DOI: http://dx.doi.org/10.14236/ewic/ICS2016.12

# Measuring the Risk of Cyber Attack in Industrial Control Systems

Allan Cook
Cyber Security Centre,
De Montfort University,
Leicester, LE1 9BH, UK
www.dmu.ac.uk
*allan.cook@my365.dmu.ac.uk*

Richard Smith
Cyber Security Centre,
De Montfort University,
Leicester, LE1 9BH, UK
www.dmu.ac.uk
*rgs@dmu.ac.uk*

Leandros Maglaras
Cyber Security Centre,
De Montfort University,
Leicester, LE1 9BH, UK
www.dmu.ac.uk
*leandros.maglaras@dmu.ac.uk*

Helge Janicke
Cyber Security Centre,
De Montfort University,
Leicester, LE1 9BH, UK
www.dmu.ac.uk
*heljanic@dmu.ac.uk*

**Cyber attacks on industrial control systems (ICS) that underpin critical national infrastructure can be characterised as high-impact, low-frequency events. To date, the volume of attacks versus the overall global footprint of ICS is low, and as a result there is an insufficient dataset to adequately assess the risk to an ICS operator, yet the impacts are potentially catastrophic. This paper identifies key elements of existing decision science that can be used to inform and improve the cyber security of ICS against antagonistic threats and highlights the areas where further development is required to derive realistic risk assessments, as well as detailing how data from established safety processes may inform the decision-making process. The paper concludes by making recommendations as to how a validated dataset could be constructed to support investment in ICS cyber security.**

*ICS, SCADA, Risk, HILF, Cyber, Security, Process Control, Deep Uncertainty*

## 1. INTRODUCTION

A US executive order signed in 2013 stated that the "cyber threat to critical infrastructure continues to grow and represents one of the most serious national security challenges we must confront" (The White House (2013)). This infrastructure is typically underpinned by industrial control systems (ICS) that automate and manage electromechanical devices to provide essential services to a nation's wellbeing and prosperity, and as such any attacks represent a significant threat to the continued security of these countries (Stouffer et al. (2011)). ICS often use operating systems, applications and procedures that may be considered unconventional by contemporary IT professionals, and have operational requirements that include the management of processes that if not executed in a predictable manner, may result in injury, loss of life, damage to the environment, as well as serious financial issues such as production losses that may have a negative impact on a nation's economy (Stouffer et al. (2011); Lopez et al. (2013); Gao and Morris (2014); Mitchell and Chen (2014); Dübendorfer et al. (2004)).

While the total number of ICS installations is unknown, a 2012 research paper by Credit Suisse (Mitchell et al. (2012)) estimated the global industrial automation market to be worth USD 152bn, suggesting a significant number of ICS facilities exist. In the same calendar year, the US ICS-CERT reported 138 incidents (ICS-CERT (2012)) to which it responded. The total number of incidents involving critical infrastructure requiring ICS-CERT to respond in the United States between October 2014 and September 2015 had increased to 295 (ICS-CERT (2016)). With the costs involved in deploying ICS security, especially within critical systems, being high (Naedele (2007)) the business case for investment must be clearly articulated.

An analysis by an insurance company in 2015 exercised a simulated scenario of malware causing an electricity blackout across 15 US states, leaving 93 million people without power, with a total impact to the US economy is estimated at USD 243bn, rising to more than USD 1trn in one version of the scenario (Lloyds and The University of Cambridge Centre for Risk Studies (2015)). However, this study, and another by the North American Electric Reliability Corporation (NERC) (NERC (2010)), defined the possible frequency of these incidents as low. The NERC report also cited cyber attack as only one of nine areas of concern facing the electricity sector in

the period 2010-2018, and as such one must assume that there will be competing demands for resources to address these issues.

Given the potential for these high-impact low-frequency (HILF) events, and the small sample of data, how do we assess the risk of cyber attack on ICS facilities? And even if data becomes available that describes the overall attack landscape of ICS, how do we translate this into attacks in a particular sector, and more parochially, attacks on an individual facility? This contrasts the requirements of industry as a whole, and the needs of security designers - an understanding of potential attacks against *an* ICS versus an understanding of potential attacks against *this* ICS.

## 2. DEFINING RISK

The term *risk* is often used when discussing the potential impact of a cyber attack on an ICS, yet risk may have many alternative meanings when taken in different contexts, such as business, national economics, or plant safety. In order to support rational decision-making to protect ICS from cyber attack it is necessary to develop a common vocabulary and definition of risk.

Risk is not uncertainty, nor is it a hazard. Rather, risk is a function of uncertainty and consequences, *Risk = f(Uncertainty, Consequences)*, and a hazard is a source of danger and exists as a source of risk. Risk to an ICS therefore includes the likelihood of converting a source of risk into damage, loss or injury. In order to mitigate the occurrence of a source of risk in an ICS, the hazard, one applies safeguards. Consequently, *Risk = Hazards/Safeguards*, but highlights that whilst we might reduce the risk through safeguards, we can never bring it to zero (Kaplan and Garrick (1981)).

*Expected Utility* theory (Bernoulli (1954)) also quantifies risk in terms of likelihood and loss: *R=Pr(c)C*, where *C* is the consequence in terms of negative impact to an ICS, and *Pr(c)* is the probability of a loss equal to *C*. When *n* independent events are possible, the risk is the sum of all expected values: $Pr(c_1)C_1+Pr(c_2)C_2+...Pr(c_n)C_n$. The consequences to an ICS can be measured in terms of lost revenue, productivity downtime, injuries and fatalities etc., but always in the same value as the risk itself (Lewis (2014)), which requires a risk analysis to consider what the key measures are out the outset. However, when considering the probability and impact of loss as a quantitative measure, one should avoid describing risk as "probability *times* consequence". This definition is misleading, as in the case of a single scenario this equation would equate a high-impact low-frequency (HILF) scenario to a low-impact high-frequency (LIHF) scenario, which in the case of an ICS are clearly dissimilar and require differing mitigation strategies (Kaplan and Garrick (1981)).

Another term often used in the context of ICS risk is 'vulnerability'. Conceptually, a vulnerability is a risk conditional on an event. Expressing an event as *A*, then *Vulnerability |A = Consequences + Uncertainty |occurrence of A* (Zio et al. (2013)).

As we have seen, the two main factors of risk are the consequences *C* and the probability of *C*, *Pr(c)*. The probability of *C* can also be expressed as the measure of uncertainty *Q*. If we define a set of consequences of interest *C'*, we can express a general description of risk as *Risk description = (C', Q, K)* or alternatively *(A', C', Q, K)* where *K* is the background knowledge upon which *Q* and *C* are based, including expert opinion, models, assumptions, datasets etc., and *A'* is the set of possible events. Consequently, a vulnerability to a given event can be expressed as: *Vulnerability = (C', Q, K |A)*. These definitions, however, are based upon the accuracy and coherence of the background knowledge *K*, on which the risk description is based (Zio et al. (2013)).

Fenton and Neil (2012) highlight the impact of *K* on risk assessment by describing the probability of an event as *P(A|K)*, demonstrating that at least some degree of subjective judgement is incorporated into an assessment, and that it is an expression of a *degree of belief* rather than an absolute value.

Cyber attacks on ICS are, to date, HILF events that lack validated datasets for analysis. This situation is similar to that faced by those responsible for quantifying the likelihood and impact of terrorist attacks, and as such it is worthwhile considering how such events are considered in terms of national security. Lewis (2014) articulates threats in the context of risk as *Threat=Intent* x *Capability*, where *intent* is the propensity of an adversary to attack, and *capability* is a measure of an adversary's ability to launch a successful attack. The National Research Council (2010) (NRC), using models based on the US Department of Homeland Security (DHS) risk practices, also incorporates threat, but uses a wider definition not limited to antagonistic actions, and cannot be considered equivalent to Lewis' description. The NRC model incorporates *T*, vulnerabilities *V*, and capabilities *C*, as *Risk=TVC*. However, in a critical analysis of this approach in light of terrorist attacks, the NRC highlight that defining the values of *T*, *V*, and *C* poses a significant challenge as there is little validated data available and poor reliable knowledge of adversary behaviours. This

data fits Zio et al's description of *K*. In this context the situation is similar to that of ICS. The NRC analysis highlights that an *intelligent adversary* who may seek to actively defeat defensive measures, causes *T*, *V*, and *C* to become interdependent, and as a consequence, risk becomes a factor of *T*, *V*, and *C*, therefore *Risk=f(T,V,C)* (National Research Council (2010)), but does not include any specific measure of antagonistic intent.

The level of uncertainty regarding both terrorist and cyber attacks requires us to consider what US Secretary of Defense, Donald Rumsfeld described as "unknown unknowns" (Rumsfeld (2002)), and how we might reduce this uncertainty. Kaplan and Garrick (1981) proposed that a risk could be described by answering three questions:

1. What can happen? (i.e., what can go wrong?)
2. How likely is it that it will happen?
3. If it does happen, what are the consequences?

In order to answer these questions it is necessary to consider a set of outcomes or scenarios, which can be expressed as a triplet $\{s_i, p_i, x_i\}$ where $s_i$ is the scenario description, $p_i$ is the probability of the scenario occurring, and $x_i$ is the measure of the consequences. A table of such risk triplets would describe the overall risk: $R=\{s_i, p_i, x_i\}$ $i=1, 2...N$. However, this approach is limited by the finite set of scenarios described in the triplet. The actual list of scenarios is infinite, and as a result, any assessment made in this manner is based on incomplete data, as the model does not account for the scenarios not included in the analysis. A method for addressing this is to account for all of the scenarios not considered in the category $s_{n+1}$. The resulting risk analysis is now the set of triplets: $R=\{s_i, p_i, x_i\}$ $i=1, 2...N+1$, which includes all of the scenarios defined, and provides an allowance for those not included. Whilst this might at first appear to be a contrived logical construct, it allows us to consider what probability we should assign to the residual category $s_{n+1}$. This allows us to contemplate the problem within a rational framework, and in particular, what elements of *K* are relevant, and what evidence exists for the scenarios of type $s_{n+1}$ that, by definition, have not occurred yet but exist within the definition of *A'*. Ostrom and Wilhelmsen (2012) list nine criteria for consequences to be considered in context of what could be viewed as $s_{n+1}$ scenarios, in order to assess their credibility. They conclude that one should "never dismiss a consequence until it is proven not to be credible."

These methods of describing risk are dependent upon qualified data or expert opinion *K*, a known set of events *A'*, an accurate quantification of probability ($Q, P_i$), and an ability to reduce the number of scenarios of type $s_{n+1}$.

We shall now explore the viability of quantifying these variables in the context of cyber attacks on ICS as a valid means of articulating risk.

## 3. RISK TECHNIQUES

The techniques discussed in this paper result from a systematic review of decision science, ICS safety and counter-terrorism research, resulting in a subjective, non-exhaustive set of risk approaches based on the literature identified.

### 3.1. Probabilistic Risk Assessment (PRA)

Probabilistic Risk Assessment (PRA), also referred to as *probabilistic safety assessment (PSA)* and *quantitative risk assessment (QRA)* (Apostolakis (2010)) is a scenario-based analysis methodology that systemises the knowledge and uncertainties about a system by answering the Kaplan and Garrick (1981) three risk questions relating to what can wrong, how likely is it, and what are the consequences? (Apostolakis (2010); Zio et al. (2013)). The process identifies a set of undesirable *end states*, then for each state it defines a set of *initiating events* that describe disturbances to normal operation that can lead to the condition. Scenarios are then generated based upon sequences of events that start with an initiating event and conclude in an undesirable end state, allowing the evaluation of the probabilities of these scenarios using all available evidence, past experience, and expert judgement. The scenarios are then ranked according to their contribution to the frequencies of the end states, as well as the systems, structures and components that also contribute (Apostolakis (2010); Ostrom and Wilhelmsen (2012)). PRA is typically employed for accident analyses in systems that are highly reliable and for which significant reliability data is available (Apostolakis (2004)), whereas we are considering wilful, malicious cyber attacks. The techniques do not preclude such an assessment, and the steps involved would accommodate the consideration of an attack against cyber-physical systems such as ICS. The methodology allows for differing probabilistic methods to be employed during the analysis, including Bayesian networks, Monte Carlo simulations etc. (Apostolakis (2010); Zio et al. (2013)), in order to generate the probabilities of the scenarios. By itself, therefore, PRA does not provide a means to mitigate the lack of available and verified ICS threat data, previously described as *K*. As a scenario-based approach neither does it address those threats defined by the type $s_{n+1}$.

### 3.2. Bayesian Networks (BN)

Bayesian networks (BN) are based on the theory that belief is conditional and depends on mounting evidence that either contribute to a belief or refute it. BN describes a belief system in terms of *conditional probabilities*, containing propositions that are either true, partially true, or false (Lewis (2014)). The subjectivity that BN supports allows us to consider the probability of an attack without a frequentist approach to representative data. Using BNs we start with a hypothesis *H* and describe a *prior belief* about *H* expressed as *P(H)*. Using observed evidence *E* we can revise our belief about *H* in light of *E* and calculate a *posterior belief* about *H* by calculating *P(H|E)* in terms of *P(E|H)* (Fenton and Neil (2012)).

BN are flexible risk analysis tools as they allow the degree of belief in a hypothesis, or a series of hypotheses, to evolve as new evidence emerges and allow greater detail to be derived over time. As such, BN offer the potential to reduce *K*, and allows for additional attack scenarios to be added to the model, thereby reducing $s_{n+1}$. However, given the lack of immediately available data regarding cyber attacks on ICS, consideration should be given to how data for BN analysis will be obtained.

### 3.3. Fault Tree Analysis (FTA)

Fault tree analysis (FTA) is a method primarily aimed at system safety and reliability, although as a technique used within ICS communities it may offer a valuable insight into the likelihood and impact of maliciously-caused equipment failure. It is a deductive analysis technique focusing on one error at a time, intended to ensure that all aspects of a system are identified and controlled in order to determine design aspects that could lead to potential failures (Kapur and Pecht (2014)). The events are generally considered as Boolean representations, insomuch as they either will or will not occur (Dillon-Merrill et al. (2008)). FTA is often used as an element of PRA, but can be used independently. Fault trees are a graphical representation of the interaction of failures and other events in a system. The process begins by supposing that a particular failure occurs, then uses deductive logic to step through a system, considering the possible direct causes that could contribute to the condition, forming a graphical tree-like structure as the analysis progresses (Ostrom and Wilhelmsen (2012)). Once a fault tree has been developed, data regarding the failure rate for individual system components can be analysed either in series or parallel, through the application of logic gates, to estimate the likelihood of the failure event (referred to as the *top event*). The process drives the production of *cut sets* - the smallest number of system components that, if they all fail, will lead to an overall system failure, then assesses the probability of such failure (Dillon-Merrill et al. (2008)).

According to Sutton (2014), the quality of the failure rate data on which the method is based are often unreliable, and subjective measures often substituted. These measures are still considered of value as the method allows experts to contribute based on their experiences in a structured manner, and enables subsequent deeper analysis if areas where the data is deemed inconclusive. The approach has a reputation for being resource-intensive, requiring significant expertise in the system under analysis. Its resource requirements aside, FTA's major advantage is that fault trees can accommodate complex logical relationships and interdependencies between system components. It is less dependent on human thought extrapolation to recognise the effects of changes in system behaviours (Dillon-Merrill et al. (2008)).

For the purposes of assessing the risk of cyber attack on ICS, the approach has some merit. By starting at a system failure, in our case a maliciously-caused failure, we can trace back through the components of an ICS, both Operational Technology (OT) and Information Technology (IT), to determine scenarios under which the events could occur. By focussing on system components and the chaining of events and connectivity to adversely affect them, as long as every system component is considered, it is likely that the set $s_{n+1}$ will be reduced, although the method has no formal mechanism for measuring this. However, the approach is still reliant on expert opinion and qualified data, as expressed in set *K*. Fundamentally, the method is focused on predicting equipment failures rather than the behaviours of malicious actors. It offers no means of determining threat actor capability or target preferences, although it can contribute to a wider threat analysis process.

### 3.4. Event Tree Analysis (ETA)

Event tree analysis (ETA) uses the same mathematical and logical techniques as FTA, but considers the impact of a failure of a particular component through inductive reasoning. Like FTA, ETA can also be used within other risk analysis techniques such as PRA. Event trees model a sequence of outcomes which may arise after a particular *initiating event*, focusing on paths or scenarios that lead to failure. As such, ETA considers the three questions posed by Kaplan and Garrick (1981) that form their risk triplet. Following each event, ETA considers the occurrence or non-occurrence of all other possible events, with probabilities calculated conditionally on all previous outcomes in the tree. The approach assumes that each event only has two outcomes; success or failure, although separate event trees can be developed

for each initiating event (Dillon-Merrill et al. (2008); Ostrom and Wilhelmsen (2012); Sutton (2014)).

ETA allows analysts to consider one path or scenario at a time, and offers some use when assessing the potential causes of an undesirable effect as it propagates through an ICS. This may facilitate the identification of unexpected system conditions. Like FTA, however, the approach is still reliant on expert opinion and qualified data *K*, and the subjective set of initiating events does not reliably reduce $s_{n+1}$.

### 3.5. Bow-Tie Analysis

The Bow-Tie analysis method combines FTA with ETA by considering an undesirable event, then analysing deductively to the left using FTA, considering what could lead to the event, then analysing inductively to the right using ETA to consider the consequences of the event (Sutton (2014)). Whilst the approach does not address the shortcomings in respect of *K*, it may generate scenarios previously not considered, and as a result may possibly reduce the set $s_{n+1}$.

### 3.6. Attack Trees

Attack trees (Schneier (1999)) consider the paths by which an antagonist would attempt to reach a particular target in a network, described as a root, with leaf nodes articulating the chain of attack surfaces by which the attacker could reach that target. As such, attack trees provide a similar analytic construct to FTA, ETA and Bow-tie analysis, and suffers from the same limitations in respect of *K* (Dillon-Merrill et al. (2008)). However, whilst FTA, ETA and Bow-tie approaches are primarily focused on failure scenarios, attack trees concentrate on malicious attempts to manipulate a system, so the combination of these methods may broaden the attack scenarios under consideration and potentially reduce the set $s_{n+1}$.

### 3.7. Monte Carlo Simulations

When assessing the potential for an attack on ICS, so far we have only considered those scenarios that we consider realistic. However, this bias does not address the set $s_{n+1}$. One option to attempt to reduce the set of unconsidered attack targets, and thereby reduce $s_{n+1}$, is to adopt a stochastic model that includes every element within the system. Monte Carlo analyses reflect the randomness of life, even in deterministic systems, by assigning each system element an initial operating condition and a probability of failure. A time sequence starts based on a given interval and a random number is generated for each element within the system. If the random number falls within a defined failure range, the system element transitions to a failed state. At the end of each iteration, the cut sets are analysed to determine the system's operability and availability and the results aggregated to produce an overall model of the system (Sutton (2014)).

Monte Carlo simulations are resource-intensive and require long run times in order to achieve stable results. By itself it cannot determine the impact of a targeted cyber attack, but the technique offers a useful model as it considers all elements within a system in a random manner, possibly introducing failure conditions not previously considered, thereby reducing $s_{n+1}$. However, the assigned probability of failure requires expert knowledge, and as such forms part of set *K*. Monte Carlo simulations alone will not immediately address the qualitative improvement of *K*, but there is potential for using Monte Carlo simulations to feed BN analyses.

### 3.8. Markov Models

Another method to perform stochastic analysis is based on Markov models. These represent all of the possible states of elements within an ICS, against which it performs a series of transitions based on a defined time interval, or step in a batch process Sutton (2014). Each transition results in a system element either remaining in its current state, or moving to a new one. As such, the probability of transition from one state into another is dependent only on the current state, and not on the history of states that preceded it. This property is usually referred to as the *Markov property*. Where the future state of the system element is dependent upon both the current state and the immediate past state, it is referred to as a second-order Markov process, and so-on for further higher-order Markov processes (Ibe (2013)).

The immediate transition between states may not accurately represent real-world considerations for ICS, where states are not binary and can introduce an indeterminate condition as large electromechanical devices execute instructions. That aside, Markov models offer similar benefits to Monte Carlo simulations, as assuming they model the complete system, may introduce failure scenarios to reduce the set $s_{n+1}$, and perhaps could be used in conjunction with BN analyses.

### 3.9. Failure Modes, Effects and Criticality Analysis

Failure Modes and Effects Analysis (FMEA) is an experience-based hazards analysis approach based on expert opinion and engineering standards. The accumulated knowledge allows hazards to be considered in light of experiential data and evaluated against

*Recognised and Generally Accepted Good Engineering Practice (RAGAEP)*. The method examines the ways in which equipment can fail and considers the consequences of such failures. Device criticality can also be analysed, in which case the method is described as Failure Modes, Effects and Criticality Analysis (FMECA). Importantly, FMEA/FMECA does not consider the causes of the failure, just the impact of its occurrence. Neither is it concerned with the sequence of events that led to the failure, or the actors or circumstances involved (Sutton (2014)).

By focussing solely on failure modes and the resulting effects, without considering the path that led to the event, FMEA and FMECA allows scenarios to be considered based purely on impact. By itself the process will not address the quantification of risk of cyber attacks within an ICS, but as it is a commonly produced engineering artefact that many ICS facilities will already possess, it offers a means to qualitatively check the background data and expert opinion, *K*, used to inform the assessment process. By not limiting the impact of failure to potential cyber access routes, we may limit the set $s_{n+1}$. FMEA/FMECA should be considered in conjunction with FTA.

### 3.10. Hazard and Operability (HAZOP) Method

The Hazard and Operability (HAZOP) method is probably the most widely used hazards analysis method in industry. Its widespread use and acceptance has led to a large number of practitioners and supporting service providers. The method divides the system under analysis into nodes, each of which represent a section of the process that undergoes a significant change or transformation. Examples of nodes include pumps, reactors, heat exchangers etc. The information is generally extracted from plant piping and instrumentation diagrams (P&ID). The size of the node is a subjective decision based on the nature of the industrial process and may group devices or system elements together in order to consider an overall process change holistically (Sutton (2014)).

A HAZOP analysis follows a consistent process whereby a system node is selected and its purpose and safe limits defined. Next, one of a set of *process guidewords* are selected, such as high flow, low/no flow, reverse flow, misdirected flow, high pressure, high temperature, polymerisation, wrong composition etc., that describe the effect that should be considered, and hazards and their causes are identified as a result. For each hazard, the process considers it will be recognised should it occur and an estimation of the consequences is reached. A set of safeguard requirements are then defined, as is the estimated frequency of the hazard's occurrence. Finally, the hazards are ranked and a set of findings and recommendations is produced.

HAZOP analyses drive a rigorous assessment of the impact of undesirable events on a process, decomposing to a detailed level. Conformance to established processes and guidewords should provide a comprehensive set of potential areas of risk. Whilst it may not be commonplace to consider these impacts from the perspective of cyber attacks, a HAZOP should identify areas where change to the process will result in adverse outcomes. The breadth of the process should reduce the size of the set $s_{n+1}$ and by adhering to RAGAEP, confidence in set *K* is arguably increased.

### 3.11. CARVER and MSHARPP

When considering the threat to an ICS, methods exist within the military community to describe an intelligent adversary's intent and capability. The US Department of Defense (DoD) use the CARVER assessment method to determine criticality and vulnerability in infrastructures. CARVER is an acronym for *Criticality, Accessibility, Recuperability* (a system's ability to recover from an attack), *Vulnerability, Effect and Recognisability*. The method focuses on an adversary's perspective of the infrastructure to enable an analysis of the weaknesses of a target, or the means by which its operations can be manipulated by an attacker (Schnaubelt et al. (2014)). In this manner, the capabilities of several threat actors can be considered. The output of the CARVER assessment is a critical-asset list that defines a prioritised set of assets that are of value to an attacker based on their importance, whereby the asset's incapacitation or destruction would have a serious impact on the military operation or facility. The use of CARVER matrices to consider threats to critical national infrastructure by civilian agencies when preparing for terrorist attacks is emerging, as it allows organisations to consider the relative desirability of targets, although its use has been limited to the assessment of physical assets (Doro-on (2014)). Another approach, encompassed in the acronym MSHARPP, describes the attractiveness of a target due to its importance to an operation (the *Mission*), the perception that a successful attack with generate (the *Symbolism*), the *History* of similar attacks, system *Accessibility*, *Recognisability* of target, impact on the local *Population*, and *Proximity* to other key assets. Like the CARVER approach, a matrix is derived using a numeric range that represents the perceived vulnerability or likelihood of attack, from the perspective of the defender. The respective numerical values are totalled to provide a relative value as a target or the overall assessment of attractiveness to an attacker, and thereby a prioritised list of assets to defend (Schnaubelt et al. (2014)).

### 3.12. Game Theory

Studies of antagonistic attacks can be conducted using game situations rather than decision models (Holmgren et al. (2007)). Game theory techniques involving sequential multi-player scenarios such as *Stackelberg Competitions* allow players to decide upon actions that result in the best possible rewards to themselves whilst anticipating the rational actions of the other participants. The outcome of the game defines the optimal allocation of resources and investment to minimise risk on the part of the defender and maximise risk for the attacker (Lewis (2014), Roy et al. (2010)). In an evaluation of physical attacks against electric power networks using game theory, Holmgren et al. (2007) illustrated the effectiveness of the technique, but highlighted the need for a greater understanding of the attacker's intent. It is possible that the use of CARVER or MSHARPP may address this issue in some way, as the attractiveness of a target may be used as a surrogate for intent. Holmgren et al. (2007) also acknowledged a wider issue; that the results of game theoretic approaches depend upon how the scenario is framed, suggesting that it is dependent upon the set $K$ and that $s_{n+1}$ is not necessarily reduced by its use.

## 4. DEEP UNCERTAINTY

If the techniques discussed thus far do not provide sufficient means to address $K$ and $s_{n+1}$, we are faced with the area of decision science referred to as *deep uncertainty*. Cox Jr (2015) describes methods of combining models and datasets in order to reduce the levels of uncertainty.

### 4.1. Using Multiple Models and Relevant Data to Improve Decisions

When no validated data set is available, a *good* decision is one that assesses clearly higher and lower probabilities of undesired outcomes based on a combination of existing models that are consistent with available, albeit not directly relevant data. This combination of alternative models is described as the *uncertainty set*.

### 4.2. Robust Decisions with Model Ensembles

Cox Jr (2015) proposes that when faced with decision-making decisions with deep uncertainty, a technique that can be employed is to generate and analyse a large number of scenarios. Of these, a set that performs well by some criterion for most scenarios is more likely to also do well in reality, if reality is well-described by at least some of the scenarios in the uncertainty set.

### 4.3. Averaging Forecasts

Simple arithmetic averaging of results from different methods is reported to usually outperform an average of any single method, and by averaging across models one can reduce the error between forecast and subsequently measured true values.

## 5. ELICITING EXPERT OPINION

If no valid data is available to inform set $K$ then Ezell et al. (2010) recommend that the probabilities of different attacker options should be elicited from experts. However, expert opinion is not necessarily a viable replacement, and care should be taken with the validity of such judgement. In particular, Wallsten and Budescu (1983) ask whether the probabilities elicited provide a valid measurement related to the frequency of the events? Merrick et al. (2015) propose that calibration is one measure of validity. When an expert assesses the probability $P$, on what basis is the judgement that proposes that the event occurs $P\%$ of the time? Those who will act on the expert advice would assume that $P$ should be close to the observed frequency of the event, $\hat{P}$. However, $P$ is simply a measure of the expert's degree of belief (Keren (1997)), and in the case of cyber attacks on ICS, no data is available to adequately describe $\hat{P}$.

Merrick et al. (2015) observe that a calibration curve formed by probability judgements is usually more extreme than the relative frequency of events, and as such, expert opinion may have a bias toward a negative perspective. Wallsten and Budescu (1983) point out that when faced with judgements based on complex events, it is natural for humans to simplify the task by using heuristics, which also introduce biases into analyses. According to Hora (2007) the quality of judgements is based on the background information used for the assessment, usually derived from their experience, and is reliant on the expert's ability to fuse this with other data sources. Merrick et al. (2015) discusses three heuristic techniques; *representativeness* in which the probability is based on the similarity to other observed events, *availability* where humans overestimate the frequency of an event due to excessive media reporting, and *anchoring* where humans adopt a starting position then adjust away from that in order to produce a probability. None of these methods serve to improve the accuracy of the expert opinion in order to inform the set $K$. However, Ravinder et al. (1988) and Howard (1989) demonstrate that decomposing complex scenarios into discrete events improves the overall calibration and reliability of probability judgements when eliciting expert opinion, and in this manner we may minimise the biases and errors introduced.

## 6. CURRENT OPTIONS FOR ASSESSING ICS CYBER RISK

PRA remains the de facto standard for risk assessment in large-scale critical facilities and provides an analysis framework that incorporates safety data and probabilistic methods such as BN and Monte Carlo simulations (Lewis (2014)). For the immediate future there does not appear to be a proven, viable alternative. Its strength lies in its scenario-based approach, which from the perspective of ICS cyber threats allows malicious activity across the electromagnetic spectrum to be included in the overall risk analysis. However, as ICS cyber attacks remain low-frequency events, details of what an ICS attack scenarios may look like are in short supply and as a consequence our ability to reduce the set $s_{n+1}$ remains limited. PRA also requires expert opinion, which given the lack of quantifiable data on ICS cyber attacks leaves the background information *K* open to the biases previously described. As such, any PRA analyses undertaken using expert opinion should require that any attack scenarios comprise a set of discrete events that aggregate to form a potentially complex attack, rather than attempting to consider the incident as a whole. Sommestad et al. (2009) have demonstrated some success in combining attack trees with Bayesian methods and expert assessment, but their analysis did not include the industrial processes under control or their safety characteristics, and as such does address the holistic impact of a cyber attack on an ICS.

Red teaming and game theoretic approaches offer possible improvements on the use of expert opinions, but their efficacy is currently limited by the cross-discipline team of IT and OT staff, and a lack of common vocabulary and understanding, especially when industrial engineering is also introduced. In order to address this issue it will be necessary to describe the industrial processes and technologies in an architectural model understood by all disciplines. Without such an analysis framework it is unlikely that a robust risk assessment could be produced.

Ultimately, however, the current options for assessing ICS cyber risk are limited by the available data, and accordingly we should consider how such a dataset can be produced.

## 7. OPTIONS FOR DEVELOPING A QUALIFIED CYBER ATTACK DATASET

Risk analysis is an established discipline, but for quantitative methods to work satisfactorily they require a validated dataset. The limited qualified reports of cyber attacks on ICS do not satisfy this requirement, and as a consequence, risk analyses on ICS are currently dependent on subjective judgements and cannot adequately consider the breadth of attack vectors. The US DoD advisory group, JASON (McMorrow (2009)), argues that predictive models for rare events are unreliable, and like Ravinder et al. (1988) and Howard (1989), recommend decomposing the rare events into smaller, well-bounded problems that can be tested. The Idaho National Laboratory (Macaulay and Singer (2011)) suspended research into quantitative analysis into this field, in favour of more subjective approaches more in line with the criteria found in the CARVER and MSHARPP models. Whilst data exists to define the risks of equipment failure in an ICS, the predicted rates of these events is based on deterioration rather than intentional direction. It would also need to be proven that safety cases have a direct relationship with security cases before using this data as the sole basis for cyber attack risk analysis.

### 7.1. Background Information

Central to the issue of providing qualified risk analyses of cyber attacks on ICS is the lack of available data, which we have referred to as the set *K*. Without validated background, those responsible for identifying vulnerable elements cannot elevate the investment priority in this area against other, more clearly perceived risks to the business, such as failure to meet regulatory targets, service outages through ageing equipment, and exposure to financial markets. As such, an essential first stage of improving the quality of *K* is to provide a means to portray the impact of HILF events in a manner recognisable to senior stakeholders. At a more detailed level, however, malicious behaviour data needs to be shared in order to allow ICS operators to gain a greater understanding of events that are indicative of potential or actual attacks. Threat intelligence sharing offers a key benefit to the whole ICS community as it allows a richer view of areas for protective analysis within an industrial operation. Bayesian networks are potentially of value in this analysis, as models for *K* could be constructed, tested and revised over time using threat intelligence data, thereby improving the quality of background information on which decisions are based.

### 7.2. Attack Vectors

The lack of available data also limits our understanding of attacker options and targets,resulting in the possibility that the set on unconsidered scenarios that we have referred to as $s_{n+1}$ is unfavourable. Decomposing the scenarios into smaller, manageable analyses offers clear benefits, but we are still largely reliant on expert opinion to define the scenarios to begin with. Introducing Monte Carlo and Markov

simulations to test all possible failure modes may add a degree of objective data to the analysis to challenge any cognitive biases introduced during the scenario construction. The wealth of safety information available, from HAZOP, FEMA, FTA and ETA, is potentially a sound basis from which to start these analyses, and would allow the consideration of the relationship between safety and security cases.

### 7.3. Intelligent Adversaries

We should also realise that any models we develop for risk analyses cannot be static. Cyber attacks are conducted by intelligent adversaries who will change their approaches dependent on the security mechanisms we deploy. Threat intelligence, Bayesian networks and game theoretic approaches will allow general attack behaviour to be modelled, not the specifics of an attack against a particular facility. In considering intelligent adversaries in the context of terrorist events, the National Research Council (2010) recommend the introduction of *red teams* to probe the defences of an organisation. This may prove problematic in an ICS, as they may not be resilient to potentially destructive testing (Stouffer et al. (2011)), and so some form of non-destructive testing is required. This does not, however, preclude red teaming as a viable concept should a safe, representative environment be made available for such activities. Sommestad and Hallberg (2012) demonstrated how cyber security exercises, conducted on dedicated infrastructure, can generate valuable data for security research.

### 7.4. Non-destructive Testing Through Simulation

Many industrial facilities utilise simulation tools to model and predict the operations of the processes under control within an ICS. This forms an essential part of the operations of the facility, based on the steady-state behaviour of the process. These models could be revised to allow boundary conditions to be introduced to predict where potential negative outcomes can be generated (Krotofil and Larsen (2015)). The ICS elements responsible for controlling the boundary conditions could then be considered as a discrete, testable scenario. This could be used to develop *synthetic environments*, testbeds, or cyber ranges on which representative architectures could be deployed using a mix of virtualised environments and non-production physical devices. Attack scenarios by intelligent adversaries could then be exercised and the results fed into background information models. In order to support this, it would be necessary to produce an architectural model of the ICS that supports the description of attack hypotheses and vulnerabilities, including security events, state transitions, dependencies, and means to describe differing consequences in various measures (financial, production loss etc.).

## 8. CONCLUSIONS

No proven, unified risk model exists for ICS that incorporates all of the elements of consequences $C$, events $A$, background information $K$, measure of uncertainty $Q$, threat $T$, vulnerability $V$, or the unconsidered scenarios of $s_{n+1}$ across all measures. Quantitative risk analysis does not provide the sole means of addressing the problem, and pragmatically, one may be forced to adopt a method best suited to the available data, or by the combination of partially-suited models, until a suitable critical mass of information can be derived.

Further work in this area is recommended to include:

1. Developing a means by which senior stakeholder awareness of HILF events can be articulated in business terms.
2. Architectural models be defined that support the testing of attack hypotheses that can feed Bayesian models, Monte Carlo and Markov simulations.
3. Guidelines be produced for the use of existing ICS process simulations to drive understanding of boundary conditions.
4. Development of synthetic environments to allow a *red team* or game theoretic intelligent adversary to test an ICS's defences in a non-destructive manner.

## REFERENCES

Apostolakis, G. (2010). *Probabilistic Risk Assessment (PRA)* (Wiley Handbook of Science and Technology for Homeland Security), 1.

Apostolakis, G. E. (2004). How useful is quantitative risk assessment? *Risk Analysis*, 24(3), 515–520.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica: Journal of the Econometric Society*, 23–36.

Cox, Jr., L. A. T. (2015). Making decisions without trustworthy risk models. *Breakthroughs in Decision Science and Risk Analysis*, 189.

Dillon-Merrill, R. L. et al. (2008). *Logic Trees: Fault, Success, Attack, Event, Probability, and Decision Trees* (Wiley Handbook of Science and Technology for Homeland Security).

Doro-On, A. M. (2014). *Risk Assessment and Security for Pipelines, Tunnels, and Underground Rail and Transit Operations*. CRC Press.

Dübendorfer, T. et al. (2004). An economic damage model for large-scale internet attacks. In: *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2004. WET ICE 2004. 13th IEEE International Workshops on*, 223–228.

Ezell, B. C. et al. (2010). Probabilistic risk analysis and terrorism risk. *Risk Analysis*, 30(4), 575–589.

Fenton, N. and Neil, M. (2012). *Risk Assessment and Decision Analysis With Bayesian Networks*. CRC Press.

Gao, W. and Morris, T. H. (2014). On cyber attacks and signature based intrusion detection for modbus based industrial control systems. *Journal of Digital Forensics, Security and Law*, 9(1), 37–56.

Holmgren, A. J. et al. (2007). Evaluating strategies for defending electric power networks against antagonistic attacks. *IEEE Transactions on Power Systems*, 22(1), 76–84.

Hora, S. (2007). 8 eliciting probabilities from experts. *Advances in Decision Analysis: From Foundations to Applications*, 129.

Howard, R. A. (1989). Knowledge maps. *Management science*, 35(8), 903–922.

Ibe, O. (2013). *Markov Processes for Stochastic Modeling*. Newnes.

ICS-CERT (2012). ICS-CERT year in review - 2012. Available form https://ics-cert.us-cert.gov/ICS-CERT-Year-Review-2012

ICS-CERT (2016, Jan.) ICS-CERT monitor November/December 2015. Available from https://ics-cert.us-cert.gov/sites/default/files/Monitors/ICS-CERT%20Monitor_Nov-Dec2015_S508C.pdf.

Kaplan, S. and Garrick, B. J. (1981). On the quantitative definition of risk. *Risk analysis*, 1(1), 11–27.

Kapur, K. C. and Pecht, M. (2014). *Reliability Engineering*. John Wiley & Sons.

Keren, G. (1997). On the calibration of probability judgments: Some critical comments and alternative perspectives. In: *Conference on Subjective Probability, Utility and Decision Making: Overconfidence: Sources, Implications, and Solutions.*, Jerusalem, Israel, John Wiley & Sons, (Aug. 1995).

Krotofil, M. and Larsen, J. (2015). Rocking the pocket book: Hacking chemical plants.

Lewis, T. G. (2014). *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation*. John Wiley & Sons.

Lloyds and The University of Cambridge Centre for Risk Studies (2015). Business blackout: The insurance implications of a cyber attack on the us power grid. Available form www.lloyds.com//media/files/news%20and%20 insight/risk%20insight/2015/business%20 blackout/business%20blackout20150708.pdf.

Lopez, J. et al. (2013). Smart control of operational threats in control substations. *Computers & Security*, 38, 14–27.

Macaulay, T. and Singer, B. L. (2011). *Cybersecurity for Industrial Control Systems: SCADA, DCS, PLC, HMI, and SIS*. CRC Press.

McMorrow, D. (2009). Rare events. DTIC Document, Technical Rep..

Merrick, J. et al. (2015). Outthinking the terrorists. *Breakthroughs in Decision Science and Risk Analysis*, 287.

Mitchell, J. et al. (2012, Aug.) Global industrial automation.

Mitchell, R. and Chen, I.-R. (2014). A survey of intrusion detection techniques for cyber-physical systems. *ACM Computing Surveys (CSUR)*, 46(4), 55.

Naedele, M. (2007). Addressing it security for critical control systems. In: *40th Annual Hawaii International Conference on System Sciences*.

National Research Council. (2010). *Review of the Department of Homeland Security's Approach to Risk Analysis*. The National Academies Press.

NERC (2010). High-impact, low-frequency event risk to the north American bulk power system. In: *A Jointly-Commissioned Summary Report of the North American Electric Reliability Corporation and the U.S. Department of Energy's November 2009 Workshop*.

Ostrom, L. T. and Wilhelmsen, C. A. (2012). *Risk Assessment: Tools, Techniques, and Their Applications*. John Wiley & Sons.

Ravinder, H. V. et al. (1988). The reliability of subjective probabilities obtained through decomposition. *Management Science*, 34(2), 186–199.

Roy, S. et al. (2010). A survey of game theory as applied to network security. In: *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, 1–10.

Rumsfeld, D. (2002, Feb.) U.S. DOD news briefing. Available form https://www.youtube.com/watch?v=GiPe1OiKQuk

Schnaubelt, C. M. et al. (2014). *Vulnerability Assessment Method Pocket Guide*. RAND Corporation.

Schneier, B. (1999). Attack trees. *Dr. Dobbs Journal*. 24(12), 21–29.

Sommestad, T. et al. (2009). Modeling security of power communication systems using defense graphs and influence diagrams. *IEEE Transactions on Power Delivery*, 24(4), 1801–1808.

Sommestad, T. and Hallberg, J. (2012). Cyber security exercises and competitions as a platform for cyber security experiments. In: *Nordic Conference on Secure IT Systems*, Springer, 47–60.

Stouffer, K. et al. (2011). Guide to industrial control systems (ICS) security. In: *NIST Special Publication*, 800–82.

Sutton, I. (2014). *Process Risk and Reliability Management: Operational Integrity Management*. Gulf Professional Publishing.

The White House (2013, Feb.) Executive order - improving critical infrastructure cybersecurity.

Wallsten, T. S. and Budescu, D. V. (1983). State of the art encoding subjective probabilities: A psychological and psychometric review. *Management Science*, 29(2), 151–173.

Zio, E. et al. (2013). *Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods*. John Wiley & Sons.