STOCHASTIC COUNTERFACTUAL ANALYSIS FOR THE VULNERABILITY ASSESSMENT OF CYBER-PHYSICAL ATTACKS ON ELECTRICITY DISTRIBUTION INFRASTRUCTURE NETWORKS

Edward Oughton, Daniel Ralph, Eireann Leverett, Raghav Pant, Scott Thacker, Jim W Hall, Jennifer Copic, Simon Ruffle, Michelle Tuveson
Stochastic Counterfactual Analysis for the Vulnerability Assessment of Cyber-Physical Attacks on Electricity Distribution Infrastructure Networks

Edward J. Oughton1*, Daniel Ralph1, Eireann Leverett1, Raghav Pant2, Scott Thacker2, Jim W. Hall2, Jennifer Copic1, Simon Ruffle1, Michelle Tuveson1

1Centre for Risk Studies, Judge Business School, University of Cambridge, Cambridge, UK.

2Environmental Change Institute, University of Oxford, Oxford, UK

* Corresponding author: e.oughton@jbs.cam.ac.uk at the University of Cambridge Judge Business School, Trumpington Street, Cambridge, UK, CB2 1AG
ABSTRACT

In December 2015, a cyber-physical attack took place on the Ukrainian electricity distribution network. This is regarded as one of the first cyber-physical attacks on electricity infrastructure to have led to a substantial power outage and is illustrative of the increasing vulnerability of Critical National Infrastructure to this type of malicious activity. Few data points, coupled with the rapid emergence of cyber phenomena, has held back the development of resilience analytics of cyber-physical attacks, relative to many other threats. We propose to overcome data limitations by applying stochastic counterfactual analysis as part of a new vulnerability assessment framework. The methodology is developed in the context of the direct and indirect socio-economic impacts of a Ukrainian-style cyber-physical attack taking place on the electricity distribution network serving London and its surrounding regions. A key finding is that if decision-makers wish to mitigate population disruptions, then they must invest resources more-or-less equally across all substations, to prevent the scaling of a cyber-physical attack. However, there are some substations associated with higher economic value due to their support of other Critical National Infrastructures, such as airports or maritime ports, which justifies the allocation of additional cyber security investment to reduce the chance of cascading failure. Further cyber-physical vulnerability research must address the trade-offs inherent in a system made up of multiple institutions with different strategic risk mitigation objectives and metrics of value, such as governments, infrastructure operators and commercial consumers of infrastructure services.

KEYWORDS

Cyber-physical attack; infrastructure; critical national infrastructure
1. INTRODUCTION

In December 2015, a power outage occurred in the Ukraine (1), where a Trojan was found on a number of electricity substations believed to be associated with a BlackEnergy Malware campaign utilising remote cyber intrusion (2). This was the first known instance where a cyber-attack caused an electricity blackout. Consequently, a key issue is how to develop risk analytics for emerging threats, such as cyber-physical attacks on energy infrastructure, where we have limited information and data on the level of risk posed. Indeed, since the Ukrainian attack multiple news outlets have reported that malicious software has been found on computers belonging to energy companies in the United States, raising concerns around the growing vulnerability of Critical National Infrastructure (CNI) (3–5).

The purpose of moving towards cyber-physical systems (including sensing, computing and communication hardware/software) is to develop intelligent monitoring and control of the physical world (6). However, the continuing shift towards smart cities, smart grids and the Internet of Things raises issues associated with increased connectivity and resilience. Indeed, the World Economic Forum’s Global Risks Report 2017 ranked the threat of cyber-attacks in the top ten for likelihood due to the growing number of physical systems connected to the internet (7). Other vulnerabilities include poor cyber security compliance, insufficient institutional training, the use of out-dated legacy software, vendor-contractor management practices and the increasingly easy access to hacking resources.

Yet, one of the largest concerns to defenders is the risk from zero-day vulnerabilities, which by definition we do not have existing information on, as the first time a vendor is made aware of an exploit is the day an attacker utilises it. Considerable time and resources must be placed into identifying and patching the vulnerability by the vendor, which can only begin once the attack has begun.

Due to the nature of rapidly evolving cyber threats, we are still some distance from undertaking effective resilience analytics of cyber-physical risks. Yet, decision-makers have highlighted the need for a quantitative framework that shows the direct and indirect socio-economic impacts of ‘what if’ scenarios. Whereas there are robust event sets of past natural catastrophes (e.g. hurricanes, flooding etc.), we have very limited information for cyber-attacks on CNI, including energy, transport, telecommunications,
water and waste, which are studied in this paper. We propose to overcome data limitations by applying stochastic counterfactual analysis as part of a new vulnerability assessment framework; see (8) for a timely introduction to counterfactual analysis. This assessment focuses on developing spatial attack footprints in the context of the direct and indirect socio-economic impacts of a Ukrainian-style cyber-physical attack taking place on the electricity distribution network serving London and surrounding regions in the South East of England. We apply both downward counterfactual analysis, where we explore the implications of a greater number of substations being affected than in the Ukrainian attack, and upward counterfactual analysis, considering what would have happened if fewer substations were affected. When we convert an attack footprint to a consumption shock for economic impact assessment, we assume a 24-hour blackout.

This counterfactual framework provides a more structured and rigorous approach to the risk analysis of emerging threats by using evidence-led scenarios, reducing subjectivity in the selection of scenario parameters by promoting benchmarking and calibration against severity (measured here as the number of electricity customers disrupted). To develop and demonstrate this counterfactual vulnerability assessment for emerging threats, we investigate the following research questions using a cyber-physical attack on the United Kingdom (UK) as a case study:

1) What is the direct impact on power consumers and how this scale with the number of substations compromised from a cyber-physical attack?

2) What is the indirect impact of a cyber-physical attack to other infrastructure users beyond electricity?

3) Does the vulnerability assessment present insights beyond the substation level, towards a systemic understanding of cyber-physical risk?

Previous research for three different types of electricity blackouts on two different continents, has found that it can be incredibly challenging for infrastructure operators to identify their own exposure (9–11). Although this may suggest a lack of defender capability, the root of the difficulty is that CNI networks have evolved over decades into very large socio-technical systems comprised of thousands of assets, technologies, networks and operator protocols. Analytics which show the potential risk to CNI are also
required by other parties such as governments, who have a responsibility to protect their citizens, but lack evidence on this matter.

It has been reported that almost half of all UK firms have been affected by a cyber breach or attack in 2016-2017 (12). Consequently, the UK has made a multi-billion-pound commitment to cyber security, as outlined in the National Cyber Security Strategy (13). Moreover, in the National Risk Register of Civil Emergencies, the UK Cabinet Office (14) identifies both (i) widespread electricity failure and (ii) the risk of cyber-attacks on CNI as having the potential to cause significant disruption. If a widespread electricity failure were to take place, current recovery plans called ‘Black Start’ could take up to five days due to a total or partial shutdown, with some potential disruption beyond this timescale (Ibid).

The rest of this paper is presented as follows. In Section 2, we undertake a literature review before presenting the methodology in Section 3. The results are reported in Section 4 before being discussed in Section 5. Finally, we conclude in Section 6.
2. LITERATURE REVIEW

2.1. Counterfactual approaches for risk analysis

‘Counterfactual’ quite literally means contrary to the facts, and usually involves a point of departure from a set of historical decisions or outcomes (15). Using counterfactuals as a method of intellectual inquiry is not new, as it has been applied by philosophers and historians for at least two millennia (16). Although there has been some focus on developing counterfactuals for risk analysis purposes (17–20), they have not been used to assess vulnerability, or ideally to develop resilience analytics for cyber-physical risk – an area where there is significant potential for application due to data limitations. Indeed, there are numerous ways that counterfactual analysis could usefully be applied to the science of emerging risks, including modelling of past events using stochastic forensics or scenario event trees (8).

After significant disruptive events, we inevitably evaluate mitigation measures and broader resilience strategies and how they could have reduced the impact, otherwise known as ‘upward’ counterfactual thinking, involving the hypothesising of how an event could have led to a more positive outcome. While this may be useful for risk practitioners, in contrast, we may rarely ask ourselves whether an event could have been much worse, otherwise known as a ‘downward’ counterfactual thinking. Figure 1 illustrates this conceptually, whereby different actions lead to more positive or negative outcomes, allowing a single event to be recreated as either an upward or downward counterfactual.
Cyber-risk is a classic example of how ‘a statistical analysis of past incidents alone does not provide a full description of the risk of future attacks’ (21). Indeed, our contention is that counterfactual analysis is a highly appropriate method for assessing cyber-physical risks to critical infrastructure, as we can borrow and combine information from the number of limited past events as well as from expert elicitation. Consequently, anchoring risk analysis in history makes the endeavour inherently more plausible and convincing than purely hypothetical scenarios, while still allowing for exploration of unknown exposure via zero-day vulnerabilities.

2.2. Cyber-risk assessments of CNI

While there are many engineering-focused assessments of direct and indirect interdependencies between power and Information Communication Technologies (22,23), with some focusing on cyber-attack risks (24,25), we still lack many appropriate methods for identifying key system interdependencies. Frameworks for cybersecurity risk assessment and management are increasingly being put forward as existing approaches (including probabilistic and risk-based decision-making techniques applied to cyber systems) do not always properly address threats, vulnerabilities and
potential consequences (26). Indeed, qualitative risk metrics are often used within current industry standards for estimating cyber security risk, as opposed to using the quantitative risk metrics commonly found in other industrial sectors such as finance and banking (e.g. Basel II) (27). Moreover, some assessment methods only focus on existing vulnerabilities, such as the Common Vulnerability Scoring System (28), without putting enough consideration to unknown vulnerabilities. In improving the risk analysis of cyber-physical systems for CNI, we need not only better cyber vulnerability assessment, but also impact analysis of how vulnerabilities may cause cascading failures to other systems and ultimately lead to different socio-economic impacts (29).

One example of a cyber-attack on the oil and gas sector in the Gulf Coast of the US presents a framework for linking cybersecurity metrics to the modelling of macroeconomic interdependencies using the Inoperability Input-output Model (IIM) (30). This methodology has also been used to quantify the risk posed by interdependent Supervisory Control and Data Acquisition (SCADA) systems vulnerable to cyber-attack (31), and fault trees have also been developed to assess the systemic risks associated with cloud-computing (32). Moreover, one method used to develop effective risk mitigation approaches involves attack and defence modelling to understand the strategic interactions between different agents (33,34). Increasingly, the risk posed by insider threats has been assessed to provide insight into the human cybersecurity factors affecting an organisation's environment during the attack of a corporate cyber network (35). Finally, in a paper presenting multiple approaches, Paté-Cornell et al. present a general probabilistic risk analysis framework for the management of critical infrastructure from cyber threats, producing loss results from cyber-attacks under different risk mitigation measures (21). A review of the potential socio-economic impacts of critical infrastructure failure will now be undertaken.

2.3. Socio-economic impacts of infrastructure failure

Although we do not have many examples of cyber-attacks on CNI, we do understand the potential socio-economic impacts of infrastructure failure due to other natural and man-made hazards. For example, major blackouts in North America took place in 1996 and 2003 (36,37), as well as in Europe in 2003 and 2006 (38) which we can learn from. For example, during the December 2015 flood-induced outage in Lancashire, UK, the loss of power led to the complete disruption of digital communications, health
care provision, retail businesses and banking, transport and essential utilities (39). To understand such disruptions across interconnected infrastructures different modelling techniques have been employed, which include, among others, agent-based models of complex adaptive systems (40), empirical analysis (41), system dynamics approaches (42), network-science based models (43), macroeconomic input-output (IO) based (30) or computational general equilibrium (CGE) models (44,45). For a detailed literature review on different modelling techniques see (46). While all the above modelling techniques provide their own advantages, increasingly network models have proven to be most convenient for suitably representing interconnected infrastructures at multiple scales (47).

This paper uses network models to quantify social impacts of infrastructure failures, in terms of customer disruptions. Investing disruption using this key metric builds on existing analysis of: (i) measuring spatial critical hotspots of total disrupted customers affected by failure to England and Wales’ interconnected electricity, transport, water, waste, telecoms infrastructures (48); (ii) flood vulnerability assessment of electricity networks and dependent water, wastewater, telecoms, and transport assets in the Thames catchment in England (49); (iii) quantifying daily passenger disruptions on Great Britain’s rail network (50); (iv) measuring customer disruptions due to flooding and drought exposures of energy, transport, water, and waste networks in China (51); and (v) analysis of customer disruption due to failures in interdependent electricity, fuels, and transport networks in New Zealand (52).

The most preferred approaches for measuring economic impacts of infrastructure failures include the CGE and IO approaches. Rose et al. use CGE models to estimate the business interruption impact of a terrorist attack on the electric power system in Los Angeles, focusing on both indirect economic effects, and the role of resilience (44). Port infrastructure has also been the focus for modelling the potential economic consequences of a 90-day disruption (45), using a supply-driven IO modelling approach. The IIM literature includes several studies on economic impact assessment of infrastructure failures (30,53). Mostly in existing CGE and IO based approaches, infrastructures are represented as macroeconomic sectors, with little link established between their socio-technical network structures and economic characteristics. Recently, some studies have established that link, by integrating customer disruptions
measured from network failure models with demand-driven IO models (50,54). This paper also integrates an infrastructure network model with an economic model, while addressing some shortcoming of existing approaches. For example, many IO models utilise data that are a number of years old due to the complexity and amount of time required to develop national account statistics. Moreover, IO and CGE models can be poor at forecasting future economic states. We overcome this by employing an industry-standard Macroeconometric Error Correction forecasting model that combines the advantages of both Vector Auto Regression and Dynamic-Stochastic General Equilibrium methods, allowing the robust testing of scenarios in future periods using the most up-to-date information. The methodology will now be outlined in Section 3.
3. METHODOLOGY

We present a methodological framework that allows risk analysts to assess the direct and indirect socio-economic impacts of a counterfactual cyber-physical attack on electricity distribution substations, as illustrated in Figure 2. In the following sections, we discuss the specific components of the framework, but first we present a generalised methodology for quantifying the different components.

As discussed previously, we are interested in counterfactuals of a historic event $HE$, which we identify as a threat to the system of study. We denote the level of cyber-physical threat manifestation of the actual event as $HE_0$, and different counterfactual events as $HE_1, \ldots, HE_z$. For example, the counterfactual scenario $HE_i$ signifies that $m$ physical components of a $n$ component system could have failed due to a cyber-attack. Specifically, in our study $HE_i$ signifies that $m$ electricity substations out of a $n$ substation network could have failed due to a cyber-attack. For an exhaustive vulnerability assessment of the counterfactual event $HE_i$ we look at all the $d = \binom{n}{m}$ combinatorial failure possibilities in the system, to obtain a set of failure scenarios $S_1, \ldots, S_d$. The failure outcomes are first measured in terms of the populations of direct electricity customer disruptions $DC_1, \ldots, DC_d$ corresponding to each failure scenario. Investigating the indirect disruptions for all failure scenarios could be infeasible, given the large number of combinatorial possibilities, so we select certain representative scenarios. This is done by assigning exceedance probabilities $EP_1, \ldots, EP_d$ to direct the customer disruptions of the failure scenarios. We then sample scenarios based on exceedance probabilities of interest.

Following the selection of a smaller set of specific scenarios $S_1, \ldots, S_h$, we estimate the indirect customers $IC_1, \ldots, IC_h$ for different sectors dependent upon electricity. We can also measure the wider macroeconomic impacts in terms of different metrics such as private consumption ($PC$), investments ($In$), capital stock ($CS$), gross value added ($GVA$) or gross domestic product ($GDP$). Hence in the end for a particular counterfactual event $HE_i$ we can sample a scenario $S_j$ and assemble different direct and indirect metrics ($DC_j, IC_j, PC_j, In_j, CS_j, GVA_j, GDP_j$).
Figure 1 Framework for assessing the socio-economic impacts of cyber-physical attack

![Diagram of Framework for assessing the socio-economic impacts of cyber-physical attack]

Figure 2 represents how the above generalised methodology was applied in the current study. The initial development of this framework involved a set of stakeholder interviews conducted with representative organisations (number of interviewees in parentheses) from energy [13], security [6], insurance [20], defence [2], government [9] and academia [4], to assess current understanding, potential exposure and analytics which could aid resilience building activities.

Research initially began in July 2015 when defining a hypothetical event similar to the Ukrainian attack, at a set of scenario development workshops, consisting of UK representatives from government [5] and academia [8], and the electricity [2], defence [1], risk management [3] and cyber security [1] industries.
Rather than regional control rooms being of greatest risk, which have significant cyber and physical security procedures, it was local substations that were identified as vulnerable assets. Traditionally, these local substations are less protected than those parts of the network which are higher up the distribution (or transmission) hierarchy. However, stakeholders identified that pinpointing which substations were potentially vulnerable was highly challenging. In December 2015, as this research was being written up into an industry white paper (9), the Ukrainian electricity substation attack took place, turning a hypothetical scenario into an actual event.

3.1. Threat identification

A threat is defined as any potential hazard implemented for malicious intent, which could interfere with normal operational conditions, causing a blackout event. This likely involves ‘hacking’ Industrial Control Systems (ICS), which include SCADA systems, distributed control systems and programmable logic controllers. These systems are found in many industrial applications including CNI systems. Indeed, electricity distribution operators are increasingly using some form of ICS to control, automate and maintain operation of their equipment and the flow of electricity (29). Since Stuxnet sabotaged Iran’s nuclear programme in 2009 and 2010, we have seen a dramatic increase in the threat to ICS and SCADA systems, raising concerns around our ability to protect CNI e.g. the Dragonfly Campaign (55). The attacker may specifically aim to spoof sensors with false data, disconnect key devices required for normal operations, and control physical components such as actuators. As well as remote attacks, there is also potential for intruders to physically connect a rogue hardware attack platforms into the Local Area Network of a substation, breaching ‘air-gapped’ networks (see Cambridge Centre for Risk Studies, 2016).

3.2. Threat manifestation

The 2015 Ukrainian attack has been documented by the US Department for Homeland Security (56). The malicious attackers managed to firstly deliver potentially malicious BlackEnergy Malware via spear phishing emails using malicious Microsoft Office attachments. Secondly, intruders conducted comprehensive reconnaissance of critical systems in advance. Finally, to commence the attack
Substation breakers were disconnected using legitimate credentials with either remote administration tools or remote ICS client software via a Virtual Private Network connection. Little about this attack was specific to Ukrainian technology or critical infrastructure, and therefore it could be replicated in similar ways in other nations. Analysis of this attack (2) indicates that (i) the attack could have been far worse, and (ii) the idea of a major cyber-physical attack on energy infrastructure is not a matter of if, but when.

Although there has been some discrepancy in the number of substations affected by the Ukrainian attack in 2015, approximately 30 substations were affected (2), consisting of seven 110 kV and the remainder 35 kV (57). However, more substations could have been affected. We replicate the Ukrainian-style threat for the UK, which is divided into electricity regions consisting of nine Distribution Network Operators (DNOs), out of which we have chosen one. To capture both upward and downward counterfactual events, we select both half and double the number of substations affected by the Ukrainian attack. We therefore test the implications of 15, 30 and 60 substations being compromised by a zero-day vulnerability. As consequently explained in this methodology, each of these three potential events is investigated separately utilising Monte Carlo Simulation. The number of substations is used to represent different scales of attack. As this analysis focuses specifically on modelling the socio-economic impact of a Ukrainian-style cyber-physical attack on the UK, we measure the direct disruption as the proportion of the local population disconnected from the electricity network.

3.3. Direct disruption: Threat scenario selection

Following the creation of the threat scenarios, we estimate the disruptive consequences associated with different cyber-physical attacks on electricity substations. Figure 3 provides a generalised overview of the electricity transmission and distribution networks, which form a hierarchy of flows, from power generation to electricity users, where the electricity is sequentially stepped down from high voltage transmission networks to low voltage distribution networks (47). ‘Stepping-down’ is performed by electricity transformers that are located within substations, which we assume are ICS operated and potentially vulnerable to a remote cyber-physical attack. Our study focusses on 132kV substations in the network hierarchy as the points of cyber-physical attack. To understand the impact, this modelling
approach assigns spatially located customers to local distribution substation assets, leading to variation in the number of customers per substation (47,48,50), and further details in Section 3.4), hence leading to ‘larger’ or ‘smaller’ substations.

Figure 3 Generalised overview of electricity transmission and distribution in England

Classical scenario analysis is a tool frequently used for risk management purposes, however it often is used in a very deterministic manner. It is arguably more beneficial to know something about the distribution of possibilities associated with a threat, given that there are a wide range of potential outcomes that could arise (particularly for zero-day vulnerability risks). There are approximately 252 132 kV substations in the London, South East and East of England region available for the simulation in this analysis. We stochastically explore the event space by randomly sampling 15, 30 or 60 substations per event, leading to a total number of $5.2 \times 10^{23}$, $6.9 \times 10^{38}$, $6.9 \times 10^{58}$ substation combinations respectively. Each substation has an equal likelihood of being selected.

This simple approach, in which we sample subsets of affected substations and calculate the population impacted in each case, is justified on two grounds. Firstly, we lack the technical details of each
substation, and their potential vulnerabilities, as well as the actual topology of the substation network that would be needed to undertake a (more sophisticated) network view of impairment. This is partly reflective of the variation in the state of knowledge of CNI operators. Secondly, the level of uncertainty embodied in the assessment of cyber-physical risks to CNI suggests that the uncertainties from zero-day vulnerabilities will outweigh any additional resolution obtained at the engineering level.

Once we have a dataset of simulation results we calculate the Exceedance Probability (EP), as is common in the modelling of catastrophe risk (58). The EP \((P)\) for event \(i\) based on the number of substations randomly attacked is as follows:

\[
P_i = \frac{m_i}{(n_i + 1)}
\]

where \(m_i\) is the rank of the total population affected by each simulation iteration (with 1 being given to the largest possible value), and \(n_i\) being the total number of simulation runs. Event simulation results are illustrated in Figure 4. For example, for an event with 15 attacked substations (Event 1), the probability of more than 1 million people being affected, i.e. the probability of exceedance of 1 million people, is approximately 50%.

Spatial and indirect disruption results can be found in Section 4.
For a given stochastic event, for example where 30 substations are attacked, the difference in the population affected by the substation selection relates to changes in the spatial attack footprint and the network topology. In contrast, differences between stochastic events relate to the scale of attack across the network and the number of infected substations. We select exceedance probabilities that relate to the mean impact (50th percentile, or median) and the tail risk (10th and 1st percentile), as these were identified as being important to the scenario stakeholders at the development workshops. Tail risk has also been identified as vital in the cyber-risk literature (21).

Figure 4 shows that an extreme scenario (10th or 1st percentile) associated with a lower number of infected substation, has a smaller disruptive impact than the mean scenario of an event with more infected substations. For example, the 1st percentile of the event with 15 infected substations affects
power supply for 1.5 million people, while the mean number of disconnections for a 30-substation event is above 2 million. The implication for the management of power distribution networks is evident: Preventing (or recovering from) cyber-attacks which can scale up from a few substations to many is the priority. By implication, protecting larger substations with greater investment of resources, may be a less effective means of prevention according to the metric of consumer disruption.

3.4. Indirect disruption: Cascading network failure metrics

We also examine other critical infrastructures, including rail, ports, airports, water, wastewater, and telecommunications, which typically derive their electrical power needs from a direct connection to the distribution network. The electricity network is connected to other infrastructures, to map their dependencies on electricity, to create a system-of-systems network model with results reported for 379 Local Authority Districts (47). The dependent infrastructures include water towers, wastewater treatment works, macrocellular base stations, airports, and ports as point assets, along with a railway network. Figure 3 already showed a graphical representation of the infrastructure types connected to the electricity network in the demand layer. For details of these infrastructure assets and networks, along with their dependencies see previous studies (47,49,50). The dependencies are derived based on multiple criteria including (i) existing data on the physical connections between networked assets; (ii) geographic proximity of assets to their nearest electricity substations of appropriate voltage, and (iii) functional understanding of the flow of electricity from substations to other infrastructures. For example, we consider instances of network redundancies where large assets such as airports are connected to multiple substations (47), and several railway stations are connected to sets of substations (50). The exact voltage level to which individual CNI assets are connected may vary. However, the majority of each asset type connects to the same voltage level (the same substation type).

We define failure as a condition of the network node (or edge) asset such that it is no longer able to perform its functional purpose. In our description, this means that the service demand satisfied by the affected node is lost and all its connections are interrupted. Based on the selected scenarios, it is assumed that all the included electricity substations have failed and subsequently the number of disrupted
electricity customers is estimated. In parallel there are also cascading failures towards the dependent assets of other sectors.

To estimate disruption, we first model customer assignments to different types of infrastructures. For all infrastructures, we create average daily customer estimates. Electricity, water, wastewater, and telecoms provide utility services over fixed areas called asset footprints (47,48,50). These assets footprints are derived using a Voronoi decomposition technique, which results in connecting customers to their nearest assets in space – an assumption which holds true in the real world (47,48,50). Assigning customer values to assets is based on a spatial union of its asset footprint with census derived population estimates. For the railway network customer demands are derived using a model of station entries, interchanges and exits, and by combining these with train frequencies along routes, to obtain daily origin-destination trip assignments for passenger flows (50). Airports customers are derived from annual flight statistics and are calculated as the total number of terminal passengers for an average day in 2009. Similarly, average daily port freight tonnage is derived using 2009 national port usage statistics.

For each threat scenario, the infected electricity substations lead to failure propagating along the whole network path where the flow of electricity via the failed substations takes place (as illustrated in Figure 3). For other critical infrastructures such as airports, ports, telecoms masts, water towers and waste water treatment works customer disruptions are estimated based on whether the connected electricity substation has failed. For the railway network disruptions, we first consider the stations disrupted due to connection to failed electricity substations. Following which we consider all origin-destination journeys that are lost even after rerouting due to disruption of the selected stations (50). The aggregated number of customers affected by each critical infrastructure sector provides the disruption estimates reported.

3.5. Indirect disruption: Economic impact metrics

We quantify economic impact given a demand-side economic shock due to reduced private consumption, from households being without power. Private consumption is affected as consumers are unable to complete daily economic transactions. The Oxford Economics Global Economic Model is
utilised which is a widely employed macroeconomic model with users including the International Monetary Fund and World Bank (59), and consists of over 26,500 interlinked equations based on historical correlations and economic theory. Multivariate forecasts are produced for many economies, but here we focus only on the UK. The modelling approach adopts Keynesian principles in the short run, where shocks to demand generate economic cycles that can be influenced by fiscal and monetary policy. While over the long run, output is determined by supply-side factors including investment, demographics, labour and productivity. The basic modelling principle of the simulation framework is that GDP output (Y) can be expressed as the summation of aggregate Consumption (C), Investment (I), Government Spending (G) and Net Exports (NX) (hence, \( Y = C + I + G + NX \)).

We use the model to see the effect of a shock directly applied to private consumption to understand the impact on GDP. A quarterly shock is parametrised using the total population affected for each scenario for a single 24-hour period. Additionally, we report a set of intermediate macroeconomic indicators including lost investment, capital stock formation, manufacturing GVA and services GVA, to quantify the economic impact by scenario.

On the one hand, this may overestimate the impact due to the potential rescheduling of consumption purchases. While on the other, we do not consider potentially much larger impacts on the supply-side due to business interruption, or indeed harm to labour productivity. Given the challenge of estimating supply-side impacts, they are of the scope of this analysis and are identified as an area of future research in the discussion.
4. RESULTS

The modelled electricity distribution network accounts for just under a third of the total UK population. In this section, the results are reported for the nine scenarios tested within the assessment framework with a spatial resolution of 379 Local Authority Districts. Firstly, the spatial direct disruption to electricity users will be reported, followed by the indirect impacts. The latter includes spatial customer disruption on non-electricity CNI and estimates of economic loss due to the consequent disruption to private consumption across all affected infrastructures.

4.1. Direct disruption metrics: Spatial distribution of disruption

In the selected scenarios, the population affected ranged from 1-5 million, reflecting between 1.6-7.6% of the total UK population. A larger proportion was affected as the number of compromised substations increased. This contrasted with changes in the exceedance probability, which only had a marginal impact on the direct level of population disruption. This pattern is illustrated within Figure 5, where a 15-substation attack with a 1% probability of exceedance (S3) had a lower severity, in comparison with a 30-substation attack with only the mean probability of exceedance (S4).
The spatial attack footprint is visibly different between the event types, in terms of the scale of the disruption. However, even in the least severe scenarios there was still a significant proportion of the population disrupted, with over 200,000 people being affected in at least one London Local Authority, in each 15-substation scenario. Having examined the spatial footprints of the direct population impacts, the indirect results will now be reported.
4.2. Indirect disruption metrics: Spatial distribution of disruption

As well as direct infrastructure disruption, a cyber-physical attack on electricity distribution substations could lead to further indirect infrastructure cascading failure. Some of this disruption will take place outside of the attack zone footprint, with rail passenger journeys being a prime example, as illustrated in Figure 6. In every scenario, the biggest impact was evident in London due to the large number of commuters relying on transportation from suburban or rural locations into urban areas. However, the spatial distribution of passenger disruptions is highly dependent on the scenario and the location of compromised substations. For example, in S3 and S7 the West Coast Mainline railway is affected with disruption to passenger journeys leading from London all the way to North Wales, North West England and Northern Scotland. In contrast, in S2 and S4 the East Coast Mainline is affected, with passenger disruption leading through the East Midlands, Yorkshire, North East England and finally to Scotland. Passenger disruptions in each 15-substation scenario were more localised to London, whereas events with more compromised substations led to a greater number of passenger disruptions in other major cities such as Birmingham, Manchester and Leeds.
Figure 6 Indirect disruption: Rail passenger disruptions
The spatial distribution of disruption to other critical infrastructures was less spread across the UK, as provision is more local. For example, with telecommunications, fresh water and waste water, services are distributed closer to the end-user’s premises, mainly confining disruption to the areas with compromised substations, as illustrated in Figure 7. Hence, the underlying topological structure of the infrastructure network has a significant impact on the spatial extent of disruption. The significance of this disruption for each CNI may manifest in different ways. For example, wireless telecommunications in theory allow users more flexibility to access different basestations, but recent evidence has suggested that during power loss it takes only a couple of hours for the system to become completely inoperable in the blackout zone (39).
Figure 7 Indirect disruption: Cascading failure across CNI

Critical National Infrastructure Disruption by Scenario
Customer disruption reported by Local Authority District

Population Disruption (Thousands)
- (>800)
- (600 - 800)
- (400 - 600)
- (200 - 400)
- (100 - 200)
- (75 - 100)
- (50 - 75)
- (25 - 50)
- (1-25)
- (0)
The total impacts are illustrated in Figure 8, for both customer disruption and economic impact by scenario. The aggregate figures show that for infrastructure cascading failure, disruption in telecoms, fresh water and waste water is highly correlated with the number of electricity customers affected in each scenario. Airport, maritime port and rail disruption is less correlated, and highly dependent on whether certain, more critical, substations are affected in the attack.

In terms of the economic impact, which we assess here based purely on consumption disruption, the estimated loss to GDP in 2018 following an event in Q4 2017, ranged from £66 million in S1, up to £276 million in S9. As illustrated in Figure 8, the quantity of lost investment and capital stock formation in each scenario was relatively similar. Lost investment amounted to £13 million in S1, to £58 million in S9, whereas lost capital stock formation ranged from £11 to £51 million respectively. The GVA impacts broadly reflected the underlying economic structure of the UK’s economy which is dominated by service-sector activities. This meant manufacturing GVA loss was relatively minor in comparison with the £38-171 million loss in services.

Having reported the results, we will now discuss their implications in Section 5.
Figure 8 Total impacts by scenario

CNI Customer Disruption Metrics by Scenario
NB Port disruption measured in thousands of tonnes

Economic Loss Metrics by Scenario
Aggregate disruption reported for the UK
5. DISCUSSION

Even the most diligent operators are exposed to zero-day exploits; indeed, the discovery and application (by attackers) of zero-day flaws is stochastic by nature. Modelling and pinpointing these exposures is therefore extremely challenging. In this paper, the analysis focused on applying counterfactual information to develop upward and downward scenarios of physical asset failure, enabling us to answer, ‘what if’ questions by quantifying impacts without getting mired in specific details of either the attackers’ modus operandi or the engineering specifications of the system attacked (which may be unknown or unavailable, respectively).

A key finding identified within this paper, pertaining to research question 1, is that the size of direct population disruption from a substation attack is better predicted by the number of substations affected, rather than by taking into account the size of population served by (larger) substations. The first order implication is that monitoring and response preparedness across the entirety of substations under the control of a distribution network operator is more valuable than increasing the resistance to attack of more important substations.

Regarding the indirect impacts, pertaining to research question 2, we find that customer disruption for different CNI sectors is correlated with power loss in two distinct ways. For example, telecoms, fresh water and waste water services are highly correlated with the spatial footprint of electricity disruption. In contrast, in transportation there are relatively few, yet very important, critical hubs. Therefore, there may be a lower probability of infecting a substation which serves one of these critical hubs. However, if this happens, there could be higher disruptive consequences from a single asset. This results from differences based on disruption metrics, and whether one uses the number of direct connections lost (as applied here), or a weighted metric which reflects the relative importance of each connection (an area for future analysis). For example, airports and maritime ports may deserve a higher weighting due to the economic disruption associated with their inoperability.

The analytics provide the following systemic information, helping to address research question 3. Customer disruptions are highly correlated in some CNI sectors with the spatial attack footprint on the
electricity distribution network. However, for other CNI sectors with critical hubs, such as commercial transportation, customer disruption is not necessarily an effective measure of societal value. Indeed, there may be three competing interests for CNI operators. Firstly, meeting regulated service standards for domestic consumers. Secondly, meeting industrial demand from high-value customers who purchase large quantities of services with high reliability requirements. Thirdly, providing security of supply for critical infrastructure assets including essential government services, hospitals, healthcare and commercial information technology for banking and payment services. We believe the next steps towards cyber-physical resilience analytics, will explicitly address the multiple objectives of these competing interests, to quantify trade-offs at the operator and policy levels.

We bring the paper to a close with a reflection, aimed at CNI operators. It is difficult to completely disable large numbers of substations, i.e scale an attack, as vulnerabilities are specific to types of substation hardware, software, and all the different components of security that need to be overcome to enable a plan to succeed. Therefore, the greater diversity between substations in terms of their software and hardware subsystems, the more difficult it is for a vulnerability of a subsystem to cause a problem at scale (across substations). Indeed, the scalability of the attack depends on the standardisation of components and systems in place.

Yet, prescribing a multiplicity of different configurations such as substation design, subsystem and implementation, may not be practical, as any increase in the number of configurations requires a comparative increase in resource for maintenance and security. Due to the uncertainty associated with the dynamic nature of cyber, a more appropriate strategy may be to invest in and actively undertake comprehensive and widespread monitoring of assets. This approach does not interfere with existing ICS, and will allow a much more efficient response if, or rather when, an attack does take place.
6. CONCLUSION

A key finding identified within this paper is that the size of direct population disruption from a substation attack is better predicted by the number of substations affected, rather than paying specific attention to larger substations. Nevertheless, certain substations are critical for the functionality of other key assets such as airports or maritime ports, for which other metrics of societal value, other than population disruption, are appropriate.

With such a small history of known cyber-physical attacks, individual organisations have struggled to justify investment in ICS cyber security measures using traditional return-on-investment thinking. We are unlikely to obtain a robust event set of cyber-physical attacks on CNI in the near future because of the rapidly changing landscape associated with this threat. This is common with low probability, high impact events, which is a key justification for using a counterfactual approach.

In many countries, such as the USA or UK, governments do not own the infrastructure – private operators do. Yet, the public will look to governments when we see another cyber-physical attack on critical infrastructure. The scenarios presented within this analysis provides further evidence on a developing area, for key private and governmental stakeholders, on how direct and indirect customer disruption takes place for different scales of cyber-physical attack. This paper sets a direction for the assessment of risks that are both systemic and emergent, by undertaking a vulnerability assessment using a stochastic counterfactual framework. Cyber-physical vulnerability assessment, as a path to developing effective resilience analytics, must quantify the trade-offs inherent in a system made up of multiple institutions with different objectives, such as governments, infrastructure operators and commercial consumers of infrastructure services.
7. REFERENCES


