works Association, 2017. **109**(2): p. 50-58.
36. Miller, B. and D. Rowe. *A survey of SCADA and critical infrastructure incidents*. in *Proceedings of the 1st Annual conference on Research in information technology*. 2012. ACM.
37. Liu, C.C., et al., *Intruders in the grid.* IEEE Power and Energy Magazine, 2012. **10**(1): p. 58-66.
38. Jones, J., *An introduction to factor analysis of information risk (fair).* Norwich Journal of Information Assurance, 2006. **2**(1): p. 67.
39. Alcaraz, C. and J. Lopez, *Wide-Area Situational Awareness for Critical Infrastructure Protection.* Computer, 2013. **46**(4): p. 30-37.
40. Vasilomanolakis, E., et al., *Taxonomy and survey of collaborative intrusion detection.* ACM Computing Surveys, 2015. **47**(4).
41. Nikolic, I. and J. Kasmire, *Theory*, in *Agent-based modelling of socio-technical systems*, K.H. Van Dam, I. Nikolic, and Z. Lukszo, Editors. 2013, Springer Science & Business Media: Dordrecht. p. 11-68.
42. Waldrop, M., *Complexity: The emerging science at the edge of order and chaos*. 1992, New York: Simon&Schuster Paperbacks.
43. Holland, J.H., *Complex adaptive systems.* Daedalus, 1992: p. 17-30.
44. Van Dam, K.H., I. Nikolic, and Z. Lukszo, *Agent-based modelling of socio-technical systems*. Vol. 9. 2012: Springer Science & Business Media.
45. Wilensky, U., *NetLogo. Evanston, IL: Center for connected learning and computer-based modeling, Northwestern University*. 1999.
46. Tisue, S. and U. Wilensky. *NetLogo: Design and implementation of a multi-agent modeling environment*. in *Proceedings of agent*. 2004.
47. Bankes, S., *Exploratory Modeling for Policy Analysis.* Operations Research, 1993. **41**(3): p. 435-449.
48. Kwakkel, J.H. and E. Pruyt, *Exploratory Modeling and Analysis, an approach for model-based foresight under deep uncertainty.* Technological Forecasting and Social Change, 2013. **80**(3): p. 419-431.
49. Augusiak, J., P.J. Van den Brink, and V. Grimm, *Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach.* Ecological Modelling, 2014. **280**: p. 117-128.

TU Delft

Appendix A: Formalised agent-based model procedures

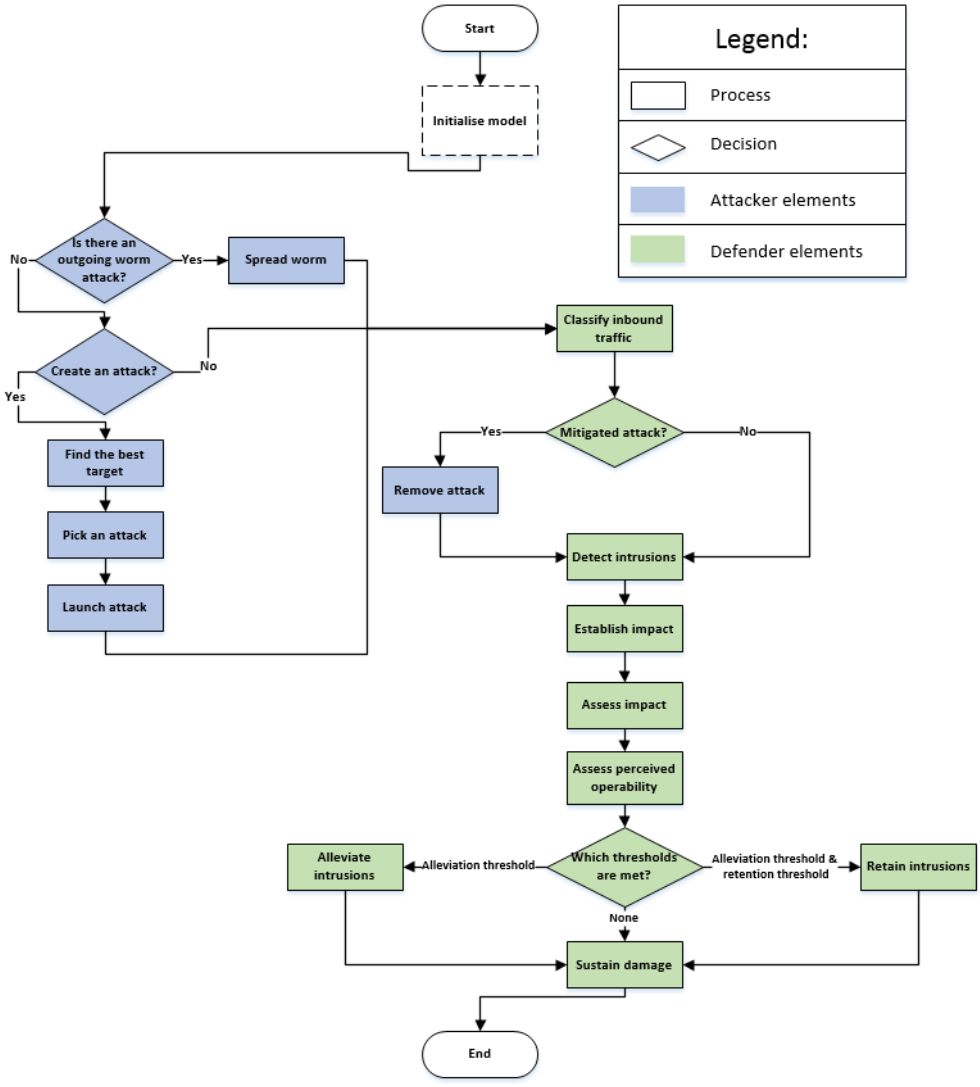

Fig. 10: Flowchart for agent-based model procedures