Journal of
Cybersecurity and Privacy

Article

Mapping Automated Cyber Attack Intelligence to Context-Based Impact on System-Level Goals

Pete Burnap [1], Eirini Anthi [1,*], Philipp Reineckea [1], Lowri Williams [1], Fengnian Cao [1], Rakan Aldmoura [1] and Kevin Jones [2]

[1] School of Computer Science & Informatics, Cardiff University, Cardiff CF24 4AG, UK; burnapp@cardiff.ac.uk (P.B.); williamsl10@cardiff.ac.uk (L.W.)
[2] Airbus, Quadrant House, Celtic Springs Business Park, Coedkernew, Duffryn, Newport NP10 8FZ, UK
* Correspondence: anthies@cardiff.ac.uk

Abstract: Traditionally, cyber risk assessment considers system-level risk separately from individual component-level risk, i.e., devices, data, people. This separation prevents effective impact assessment where attack intelligence for a specific device can be mapped to its impact on the entire system, leading to cascading failures. Furthermore, risk assessments typically follow a failure or attack perspective, focusing on potential problems, which means they need to be updated as attacks evolve. This approach does not scale to modern digital ecosystems. In this paper, we present a Data Science approach, which involves using machine learning algorithms and statistical models to analyse and predict the impact of cyber attacks. Specifically, this approach integrates automated attack detection on specific devices with a systems view of risk. By mapping operational goals in a top-down manner, we transform attack intelligence on individual components into system success probabilities.

Keywords: cyber risk; machine learning; statistical modelling; risk management; industrial control systems

Citation: Burnap, P.; Anthi, E.; Reineckea, P.; Williams, L.; Cao, F.; Aldmoura, R.; Jones, K. Mapping Automated Cyber Attack Intelligence to Context-Based Impact on System-Level Goals. *J. Cybersecur. Priv.* **2024**, *4*, 340–356. https://doi.org/10.3390/jcp4020017

Academic Editors: Chrysostomos Stylios, Kyriakos Stefanidis and Vasiliki Liagkou

Received: 27 March 2024
Revised: 21 May 2024
Accepted: 31 May 2024
Published: 6 June 2024

1. Introduction

Industrial Control Systems (ICS) sit at the heart of many Critical National Infrastructure (CNI) facilities, such as transportation, water, and energy. Significant numbers of devices within ICS systems are connected to the Internet and are, therefore, vulnerable to cyber attack. Attacks such as data leakage, spoofing, denial of service (DoS/DDoS), energy bleeding, insecure gateways, etc., can cause system blackouts and other indiscriminate and long-lasting damage. The effects of these security issues may cause major interference to the operation of services (e.g., public transportation networks can be targeted to cause chaos during peak travel periods, attacks to manufacturing plants can lead to downtime within the system, etc.) and therefore require appropriate risk management solutions.

The UK Government's National Cyber Security Centre (NCSC) [1] categorises cyber risks into: Component-driven risk management, which focuses on the threats and vulnerabilities that, for example, ICS devices may face; and System-driven risk management, which takes a whole-system view and focuses more on the systemic risks likely to be faced in the context of goals and objectives of the digital environment—requiring the definition of a higher level purpose and subsequent understanding of sub-systems and how various parts interact. Rasmussen's work [2] proposes a hierarchy of abstraction that enables us to consider how both perspectives are complimentary—with the systems view (goals and objectives) being fundamentally underpinned by the component view of risk (devices, data, people, etc.).

In this paper, we propose a novel way to link these views by investigating how Data Science can support the detection of risks to components (with machine learning for the detection of cyber attacks on devices) and understanding how this risk impacts the entire system's goals and objectives. Jaquith provides a perspective on what constitutes good and bad risk metrics [3], including the need for consistent measurements that are collected in an automated way and contextually relevant enough for decision makers to take action. Previous research has focused on developing systems and methods for measuring malicious network activity in evolving digital ecosystems (e.g., [4,5]). Although these methods successfully identify cyber attacks, they do not necessarily produce contextually relevant information in order to enable decision makers to take action. They focus mostly on detecting the attack rather than the relationship between the possible attack and its impact on the wider system. In other words, the research field currently lacks a methodology to integrate metrics and measurements relating to cyber attacks with systems and component views of risk assessment. Tools such as Symantec and Metasploit (metasploit.com) have developed a severity scale to rate the 'impact' of various attacks. However, this rating is a general score, which refers only to the attack and does not consider its risk in the context of a specific digital ecosystem. Advancing the domain and enabling the measurement of impact of cyber attacks in rapidly evolving environments, for instance, converged IoT and industrial control systems, is crucial in order to be able to plan countermeasures and understand the hardware/system implications and financial impact of cyber attacks. Context is everything.

This paper introduces a novel methodology that determines the severity of a cyber attack in a digital environment. While the methodology is exemplified in a converged IoT/ICS context for the purposes of this paper, it is designed to be broadly applicable to various digital environments, including, but not limited to, IoT contexts. The approach can be adapted to different digital ecosystems by considering the specific operational goals and interdependencies present in those environments. It achieves the above by linking: (i) measurements automatically collected from machine learning classification methods to detect attacks on IoT devices; (ii) systems-level goals and their interdependencies with component-level devices; and (iii) statistical distributions to calculate the reduction in probability of achieving system-level goals when component-level attacks occur on (IoT) devices. As far as we are aware, this is the first proposal for an integrated Data Science-based attack detection system coupled with a harmonised systems-component risk assessment method, giving 'real-time' contextually relevant updates on the cascading impact of cyber attacks on devices across a digital ecosystem—paving the way for *trusted* (industrial control) systems.

2. Related Work

A considerable amount of work has been reported on cyber security risk assessment methodologies in various domains: traditional IT, industrial control systems (ICS), IoT, etc. The majority of cyber risk assessment methods and frameworks fall under two main high-level measurement methodologies—qualitative and quantitative. The former refers to methods that rate the risk as low, medium, or high, and the latter attempts to measure the severity of the risk numerically.

Risk Assessment Methods for Traditional IT

Dantu, Loper, and Prakash [6] designed attack graphs (which represent all the possible attacker actions on a specific system) based on attacker behaviour and used Bayesian methodology to identify vulnerabilities and estimate the risk level of a critical resource in a given network. Similarly, Kotenko and Chechulin [7] proposed a risk assessment framework that employs attack graphs and monitors the timeline of events in order to measure the risk of a cyber attack in real-time. Hariri et al. [8] designed a tool that monitors various network vulnerability metrics and measures the impact of faults and attacks. Ryan and Ryan [9] propose a quantitative approach for risk management, which suggests calculating expected losses and other failure-related metrics. Wang et al. [10,11] developed probabilistic methodologies based on attack resistance metrics to measure the security risk in networks. Asosheh et al. [12] implemented a quantitative approach based on the popular Microsoft and Callio Secura methodologies in an attempt to measure the the security risk in a business environment. Factor Analysis of Information Risk (FAIR) [13] is a model that promotes a quantitative, risk-based, acceptable level of loss exposure. RiskLens [14] is another quantitative assessment method based on FAIR and provides a model for understanding, analysing, and quantifying cyber risk in financial terms. Tsakalidis et al. [15] presented a decision support system that qualitatively evaluates the severity of cyber security threats regarding frequency of appearances/references and number of incidents. Wynn et al. [16] developed a Threat Assessment and Remediation Analysis (TARA) methodology, which identifies and assesses cyber vulnerabilities and further suggests appropriate countermeasures. Common Vulnerability Scoring System (CVSS) [17] is a framework that captures the principal characteristics of software vulnerabilities and produces a numerical score that reflects their severity.

Although extensive work in this area exists, it is largely based on pre-defined or known attack paths and failure modes, which means new attacks require the updating of large and often complex models. Also, existing work often treats risk on an asset-by-asset basis, not capturing the interdependencies between them, and is thus not able to model cascading failure. The methodology proposed in this paper aims to capture system goals (as opposed to possible failure modes or attack paths) and measure the impact of attacks in the context of inter-dependencies between goals. Thus, as intelligence emerges of new attacks we need only map these to operational goals, where interdependencies are already captured, enabling risk managers to be able to determine the impact of unknown or unexpected attacks as they occur. This overcomes existing limitations of needing to update risk models when new attacks occur and providing impact metrics based on the context in which the attack is occurring.

In an industrial control systems context, various approaches of qualitative and quantitative risk management have been proposed from an attack-oriented security and safety perspective that exhibit the same limitations (detail omitted for brevity) [18–25].

3. Proposed Method

The overarching aim in this paper is to advance the way we link automated alerts and detection of cyber attacks with systems and component-level risk assessment methods in a contextually relevant manner. We aim to measure the direct impact on a component, alongside the broader impact of a successful attack on the processes and operational goals that depend on it (a systems view). The methods in the paper are established to propose a defensible approach of mapping the outputs of an automated (e.g., rule or machine learning-based) component-view attack detection tool, to probabilities that system-level or organisational goals will be achieved. To achieve this, we first set out a range of statistical distributions, then take input relating to the confidence that an attack is occurring (in this case, the output from a machine learning classifier of the range 0 to 1), and map this to a goal success probability score (of the range 0 to 1). As an example of a machine learning attack detection tool, [26] developed an approach that was highly accurate at detecting attacks on popular commercially available IoT devices, while deploying 12 attacks from 4 main network-based attack categories, such as: Denial of Service (DoS), Man-In-The-Middle (MITM)/Spoofing, Reconnaissance, and Replay. Each output from the classifier includes the attack type and a confidence of the network features it has analysed that relates to attacks of this type (between 0 and 1).

Not all attacks are equal. For instance, a spoofing attack is much less likely to achieve a massive impact on goals immediately than a DDoS attack that has an immediate and significant impact on a set of goals. Thus, the rationale behind the statistical distributions is to capture this variance. The paper is intended as a first step towards using cyber attack intelligence in a contextually relevant model to produce actionable risk management information relating to its impact. To help develop the framework, we ground the work in cyber attacks on IoT devices. The reason for this is because widespread adoption of IoT (for example, in the manufacturing sector) provides an excellent example to highlight the limitations we presented in related work—that new attacks require a rethink of an entire risk model when considering attack/failure modes and interdependence between assets. The addition of IoT (proposed as 50 billion new devices by 2050 in some cases [27]) highlights the scale at which such adaptation would need to be undertaken. This is not feasible—hence the need for our proposed method.

There are several dependency modelling techniques available, each with its own strengths. Bayesian networks are graphical models that represent probabilistic relationships among variables and are particularly useful for modelling uncertainty and updating probabilities based on new evidence. Fault Tree Analysis (FTA) is a top-down, deductive failure analysis technique that uses Boolean logic to combine a series of lower-level events to analyse an undesired state of a system. Markov chains are stochastic models that represent systems with states and transitions, useful for modelling randomly changing systems where future states depend only on the current state. However, to make use of the proposed statistical distributions, we integrate the Dependency Modelling methodology proposed in the Open Dependency Modelling (O-DM) standard, published by the Open Group [28]. The concept of Dependency Modelling is explained and contrasted with standard failure-oriented, asset-based risk modelling in [29]. The advantages of this include its flexibility, which allows for modelling complex interdependencies within a system. This is crucial for accurately assessing cyber risks in converged IoT/ICS environments; it supports dynamic updates to the dependency model as new attack intelligence is gathered, making it highly adaptable to evolving cyber threats. It employs a Bayesian framework to propagate probability changes through the dependency tree, enabling a comprehensive understanding of how individual component risks impact overall system goals. These features make ODM particularly well-suited for our objective of linking automated attack detection to system-level risk assessments, thereby providing actionable insights into the cascading impacts of cyber attacks on digital ecosystems.

Furthermore, the Dependency Modelling method develops a top-down model of a system from an overall goal to its first-level dependencies by asking the question 'what does this goal depend on?'; and then to next-level dependencies, and so on, until a tree of dependencies is built. 'Leaf' nodes sit at the bottom of the tree and do not have explicit dependencies in the model. These are 'uncontrollables' and require only a probability to be added (a value between 0 and 1 representing the probability of success—where 0 is guaranteed failure and 1 is guaranteed success). The tree can be treated as a directed graph, meaning changes in probability at the leaf nodes are used (in a Bayesian model) to propagate probability changes back up the tree. Thus, by assuming certain dependencies can be underpinned by cyber-physical components (e.g., IoT devices) to partially fulfil goals (e.g., in converged IoT/OT system), it is possible to suggest that intelligence about attacks detected on such components could be mapped onto the dependency model to adjust the probability value for leaf nodes and therefore update the whole model to show the impact the attack could have on the system, in real-time and in the context of all other operational goals. We demonstrate examples of this in action later in this paper. The open research question is how to turn cyber attack intelligence into a probability; a goal underpinned by a cyber-physical component will succeed or fail.

This is our proposed contribution. There are two major step-changes in cyber risk modelling in this achievement. First, we move away from attack- or failure-based models that are vulnerable to incomplete and ever-changing knowledge about 'what could go wrong?'. Coming at the problem from the perspective of 'what needs to go right?' is more intuitive and likely to be readily accessible through individuals responsible for managing risk to these systems. As we move towards more intelligent ways of detecting cyber attacks, this information can be utilised, regardless of the attacker methodology. Second, the impact of a changing probability shows the impact on other operational goals in the context of a systems view. Previous approaches may use financial or qualitative (low, medium, high) measure for impact. But where does this information come from? Without a clear understanding of the impact on interdependent component parts of the system, these figures often relate only to the failure of a single asset (e.g., what if this server goes down?).

The reality is, this failure will likely have a knock-on effect. Our proposed method captures this to provide more context for financial and qualitative decisions to be made.

3.1. Mapping Attack Types to Distributions

Figure 1 represents a sample distribution, which demonstrates how the probability of the attacked device being operational drops as the confidence of the occurring attack increases over time.

Identifying the correct type of distribution that fits different attack types can be challenging, as there are many parameters that may need to be taken into consideration. DREAD [30] and CVSS [17] take a qualitative approach to assessing the severity of vulnerabilities. We include two qualitative metrics (see Table 1): (i) the difficulty in detecting the attack—this may be based on the attack itself being a known surreptitious attack (hard) or something more obvious, such as DoS (easy); and (ii) the damage potential—this is based on the device itself, not the wider damage potential. We need to make a decision on whether the specific attack on a specific device will completely prevent the device from fulfilling its task (high) or just have a minor impact (low). Context is required here and therefore requires input from the process owner. For instance, interfering with input/output data flow on a safety critical device may have high damage potential, while a printer i/o may be low. Note the focus of damage is on the device itself (component), not the wider system.

Table 1. Elements and evaluation in the proposed model.

Damage Potential	Low = attack has minor impact on the device; Medium = attack could lead to significant disruption on the device; High = attack completely prevents the device from fulfilling its task.
Detection Difficulty	Easy = attack is detected at an early stage or can be predicted; Medium = attack can be detected but requires well-curated and managed methods for detection; Hard = attack is only likely to be detected once it begins to impose a major impact.

3.2. Classification of Distributions

We use the qualitative damage potential and detection difficulty labels to select goal probability–attack confidence distributions using three models: Exponential Correlation, Logarithmic Correlation, and Perfect Negative Correlation.

3.2.1. Exponential Correlation

We propose exponential distributions to be used when the detection of an attack has an easy or medium difficulty. This choice is grounded in the properties of exponential distributions, which are suitable for modelling time-to-failure due to their 'memory-less' nature [31], making them appropriate for certain IoT devices that do not accumulate damage over time. However, we acknowledge that different types of cyber attacks and operational contexts might benefit from alternative distributions. For instance, Weibull [32] distributions could model scenarios with varying failure rates, and Gaussian [33] distributions might be suitable for modelling normally distributed operational abilities. The exponential distribution was selected because it effectively captures the rapid changes in operational status under attack scenarios.

As Figure 1 (Equation (1)) and Figure 2 (Equation (2)) show, the curve remains high on the probability of being operational until we are highly confident that the attack is occurring. That is, the machine classification algorithms used to detect these attacks are confident in their output. Given the attack is likely to be detected early, we can wait until we are highly confident before reducing the probability that the goal will succeed. Notice that there are three lines on the chart. These relate to decay rates based on damage potential. Slow decay (top line) is for low-damage attacks, rapid decay (bottom line) is for high-damage attacks. The decay variance captures the increased reduction in probability in correlation with increased confidence that the attack is happening and the extent of damage potential. Also notice that there are two ways to map the change in probability. Figure 1 reduces the probability to zero for the high-damage decay line when the confidence is 100%, while the slow-decay line only ever reaches around 60%. This indicates that some attacks may not ever reduce the probability of success to zero—for instance, low-damage probing attacks. Figure 2 is used for cases when all forms of decay will eventually reduce the probability of device success to zero.

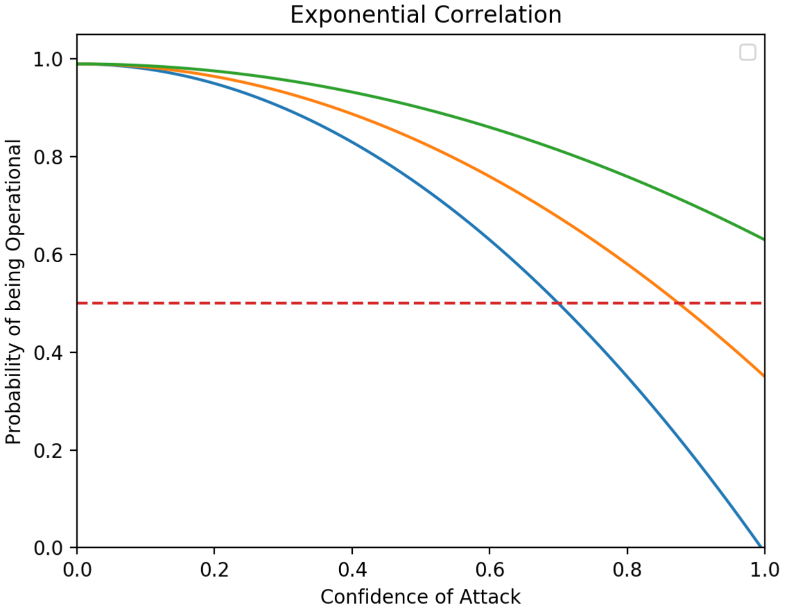

Figure 1. Exponential correlation by different Damage damage potential—not to zero (low-damage attacks (top line (green)) to high-damage attacks (bottom line (blue)).

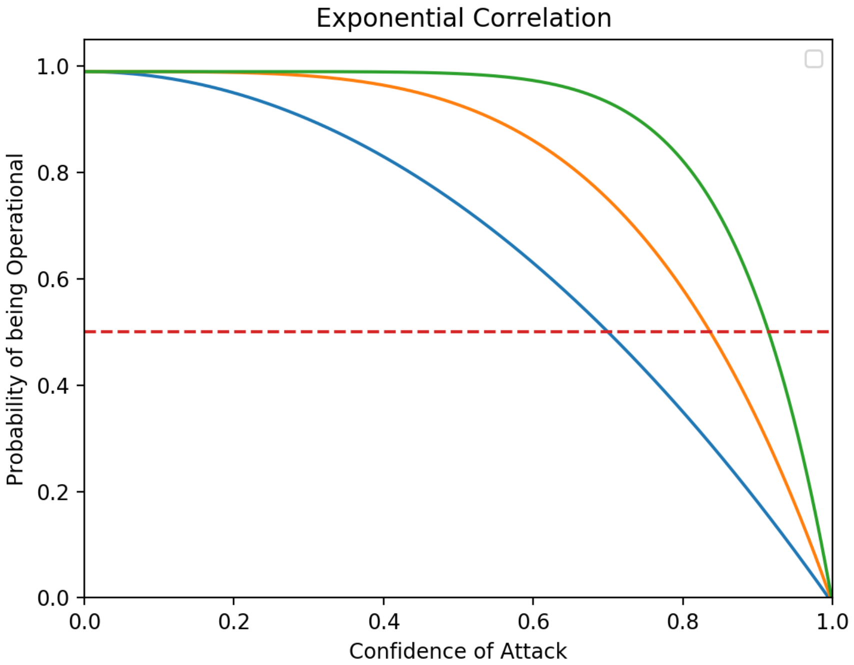

Figure 2. Exponential correlation by different damage potential—to zero (low-damage attacks (top line (green)) to high-damage attacks (bottom line (blue)).

$$f(x) = -kx^2 + 0.99, x \in [0,1], f(x) \in [0,1], k = 1, 0.8, 0.6 \quad (1)$$

$$f(x) = -x^k + 0.99, x \in [0,1], f(x) \in [0,1], k = 2,4,8 \tag{2}$$

3.2.2. Logarithmic Correlation

For hard-to-detect attacks, we propose a logarithmic distribution. As Figure 3 (Equation (4)) and Figure 4 (Equation (5)) show, the probability of being operational decreases much more quickly, even when we are less confident in an attack. This is because indicators of compromise for more difficult-to-detect attacks may have to be taken as early warning signs of significant impact, and we need to be able to respond to these. The same principles as exponential correlation apply, regarding decay speed and damage potential (three lines) and whether the attack reduces the probability to zero or a number greater than zero.

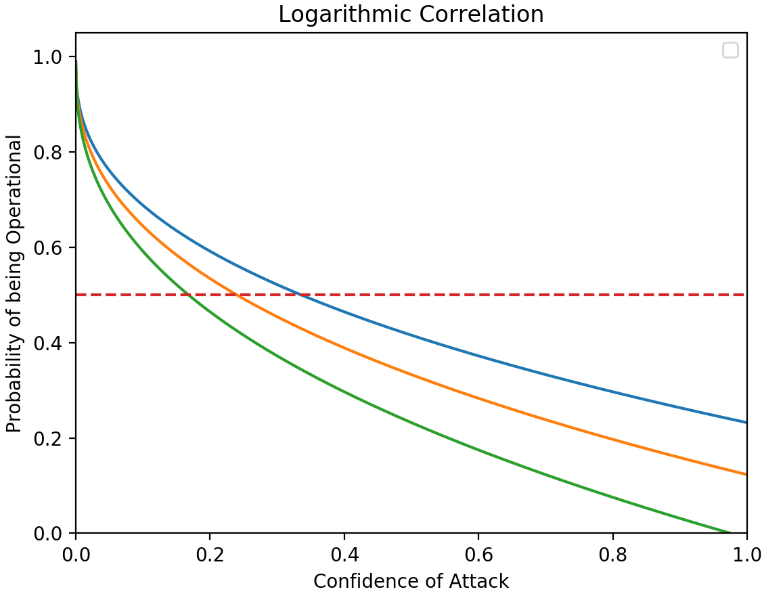

Figure 3. Logarithmic correlation by different damage potential—not to zero (low-damage attacks (bottom line (green)) to high-damage attacks (top line (blue)).

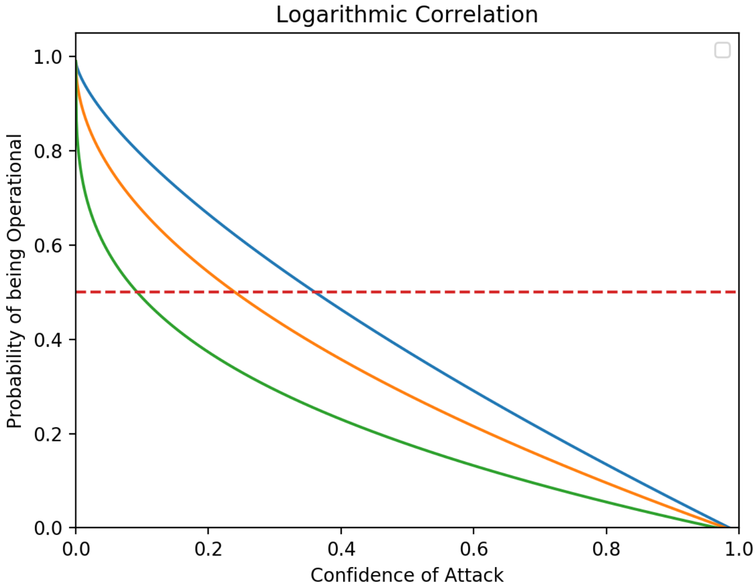

Figure 4. Logarithmic correlation by different damage potential—to zero (low-damage attacks (bottom line (green)) to high-damage attacks (top line (blue)).

$$\begin{aligned} 0.5 &= \log_x (0.99 - y)/k, y \in [0, 0.99] \\ &=> f(x) = -k\sqrt{x} + 0.99, x \in (0, 1], \\ &f(x) \in [0, 1], k = 0.5, 0.7, 1 \end{aligned} \tag{3}$$

$$\begin{aligned} k &= \log_x 0.99 - y, y \in [0, 0.99] \\ &=> f(x) = -x^k + 0.99, \\ &x \in (0, 1], f(x) \in [0, 1], k = 0.7, 0.5, 0.3 \end{aligned} \tag{4}$$

3.2.3. Perfect Negative Correlation

$$f(x) = -kx + 0.99, x \in [0, 1], f(x) \in [0, 1], k = 1, 0.7, 0.5, 0 \tag{5}$$

Finally, there will be cases where we may not be able to detect the attack at all. For this, we propose a Perfect Negative distribution (Equation (5)), where x ranges from 0 to 1, and $f(x)$ also lies within the interval [0,1]. The parameter k can take on the values 1, 0.7, 0.5, or 0, influencing the slope of the linear function. As Figure 5 shows, these take the form of a skewed linear decay where the range of gradients can have no effect (top line) or a complete failure effect (bottom diagonal line), and ranges in between depending on damage potential. Figure 6 provides a logical flowchart to support the mapping of attack detection difficulty and damage potential to the choice of probability–confidence distribution, as well as the decay speed distribution.

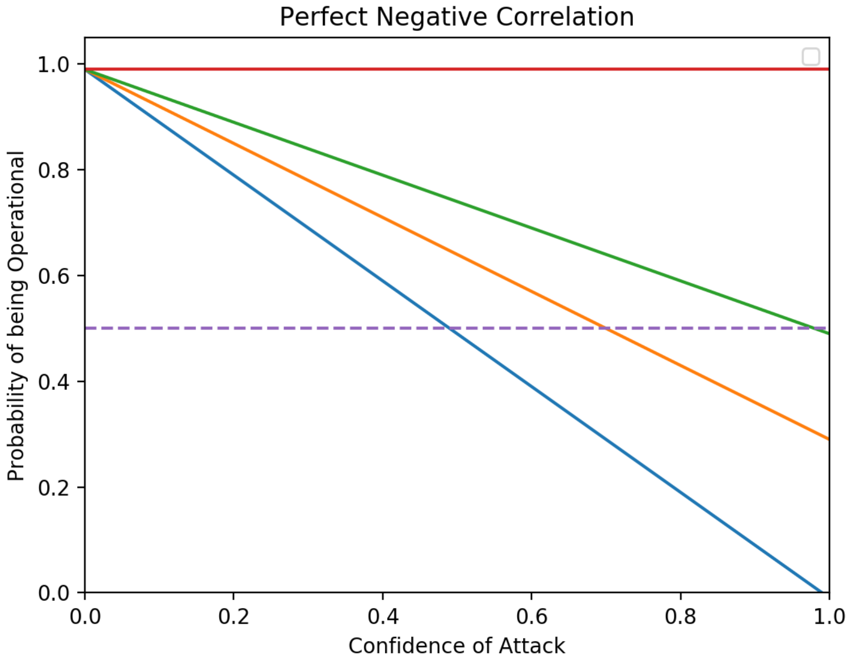

Figure 5. Perfect negative correlation by different damage potential (gradients range from no effect (top line (red)) to a complete failure effect (bottom diagonal line (blue)).

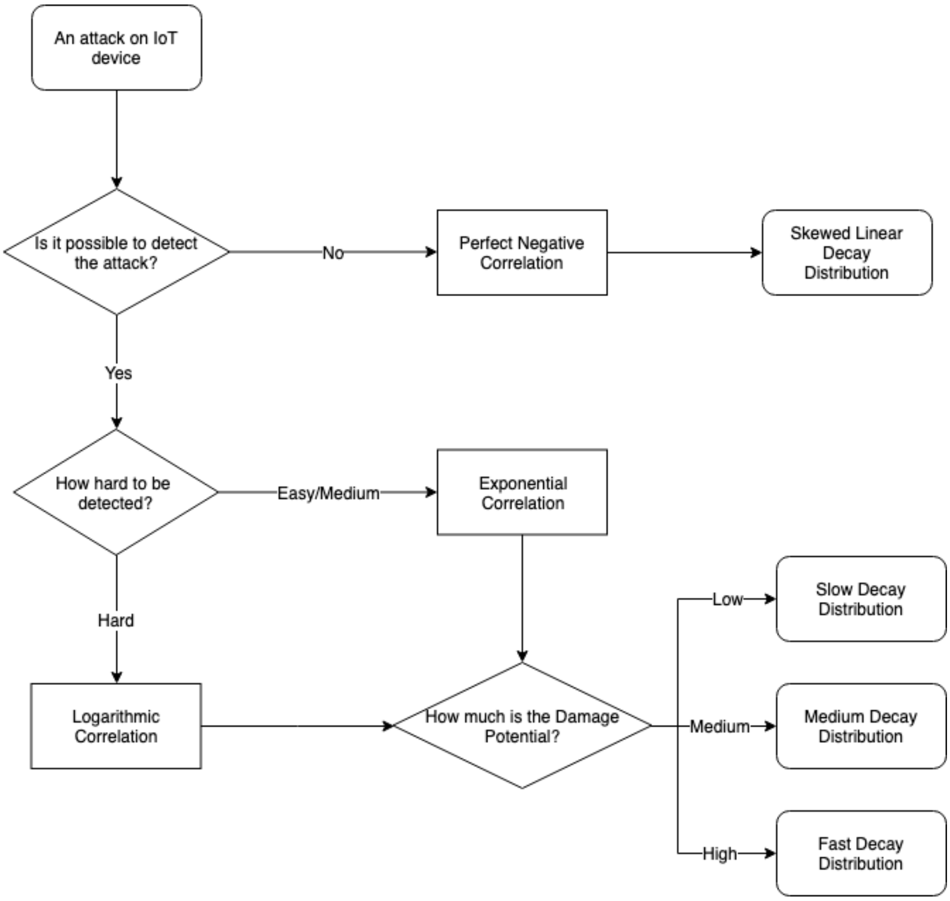

Figure 6. Procedure of mapping attacks to distributions.

4. Mapping Cyber Attacks to Statistical Distributions

In this section, we use IoT device attacks to walk through examples of selecting the different distributions presented in the previous section. While IoT provides a useful case for evolving technology and associated risk of attack, the distributions could be used for any digital environment within which the difficulty of detecting attacks and damage potential for components can be reliably assessed. Studies [34–36] have shown that IoT devices are vulnerable to a range of cyber attacks that fall under four main categories: physical, network, software, and encryption attacks [36]. In this section, we provide attack examples for each type of distribution presented in the previous section.

4.1. Network Attacks

4.1.1. DoS Attack

DoS attacks aim to disrupt services of IoT devices temporarily or indefinitely [36]. As Figure 7 shows, and as explained in Table 2, the distribution belongs to a logarithmic distribution because the attack is hard to detect prior to its impact (it is fast-acting). It could be effective even when the confidence is small, so it may impose an impact on the probability of an affected device being operational. The attack has high damage potential, so it is mapped to a fast-decay logarithmic distribution. A DoS attack may severely disable the device, so the distribution takes the probability of being operational to 0% when confidence in the attack is 100%.

Table 2. Denial of Service Attack (DoS)—qualitative summary.

Attack	Denial of Service (DoS)
Damage Potential	High—DoS attack can severely impact the internet of things devices [37]; can completely interrupt the communications in the network [38]; and can severely disable the device by blocking communication.
Detection	Hard—DoS attack is very hard to detect before the network is unavailable [39].

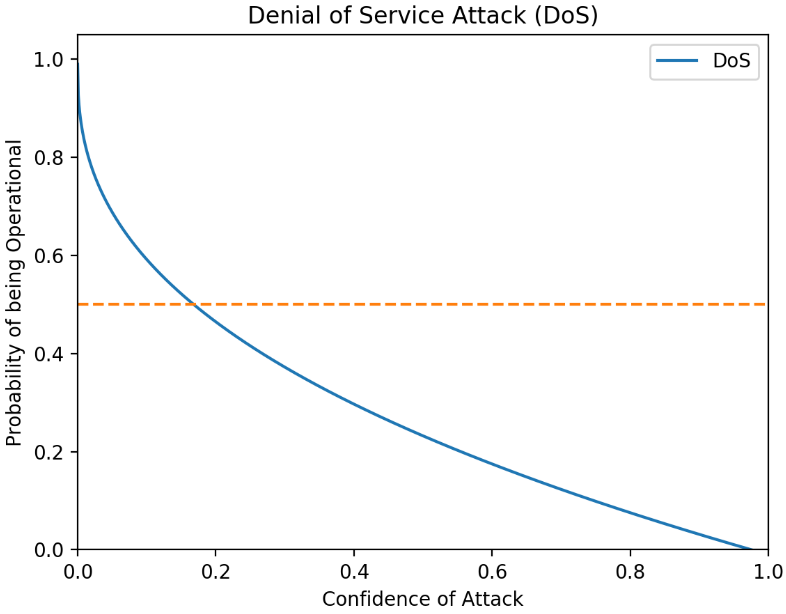

Figure 7. Denial of Service attack (DoS).

4.1.2. Remote Control/Packet Alteration

This kind of attack targets devices that lack authentication mechanisms (e.g., TP-Link Smart Plug). As a result, the IoT devices accept a range of commands that can either change their state or provide sensitive information. Examples of such commands include: turning the device on/off, returning device information, returning cloud connectivity information, scanning for nearby access points, resetting the device to factory settings, returning real-time voltage/current/power, and returning system time [40]. Evidently, this attack can directly impact the IoT device and make it behave in an irregular manner, whilst it can also provide the attacker with crucial system and device information.

As Figure 8 shows, and as explained in Table 3, the attack is of medium difficulty to detect, as control packets will stand out in the netflow traffic, so we map this to an exponential distribution. The attack has high damage potential, possibly completely disabling the device, so it is mapped to a rapid decay.

Table 3. Remote Control attack—qualitative summary.

Attack	Remote Control
Damage Potential	High—The attack can change the state of the device, making it behave abnormally, and can also provide the attacker with sensitive information [41].
Detection	Medium—The probability of detecting the attack is higher if the attacker sends more than one request and gets lower if the attacker sends one request.

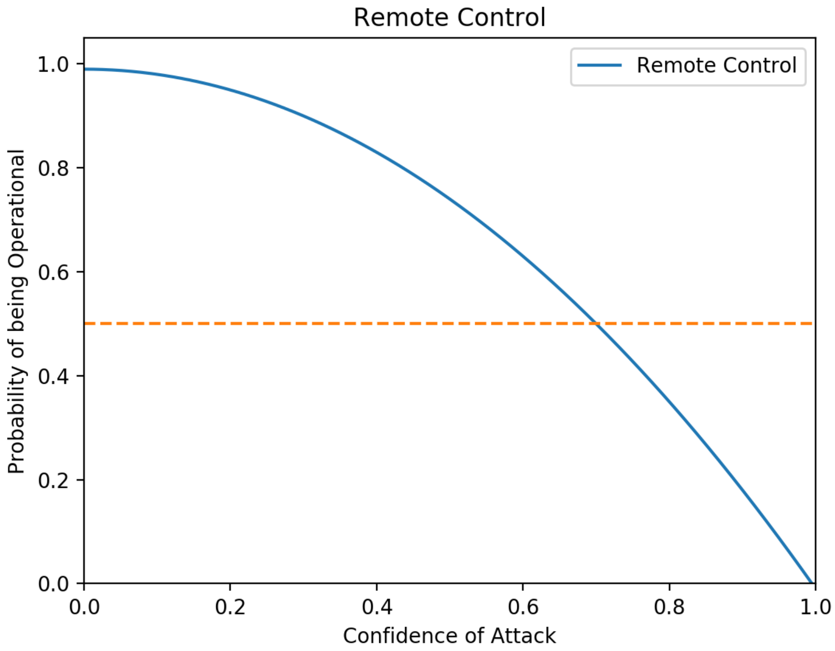

Figure 8. Distribution of Remote Control.

4.2. Encryption Based Attacks

A key feature in the majority of IoT devices is their ability to communicate using wireless technology. This introduces several risks, including data leakage, as an attacker could potentially eavesdrop on information sent and received from the IoT. Specifically, if the Secure Sockets Layer (SSL)/Transport Layer Security (TLS) protocol is not employed, data transmitted from across the IoT are not encrypted. As a result, an unauthorised party could intercept it by performing passive network sniffing on the operating channel [2,19]. This can have serious consequences, as an attacker can compromise sensitive personal and identifying information, such as passwords, location, name, etc. Therefore, the damage potential in this scenario is high. On the other hand, if the SSL/TLS protocol is employed, the transmitted data are encrypted. Thus, an eavesdropping attacker will not be able to intercept information in clear-text, in which case the damage is low.

For wireless communications, an attacker can attempt to perform a MITM attack. Via this attack, the original connection between the two parties gets split into two new ones: one connection between the first party/device and the attacker and another one between the attacker and the second party/device. When the original connection is finally compromised, the attacker is able to act as a proxy and therefore read, insert, and modify data in the intercepted communication. Given the varying damage potential based on system design choices (SSL vs No SSL), a mixed probabilistic distribution has been allocated to each one. Table 4 captures the damage potential and detection summary, and Figure 9 visualises how the risk of these attacks changes over time, as the confidence in an MITM attack increases. This example shows that our proposed methodology can take into account specific system design aspects in this way. MITM attacks are easy to be detected using traffic flow analysis, so they are mapped to an exponential distribution. The probability of a device being operationally impacted is medium damage (as the instructions/actuation may change based on changing payloads), so they follow a medium-decay exponential distribution.

Table 4. Encryption attack—qualitative summary.

Attack	Passive Sniffing (SSL/TLS Not Employed)	MITM (SSL/TLS Is Employed)
Damage Potential	High—If SSL/TLS is not employed, the attacker can intercept the channel and perform the attack [42,43].	Low—If SSL/TLS is employed, the attacker would sniff traffic and gain only encrypted information [42,43].
Detection	Easy—Using timestamps of TCP packet headers, finding unusually long delays, the attack can be detected accurately once the attack happens.	

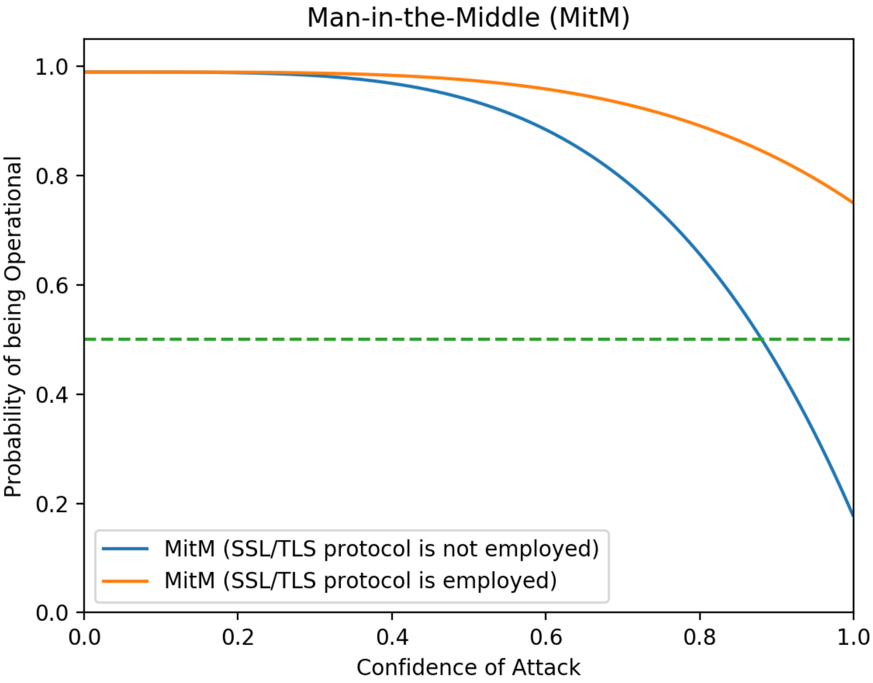

Figure 9. MITM attack.

4.3. Physical Attack

Table 5 provides a qualitative summary of the risk potential for the physical attack applied to the sensor. As Figure 10 shows, the distribution for a physical attack belongs to a perfect negative distribution because the attack is out of the range of detection until it becomes damaged and network monitoring tools flag it as offline. Once the attack is flagged, a linear line connects the probability of being operational to confidence in the attack, such that a direct relationship from confidence = 1 to probability = 0 exists.

Table 5. Physical attack-qualitative summary.

Attack	Physical Damage
Damage Potential	High—The attack can harm hardware and components in an IoT system physically, making them completely unavailable [41].
Detection	Impossible—The attack is implemented by physical approaches, so it is undetectable until the device fails.

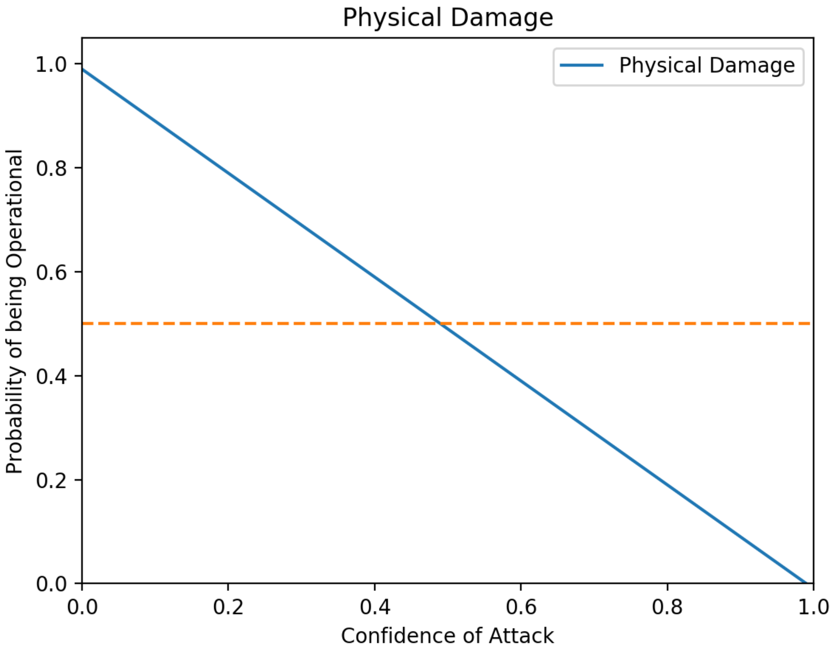

Figure 10. Distribution of physical damage.

5. Mapping Probabilities to System-Level Failure

As we now have a methodology for mapping attacks on devices at a component level to a probability that these devices will be operational, we can now utilise this information to understand the impact at a system-level view. Burnap et al. [29] present Dependency Modelling (DM) as a method of determining the risk to an enterprise through system-level analysis. A top-down model of the system is built from an overall goal to its first-level dependencies, and then to next-level dependencies, and so on. Dependencies are assigned a number of possible states (failure or success being the simplest) and a conditional probability of being in those states. Repeating this for sub-processes produces a tree or graph-based model. This model can be interpreted both as a Bayesian network and as a failure-mode model (see the paper for more detail). The probabilities are used by a Bayesian analysis engine to determine a range of sensitivities or vulnerabilities, illustrating which elements in the model are most pivotal to its success, while a mode-analyser intuitively illustrates the most likely cause of failure in a model, given that a hypothetical (for simulation purposes) or actual (for response purposes) failure has occurred. We propose this modelling method can be enhanced by incorporating real-time *contextually relevant* information produced through the methodology proposed in this paper. For instance, Table 6 captures the metrics from the examples in the previous section where IoT devices were under cyber attack. As confidence in the attack grows, the probability that the device is operational decreases. If these devices were underpinning a goal in a dependency model, we could use this information to translate the outcome of an automated attack detection method (such as machine learning methods for attack detection in IoT networks) into useful input for the dependency model.

As an example, Figures 11 and 12 present two dependency models, representing hypothetical system-level dependencies in a water treatment plant. The plant contains a system-level goal (monitoring pH of water) that depends on a specific IoT device (a sensor). If the device fails, the goal fails. Using the methodology proposed in [29] for top-down Dependency Modeling, we show how intelligence mapped to probabilities of device success or failure changes the likely success of this system-level goal. In this scenario, an automated attack detection tool has identified a remote control attack with 80% confidence as an early warning alert. We push this information in the distribution, as suggested in

Section 4.2 and Figure 11. The resulting output from the mapping of confidence in this attack to the probability of this goal being successful is 35% (as shown in Table 6). Figure 12 shows the probability of this goal being successful being reduced from 99% (normal state) to 35% (under attack state). As a result, the Bayesian probabilistic calculations demonstrate the cascading impact across the whole model. This shows a context-specific impact as opposed to a pre-defined impact based on attacking a device. The context is provided by the dependencies that depend on this device. The same attack on the same device may have a completely different impact when the device is deployed in a different scenario.

Table 6. Mapping of different attack confidence levels to changes in goal probability.

	Confidence					
	0	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
DoS attack	0.99	0.46	0.30	0.17	0.08	0.00
Remote control attack	0.99	0.95	0.83	0.63	0.35	0.00
MITM (no SSL/TLS)	0.99	0.98	0.97	0.88	0.66	0.18
MITM (with SSL/TLS)	0.99	0.98	0.98	096	0.89	0.75
Physical Damage	0.99	0.79	0.59	0.39	0.19	0.00

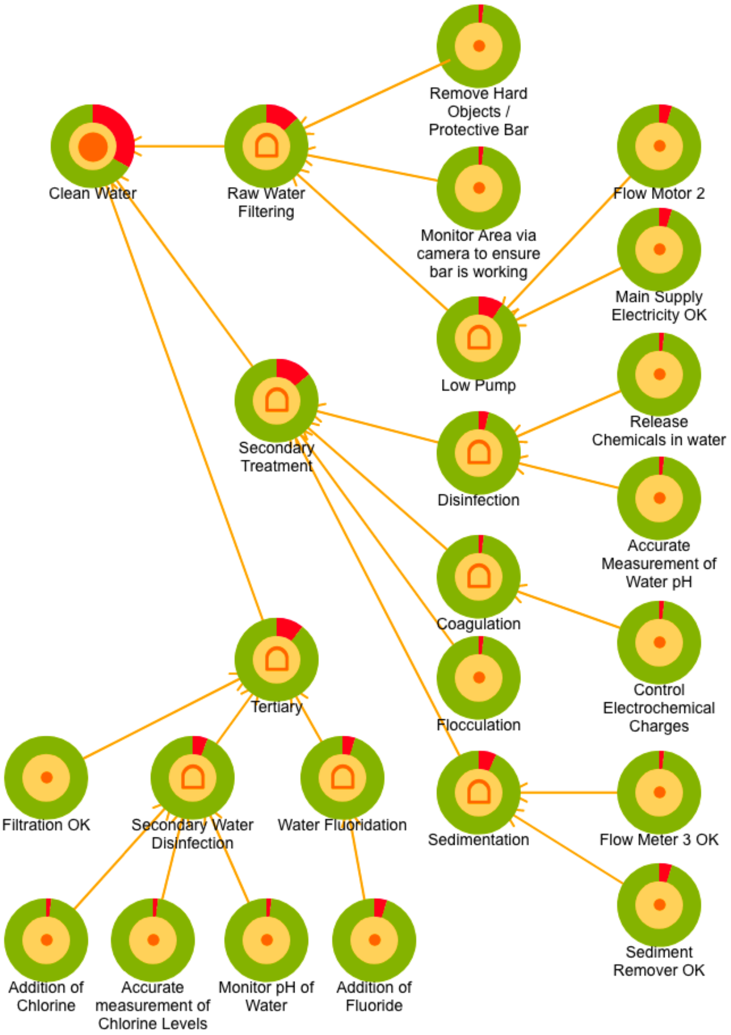
Remove Hard Objects / Protective Bar
Clean Water
Raw Water Filtering
Flow Motor 2
Monitor Area via camera to ensure bar is working
Main Supply Electricity OK
Low Pump
Release Chemicals in water
Secondary Treatment
Disinfection
Accurate Measurement of Water pH
Coagulation
Control Electrochemical Charges
Tertiary
Flocculation
Filtration OK
Secondary Water Disinfection
Water Fluoridation
Sedimentation
Flow Meter 3 OK
Sediment Remover OK
Addition of Chlorine
Accurate measurement of Chlorine Levels
Monitor pH of Water
Addition of Fluoride

Figure 11. Dependency model pre-attack (red = probability of a cyber attack occurring, green = probability of a healthy system).

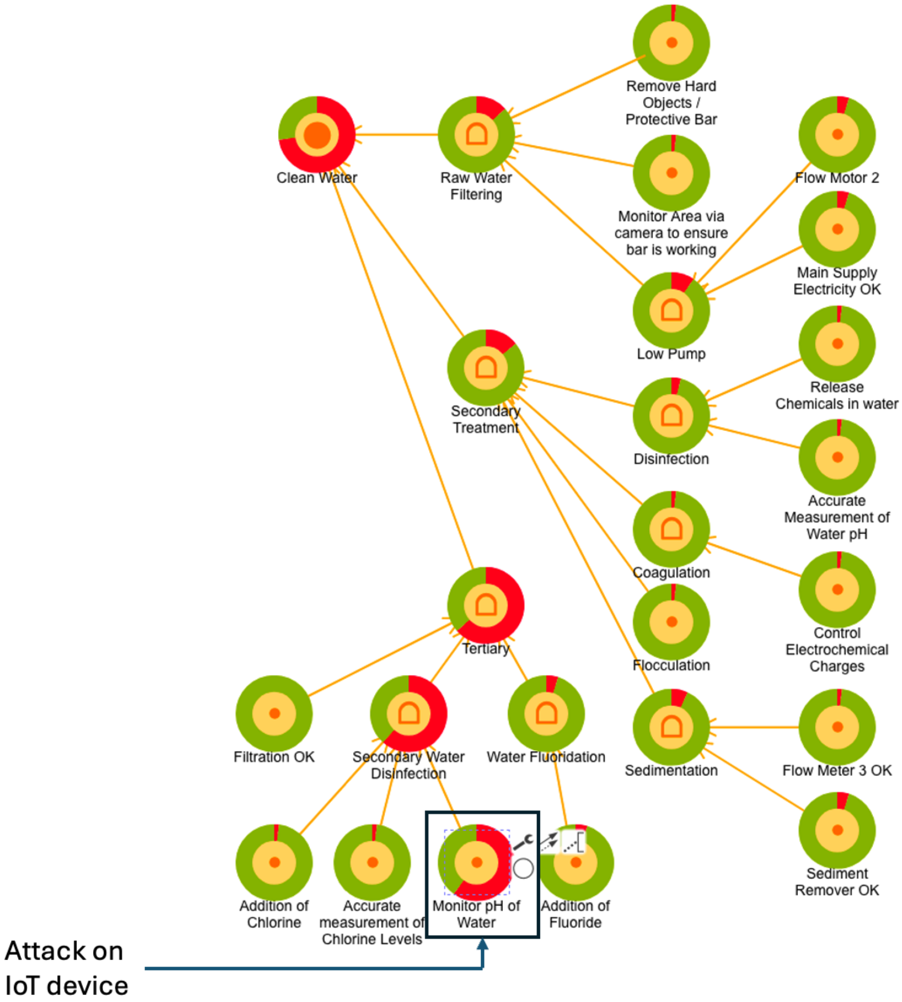

Figure 12. Dependency model after attack (red = probability of a cyber attack occurring, green = probability of a healthy system).

6. Conclusions

In this paper, we have presented a novel methodology to map intelligence gathered automatically about cyber attacks on digital environments—in this case IoT devices—to changes in success probabilities of goals required to maintain system-level operations. The rationale behind this was two-fold. First, by enacting Jaquith's call for consistent measurements that are collected in an automated way and contextually relevant enough for decision-makers to take action [3]. We have proposed a consistent mapping from attack intelligence to probability change while providing flexibility depending on contextual damage caused by the attack on a specific device and difficulty in detecting the attack in an automated way. Second, by enhancing the capability to deal with complex and/or unknown attacks impacting digital components, and determining the impact of this in the context of a system-level dependency model, realising the vision presented in [29]. Previous work focused mostly on detecting attacks and judging their impact on components in isolation, rather than the relationship between the possible attack and the component's impact on the wider system. We provided the first steps towards a methodology to integrate metrics and measurements relating to cyber attacks with systems and component views of risk assessment. In doing so, we can now consider risk in the context of a specific digital ecosystem, such as converged IoT and safety critical systems. This is crucial in order to be able to plan countermeasures and understand the hardware/system implications and financial impact of cyber attacks.

By linking (i) measurements automatically collected from machine learning classification methods to detect attacks on digital components; (ii) systems-level goals and their interdependencies with component-level devices; and (iii) statistical distributions to calculate the reduction in probability of achieving system-level goals when component-level attacks occur on devices, this is the first proposal for an integrated AI-based attack detection system coupled with a harmonised systems-component risk assessment method, giving 'real-time' contextually relevant updates on the cascading impact of cyber attacks on IoT devices across a digital ecosystem—paving the way for *trusted* digital ecosystems.

Author Contributions: Conceptualization, P.B., E.A. and P.R.; Methodology, P.B., F.C. and R.A.; Validation, F.C. and R.A.; Investigation, F.C. and R.A.; Writing—original draft, P.B., E.A., P.R., L.W. and K.J.; Writing—review & editing, P.B., E.A. and L.W.; Funding acquisition, P.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analysed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Risk Management Guidance. Available online: https://www.ncsc.gov.uk/collection/risk-management-collection?curPage=/collection/risk-management-collection/essential-topics/introduction-risk-management-cyber-security-guidance (accessed on 15 April 2019).
2. Rasmussen, J. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Trans. Syst. Man Cybern.* **1983**, 3, 257–266.
3. Jaquith, A. Security Metrics: Replacing Fear, Uncertainty, and Doubt. *J. Inf. Priv. Secur.* **2007**, *4*, 62–63.
4. Doshi, R.; Apthorpe, N.; Feamster, N. Machine learning ddos detection for consumer internet of things devices. In Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 24 May 2018; pp. 29–35.
5. Amouri, A.; Alaparthy, V.T.; Morgera, S.D. Cross layer-based intrusion detection based on network behavior for IoT. In Proceedings of the 2018 IEEE 19th Wireless and Microwave Technology Conference (WAMICON), Sand Key, FL, USA, 9–10 April 2018; pp. 1–4.
6. Dantu, R.; Loper, K.; Kolan, P. Risk management using behavior based attack graphs. In Proceedings of the International Conference on Information Technology: Coding and Computing, 2004; Proceedings; ITCC 2004, Las Vegas, NV, USA, 5–7 April 2004; Volume 1, pp. 445–449.
7. Kotenko, I.; Chechulin, A. A cyber attack modeling and impact assessment framework. In Proceedings of the 2013 5th International Conference on Cyber Conflict (CYCON 2013), Tallinn, Estonia, 4–7 June 2013; pp. 1–24.
8. Hariri, S.; Qu, G.; Dharmagadda, T.; Ramkishore, M.; Raghavendra, C.S. Impact analysis of faults and attacks in large-scale networks. *IEEE Secur. Priv.* **2003**, *99*, 49–54. [CrossRef]
9. Ryan, J.J.; Ryan, D.J. Performance metrics for information security risk management. *IEEE Secur. Priv.* **2008**, *6*, 38–44. [CrossRef]
10. Wang, L.; Jajodia, S.; Singhal, A.; Noel, S. k-zero day safety: Measuring the security risk of networks against unknown attacks. In *Proceedings of the European Symposium on Research in Computer Security*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 573–587.
11. Wang, L.; Islam, T.; Long, T.; Singhal, A.; Jajodia, S. An attack graph-based probabilistic security metric. In *Proceedings of the IFIP Annual Conference on Data and Applications Security and Privacy*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 283–296.
12. Asosheh, A.; Dehmoubed, B.; Khani, A. A new quantitative approach for information security risk assessment. In Proceedings of the 2009 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, China, 8–11 August 2009; pp. 222–227.
13. Quantitative Information Risk Management | The FAIR Institute. Available online: https://www.fairinstitute.org/ (accessed on 4 April 2019).
14. Cyber Risk Management Software and Solutions | RiskLens. Available online: https://www.risklens.com/ (accessed on 4 April 2019).
15. Tsakalidis, G.; Vergidis, K.; Madas, M.; Vlachopoulou, M. Cybersecurity threats: A proposed system for assessing threat severity. In Proceedings of the 4th International Conference on Decision Support System Technology–ICDSST 2018 & PROMETHEE DAYS 2018, Heraklion, Greece, 22–25 May 2018.
16. Wynn, J. *Threat Assessment and Remediation Analysis (TARA)*; Technical Report; MITRE Corporation: Bedford, MA, USA, 2014.
17. Mell, P.; Scarfone, K.; Romanosky, S. *A Complete Guide to the Common Vulnerability Scoring System Version 2.0*; FIRST-Forum of Incident Response and Security Teams: Cary, NC, USA, 2007; Volume 1, p. 23.
18. Byres, E.J.; Franz, M.; Miller, D. The Use of Attack Trees in Assessing Vulnerabilities in SCADA Systems. In Proceedings of the International Infrastructure Survivability Workshop, Citeseer, Lisbon, Portugal, 5–8 December 2004; pp. 3–10.

19. McQueen, M.A.; Boyer, W.F.; Flynn, M.A.; Beitel, G.A. Quantitative cyber risk reduction estimation methodology for a small SCADA control system. In Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), Kauai, HI, USA, 4–7 January 2006; Volume 9, p. 226.
20. Gertman, D.I.; Folkers, R.; Roberts, J. Scenario-based approach to risk analysis in support of cyber security. In Proceedings of the 5. International Topical Meeting on Nuclear Plant Instrumentation Controls, and Human Machine Interface Technology, Albuquerque, NM, USA, 12–16 November 2006.
21. Patel, S.C.; Graham, J.H.; Ralston, P.A. Quantitatively assessing the vulnerability of critical information systems: A new method for evaluating security enhancements. *Int. J. Inf. Manag.* **2008**, *28*, 483–491. [CrossRef]
22. Henry, M.H.; Haimes, Y.Y. A comprehensive network security risk model for process control networks. *Risk Anal. Int. J.* **2009**, *29*, 223–248. [CrossRef]
23. Baiardi, F.; Telmon, C.; Sgandurra, D. Hierarchical, model-based risk management of critical infrastructures. *Reliab. Eng. Syst. Saf.* **2009**, *94*, 1403–1415. [CrossRef]
24. Ten, C.W.; Liu, C.C.; Govindarasu, M. Cyber-vulnerability of power grid monitoring and control systems. In Proceedings of the 4th Annual Workshop on Cyber Security and Information Intelligence Research: Developing Strategies to Meet the Cyber Security and Information Intelligence Challenges Ahead, Oak Ridge, TN, USA, 12–14 May 2008; p. 43.
25. Woo, P.S.; Kim, B.H. A study on quantitative methodology to assess cyber security risk of SCADA systems. *Adv. Mater. Res.* **2014**, *960*, 1602–1611. [CrossRef]
26. Anthi, E.; Williams, L.; Malgorzata, S.; Theodorakopoulos, G.; Burnap, P. A supervised intrusion detection system for smart home IoT devices. *IEEE Internet Things* **2019**, *960*, 1602–1611. [CrossRef]
27. Statista, G.; The Internet of Things (IoT)* Units Installed Base by Category from 2014 to 2020 (in Billions). 2002. Available online: https://www.statista.com/statistics/370350/internet-of-things-installed-base-by-category (accessed on 15 April 2019).
28. Dependency Modeling. Available online: https://publications.opengroup.org/c133 (accessed on 15 April 2019).
29. Burnap, P.; Cherdantseva, Y.; Blyth, A.; Eden, P.; Jones, K.; Soulsby, H.; Stoddart, K. Determining and Sharing Risk Data in Distributed Interdependent Systems. *Computer* **2017**, *50*, 72–79. [CrossRef]
30. Howard, M.; LeBlanc, D. *Writing Secure Code*; Pearson Education: London, UK, 2003.
31. Ross, S.M. *Introduction to Probability Models*; Academic Press: Cambridge, MA, USA, 2014.
32. Dubey, S.Y.D. Normal and Weibull distributions. *Nav. Res. Logist. Q.* **1967**, *14*, 69–79. [CrossRef]
33. Williams, C.K.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 2.
34. Wichers, D. Owasp top-10 2013. OWASP Found 12 February 2013. Available online: https://wiki.owasp.org/images/1/17/OWASP_Top-10_2013--AppSec_EU_2013_-_Dave_Wichers.pdf (accessed on 15 April 2019).
35. Miessler, D. HP Study Reveals 70 Percent of Internet of Things Devices Vulnerable to Attack. Retrieved 30 June 2014, 2015. https://www.scirp.org/reference/ReferencesPapers?ReferenceID=2024873 (accessed on 15 April 2019).
36. Andrea, I.; Chrysostomou, C.; Hadjichristofi, G. Internet of Things: Security vulnerabilities and challenges. In Proceedings of the 2015 IEEE Symposium on Computers and Communication (ISCC), Larnaca, Cyprus, 6–9 July 2015; pp. 180–187.
37. Suo, H.; Wan, J.; Zou, C.; Liu, J. Security in the internet of things: A review. In Proceedings of the 2012 International Conference on Computer Science and Electronics Engineering, Hangzhou, China, 23–25 March 2012; Volume 3, pp. 648–651.
38. Perrig, A.; Stankovic, J.; Wagner, D. Security in wireless sensor networks. *Commun. ACM* **2004**, *47*, 53–57. [CrossRef]
39. Kasinathan, P.; Pastrone, C.; Spirito, M.A.; Vinkovits, M. Denial-of-Service detection in 6LoWPAN based Internet of Things. In Proceedings of the 2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Lyon, France, 7–9 October 2013; pp. 600–607.
40. TP-Link WiFi SmartPlug Client and Wireshark Dissector. Available online: https://github.com/softScheck/tplink-smartplug/blob/master/README.md (accessed on 13 May 2019).
41. Deogirikar, J.; Vidhate, A. Security attacks in IoT: A survey. In Proceedings of the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 10–11 February 2017; pp. 32–37.
42. Barcena, M.B.; Wueest, C. Insecurity in the Internet of Things. *Secur. Response Symantec* **2015**, *20*. Available online: https://candid.ch/cv/insecurity-in-the-internet-of-things-15-en.pdf (accessed on 13 May 2019).
43. Anthi, E.; Javed, A.; Rana, O.; Theodorakopoulos, G. Secure Data Sharing and Analysis in Cloud-Based Energy Management Systems. In *Cloud Infrastructures, Services, and IoT Systems for Smart Cities*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 228–242.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.