frontiers in Energy Research

ORIGINAL RESEARCH
published: 30 October 2020
doi: 10.3389/fenrg.2020.591999

Using Design Thinking to Understand Cyber Attack Surfaces of Future Smart Grids

Stephen Snow[1]*, *Jassim Happa*[2], *Neil Horrocks*[1] and *Mashhuda Glencross*[1]

[1]Information Technology and Electrical Engineering, University of Queensland, Brisbane, QLD, Australia, [2]Information Security Group, Royal Holloway, Egham, United Kingdom

The success and proliferation of smart enabled electricity grids depends on our ability to reason predict and prevent adversarial behavior. This paper details a novel application of design thinking to smart grid cyber security, presenting a scalable framework for defining, ranking and externally validating future smart grid threats and then modeling adversarial behavior. Using an expert panel for external validation, this paper prioritises three salient threats to smart grid security in the near future: 1) malicious entry to a network operator's control room allowing remote shutdown of grid infrastructure, 2) distract and decoy tactics as a means of diverting resources away from the site of an attack, and 3) manipulation of demand attacks using widespread commandeering of household IoT technology. Smart grids represent a salient test deployment for this framework, given the near complete lack of successful existing attacks from which empirical evidence can be leveraged. Our framework for reasoning about potential future threats is scalable from company-specific to sector-wide threats and enables risk owners to make well-informed decisions and better prepare against future threats.

Keywords: smart grids, design thinking, cyber security, human behavior analysis, smart meters, privacy

OPEN ACCESS

Edited by:
Chau Yuen,
Singapore University of Technology and Design, Singapore

Reviewed by:
Wen-Tai Li,
Singapore University of Technology and Design, Singapore
Hwei-Ming Chung,
University of Oslo, Norway

***Correspondence:**
Stephen Snow
s.snow@uq.edu.au

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *06 August 2020*
Accepted: *16 September 2020*
Published: *30 October 2020*

Citation:
Snow S, Happa J, Horrocks N and Glencross M (2020) Using Design Thinking to Understand Cyber Attack Surfaces of Future Smart Grids. Front. Energy Res. 8:591999. doi: 10.3389/fenrg.2020.591999

1. INTRODUCTION

Smart grids combine machine intelligence, automation and computation with existing physical energy-grid infrastructure to achieve more intelligent energy distribution, delivery and management (Mrabet et al., 2018; Otuoze et al., 2018). They offer benefit to stakeholders and energy networks through interfaces and mechanisms of engaging with each other to support transactions (Farhangi, 2009). Yet the transition from physical control to automated or remote control underscores the importance of smart grids being secure and resilient against system anomalies (Mrabet et al., 2018) and threats posed to the system. Cyber attacks have already exposed over three billion Yahoo account holders' personal information (Lorio, 2017), shut off power to over half a million Ukrainian households (Case, 2016). With over 35 billion internet connected (iot) devices online (Yu et al., 2015) and widely reported poor security practices among users (Wash et al., 2016), modeling suggests even greater threats to energy stability exist through malicious commandeering and simultaneous activation of high powered iot devices (Soltan et al., 2018).

The security of smart-enabled electricity grids rests on our ability to reason, predict, model and hence prevent adversarial behavior. Numerous reports have been published on the state and challenges of technical smart-grid security (Khurana et al., 2010; Santacana et al., 2010; Yan et al., 2012), and robust security assessments exist for the network communications protocols which have formed components of smart grids in the past, e.g., network protocols such as DNP3

and home/local area network protocols such as Zigbee, Z-Wave, 802.11, etc. (Hu et al., 2015; Mahmood et al., 2015). Yet detailed, empirical analyses of real-world smart grid attacks are lacking or commercially sensitive. Further, while attacks are typically analyzed in terms of their technical accomplishment, entry point and attack surface, behind each successful attack is a human (or group of people) with specific motivations, goals and characteristics that inform their decisions (Sliva et al., 2017). Van Ruitenbeek et al. (2010) argues that "meaningful measurements of system security cannot occur in a vacuum void of information about the system's adversaries."

Design thinking is increasingly incorporated into business practices as an strategic means of engaging with wicked problems (Fonseca Braga, 2016). This approach emphasises problem definition and human-centricity in both problem definition and problem solving (Plattner et al., 2009). We postulate it is a theory that can be applied to threat analysis for smart grid systems, emphasising the understanding and prediction of adversary behavior. To the best of our knowledge, this paper represents the first application of design thinking to cyber threat modeling.

1.1. Paper Contributions

In this paper we apply design thinking to cyber threat modeling through a process of empathising, defining, ideating, prototyping and testing. First, potential attack scenarios are appraised and ranked by an panel of three experts, comprising two energy sector professionals with experience in cyber threat analysis and a cyber security expert with experience consulting in the Australian energy sector. We identify 1) remote access to network control systems, 2) deception and distraction of command center operations, and 3) manipulation of demand attacks involving distributed household IoT devices, as key threats to smart grid security. Second, from this process we identify how smart-grid architectures can be exploited by attackers, defining a human-centric behavioral model for adversary behavior, building on Fogg (2009) and Van Ruitenbeek et al. (2010), understanding adversary behavior as dependent on adversary motivation, ability and triggers for a given behavior. The three key contributions of our work are:

- **A scalable application of design thinking to identify and reason about security threats in cyber-physical systems**, using smart grids as the primary use case because their attack surfaces (in particular inter-dependencies between systems) are poorly understood today.
- **An analysis of current and near-future smart grid security threats**. Based on feedback from energy sector and security sector experts, we present an externally validated and ranked list of smart grid vulnerabilities that future work should prioritise.
- **A framework for adversarial modeling** [using Fogg's behavioral theory (Fogg, 2009)] to define, preempt and characterise threats on energy systems to examine where and how attacks may unfold in the grid.

Our paper is organised as follows: Section 2 overviews smart grid security and privacy issues. Then, we present a background in design thinking and behavior theory in Section 3. In Section 4 we discuss a set of smart grid security scenarios using design thinking. These were presented to a smart grid expert panel to rank potential attacks in terms of their a) feasibility, b) likelihood, c) consequence. Section 5 summarises our results before Section 6 describes how our approach could be used to identify likely attacks. We discuss our findings in Section 7 and conclude the paper in Section 8.

2. SMART GRIDS AND CYBER ATTACKS

Smart grids are defined by the deployment of distributed computing into energy networks and grid operations (Otuoze et al., 2018). Distributed computing facilitates two-way communications between multiple grid components, providing for greater control, intelligence, management and higher frequency energy information allowing for alternative tariff structures (Otuoze et al., 2018). The US National Institute of Standards in Technology (NIST) identifies a smart grid as composed of seven logical domains: bulk generation, transmission, distribution, customer, markets, service provider, operations (FitzPatrick and Wollman, 2010). Smart meters represent a vital component of smart grids, capable of frequent measurement and transmission of energy data enabling two-way communication with an electricity provider or network (Pitì et al., 2017). Smart meter adoption rates vary widely between countries and between states within countries, e.g., in the US, four states have less than 1% smart meter adoption, while six states have over 80% adoption (Alexander Mey, 2017).

2.1. Security-Related Smart Grid Vulnerabilities

It is difficult to generalise security vulnerabilities of smart grids as metering infrastructure, communication protocols, security protocols, physical network architecture and other factors vary widely between countries and between energy networks. In 2013, Aloul et al. (2013) provided a list of eight smart grid vulnerabilities, which include:

(1) **Customer security.** Smart meters collect large amounts of potentially sensitive data (depending on the granularity recorded), from which it may be possible to infer customer activities in the home. Threats include interception of transmission and the storage of this information by collectors of this information such as utilities or network operators.
(2) **Lifetime of power systems**. The integration of IT systems with legacy infrastructure means it is likely that outdated equipment will be in service alongside modern equipment. Older equipment and/or older communications protocols may serve as weak points.
(3) **Physical security**. Strength of security in existing power systems is dependent partly on their age, necessitating physical access (e.g., to sub-stations or transformers) in order to access information and service equipment.

Introducing remote access also introduces potential for unauthorised access.

(4) **Implicit trust between devices.** Owing to the lack of possibility for remote access, legacy systems involve implicit trust between components, meaning signals can easily be spoofed. While previously not an issue, the addition of remote access means spoofing signals between legacy systems may cause false alarms or misuse.
(5) **Greater number of intelligent devices**. The size of smart grids, both physical size and the sheer number of connected devices makes for multiple entry points and makes monitoring challenging.
(6) **Different team's backgrounds/Human error**. A large number of cyber attacks involve some form of social engineering. Unorganised communication between employees increases risk of attack. 52% of businesses list employees as their biggest weakness in IT security (Kapersky Lab, 2017), yet a survey reported less than 20% of UK businesses required their staff to receive cyber security training (Klahr et al., 2016). Human error, carelessness and inadequate training remain as security vulnerabilities, with potential for phishing attacks to target households or employees of energy utilities or network operators.
(7) **Using Internet Protocol (IP)**. Using IP standards in smart grids increases compatibility between components, but IP is inherently vulnerable to IP-based network attacks.
(8) **More stakeholders**. An increasing number of stakeholders involved in smart grids relative to traditional grids raises the potential for insider attacks.

Mrabet et al. (2018) provides a generalised overview of the anatomy of a smart grid attack based on the sequential attacking cycle which follows these steps: 1) Social engineering and traffic analysis for reconnaissance and entry, 2) Vulnerability scanning through port scanning, IP scanning, etc., 3) Exploitation, using a combination of one or more of; virus/worm, denial of service, man-in-the-middle, replay attack, etc., and 4) Establishing a backdoor to maintain access after the attack has been identified.

2.1.1. History of Smart Grid Related Attacks to Date

To date there have been no direct large-scale attacks on smart-enabled grids. However energy companies have been targeted by previous cyber security attacks:

- **Stuxnet** targeted Iran's nuclear enrichment program in 2010. A consultant inserted an infected removable device into a company computer. Malware then propagated throughout the computer network instructing company computers to connect to an external command and control center. The virus targeted programmable logic controllers operating centrifuges, causing the centrifuges to malfunction and self-destruct, reportedly damaging close to a fifth of Iran's nuclear centrifuges (Langner, 2011).
- **Ukrainian power grid** Illegal third-party entry was gained into the supervisory control and data acquisition (SCADA) systems of a large Ukrainian electricity distribution company. The attack was initiated with a combination of spear phishing and malware and resulted in the de-activation of 30 substations, affecting power to 225,000 households (Case, 2016).
- **NightDragon** gained entry to a number of energy companies' systems using spear-phishing. An exploited company server was then established as a command and control server. NightDragon was used to harvest sensitive data related to financial details, bids and operations, but did not seek to disrupt grid operation (Miller and Rowe, 2012).
- **Brown Ferry** A nuclear plant in the US named Brown Ferry experienced failure of circulation pumps in 2006 necessitating a full shut down. A later investigation of the incident identified the attack as a cyber attack (Goel and Hong, 2015).

2.1.2. Energy Network Preparedness

Interviews with six operators of Distributed Systems Operators in Norway's power grid highlighted the relatively low level of preparedness for large-scale attacks, and the current reliance on the physical security of network infrastructure (Line et al., 2014a). The interviews revealed that although extremely unlikely, gaining access to control systems (e.g., the network's SCADA systems) could enable hackers to cause widespread power outages in minutes. Such a breach would require navigation of layers of security mechanisms and extensive technical knowledge of protocols and software used. The interviews found that: 1) all but one of the six Distributed Systems Operators lacked written procedures for incident response in the control room, 2) three had never ran training exercises involving a cyber security incident, 3) all suffered from a lack of intrusion response policy and practice (Line et al., 2014a). In control rooms specifically, a challenge for operators is the initial identification that an incident is occurring, and determining an appropriate response. It was found that: control room operators have substantial technical knowledge of the systems, but are not well training in cyber security issues, and while emergency preparedness training is common, the use of IT-based scenarios in these exercises is rare. This line of work suggests that competence in identifying possible symptoms of a cyber attack needs to be strengthened (Line et al., 2014a; Line et al., 2014b). The limited preparedness for attacks represents a salient threat, as networks transition from hard-wiring and physical mechanisms, to more dynamic distributed control and remote access associated with smart grids.

During the COVID-19 pandemic the move from Australian networks to require large numbers of employees to work from home (EnergyAustralia, 2020) represents a separate threat, owing to potentially less secure home WIFI and reduced physical home security relative to offices. Given the history of past large-scale attacks on energy infrastructure, coupled with these network vulnerabilities (Line et al., 2014a) and the progressive transition in workplaces to embrace greater degree of working from home, the question of further cyber attacks on energy infrastructure are now arguably a question of "when and how" rather than "if." Several authors provide evidence for how future attacks on smart energy infrastructure may be carried out.

2.1.3. Manipulation of Demand Attacks

The estimated 35 billion internet connected (iot) devices (Yu et al., 2015) online around the world are considered a serious threat to future energy network cyber security (Soltan et al., 2018). This is due to 1) the trend toward the IOT-enabling of high wattage devices such as thermostats, air conditioners, dishwashers, space heaters and pool pumps etc. (Soltan et al., 2018), 2) the carelessness of users with their personal cyber security practices and permission granting (Wash et al., 2016), and 3) the length of ownership of household appliances; the security of smart devices is likely to deteriorate over time, given part of the strength of security on smart phones, routers and laptops comes from replacing them frequently. Yet fridges, solar inverters, heating/air-conditioning systems or dishwashers are typically replaced less than every 10 years (Foulds et al., 2016). Soltan et al. (2018) identifies the manipulation of multiple IoT-enabled high power devices (air conditioning, space heaters etc.) may be capable of seriously disrupting a grid's stability, without the need to hack an energy network company's systems directly.

Energy grids operate on a relatively fine balance of supply and demand and are vulnerable to attacks that drastically affect demand on a short term basis (Otuoze et al., 2018). Failures in a very small proportion of the entire grid can cause frequency disruptions which may result in large scale outages due to voltage collapse. In simulations from the Polish power grid, Soltan et al. (2018) estimated that the simultaneous activation of 90,000 air conditioners or 18,000 hot water heaters (90 MW) would be enough to cause a significant frequency disruption before the grid's primary controllers could react. A frequency disruption of this magnitude would be sufficient to cause large-scale blackouts and potentially trigger a black start. An adversary with a large number of distributed high-wattage iot devices commandeered could then potentially hinder the re-start process, given each small island of a grid that is re-activated, is more sensitive to frequency disruptions than the grid as a whole due to their smaller size (Soltan et al., 2018). This scenario compares to the Mirai Botnet, which commandeered 600,000 devices in only a few months (Antonakakis et al., 2017). The growth of electric vehicle ownership and the remote control over high-power domestic EV chargers (currently up to 22 kW) (Ashok et al., 2016) and discharge behavior of household batteries in the future would require a far smaller number of malicious remote IoT acquisitions to cause a powerful MadIoT attack.

2.1.4. Smart Meters

Meter-based attacks in the past have involved physical access to individual devices, e.g., meters using optical converter devices or magnets (Krebs, 2012), however tamper-detection systems and tamper-resistant hardware in newer smart meters now limit these kinds of attacks (Hunn, 2018). However, despite their relative security in operation, smart meters may be vulnerable to coordinated attacks on software or firmware through an adversary in the production process (Skopik et al., 2013). Manufacturing security has been flagged as a potential issue, where a malicious programmer (or programmers) working for a smart meter manufacturer may be able to insert malicious code into the smart meter bootloader or Read-Only-Memory or in firmware updates (Skopik et al., 2013; Hunn, 2018). In large smart meter roll-outs, malicious code could be copied to the memory of millions of smart meters, eg programming all to disconnect from the grid at a specific time to affect grid stability Soltan et al. (2018).

A range of standard measures exist to ensure the privacy of individuals' data within a smart grid system (we refer the reader to Aloul et al., 2013; Otuoze et al., 2018 for in-depth reviews). Yet privacy concerns remain a limiting factor constraining the proliferation of smart metering. In 2018 a court in Naperville Illinois ruled that the inference available through 15-min frequency smart-metered data constitutes a "search" under the US Fourth Amendment ("The right of the people to be secure in their persons, houses ... against unreasonable searches and seizures,") if residents could not opt-out of the installation of a smart meter (United States Court of Appeal, 2018).

Sub-second household-level energy use information collected by the growing range behind-the-meter energy monitoring hardware is arguably a greater threat to privacy than 15 min frequency smart meter data. Sub-minute energy use data is increasingly valuable to power networks, given the benefits of improved network management (EnergySafe Victoria, 2016) and private companies for customer profiling, e.g., insurance companies wishing to assess risks (Zhang et al., 2019), and targeted marketing. Companies such as Bidgely.com already collect and offer energy utilities insights and monetisation opportunities from customers' data. Thus a tension exists between the current availability of low-granularity data from smart metering, and the desire from a number of stakeholders for much higher-granularity energy use information.

This paper builds upon the important existing work concerned with illustrating the technical fundamentals of smart grid attacks (Mrabet et al., 2018), more generalised overviews of smart grid vulnerability (Aloul et al., 2013) and work detailing energy sector preparedness for cyber attacks (Line et al., 2014a) with a specifically human-centred approach. To assist energy sector stakeholders better anticipate potential cyber vulnerabilities, we use multi-disciplinary expert panel to rank potential threats and from this, produce a behavioral framework for better understanding and anticipating adversarial behavior.

3. DESIGN-THINKING AND BEHAVIOUR THEORY

3.1. Design Thinking

Design thinking is an iterative process to facilitate cognition, strategic and practical thinking about design concepts such as new systems being developed. It is often used in user-interface design, but is increasingly becoming a mainstream component of business operation with which to tackle poorly-defined "wicked" problems. Design thinking places humans at the center of enquiry (Ney et al., 2019). "...It is a collaborative methodology that involves iterative prototyping. It involves a series of divergent and convergent phases" (p. 14) combining creative and analytical thinking approaches (Curedale, 2013). Plattner et al. (2012) break

design thinking into five main stages: 1) **empathise**, i.e., understand the users; 2) **define**, the scope of the problem and needs; 3) **ideate**, i.e., generate ideas; 4) **prototype**, through experimentation and; 5) **test**, in laboratory and production environments where appropriate.

The likelihood of attacks to future grids is high, given how dependent we are expected to be on them. We envisage that attacker motivations will vary greatly from petty theft of energy, organised energy theft, through to denial of (energy) services by individuals, organisations and terror groups. The complexity of this kind of system with multiple points of entry and multiple motivations is an example of a wicked problem which is well suited to design thinking (Buchanan, 1992). We re-purpose design thinking to smart grid architecture design, and also to determine how attackers might exploit these smart grid infrastructures.

By employing design-thinking (Plattner et al., 2012) to reason about cyber threats and attacks upon future smart grids, we are able to provide a framework for reasoning about their potential attack surface. Smart grids are a clear use case for design thinking, because they are not fully implemented today. It is therefore necessary to make use of a structured methodology to reason about future attacks on the smart grid in order to plan for and combat them. By making use of design thinking in a smart grid setting we can reason about the types of physical harms possible on real-world infrastructure (Applegate, 2013), by creating speculative narratives facilitating exploration of attack surfaces and evaluation of exposure.

3.2. Behavioral Theory

Cyber attacks literature typically focuses on the technical aspects of attacks, rather than the human motivations, goals and characteristics that inform their decisions (Sliva et al., 2017). Behavioral profiling of successful prosecutions from "computer criminals" in the early 2000s found support for categories of adversaries including Spies, Saboteurs, Thieves and Abusers (Nykodym et al., 2005). Yet, such profiles have not gained widespread traction in academic literature and rely on post-hoc information on perpetrators, which is difficult in an environment where less than 1% of successful hacks lead to prosecutions.[b]

Van Ruitenbeek et al. (2010) characterise the behavior of cyber adversaries, according to means, motive, and opportunity as key tenets of a successful cyber attack. In this model, opportunity represents the pre-condition of a minimum level of systems knowledge and technical skill to enable an attack to be attempted. Motivation represents the probability of an attack attempt based on an adversaries' perception of the attractiveness of a specific attack vs. the attractiveness of other attacks, and the risk of getting caught. Means represents the probability of success based on the skill of the adversary and means of accessing a given system.

Fogg's behavior theory (Fogg, 2009) asserts that any specific behavior in question is a product of three factors: motivation (M), ability (A), and triggers (T) (each of which has standalone sub-components), $B = MAT$. The model assumes these factors must happen at the same time for the behavior to happen. Motivator sub-components include: positive and negative expectations, such as: pleasure/pain, hope/fear/, acceptance/rejection. Ability sub-components include: time, money, effort (physical/mental), social deviance and non-routine. Trigger sub-components include: spark, facilitator and signal. The model was created for analysis and design of persuasive technologies. Since it is a generic psychology model to express behavior, we assume it can be used to describe both adversaries and stakeholders.

The similarities between Van Ruitenbeek's adversary behavior model and Fogg's user behavior model serves to highlight: 1) cyber attacks are an extension of human behavior, and 2) the similarity of models of adversary behavior with behavioral theories validates the application of customer behavioral theory to cyber security challenges, despite that (to the best of our knowledge) this has not been attempted to date. We postulate that behavior models, such as Fogg, can be used to describe and reason about attacker behavior for poorly understood attack surfaces such as smart grids.

4. METHODOLOGY

Our methodology follows design thinking from the outset, following the steps of 1) empathise, 2) define, 3) ideate, 4) prototype, and 5) test, employing divergent and convergent thinking (Curedale, 2013).

- **Empathise**: After a substantial literature review of threats and previous attacks on energy systems, the research team met online to *empathise* with potential adversaries wishing to target the energy sector, brainstorming motivations for attacks, potential targets and points of entry for near future electricity grids.
- **Define**: The research team scoped the exercise to cover security and privacy threats to *near future* energy systems, given the sufficiency of existing literature describing threats to current energy systems (Skopik et al., 2013; Line et al., 2014a; Goel and Hong, 2015; Case, 2016; Hunn, 2018). We defined near future energy systems as within the next 10–15 years, assuming an increased proliferation of IoT consumer devices and the continued trajectories of both smart meter roll-outs and renewables integration in the grid.
- **Ideate**: The research team ran two collective brainstorming sessions with the aim of generating as many threat scenarios as possible. This process was continued to the point of saturation where any further scenarios began to duplicate existing scenarios or were considered too far-fetched. Drawing from the authors' diverse expertise of human-centred design and design thinking (Snow), cyber security (Happa) and business management/computer graphics (Glencross), a specific intent was to envisage the widest possible range of threats, including and beyond those already documented in the literature and in the widest variety of smart infrastructure including smart grids, smart meters and smart home appliances. Through this stage

[b]https://www.rpc.co.uk/press-and-media/65-hacking-prosecutions-last-year-up-from-47-percent/.

of the process, no member of the research team could discount a threat, they could only generate threats. This process resulted in a total of 34 unique scenarios produced by the research team, covering a wide range of adversary motivations, attack targets and points of entry. These 34 scenarios were then reduced using a process of convergence, involving further discussions where: a) scenarios that were too speculative, unlikely or of limited consequence discarded and b) similar scenarios were merged, leaving a total of 16 threat scenarios which were presented to the expert panel.

- **Prototype and Test**: The final 16 scenarios were then collated and presented as part of a larger survey which we sent to a panel of three selected specialists (described in detail below) whom we tasked with ranking the feasibility, likelihood and consequence to enable us to rank all 16 in terms of overall threat and use these insights to prioritise future work.

It should be noted that the purpose of this paper is to analyze possible attacks and vulnerabilities, prioritised by the integration of a panel of experts (below). Attempting to cover defense planning for each of the 16 scenarios is beyond the scope of this research and represents a basis for future work based on the results of this present paper.

4.1. Panel of Experts

A panel of experts was created through purposive sampling. Criteria for inclusion was either energy sector professionals with over 10 years' experience in the energy sector with direct experience of security or threat management OR security sector specialists with over 10 years in security, with substantial experience consulting for -or employed within- the energy sector. Contacts were approached from Australian distributed network service providers, generators, distribution companies, security companies and consultancies. The final panel included:

- **P1**: Male, 35–44 years of age, Manager. 20+ years of continuous employment in an Australian energy network service provider. Roles incl. network management systems and threat exposure and risk (man-made and natural).
- **P2**: Male, 55–64 years of age, Director. 20+ years working within and consulting for the Australian energy industry, roles related to major project management, strategic management, including security and risk analysis (man-made and natural).
- **P3**: Male, 35–44 years of age, Academic and cyber security professional. 10+ years in cyber security, including four years' consulting in energy sector cyber security challenges.

4.1.1. Process, Analysis, and Scenarios

The speculative threat scenarios were presented in a survey sent to the three panel members, which additionally requested demographic information, their self-reported biggest threat to smart grid security, before each panel member was asked rate the feasibility, likelihood and consequence of each of the 16 threats ideated through the design thinking process between 0 and 10, 0 = extremely unlikely/inconsequential to 10 = almost certain or already occurred/catastrophic (refer to **Supplementary Material**). Feasibility was added to the traditional risk matrix factors of likelihood and consequence because our intent is to assess risks to near future smart grids, and we wished to differentiate between panel member's considerations of likelihood (implying current likelihood) with future feasibility. Following the risk matrix each panel member was invited to self-identify any other threats which have not been covered in the table.

Demographic and self-reported threat responses were collated to a master spreadsheet and averaged. From these averages, the 16 threats were ranked in terms of overall risk (Overall risk = average feasibility + average likelihood + average consequence). The top three threats identified are discussed below.

At the most abstract level, cyber attacks fall into three broad categories (also known as the CIA triad): 1) Confidentiality (data can be stolen or leaked), 2) Integrity (data can be modified), 3) Availability (a service or access to data, can be denied) (Anderson, 1972; Mrabet et al., 2018). NIST applies the CIA triad to smart grid security planning (U.S. National Institute of Standards and Technology, 2014):

- **Confidentiality**: "Preserve authorised restrictions on information access and disclosure," customer metering ad billing information sent between a customer and further entities must be protected and remain confidential. Confidentiality is lost if information is accessed or disclosed by unauthorised entities or processes.
- **Integrity**: "Guarding against improper information modification or destruction." Information can be modified (e.g., smart meters measurement algorithms altered to enable energy theft).
- **Availability**: "Ensuring timely and reliable access to and use of information." Attacks can deny the availability of data or a service (e.g., Denial of Service attacks).

In the following, we provide the 16 speculative threats precipitated through the empathise, define and ideate stages (refer above) which the expert panel were asked to rate and comment upon, categorised according to Condifentiality (C), Availability (A) and Integrity (I).

5. RESULTS

Table 1 shows the expert panel's opinions on risks per-scenario, while **Figure 1** shows the relationships between feasibility, likelihood and consequence. Any blank cells in the returned surveys were entered as zeros rather than blanks, in order to ensure that the overall highest ranked threats were those with consensus among all the experts, and not any where the Average Total Risk (**Table 1**) was based on less than all three of the experts' opinions.

Our application of design thinking to cyber threat identification, including a) external literature review, b) ideation of further threats, c) refining of potential threats based on feasibility, and d) further refining achieved through external expert validation suggests that many cyber threats to

Risk matrix of potential cyber-physical smart grid attacks

Scenario	C/I/A	Technology/attack	Entry point
(1) Energy retailers sell high granularity energy data to insurance companies for improved insurance risk analyses	C	Privacy breach and/or uninformed consent	N/A
(2) Use of deception to distract command center and technician resources to allow for further cyber crime, or physical attacks such as forcing access to a substation or similar	A	Hack/Spoof	Energy network operator's' computer systems
(3) A security weakness in an IoT smart home device is used to gain full admin access to the household's smart meter or other behind-the-meter measurement technology; and from there, gain access to the computer systems of the energy utility	C	Privacy breach and/or uninformed consent	Compromised smart home technology
(3a) With access obtained to an energy utility's computer systems (see question 3): Large-scale data theft of personal data, account details and energy use details (similar to the Yahoo attacks of 2013, 2014, and 2016)	C	Privacy breach and/or uninformed consent	Compromised smart home technology
(3b) With access obtained to an energy utility's computer systems (see question 3): Hackers use the compromised energy utility systems to gain access to network providers systems to shut off a network sector's electricity, other large-scale attacks	C	Privacy breach and/or uninformed consent	Compromised smart home technology
(4) Automotive electric charging infrastructure is hacked overnight, preventing cars from being charged and causing wide scale disruption to commuters the next morning	A	Hack/Attack	Energy network operator or automotive company's computer systems
(5) Disruptors hack the Bureau of Meteorology's industry API arm, sending erroneous weather data to energy networks, causing them to greatly under-estimate the expected load of solar and wind, resulting in multiple transformers tripping and other safety issues associated with over-voltage on the grid	A	Hack/Spoof	Meteorological bureau's computer systems
(6) Historic energy use data used as evidence in court, e.g., proof someone was at home/on their computer at a given time and date	C	Privacy breach and/or uninformed consent	N/A
(7) A household in 2019 agrees to sharing sub-second energy use information based on knowledge of what can be gleaned from it in 2019. Re-analysis of the data in 2029 with greatly improved disaggregation techniques (to the level of individual LED energy signatures on OLED TV's) finds evidence of household member streaming illegal on the TV 10 years ago and info is sent to police to prosecute	C	Privacy and/or uninformed consent	N/A
(8) Multiple burglaries occur when a security weakness/ vulnerability is identified in a commercial smart home device, allowing criminals to remotely disable security alarms and unlock IoT door locks	A	Attack- household level	Broad range of consumer devices
(9) Unauthorised shutdown of the smart grid. Hackers gain access to network command center	A	Attack	Network control center
(9a) [Answer only if likelihood for Q9 is not rated as 0- "impossible"] having gained access to network command center, hackers shut off power to suburbs. (note this may enable terrorist activity/mass burglary/other crime)	A	Attack	Network control center

(Continued on following page)

(*Continued*)

Risk matrix of potential cyber-physical smart grid attacks

Scenario	C/I/A	Technology/attack	Entry point
(10) Online/remote theft of power (individual(s) taking power without being billed)- enabled by hacked access to energy networks	I/A	Attack&Theft	Phishing and malware, targeting energy network operator
(11) Physical theft of power by using street-based, super-high-voltage EV chargers to charge multiple batteries which are then transported elsewhere, so the theft cannot be traced to a specific house once the authorities realise there has been a theft	I/A	Theft only	On-street EV chargers
(12) Supply chain security: Overseas- manufactured smart meters have malware or nefarious hardware modifications inserted in them during manufacture. This would mean 10,000'|s of compromised smart meters rolled out over a city. Could be used for (1) simultaneous deactivation causing manipulation of demand frequency attack, or (2) theft of data	A	Adversary in manufacture process	Manufacturing company
(13) Manipulation of demand attack: Botnet-style malicious commandeering and simultaneous activation/de-activation of thousands of vulnerable high-power smart home devices (eg iot enabled air conditioners or water heaters) to cause frequency disruption	A	Hack/Attack	Multiple iot devices
(14) Attacks on the hardware/firmware of a smart meter: Computer worm phlash (permanently disable or "brick") smart meters, meaning power cannot be restored until the smart meter is replaced	I/A	Hack/Attack	Hardware
(15) Attacks on the metrology system of a smart meter: Energy theft, or cover-up of hydroponic labs, etc.	I/A	Hack/Attack	Multiple iot devices
(16) Hackers gain full admin access to many household's smart meters	A	Hack/Attack	Consumption information (either smart meter or CT clamp)
(16a) Using NILM on the smart metered data, hackers determine the use of Sleep Apnea machines or other life support equipment	A	Hack/Attack	Smart meter data interception or hack of energy retailer)
(16b) Hackers commit remote murder through turning off power to life support customers remotely through their (hacked) smart meter during the night	A	Hack/Attack	

TABLE 1 | Expert panel opinions on risks per scenario as rated in a likert scale.

#Scenario	Avg. feasibility	Avg. likelihood	Avg. consequence	Avg. total risk
1: Energy data sold to insurance companies	4	5.7	5.3	5
2: Deception to enable cyber crime	8.7	9.7	9.3	**9.2**
3: Full admin access to smart meter	4.3	4	4.3	4.2
3a: Theft of personal data	7	6	5.7	6.2
3b: Shut off a sector's electricity	5.3	6	8.3	6.5
4: Preventing car charging	4	5	5.5	4.8
5: Energy under/over-estimation hack	6.7	7.7	6.7	7
6: Evidence in court	7.7	4.7	8	6.8
7: Improved disaggregation	6	5	5.5	5.5
8: Remotely disable security alarms	7.7	8	7.3	7.7
9: Unauthorised shutdown of smart grid	10	9.3	10	**9.8**
9a: Shut off power to suburbs	10	10	9	**9.7**
10: Online/remote theft of power	4.7	1.7	2.3	2.9
11: Physical theft of power	6	4	3	4.3
12: Supply chain security concerns	7	7.3	8.7	7.7
13: Manipulation of demand attack	7.3	8.7	7.3	**7.8**
14: Smart meter hard/firmware attack	7	8.7	4.3	6.7
15: Smart grid phlashing	7	5.7	2.7	5.1
16: Widespread compromise of smart meters	7	5	5	5.7
16a: Life support attack	4.7	3.7	5.7	4.7
16b: Remote murder	5	3.3	6.3	4.9
Avg	6.5	6.1	6.2	6.3

We see feasibility [ranked between 1 (impossible) and 10 (having already happened)], likelihood [likelihood in next 20 years, ranked between 1 (extremely unlikely) and 10 (almost certain)] and consequence [Ranked between 1 (harmless) and 10 (catastrophic)] are averages. Average total risk is the average three previous column values. Bold values shows the highest quartile.

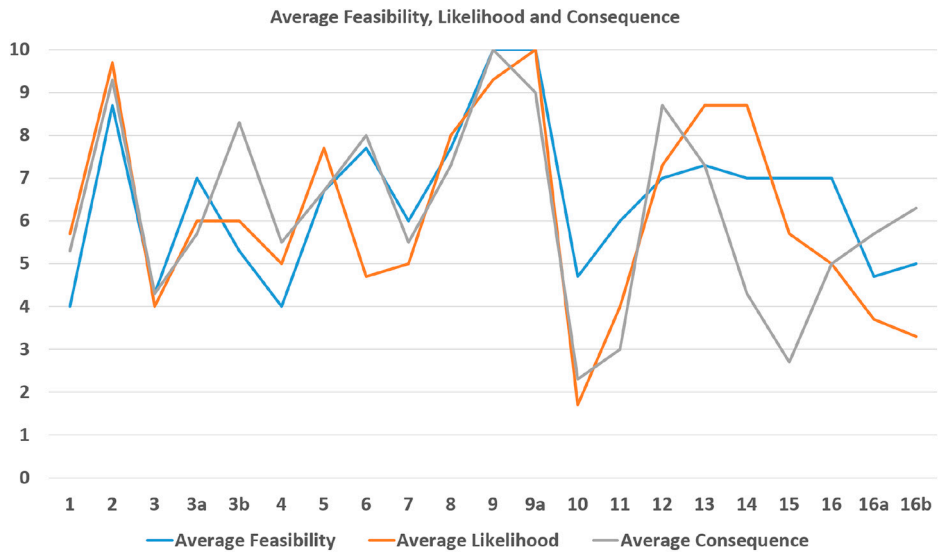

FIGURE 1 | Feasibility, likelihood, and consequence plotted against each other.

future smart grids exist, yet many remain unrealised. The top three threats identified by this paper include:

(1) **Unauthorised shut down of smart grid**. Hackers gain access to an energy network control room, shut off power to suburbs to enable terrorist activity/mass burglary or other crime.

(2) **Unauthorised access to power**. Use of deception to distract command center (control room) and technician resources to allow for further cyber crime or physical attacks, eg forcing access to substations or other infrastructure.

(3) **Manipulation of demand attack**. Botnet style malicious commandeering and simultaneous activation/deactivation of thousands of vulnerable high powered iot appliances.

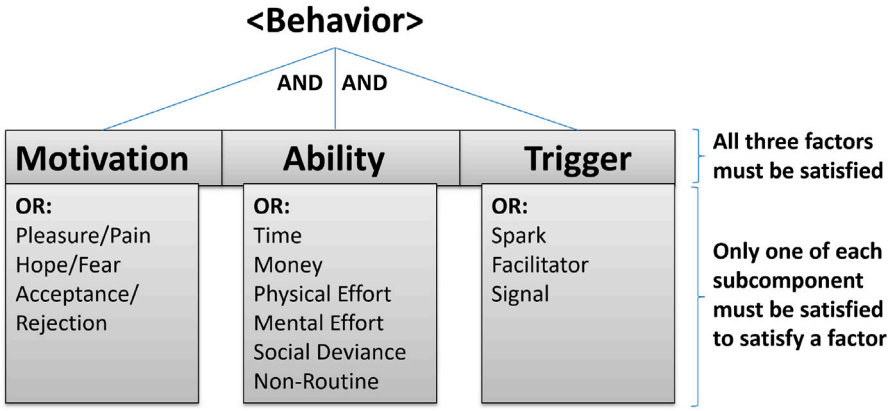

FIGURE 2 | For a behavior to occur during an attack, it is necessary to have the right motivation, ability and trigger in place. Fogg's model as a tree. The top level has AND dependencies, whereas the next level down has OR dependencies. The specific sub-components to include is not fixed, but should be determined by the analyst.

These top three threats were ranked highly in terms of likelihood, consequence and feasibility by each panel member. The threats also map closely to the greatest self-reported threats to smart grid security, which included: "*Targeting of critical infrastructure by an increasing range of threat actors*" (P1), hacking of network SCADA systems (P2) and vulnerabilities in industrial control systems (P3). APTs as a threat to network systems (P3), while P1 noted the increasingly shorter times in these types of attack between discovery of a vulnerability and exploitation. P3 noted that threat #1 and #2 had already been used in combination in the Ukrainian energy sector attack in 2015 (Case, 2016) and that #2 can be employed by both hackers and in penetration testing. P1 noted the consequences of #2 depend on a hacker's intentions, yet the technique implies a high degree of premeditation or planning and could be catastrophic if used for a large scale control room attack as in the Ukraine. P2 noted that electric vehicle charging represents a further potential target with respect to threat #3.

Overall there was very good agreement between panel members' rankings, however some variance in scoring occurred. For example questions concerned with data theft or denial of service following entry gained into an energy utility's computer systems (Q3a, Q3b) received feasibility scores of 3 and 2 from P1 (network service provider representative), but scores of 10 and 10 from P3 (security consultant). The variance may be due to lived experience, e.g., P1 confidence in his own company's cyber security relative to P3's experience consulting in the energy sector. P1 also scored the likelihood of supply chain attacks eg nefarious modification of smart meters in manufacturing (Q12) as 10, relative to 7 and 5 from P2 and P3. Despite this high likelihood, P1 rated the consequence lower than P2 and P3, perhaps indicative of thinking that compromised smart meters do not pose a strong threat to overall Australian energy security.

6. FRAMEWORK

So far we have provided a proof of concept for using design thinking as a means of ideating, refining and externally validating potential threats and prioritising threats according to likelihood and consequence. Yet Van Ruitenbeek et al. (2010) argues that to most accurately predict how, when and why a given attack may occur, it is necessary to consider not only what cyber attack may occur, but more important *why* it may occur; underscoring a need to understand the human at the other end of any attack in terms of their decision making processes. Here, we describe a framework for profiling adversaries in terms of their motivation, ability, means and triggers (Fogg, 2009; Van Ruitenbeek et al., 2010). This framework deals specifically with prioritising cyber threats for smart grids and is intended as a tool for those who are potentially exposed to—or tasked with the defense of—cyber-attacks. Accordingly, we do not attempt to speculate on possible defences to these attacks, but empower those tasked with the defense—and others—in better anticipating, reasoning about and predicting potential attacks.

Our framework takes a behavioural-model approach to evaluate whether a "step" (behavior) in an attack is successful or not, with each step being assigned a probability. We codify this behavior as a tree to predict risk of a particular attack (**Figure 2**; Listing 1).

A specific challenge of defining, preempting and characterising potential cyber threats on future energy systems is simply not having sufficient empirical evidence to make robust assumptions about smart grid attacks (Skopik et al., 2013). Our framework systematically posits realistic smart-grid security narratives, from which we can determine likely future security challenges, threats, and mitigation to limit potential harms. It relies on several components that examine where and how attacks may unfold in the smart grid, and uses design-thinking and behavioral theory to reason about attacker behavior in smart grids.

The Fogg (2009) behavioral model states that a target behavior is enacted when users have—simultaneously—the motivation and ability to perform the behavior and a trigger for that behavior. Similarly, Van Ruitenbeek models cyber adversary behavior according to an adversary's motive, means and

Algorithm 1: EvalbehaviourStepSuccess();

```
input: attackerProfile p, behaviourModelTree t;
output: bool behaviourStepSuccess;
foreach factor in t do
    foreach (subcomponent in factor) do
        If (subcomponent in p == 1) then
            factor = true;
        else
            factor = false;
        end
    end
end
if (all(factor in t) == true)) then
    behaviourStepSuccess = true;
else
    behaviourStepSuccess = false;
end
```

opportunity for carrying out a given attack (Van Ruitenbeek et al., 2010). There is a high degree of similarity and equivalence of meaning in Van Ruitenbeek's and Fogg's models. We choose Fogg because "Opportunity" (i.e., financial and computational capacity) in Van Ruitenbeek can be assumed as a given for the purpose of this framework, in that any successful cyber attack relies on an appropriately financed and equipped adversary. "Trigger" relates to a prompt that triggers an attack once other preconditions are met, which may be personal, social or related to a specific event, which we class as a more salient consideration.

We consider the design-thinking process as an iterative feedback loop, with behaviors feeding into each of the individual steps. At each stage in the design thinking process we consider how Fogg's model can influence each of the steps. We make use of this feedback loop to propose how users will make use of the smart grid infrastructure, as well as how attackers might exploit it. First, we emphasise with attackers to identify what motivates them to use or attack smart grids, what abilities they have to use or exploit the smart grid infrastructure, and what triggers their behavior.

As previously mentioned, NIST identified a smart grid to be composed of seven logical domains: bulk generation, transmission, distribution, customer, markets, service provider, operations (FitzPatrick and Wollman, 2010). Distilling this to topological networks, we assume that behavior can affect attack surfaces belonging to one or more of the following smart-grid layers:

- **Generation layer** including power utilities and power providers responsible for creating electricity to be used by all layers below.
- **Transmission layer** including transmission units and substations.
- **Distribution layer** including distributors and aggregators of electricity.
- **Consumption layer** including users, sensors and assets that ultimately make use of the power generated, transmitted and distributed across the smart grid.

Between each layer, we have communication channels in addition to power-line channels. The attack surface spans all of these dependencies. It is possible to use dependency modeling to compute propagation of risks across the entire smart grid infrastructure. However, for the purpose of the framework (instead of computing various dependencies) we assert which assets are affected as part of the modeling exercise to predict risk propagation for poorly understood infrastructures (such as smart grids). We do this to reduce complexity of ideation (i.e., prevent state explosion), and to ensure the models can be reasoned with and discussed by (human) domain experts first. As our framework is intended to promote design thinking where little empirical evidence exist today, we are exploring the real of possibility, rather than examining the low-level details of risk propagation. Instead, we limit our approach to a risk registry that contains an enumerated list of assets and risks posed to them. An adversary's behavior can greatly differ depending on the attack surface as well their means, motivations and ability. Here, we provide assumptions about the adversaries as well as the attack surface using the information from our expert panel.

As shown in Listing 1, we assume that each adversary has different *motivations*, *abilities* and *triggers*. We also assume it necessary to create several attacker profiles populated with different sub-components probabilities (or evaluation functions) of being successful at a time of evaluation (when the behavior/attack step happens). Each factor is binary. If all factors equal 1, the behavior (attack step) is successful. We can model the sub-component as a function that evaluates (probabilistically or from observation) whether a sub-component is satisfied. In other words, If all three factors have at least one sub-component that equals 1, we get: MAT = 1·1·1 = 1. If one or more of the factors equal 0, the behavior is unsuccessful.

We argue that this approach can be automated and help generate narratives that analysts and risk-owners can use to improve their own smart grid system security. This can be achieved by providing a programmatic set of rules that can be used explore attacker behavior. Probabilities can be assigned based on heuristics, empirical evidence or be experimentally added as part of ideation.

6.1. Behaviour-Theory Attack Matrix

Assuming Fogg's behavior model, we can codify behavior into a Behaviour-theory Attack Matrix (BAM) as shown in **Table 2**. The table is comprised of factors and sub-components. We assign probabilities to each sub-component, and assume that for an attack behavior to be successful, the adversary needs to have at least one motivation sub-component, at least one ability component, and at least one trigger component successful at the time of evaluation. The generic table below can be regarded as a list of lists of requirements to satisfy any attack behavior. The behavior is a descriptor of what the attack step entails, all other fields can be functions or associative arrays containing either: observations or probabilities. We can then use this insight to compute the exposure of risk to assets in a smart grid. Using Fogg's behavior model we can build attack exposure graphs from design-thinking narratives.

Table 2 shows how the framework contains a detailed description of attacker inputs necessary to execute any behavior. For the

TABLE 2 | Following Fogg's model, we can enumerate all parts of attacker behavior and develop reasoning tables that we can use to generate attack exposure graphs.

Behavior	Motivation	Ability	Trigger	Impact valuation
P(behavior step 1)	P(Pleasure/Pain) P(Hope/Fear) P(Acceptance/Rejection)	P(Time) P(Money) P(Physical effort) P(Mental effort) P(Social deviance) P(Non-routine)	P(Spark) P(Facilitator) P(Signal)	High/Med/low
P(Behavior step 2)	⟨···⟩	⟨···⟩	⟨···⟩	⟨···⟩
P(Behavior step *n*)	⟨···⟩	⟨···⟩	⟨···⟩	⟨···⟩

The values can be estimated to be a simple probability value between 0 and one or have probabilities assigned per sub-component [e.g., P(Time)]. Steps can be stacked on top of each other as shown with Step 2 and Step n.

adversary, provided we store assumptions about their motivations, abilities and triggers, we can compute whether each step (behavior) in an attack graph is successful. A table like **Table 2** can also be used to store an attacker profile. This means that to check whether an attack step (behavior) is successful in their attack step, we simply compare two tables to check whether the attacker meets all the requirements as outlined by the BAM. Provided we generate BAMs that also store assets and an estimate of risks, we can also generate graphs that communicate attack exposure to stakeholders. By enumerating which assets can be impacted by this step, we can compute a graph of exposure of assets at risk: a visual representation of the adversary model that risk-owners can use. First, we present a generic example, then we detail the top threat identified in the smart grid (unauthorised shutdown of the smart grid).

6.2. Attack Exposure Graphs

We can express attacks as a series of behavior steps in an attack graph or tree. The tree represent the number of steps an attacker needs to execute from the leaf node to reach the root note (the overall goal of the attack). Our attack patterns (used here) take a form similar to Agrafiotis et al. (2015)'s work: attack patterns are expressed as sequential steps, with many routes being possible to reach the end goal. Unlike their work, for any step to be successful, we require an attacker to be successful in all three factors of behavior: motivation, ability and trigger. The likelihoods are determined by the analyst creating this diagram, and should be driven by empirical evidence where available, but design-thinking narratives if no evidence to estimate probabilities are available.

Our example of a generic tree in **Figure 3** consists of seven behavior steps, in these steps there are four distinct routes. In all four routes, the attacker only needs to be successful in two steps. Each step consists of a table to describe the motivation, ability and trigger probabilities necessary. The coloring shows the risk value involved. Encoded in the BAM is one or more asset IDs and the overall risk posed to the asset. The final color of the node in the graph is then colour-coded by low (yellow), medium (orange) and high (red) risk. We suggest 33% to be the threshold, but any consistent threshold can be used by analysts and risk-owners as they deem appropriate.

6.3. Use Case: Unauthorised Shutdown of the Smart Grid

Unauthorised shutdown of the smart grid was deemed the most feasible, likely and having the most consequence overall

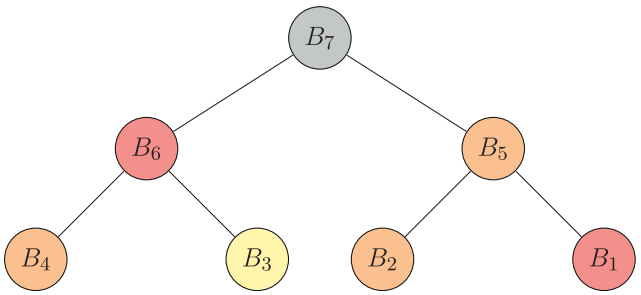

FIGURE 3 | An example of a tree consisting of seven behavior steps. Nodes signify an evaluation of a behavior step, while edges only signify connectivity between each attack step. This means that if a node is not connected with another node, they are unrelated attack steps. Each step consists of a table to describe the motivation, ability and triggers necessary. The coloring shows the risk value involved. This particular example shows our graph as a tree (a tree is considered as a minimally connected graph).

of all the 16 speculative scenarios put before the panel of experts (Section 5). Here, we will first review the scope of possible types of unauthorised shutdowns before showing our framework with one of them.

Going back to the four aforementioned smart grid layers, we make note that there are four key attack surfaces that the attacker may wish to target to shutdown the smart grid: generation, transmission, distribution, and consumption. Unauthorised shutdown would then involve rendering any of these systems (and others not detailed here) unavailable. In our use case, we consider this to be the prevention of access to energy from being: generated/transmitted/distributed or consumed. These may relate to attacks on:

(1) **Power plants**. The types of attacks may include, but are not limited to:
- Air-gapped access to sensitive systems, which at the time of attack forces the system to shut down (eg via an insider or contractor, either maliciously or accidentally).
- Sabotage system access using legitimate credentials, i.e., an insider with legitimate access shuts off access, either in the form of blackmail or reward for doing so.
- Physical sabotage (e.g., physically damage solar panels, wind turbines or dams) would require physical access to the devices in question, either as a guest or undetected.

- Identifying and exploiting vulnerabilities from the outside going in.

(2) **Power lines and pipes**.
- Physical harm to the infrastructure (such as substations).
- Compromise infrastructure such as substations and ability to switch off critical SCADA systems and programmable logic controllers (PLCs) by remotely accessing them or misusing legitimate credentials.
- Physical redirection of power, i.e., tapping into the physical infrastructure and send the power elsewhere. This may be perceived as a shutdown from other parts of the network.

(3) **Distribution platforms**
- Tampering of assets that direct the energy to users.
- Manipulation of the market to cause demand of access to power greater than its capacity.

(4) **Smart homes** (including IoT, local area networks (LAN) and home solar panels).
- Manipulation of demand attack through multiple commandeered high power smart home (IoT) devices.
- Compromising mobile phone app controlling the smart home, and restricting access to power. Existing smart meters do not make use of the home WiFi in order to ensure the integrity of the information sent to the supplier, but instead set up their own mobile network not visible to the consumer (decoupling the attack surface from non-accredited engineers). In our design-thinking narrative, we envisage attackers being able to invoke power requests to the point it would shut off the smart home power access as a safety measure.
- Tampering with smart meter infrastructure manually (e.g., pretending to be an engineer and having malicious intent would allow attackers access to smart home power supplies).

In line with design thinking which emphasises working within- rather than against complexity-we do not create behavior steps that can be considered "atomic actions." Instead, a degree of abstraction is necessary to limit the number of steps from increasing to an unmanageable amount. As a rule of thumb, we regard each behavior as an event executed by one person that has the potential to lead to another trigger of the next behavior. Similarly as per design thinking, the probabilities we assign aim to support ideation, rather than being values that should be employed in any production environment. Any probabilities assigned, ought to be derived from empirical insights where possible. However, where none exists, we advocate the application of heuristics, common sense and a divergence convergence process as we have applied above. Fogg's model simply pushes the narrative forward through behavior.

Table 3 and **Figure 4** shows the values filled in for the highest ranked threat: unauthorised shutdown of the smart grid. We have selected the shutdown of a substation sub-scenario which we elaborate further. For the purpose of readability, we compressed all probabilities to a single number. We regard the last behavior (gray) to be achieving the goal we set out in the attack. This is where we consider overall motivation, ability and trigger of the attack. Here, we may consider sub-components of motivation of the attack as a whole to be P(blackmail), P(reward), P(revenge), P(accident), P(curiosity), P(political) or P(competition). Note that in most cases the trigger probability is low. Each time we evaluate the behavior function, we may get $B = 1$, but the probability of the trigger happening makes the evaluation sufficiently unlikely to yield in one each time the function is called. We score the impact valuation risk to be high when the average $B \geq 0.7$, medium when the average $B \geq 0.4$ and $B < 0.7$, and low if the average $B < 0.4$.

Using this table, we can infer a number of routes the attacker might take to conduct their attack to shutdown a component of the

TABLE 3 | Behaviour-theory Attack Matrix for unauthorised shutdown of a substation on the grid.

Attacker profile				
Behavior: Unauthorised shutdown	**Motivation**	**Ability**	**Trigger**	**Impact valuation**
Reconnaissance				
B_{01}: Study attack surface	0.5	0.3	0.1	Low
B_{02}: Enumerate assets	0.7	0.2	0.1	Low
B_{03}: Identify vulnerabilities	0.7	0.7	0.2	Medium
Weaponisation				
B_{04}: Download malware	1	1	0.1	High
B_{05}: Buy malware	0.5	1	0.2	Medium
B_{06}: Craft malware	0.9	0.9	0.3	High
B_{07}: Take physical object intended for shutdown	0.1	0.9	0.05	Low
Exploit				
B_{08}: WiFi access point	0.7	0.6	0.1	Low
B_{09}: Workstation/Server	0.9	0.9	0.4	High
B_{10}: Firewall	0.9	0.9	0.2	Medium
B_{11}: Sensor (heat, pressure)	0.6	0.6	0.1	Medium
B_{12}: PLC/SCADA	0.7	0.6	0.1	Medium
B_{13}: Lateral movement/maintain foothold	0.9	0.9	0.2	Medium
	Unauthorised shutdown			
B_{01}: Malware shutdown of substation	?	?	?	?
B_{01}: Physical shutdown (violent act or using legitimate credentials to switch off PLCs)	?	?	?	?

After having reviewed a selection of probabilities, the total probability of motivation, ability and trigger are added, assuming an external attacker.

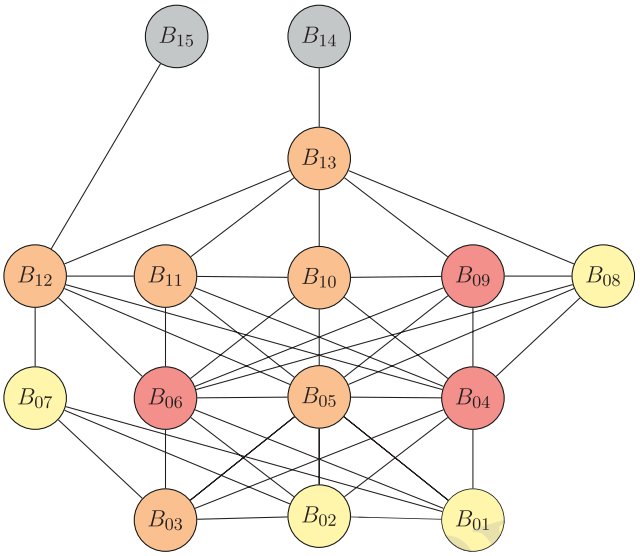

FIGURE 4 | An Attack Exposure Graph of unauthorised shutdown of the smart grid. Edges show the path of behavior that an attacker might take. Red nodes indicate high risk impact valuation in which we expect attackers to buy or craft phishing emails to gain remote access. A scenario with less risk is the physical harm approach in which attackers have less motivation because there is a higher chance of getting caught.

grid (in our example a substation). An external attacker may, study the attack surface of their environment (for instance, using tools such as Shodan (https://www.shodan.io/) to identify any connected devices at the substation, or assets that remotely connect to the substation); *then* enumerate assets they have access to the environment; *then* take a physical object intended for shutdown (e.g., a object for violent measures such as a crowbar to hit sensitive equipment, OR stealing access cards necessary to gain access); *then* execute an exploit via a phishing attempt and remotely access and seize control of a sensitive PLC, SCADA asset to shut it down (akin to the aforementioned 2015 Ukraine attack), see **Figure 4**. Alternatively, physically access a substation and access power cables/cabinets and shut down the grid using stolen credentials (which we deem to have lower chance of happening as the attacker has a higher chance of getting caught).

Narratively (and more specifically), risk owners can follow individual routes and consider them in isolation, or altogether as one set of possible attack steps. For instance, we see the following connections:

- $B_{01}\leftrightarrow B_{02}/B_{04}/B_{05}/B_{06}/B_{07}$ states that from the first node (studying the attack surface), we can reach any of the following steps:
  - B_{02}: enumerate the assets to potentially attack (note in our example, we do assume that enumerating assets to some degree is a requirement prior to identifying the vulnerabilities of the assets).
  - B_{04}: malware is downloaded from an online resource.
  - B_{05}: malware is purchased from a malicious vendor.
  - B_{06}: malware is crafted by the attacker.
  - B_{07}: a physical object is taken with the intention for shutdown (e.g., this could be an object for violent purposes, or an access card).

This evaluation process is recursively reviewed, and while a directed graph is certainly possible, we do not specify the direction of the edges, and instead leave this to the risk owner. This means that a behavior step can be executed several times in the same attack. We start at the leaf nodes level, and evaluate our way upwards to the root node.

In this use case, we make note of the following observations:

- **Behavior that is fundamentally different in nature, allows for more distinct paths (to explore) in our scenarios**. In the attacker's path to shutdown the substation, they may wish to do all three reconnaisance activities before looking at weaponisation. However, it is not a requirement that they do so. Instead, the *two key distinct paths* we have outlined, illustrate that there is a malware-based approach, and a physical violence approach (destroying the PLC and achieving the goal, B_{14} directly, instead of relying on B_{13}). While many edges exist each of these are used to show various path of behavior possible to take—ensuring path space is fully explored, given any assumptions stated about the attacker and the system they intend to attack.
- **Multiple end-nodes are possible**. In both B_{15} and B_{14} the same goal has been achieved (unauthorised shutdown), but the behavior to reach them are nuanced enough to be distinct. B_{14} demonstrates a malware-based approach (e.g., using phishing to compromise an asset in a network command center and then gain remote access to a substation), whereas B_{15} highlights that a physical activity has taken place (eg violence or simply switching off the system using stolen legitimate credentials).
- **Assigning probabilities is an iterative refinement process**. We see that if an attacker wish to shutdown the substation, there is a high degree of confidence that provided enough motivation and a sufficient trigger, they would succeed. Programmatically, we can see that creating a variety of different attacker profiles, we can identify mitigation tactics across a wide range of threats through straightforward simulation (with probability distributions that can be refined over time). We make note that the degree to which nuance matters should be up to those responsible for the design-thinking narratives.
- **Our approach can, and should be extendable**. enabling the encoding of other pieces of information relevant for the scenario. Other columns can be appended to the table if deemed appropriate such as additional notes or affected assets (should we wish to examine the particular infrastructure/network topology in question).
- **This type of attack graphs describes neither attack surface topography or attacker capability**. Instead, in our current iteration, when we state behaviors, as conducted by the attacker. For instance, with "Exploit Firewall" (e.g., B_{10}), we simply mean a firewall with suitable connectivity to other

devices (hence the addition of a lateral movement behavior, which effectively means to recursively conduct any of the previous behaviors again in a catch-all statement), rather than a particular firewall.

- **Attacker behavior is modeled in generic manner to allow for peer-discussions**. To effectively communicate our use case, we keep our tables and graphs generic. This is by design to allow multiple stakeholders to understand the actions taken by the attacker rather than providing specificity which may alienate certain audiences.
- **Verisimilitude (or lack thereof) to provoke discussions**. Our representation is a significantly simplified view of reality. We make note that real-world systems that are likely more robust and resilient with fault-tolerance and redundancy mechanisms in place (which here will be specific to the instance of a substation). Fault tolerance and redundancy ensure external sources of power provide energy during a shutdown (e.g., an uninterrupted power supply in hospitals, or sophisticated system may query other parts of the smart grid for access to power). As such, to ensure successful unauthorised shutdown of substation, we believe behavior would be required to consider load balancing of energy in light of attacks. For simplicity (and readability), we only show the substation aspect to communicate our idea. We envisage many tables and graphs being created, either manually or programmatically to ideate possible attacks.

7. DISCUSSION

Design thinking can be usefully leveraged as a means of researching, ideating, refining, reasoning and prototyping (in our case used for external validation of) potential future smart grid security and privacy threats. Here we have focused on broad-scale threats to smart grid security which are necessarily speculative in nature given the rapidly developing technology in this space. Yet the same framework is scalable to the organisation level, where individual energy sector organisations could use this framework with a cross-section of employees to identify their own individual security vulnerabilities.

Using a design thinking framework we have prioritised three salient threats to smart grid security in the near future, namely 1) malicious entry to a network operator's control room allowing remote shutdown of grid infrastructure (as per Case, (2016)), 2) distract and decoy tactics as a means of diverting resources away from the site of an attack, and 3) manipulation of demand attacks using widespread commandeering of household IoT technology (as per Soltan et al. (2018)). We have also presented a framework for profiling an adversarial actor in terms of their motivation, ability and trigger for a specific attack according to Fogg's behavioral model to inform attack graphs. The intention is for these methods to be used in conjunction, first 1) to use the design thinking framework to ideate and rank (through external validation) potential cyber security vulnerabilities, before 2) applying Fogg's model to better understand, empathise and hence predict adversarial behavior. This can be used to predict risks posed to the organisations, enabling risk-owners to harden their attack surfaces prior to attacks actuating.

Encouraging design thinking and systemic change can make security analysts and risk owners more agile in their approach to security. Design thinking deals excellently with managing uncertainty and wicked problems (Curedale, 2013). Smart grids represent a salient deployment for this proof of concept, given the near complete lack of successful existing attacks from which empirical evidence can be leveraged. Yet the two-part framework is designed to be scalable, and can be discussed at different levels of abstraction, providing a foundation from which stakeholders can more meaningfully discuss and respond to cyber attacks (Happa and Fairclough, 2017). We intend that the same framework could be used by a single organisation to ideate, rank and prioritise action toward more defined organisation-specific threats, or to other increasingly digital sectors, e.g., telehealth or online retail.

We believe having a framework to reason about potential future threats, enables risk owners to make well-informed decisions and better prepare against future threats. Further, using such an approach is also likely to net other benefits such as improving resilience of smart grid systems. However, a significant amount of future work it necessary.

7.1. Future Work

7.1.1. Tool Development and Validation

Currently our framework is a work in progress. We expect significant work will be necessary to expand, test and validate our framework. To support analysts and risk-owners, it will be necessary to develop tools to support and automate much of the computation discussed here, allowing our framework to be facilitated by technology. Our work has relied on previous work in psychology literature, and we make a number of assumptions that this approach will also succeed in real-world settings. This remains to be tested and validated by collecting empirical and experimental evidence.

7.1.2. Stakeholder Perspective to Promote Behavior Change

Our approach examines design thinking and behavior theory in adversaries of smart grids. It might be useful to examine the stakeholder's perspective. It is possible to make use of these approaches to investigate stakeholders. This means to look at the defender as opposed to the attacker's side of scenarios. For instance, while we can use it to understand attacker behavior, we can also use it to promote behavior change with the incident response team or users of smart grids. Indeed, this is aligned with Fogg's behavior theory. As it is a generic psychology model to express behavior, we assume it can be used to describe both adversaries and stakeholders, and expect to explore this further in future research.

7.1.3. Smart-Grid Security During Societal Disruptions

In a post-Covid-19 world, we believe such security considerations will also directly influence smart grid security as work paradigms may switch overnight disrupting and having a long-lasting effect on the smart grid attack surface. The salience of these tools and frameworks for ideating threats and preempting and modeling adversary behavior could not be stronger. Not only is the energy sector in Australia and worldwide entering a new era of computation, automation and intelligence, but COVID-19 has

precipitated a rapid progression toward increased working from home, which highlights the potentially increasing vulnerability of energy networks and distribution companies to cyber attack.

Anecdotal reports suggest that during COVID-19, energy market operators have undertaken testing to determine the feasibility of operating entire energy grids from home, via a digital twin, as a precaution against a worst-case COVID-19 scenario (personal communication). The pandemic (as of September 2020) has shifted how organisations in society operate. Many employees now work from home, have been furloughed or have lost their jobs. Attack surfaces have changed and the protection of home networks should become further prioritised: as it is now a place where people both live and work. From a security perspective, this change can have both positive and negative effects. Positive effects include: 1) Distribution of employees (who now work from home) may make direct attacks on organisations more difficult to achieve; 2) Furloughed employees make an organisation's attack surfaces leaner. Negative effects include: 1) Certain sectors may get stretched for resources and may neglect or de-prioritise security concerns; 2) IT infrastructure is not actively monitored by IT security professionals (as people work from home, and home networks are likely less secure as they are not managed by IT professionals). Future work should examine the relationship between non-cyber attack threats disrupting and affecting the smart grid attack surface.

8. CONCLUSION

Energy grids are becoming increasingly digitised at a time when workplaces are becoming more de-centralised: never before has cyber threat analysis been more salient, or more complex in nature. In this paper, we have presented and demonstrated a two-part framework for reasoning about smart grid attack surfaces by leveraging design thinking (Plattner et al., 2009) as a means of ideating, prototyping and cross-validating threats, and then behavior theory (Fogg, 2009) as a cornerstone of anticipating and modeling adversarial behavior. Given the uncertain nature of what smart grids may look like in the future, we assume that viewing the system as a wicked problem facilitates creative reasoning about attacker assumptions, motivations and behavior to help facilitate the development of countermeasures. We provide an overview of key privacy and security risks in the smart grid from a survey of recent literature and identify how smart grid architectures could be exploited by attackers using our model. We demonstrate the validity to our framework by giving examples of several worked-through attack surfaces, based on a set of use-cases validated by an expert panel of practitioners.

DATA AVAILABILITY STATEMENT

The survey used for this study is included in the **Supplementary Material**. The raw data supporting the conclusions of this article can be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SS contributed more than half the writing, background research and conduct of the expert panel. JH contributed the application of the behavioural framework and contributed to writing. NH contributed to writing and proof reading. MS contributed to writing, proof reading and organisation / project management.

FUNDING

AQPTP01216-17RD1 Redback Smart Monitoring Platform (Advance Queensland Platform Technology Program grant). NS-1906 New Staff Research Start-up Funds Dr Glencross The University of Queensland.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenrg.2020.591999/full#supplementary-material

REFERENCES

Agrafiotis, I., Nurse, J. R., Buckley, O., Legg, P., Creese, S., and Goldsmith, M. (2015). Identifying attack patterns for insider threat detection. *Comput. Fraud Secur.* 2015, 9–17. doi:10.1016/S1361-3723(15)30066-X

Alexander Mey, S. H. (2017). *Nearly half of all U.S. electricity customers have smart meters.* Available at: https://www.eia.gov/todayinenergy/detail.php?id=34012 (Accessed July 07, 2020).

Aloul, F., Al-Ali, A. R., Al-Dalky, R., Al-Mardini, M., and El-Hajj, W. (2012). Smart grid security: threats, vulnerabilities and solutions. *Int. J. Smart Grid Clean Energy* 1, 1–6. doi:10.12720/sgce.1.1.1-6

Anderson, J. P. (1972). Computer security technology planning study US Defense technical information centre, Technical Report, 02 Feb 1972.

Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., et al. (2017). "Understanding the Mirai Botnet," in Proceedings of the 26th USENIX security symposium, Vancouver, BC, Canada, August 16–18, 2017.

Applegate, S. D. (2013). "The dawn of kinetic cyber," in International conference on cyber conflict (CYCON), Tallinn, Estonia, June 4–7, 2013 (IEEE), 1–15.

Ashok, V. V., Chandra Mouli, G. R., Van Der Burgt, J., Vera, S. P., Huibers, M., Ramirez Elizondo, L., et al. (2016). "Using dedicated EV charging areas to resolve grid violations caused by renewable energy generation," in 2016 IEEE transportation electrification conference and expo, Dearborn, MI, June 27–29, 2016. ITEC. doi:10.1109/ITEC.2016.7520231

Buchanan, R. (1992). Wicked problems in design thinking. *Des. Issues* 8, 5–21. doi:10.2307/1511637

Case, D. U. (2016). *Analysis of the cyber attack on the Ukrainian power grid.* Washington, DC: Electricity Information Sharing and Analysis Center (E-ISAC). 388.

Curedale, R. (2013). *Design research methods: 150 ways to inform design* (Los Angeles, CA: Design Community College, Inc.).

EnergyAustralia (2020). *Energy industry managing COVID-19 pandemic*. Available at: https://www.energynetworks.com.au/news/media-releases/2020-media-releases/energy-industry-managing-covid-19-pandemic/ (Accessed July 7, 2020).

EnergySafe Victoria (2016). *Safety performance report on Victorian energy networks*. Available at: https://www.esv.vic.gov.au/wp-content/uploads/2016/10/SPR-Electricity-2016.pdf. (Accessed June 13, 2020).

Farhangi, H. (2009). The path of the smart grid. *IEEE Power Energy Mag.* 8, 18–28. doi:10.1109/mpe.2009.934876

FitzPatrick, G. J., and Wollman, D. A. (2010). "NIST interoperability framework and action plans," in IEEE PES General Meeting, Providence, RI, July 25–29, 2010, (PES). doi:10.1109/PES.2010.5589699

Fogg, B. J. (2009). "A behavior model for persuasive design," in Proceedings of the 4th international conference on persuasive technology, Claremont CA, April 2009, 1–7. doi:10.1145/1541948.1541999

Fonseca Braga, M. (2016). "The value of design: an issue of vision, creativity and interpretation," in Proceedings of DRS2016: Design+ Research + Society, Brighton, UK, June 27–30, 2016, 1865–1881. doi:10.1080/14606925.2017.1353011

Foulds, C., Powell, J., and Seyfang, G. (2016). How moving home influences appliance ownership: a Passivhaus case study. *Energy Effic.* 9, 455–472. doi:10.1007/s12053-015-9364-0

Goel, S., and Hong, Y. (2015). "Security challenges in smart grid implementation," in *Smart grid security*, London: Springer London, 1–39. doi:10.1007/978-1-4471-6663-4_1

Happa, J., and Fairclough, G. (2017). "A model to facilitate discussions about cyber attacks," in *Ethics and policies for cyber operations*. New York, NY: Springer, 169–185. doi:10.1007/978-3-319-45300-2_10

Hu, H., Kaleshi, D., Doufexi, A., and Li, L. (2015). "Performance analysis of IEEE 802.11af standard based neighbourhood area network for smart grid applications," in IEEE vehicular technology conference, Glasgow, UK, May 11–14, 2015 (IEEE). doi:10.1109/VTCSpring.2015.7146000

Hunn, N. (2018). *How to hack a smart meter*. Available at: http://www.nickhunn.com/how-to-hack-a-smart-meter-and-kill-the-grid/ (Accessed May 5, 2020).

Kapersky Lab. (2017). *Report: human factors in IT security*. Available at: https://media.kasperskycontenthub.com/wp-content/uploads/Ωsites/100/2017/11/10083900/20170710{_}Report{_}Human-Factor-In-ITSec{_}eng{_}Ωfinal.pdf (Accessed June 15, 2020).

Khurana, H., Hadley, M., Ning Lu, N., and Frincke, D. A. (2010). Smart-grid security issues. *IEEE Secur. Privacy Mag.* 8, 81–85. doi:10.1109/MSP.2010.49

Klahr, R., Amili, S., Shah, J. S., Button, M., and Wang, V. (2016). Cyber Breaches Survey 2016. Ipsos MORI Tech. Rep. doi:10.13140/RG.2.1.4332.6324.

Krebs, D. (2012). *FBI: smart meter hacks likely to spread*. Available at: https://krebsonsecurity.com/2012/04/fbi-smart-meter-hacks-likely-to-spread/ (Accessed May 02, 2020).

Langner, R. (2011). Stuxnet: dissecting a cyberwarfare weapon. *IEEE Secur. Privacy Mag.* 9, 49–51. doi:10.1109/MSP.2011.67

Line, M. B., Tondel, I. A., and Jaatun, M. G. (2014a). "Information security incident management: planning for failure," in Proceedings—8th international conference on IT security incident management and IT Forensics, IMF, Munster, Germany, May 12–14, 2014 (IEEE). doi:10.1109/IMF.2014.10

Line, M. B., Zand, A., Stringhini, G., and Kemmerer, R. (2014b). "Targeted attacks against industrial control systems," in Proceedings of the 2nd workshop on smart energy grid security, Scottsdale, Az, November 2014 (ACM), 13–22. doi:10.1145/2667190.2667192

Lorio, P. (2017). Access denied data breach Litigation, Article III standing, and a proposed statutory solution. *Colum. JL Soc. Probs.* 51, 51–79. doi:10.2139/ssrn.2996533

Mahmood, A., Javaid, N., and Razzaq, S. (2015). A review of wireless communications for smart grid. *Renew. Sustain. Energy Rev.* 41, 248–260. doi:10.1016/j.rser.2014.08.036

Miller, B., and Rowe, D. (2012). "A survey SCADA of and critical infrastructure incidents," in Proceedings of the 1st Annual conference on research in information technology, Calgary, AB, Canada, October 2012 (New York, NY: ACM), 51–56. doi:10.1145/2380790.2380805

Mrabet, Z. E., Kaabouch, N., Ghazi, H. E., and Ghazi, H. E. (2018). Cyber-security in smart grid: survey and challenges. *Comput. Electr. Eng.* 67, 469–482. doi:10.1016/j.compeleceng.2018.01.015

Ney, S., Meinel, C., Meinel, C., Meinel, C., Meinel, C., Meinel, C., et al. (2019). *Putting design thinking to work*. Berlin: Springer International Publishing.

Nykodym, N., Taylor, R., and Vilela, J. (2005). Criminal profiling and insider cyber crime. *Comput. Law Secur. Rev.* 21, 408–414. doi:10.1016/j.clsr.2005.07.001

Otuoze, A. O., Mustafa, M. W., and Larik, R. M. (2018). Smart grids security challenges: classification by sources of threats. *J. Electr. Syst. Inform. Technol.* 5, 468–483. doi:10.1016/j.jesit.2018.01.001

Pitì, A., Verticale, G., Rottondi, C., Capone, A., and Lo Schiavo, L. (2017). The role of smart meters in enabling real-time energy services for households: the Italian case. *Energies* 10, 199. doi:10.3390/en10020199

Plattner, H., Meinel, C., and Leifer, L. (2012). *Design thinking research*. Berlin: Springer.

Plattner, H., Meinel, C., and Weinberg, U. (2009). *Design-thinking research*. Landsberg am Lech, Mi-Fachverlag:Springer.

Santacana, E., Rackliffe, G., Tang, L., and Feng, X. (2010). Getting smart. *IEEE Power Energy Mag.* 8, 41–48. doi:10.1109/mpe.2009.935557

Skopik, F., Ma, Z., Bleier, T., and Grüneis, H. (2012). A survey on threats and vulnerabilities in smart metering infrastructures. *Int. J. Smart Grid Clean Energy* 22, 22–28. doi:10.12720/sgce.1.1.22-28

Sliva, A., Guarino, S., Weyhrauch, P., Galvin, P., Mitchell, D., Campolongo, J., et al. (2017). "Hybrid modeling of cyber adversary behavior," in *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation* (New York, NY: Springer), 133–138.

Soltan, S., Mittal, P., and Poor, H. V. (2018). "BlackIoT: IoT Botnet of high wattage devices can disrupt the power grid slides," in Proceedings of the 27th USENIX security symposium, Baltimore, MD, August 15–17, 2018 (USENIX). doi:10.1109/pesgm.2018.8586142

United States Court of Appeal (2018). Naperville smart meter Awareness vs City of Naperville. Case 16-3766. Document 65. Technical Report, Naperville IL.

U.S. National Institute of Standards and Technology. (2014). Guidelines for smart grid Cybersecurity NISTIR 7628 Revision 1. Technical Report, NIST, Gaithersburg, MD: NIST. doi:10.6028/NIST.IR.7628r1

Van Ruitenbeek, E., Keefe, K., Sanders, W. H., and Muehrcke, C. (2010). "Characterizing the behavior of cyber adversaries: the means, motive, and opportunity of cyberattacks," in 40th Annual IEEE/IFIP international conference on dependable systems and networks supplemental (DSN 2010), Chicago, IL, June 28–July 1, 2010 17–18.

Wash, R., Rader, E., Berman, R., and Wellmer, Z. (2016). "Understanding password choices: how frequently entered passwords are re-used across websites," in Twelfth symposium on usable privacy and security, (SOUPS), Denver, CO, June 22–24, 2016 (USENIX) 175–188.

Yan, Y., Qian, Y., Sharif, H., and Tipper, D. (2012). A survey on smart grid communication infrastructures: motivations, requirements and challenges. *IEEE Commun. Surv. Tutorials* 15, 5–20. doi:10.1109/surv.2012.021312.00034

Yu, T., Sekar, V., Seshan, S., Agarwal, Y., and Xu, C. (2015). "Handling a trillion (unfixable) flaws on a billion devices: rethinking network security for the Internet-of-Things," in Proceedings of the 14th ACM workshop on hot topics in networks—HotNets-XIV, 24 February, 2012 (IEEE), 5–20. doi:10.1145/2834050.2834095

Zhang, Z., He, J., Zhu, L., and Ren, K. (2019). "Non-intrusive load monitoring algorithms for privacy mining in smart grid," in *Advances in cyber security: principles, techniques, and applications*. Editors K.-C. Li, X. Chen, and W. Susilo (Singapore: Springer Singapore), 23–48. doi:10.1007/978-981-13-1483-4_2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.