Open Source Intelligence for Energy Sector Cyberattacks

Anastasis Keliris, Charalambos Konstantinou, Marios Sazos and Michail Maniatakos

Abstract In March 2018, the U.S. DHS and the FBI issued a joint critical alert (TA18-074A) of an ongoing campaign by Russian threat actors targeting U.S. government entities and critical infrastructure sectors. The campaign targets critical infrastructure organizations mainly in the energy sector and uses, among other techniques, Open Source Intelligence (OSINT) to extract information. In an effort to understand the extent and quality of information that can be collected with OSINT, we shadow the threat actors and explore publicly available resources that can generate intelligence pertinent to power systems worldwide. We undertake a case study of a real, large-scale power system, where we leverage OSINT resources to construct the power system model, validate it, and finally process it for identifying its critical locations. Our goal is to demonstrate the feasibility of conducting elaborate studies leveraging public resources, and inform power system stakeholders in assessing the risks of releasing critical information to the public.

1 Introduction

Electric power systems have significantly evolved over the years and grew to become essential in our everyday life. Our expectation of uninterrupted power supply in everyday life is further exemplified by the far-reaching impact of power outages, also known as *blackouts*. Table 1 lists notable power outages of the 21st century. The examples showcase the diversity of possible causes, and are sorted by their impact measured in millions of people affected.

Anastasis Keliris, Charalambos Konstantinou
New York University Tandon School of Engineering, 6 MetroTech Center, Brooklyn, NY 11201, USA, e-mail: anastasis.keliris@nyu.edu, ckonstantinou@nyu.edu

Marios Sazos, Michail Maniatakos
New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates e-mail: marios.sazos@nyu.edu, michail.maniatakos@nyu.edu

Table 1 Notable power outages of the 21st century

Year	Country	People affected	Cause
2012	India	620 million	Misoperation [21]
2015	Pakistan	140 million	Malicious destruction [13]
2014	Bangladesh	100 million	Equipment failure [1]
2009	Brazil & Paraguay	87 million	Adverse weather conditions [3]
2015	Turkey	70 million	Maintenance and oversupply [50]
2003	U.S. & Canada	55 million	Shortcircuit because of trees [19]

Most blackouts observed to date are the result of equipment faults, natural phenomena, animals, or human errors. However, there is increased concern in the international community regarding *cyberattacks* that target power grids [38]. When it comes to cyberattacks against cyberphysical systems and critical infrastructure, Pandora's box was opened in 2010 with Stuxnet, a worm targeting equipment in a nuclear plant in Iran [41]. The first cyberattack targeting power systems is an incident in Ukraine reported in December 2015. The attack targeted computer systems of three energy distribution utilities and is believed to be the work of a nation-state actor [6]. The outage mainly affected the Ivano-Frankivsk region, leaving about 230,000 end consumers without power for hours [42]. A second, smaller scale cyberattack against the Ukrainian power grid hit Kiev a year later, this time targeting the transmission network [16].

A large scale, ongoing Advanced Persistent Threat (APT) campaign by Russian actors targeting U.S. critical infrastructure sectors was jointly reported in March 2018 by the U.S. Department of Homeland Security (DHS) and the Federal Bureau of Investigation (FBI) in US-CERT Technical Alert 18-074A [61]. According to Symantec, who has been closely monitoring the group behind this campaign, the energy sector is the main target of the campaign, and the attack focus is not limited to the U.S. [58]. For gathering information during the reconnaissance phase, the threat actors are believed to employ several techniques, including open source reconnaissance, also known as *Open Source Intelligence* (OSINT).[1]

In this chapter we undertake a study of publicly available resources that are pertinent to power systems across the globe, in an effort to understand the extent and quality of intelligence that can be generated with OSINT, as well as the feasibility of constructing exploitation vectors based on these resources. We showcase the practical applicability of OSINT-based studies by modeling a *real national power grid* using OSINT resources, cross-validating our model using secondary resources, and identifying the model's critical operational points through power security studies. Our contributions can be summarized as follows:

- We provide a corpus of publicly available resources pertaining to power systems. These can be leveraged using OSINT techniques to extract intelligence, identify critical operational points, and construct attack vectors against target systems.

[1] OSINT refers to data and information passively collected and analyzed from publicly available sources. It is not related to open source software.

- We demonstrate the significance of OSINT-based intelligence by undertaking an in depth case study of a real, large-scale power system. We build and successfully validate the model of this system from the ground up and analyze it to identify its critical operation points.

To the best of our knowledge this is the first broad study of publicly available resources regarding power systems. Moreover, this is the first work where a model of a real power system was built from the ground up, leveraging and fusing publicly available information using OSINT techniques. Our motivation for this study is the uncertainty currently observed among the various stakeholders of the power industry, including governments, vendors, and power utilities, regarding the real threat cyberattacks pose to power systems. Our study can assist stakeholders and regulators take informed decisions by raising awareness relating to the dangers of divulging more-than-required information to the public, and showcasing potential implications of public dissemination of sensitive information. Due to the sensitive nature of the extracted information, we anonymize certain critical parts of the study.

The remainder of the chapter is structured as follows: We present our threat model, target analyses, and methodology in Section 2. Publicly available resources for modeling a power system are presented in Section 3. Section 4 provides details on contingency analysis, a technique that can be used to derive the critical locations of a system. Section 5 outlines OSINT resources for constructing attack vectors. We evaluate the practicality of an open source campaign in Section 6 by constructing, validating, and analyzing the model of a real power system. A discussion on the significance of cyberattacks against the power grid and possible mitigation strategies are presented in Section 7, and we conclude the chapter in Section 8.

2 Target analyses and methodology

The threat model we assume considers adversaries with power systems expertise, whose objective is to cause large scale power outages. We assume that adversaries do not necessarily have footholds inside target organizations, nor physical access to the Supervisory Control And Data Acquisition (SCADA) center or power substations. Since we focus on publicly available resources for generating intelligence and formulating attack vectors, adversary groups represented in our threat model are not limited to nation-states or heavily funded private/government organizations.

2.1 Strategic target analysis

One of the first steps of a campaign is to study the target system at a high level of abstraction in order to understand and identify the strategic assets of interest [37]. For campaigns against the energy sector, this step concerns identifying which power systems stage or combination of stages are more suitable for achieving the

required objectives. In general, power systems are comprised of four stages: generation, transmission, distribution, and consumption. The first stage is *generation*, where electricity is produced. It is then transferred near consumption centers in the *transmission* stage, and distributed to end consumers in the *distribution* stage. Finally, electricity is utilized in the *consumption* stage by industrial, commercial, or residential consumers.

For causing a large scale power outage, the consumption stage is not particularly attractive because a very large number of consumers would need to be compromised to achieve the required outcome. Employing a similar rationale, adversarial campaigns would likely not target the distribution network, because the attack would require the compromise of a large number of distribution substations, possibly controlled by several power utility companies. The generation stage, at which electricity is produced could be a promising target. However, power plants are manned 24/7 and typically employ a variety of protection mechanisms, including physical security, rendering attacks against them significantly more difficult. In addition, restoration of lost capacity from generator losses can be very quick, as demonstrated by the 2011 Cyprus explosion which destroyed 60% of the state-island's installed capacity [62]. Although the island's grid is not interconnected to other national grids, demand was met with distributed generation in a short period of time.

Considering attack difficulty and attack impact tradeoffs, the most attractive target is the *transmission network*. It fulfills the requirement of large scale impact, while at the same time reduces the difficulty of launching an attack. Several transmission substations are unmanned and situated in remote, not populated areas. This finding is also supported by the fact that the majority of impactful blackouts were caused by transmission stage failures.

2.2 Tactical target analysis

Following the strategic target analysis, tactical target analysis can identify *specific* targets in a power system (e.g., transmission lines and substations) that could fulfill the requirement of causing a large scale power outage, and how to attack these targets. For successful target selection we identify two prerequisites. First, adversaries need to create a *model* of the target system that enables power studies and can generate intelligence regarding the entire power system. Second, the constructed model must be processed and analyzed towards identifying the specific points of interest whose compromise could lead to a large scale power outage. To that end, well known tools and techniques from power system research, such as *contingency analysis*, can be used.

Once the specific critical points of interest are identified, *attack vectors* for exploitation of these points must be constructed. In general, Circuit Breakers (CBs) and their corresponding relay signals constitute attractive targets when attacking specific locations of a power system. The operation of CBs guarantees normal service of grid equipment due to system separation into protective zones, and the iso-

lation of faulty zones as necessary to change load routing. In addition, the control of CBs via relay signals allows the control circuity to command the various CBs to open and interrupt or re-route the flow of electricity. Testifying to their criticality, 70% of the major disturbances in the U.S. are associated with faulty operation of relay controllers [45]. For constructing attack vectors, OSINT techniques can be leveraged towards forcing CBs to open/close connections in a target system.

2.3 Methodology

In fulfilling the steps against targets in the transmission stage as they are identified in the tactical target analysis, adversaries need to create a representative model of the power system, conduct power studies on the model to derive its critical operational points, and finally construct attack vectors against these specific points. In this work, we focus on OSINT-based intelligence that can be leveraged towards achieving these steps. We shadow threat actors seeking to cause a large scale blackout, investigating the feasibility of utilizing publicly available resources to achieve this objective.

We begin by carrying out extensive research on the sources of information on power systems that are available to the public. We provide a corpus of our findings in Section 3. Such sources can provide fragmented information that can be then combined towards creating models of power systems. Subsequently, we explore the types of power system studies that are necessary for identifying critical operational points of a target system. Contingency analyses are particularly relevant in achieving objectives such as large-scale power outages. An investigation of contingency analyses is provided in Section 4. We then examine OSINT resources that enable or can assist the construction of attack vectors against power systems. We consider a broad spectrum of possible attack-enabling resources, and provide a corpus for OSINT exploitation sources in Section 5.

Finally, in investigating the feasibility of leveraging OSINT intelligence and assessing the quality of analyses enabled by it, we undertake an in depth case-study of a real, large-scale power grid. We rely solely on OSINT intelligence, and fuse information from a variety of sources to first build the entire model, and then validate it using secondary OSINT sources. Using contingency analysis tools on the constructed model, we additionally identify the model's critical points.

3 OSINT resources for modeling a power system

In this section we provide representative, but non-exhaustive publicly available resources, which can generate intelligence and provide sensitive information for power systems and their components. Leveraging OSINT, it is possible to obtain the information required to model a power system, enabling tactical target analyses through power studies on the constructed model.

Power system databases: Several power system databases are publicly available, providing access to data relating to real systems across the globe. For example, the Open Power System Data platform provides data regarding power plants, generation capacities, and loads for several European power systems [12]. Another source of data is the European Network of Transmission System Operators for Electricity (ENTSO-E), which was established in an effort to ensure the optimal functioning of the EU internal energy market. Data include maps of transmission networks, grid interconnection details, real time cross border flows, historical and forecast loads and generation statistics, as well as development plans [5]. For the U.S., the Open Energy Information (OpenEI) is a government collaborative website that provides public access to energy data [11]. On a worldwide scale, the Enipedia semantic database features a plethora of load, generation, topology and line characteristics data for power systems across the globe [25].

Geographic Information Systems: The topology of a power system can be constructed or validated by observing the physical components of the system and their interconnections. Instead of on-the-ground tracking of physical structures, it is possible to generate the network topology of a system using satellite imagery from Geographic Information Systems (GIS). This enables a bird's-eye-view analysis of a system that can be performed remotely. Examples of power system components as they appear in a GIS are presented in Fig. 1. The ability to map a power system with GIS can be likened to the analog loophole problem in digital rights management. Location information of physical power structures cannot be hidden from orbiting satellites or aerial photography aircraft, except in the case of underground cables. However, the vast majority of transmission level power lines are overhead [20]. Finding the topology of a power system with GIS can be possibly automated through machine learning and image processing techniques. In addition, tagging power structures and power networks can be crowd sourced. The Power Networks subproject of OpenStreetMap follows this exact approach [47].

Public reports: Oftentimes, power utilities, Transmission System Operators (TSOs), or government agencies release reports to the public that contain operational details and information regarding their power system. These reports may be in the form of reports required by law, annual financial reports to shareholders, and statements that outline future requirements and how they will be met (e.g., [29, 28, 15]). Somewhat ironically, blackout reports may also contain sensitive information regarding a power system. Blackout reports are released to the public, typically for transparency reasons, and usually contain technical details in an effort to pinpoint and explain the source of the blackout (e.g,. [50]). In addition, reports from initiatives that aim to enhance the resiliency of the power grid and accelerate grid modernization can contain information pertinent to specific power systems. For example, the North American SynchroPhasor Initiative (NASPI) website includes a report listing the geographical location of PMUs in the U.S. [9]. Information may be publicly available by design, as is the case with the Northern Regional Load Despatch Centre (NRLDC) of India. The NRLDC shares real-time data regarding frequency, scheduled and available capacity, and next-day load forecasts for operational reasons [7].

Fig. 1 Images of power system components from GIS. Top left: Transmission substation, Top right: Power plant, Bottom: Power lines.

Press releases: Adversaries can extract information regarding a power system from information released to the press. Such information can be in the form of newsletters, press releases by power system operators, success stories by the vendors who installed components of the system, corporate presentations, etc. Some examples include the media centers of vendors that include references to awarded, completed and ongoing projects [55], reports from turnkey solutions providers where information on voltage levels and transmission lines is listed [54] and vendor success stories that reveal communication protocols and system topology [51].

The resources presented here, as well as additional information that can be collected through other OSINT channels or through other more invasive techniques and non-public sources, can be fused to create a model of a power system. The information can be used to derive the topology of the system, as well as to extract, or estimate operational characteristics for performing power studies. Furthermore, information from different sources can be compared to evaluate the accuracy of the constructed model, a process we undertake in our experimental evaluation section.

4 Power system studies: Contingency Analysis

Modeling a power system enables carrying out power studies on the system. Different power studies may be necessary for different campaign objectives. For campaigns that aim to disrupt the system and cause blackouts, power system security

studies can provide valuable information to adversaries as to which specific targets are necessary and sufficient for destabilizing the entire system. Power system security is defined as the probability of the system's operating point to remain within acceptable ranges given the system constraints, the probabilities of changes (contingencies), and its environment [44].

Contingency analysis is a well known operation in modern Energy Management Systems (EMS), which provides necessary information to the system operator about the static security of the system. In contrast to state estimation, which is considered an online application, contingency constrained analysis is an offline application for power system planning and operation [49]. Abstractly, contingency analysis can be viewed as a "what if" scenario simulator that assesses, produces and ranks the impact of unscheduled events on a power system. For example, a contingency can be the failure, or loss of an element of the system (e.g., generator, transmission line, transformer), or the unplanned opening of a CB. These events form the contingency list, which is then used by contingency analysis algorithms to evaluate effects on the overall system.

In its basic form, contingency analysis generates a power flow solution for each event specified in the contingency list. The objective of the power flow analysis is to obtain a set of voltage magnitudes and angles for each bus in the power system corresponding to a specified load and generation condition. Subsequently, active and reactive power flows on each branch and generator are analytically determined. The loss or failure of each contingency event is simulated in the network model by removing that part from the simulated power system. The resulting network model is solved to compute the corresponding power flows, currents, and voltages for the remaining elements. The outcomes from each contingency test are then compared with the operational limits for every element (e.g., thermal ratings of transmission lines) to determine if a limit violation occurs.

Since contingency analysis relies on the execution of a power flow study, the first step is acquiring the required data to develop a power flow model of a power system. Specifically, for a power system model to be sufficient for contingency analyses, the following data are required [32]:

- System topology (Edges: Transmission lines/Transformers, Nodes: Buses).
- Transmission line parameters.
- Tie-line locations and ratings.
- Transformer and phase shifter parameters.
- Location, ratings, and limits of generators.
- Load location and load compensation.

In general, power systems must be able to sustain a single contingency condition ($N-1$) to enable maintenance operations, where N is the number of components (typically the N branches of a network). North America Electric Reliability Corporation (NERC) and other regulatory agencies around the world enforce strict power security standards that require power system operators to satisfy the $N-1$ security constraint [46]. NERC standards also necessitate that operators ensure sufficient system performance in the event of multiple outage contingencies. Nevertheless,

the problem of contingency identification remains computationally challenging due to the total number of possible initiating events: it increases exponentially with k, where k is the number of outaged elements. The complexity is further exaggerated if outage scenarios are analyzed with a full AC power flow technique, which requires significant computational resources.

Instead of using full non-linear AC power flow analysis, approximate, but much faster techniques based on DC approximation can be used to estimate post-contingency values of interest [35]. In general, DC power flow analyses are commonly used in contingency studies where approximate real power flows are more important than voltage limits on buses [60]. The DC formulation is based on the same parameters as the AC problem, with additional simplifying assumptions: the voltage profile is flat, meaning that all bus voltage magnitudes are close to 1 p.u., line resistances and charging capacitances are considered negligible, and voltage angle differences between neighboring nodes are small enough such that $\sin(\theta_{ij}) \approx \theta_{ij}$.

Regarding algorithmic approaches that address the complexity of calculating $N-k$ contingencies, multiple techniques have been proposed in literature. Ranking and selection methods are traditional techniques that rank configurations of outages based on a heuristic index [56]. Advancements of such methods study contingencies based on Line Outage Distribution Factors (LODF), which are used to approximate the change in the flow on one line caused by the outage of a second line [24]. For $N-2$ contingency screening in particular, recent work on LODF based approaches can mathematically guarantee identification of all the dangerous $N-2$ contingencies with low computational costs [60].

For every system, given its topology and flows, there *always* exist a number p of multiple contingencies, which cannot be sustained and will lead to cascading failures. Adversaries can leverage contingency analysis techniques for tactical target analyses, towards identifying these p contingencies. Having constructed a model of the target system, they can identify which *specific* p locations are critical, and target them explicitly to materialize an attack.

5 Constructing attack vectors with OSINT

With knowledge of the critical points of a power system from the modeling and analysis steps, adversaries need to find attack entry points and construct attack vectors against the system. More specifically, they need to devise means towards disconnecting the critical transmission lines capable of a non-sustained contingency scenario, as they are identified using techniques outlined in Section 4. We provide representative, but non-exhaustive sources of public information that can be leveraged towards this objective through OSINT analysis.

Network: One possible entry point is over the network. Direct network channels to industrial devices in the identified target locations over the public internet may be available. To this end, Shodan, a "search engine for internet-connected devices", can

Table 2 Internet connected power grid devices indexed by Shodan

Protocol	Port	Indexed devices
DNP3	19999/20000	341
Modbus	502	13575
IEC 104	2404	445
IEC 61850	102	161

be employed [14]. Shodan uses crawlers that periodically index the web, searching for open ports and a wide variety of protocols including several industrial protocols. Table 2 is a snapshot of indexed devices by Shodan for the top four most commonly used protocols in the power industry taken at March 30, 2018. Moreover, to ensure non-stale results it is possible for attackers to launch their own crawlers. The release of efficient open source scanning tools such as ZMap, which can scan the entire IPv4 range in a matter of a minutes, have enabled large scale scans of the internet with limited resources [26]. Network telescope studies focusing on industrial protocols have shown that several scanning campaigns specifically target industrial protocols employing these tools [30]. The situation is exacerbated given the poor security the majority of industrial protocols employ, allowing unauthenticated access [33]. To extract the IPs of interest from the set of results, attackers can identify the organization in control of the target locations (e.g., power utility or governmental organization) and find IP addresses relating to the organization through reverse WHOIS searches. All IP addresses belonging to the organization are possible entry points; their compromise could enable lateral movement within the organization's network. The most promising results lie at the intersection between IPs owned by the company and industrial devices indexed by scanning campaigns. With the public IP address of a device known, attackers can employ remote fingerprinting techniques to identify the specifics of the industrial device [36].

Supply chain: When specific make and model information for target devices are known, attackers can carry out device-specific studies. To that end, a possible option for an attacker is to obtain a physical copy of the target device for further hands-on experimentation. With access to physical copies of power system devices, attackers can validate known vulnerabilities and test their developed attack strategies to increase the success probability of the final attack. Besides official vendors, online marketplaces such as eBay, Amazon, Alibaba and other third party companies offer used or new industrial equipment for sale. The majority of listings concern surplus or decommissioned equipment, typically sold at a fraction of the original price. Table 3 contains eBay listing statistics regarding microprocessor-enabled power grid devices from the four vendors with the largest market share [59], gathered on March 30, 2018.

Vulnerability reports: Publicly available vulnerability databases, such as the Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) advisories and alerts, and the National Vulnerability Database (NVD) can provide a public source of vulnerabilities for target devices [8, 10]. Such databases are constantly

Table 3 Microprocessor-enabled power grid devices from top vendors listed in eBay

Vendor	Listings
ASEA Brown Boveri (ABB)	216
General Electric (GE)	458
Schneider Electric	373
Siemens	271

updated with vulnerabilities discovered in industrial devices, including power system devices. If no vulnerabilities in the specific target device are published publicly, attackers can investigate vulnerabilities in the same family of products which will likely affect the target device because of intellectual property and code reuse between products in the same product line. Even if a patch was issued to address a known vulnerability, there is high probability that the target system is still vulnerable. Beyond reluctance and financial reasons, a major prohibitive factor for timely updates is that power systems must be available at all times and thus any modifications can only take place at prescheduled maintenance times [31].

Vulnerability development: In case no known vulnerabilities exist, or the objectives of an attack cannot be fulfilled with known vulnerabilities, adversaries can develop their own zero-day vulnerabilities. This approach is more resource demanding but has a higher probability of success and lower probability of detection. To achieve this, attackers can follow several strategies. With access to a physical device, they can extract the firmware of the device and analyze it for vulnerabilities, monitor the network traffic exchanged looking for vulnerabilities in the network stack, and examine the configuration software for attack entry points. Several publicly available blog articles outlining step-by-step approaches and techniques for hacking embedded systems and Internet of Things (IoT) devices contain knowledge that is directly transferable to power system devices [4]. For example, copies of firmware images may be obtained from vendor websites. If that is not possible, or if the firmware is encrypted, it can be directly dumped from the physical device via debug ports, or extracted from the device's flash memory using chip-off forensics techniques [39]. Reverse engineering of firmware images can be accelerated with the use of open source tools such as binwalk [2]. Fuzzing, a black box technique for testing software, can also yield exploitable results. As regards to the network stack, information exchanges between the target device and the configuration software can be intercepted and analyzed using open source tools such as Wireshark [48].

Open source projects: An observation that can be drawn from the analysis above is that there exists an abundance of open source software that can play an enabling role when designing an attack. The majority of this software was designed for benign uses (e.g., gathering statistics, education, penetration testing), but it can be misused by adversaries with a malicious agenda. This is the case for the ongoing campaign against U.S. critical infrastructure, where open source tools are employed [61]. In addition to the direct enabling role of open source projects, open source tools and libraries can become attack entry points. Adversaries can contribute to

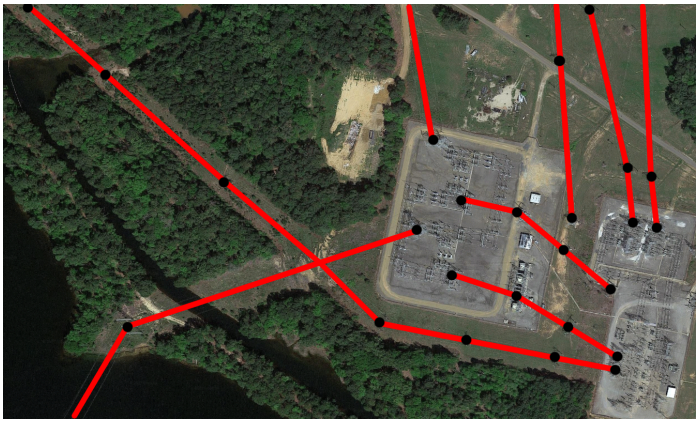

Fig. 2 Tracing transmission lines on GIS services.

open source projects they know are used by the target organization (e.g., a widely-used open source project in the power industry is openSCADA [52]). Hidden within updates, they can inject malicious code and backdoors.

6 Experimental evaluation of OSINT techniques

For experimentally evaluating the impact and quality of information generated with OSINT techniques we use an OSINT approach to construct a model of a large, real system, and analyze it with contingency analysis techniques to identify its critical operation points. Compromise and adversary control of these points is sufficient to create a system wide blackout.

6.1 Modeling a real power system

In testing the feasibility of modeling an entire real complex power grid using publicly available information, we select a real system and set out to find the required information outlined in Section 4. Because of the sensitive nature of this study, we refrain from identifying the system and present only anonymized and non-identifying information. For the remainder of this section, the system under study is referred to as *Outage Land*.

We employ OSINT techniques to generate the model of Outage Land. We identify several sources of information for the power system including publicly available corporate presentations, expansion planning reports, and vendor success stories. From these public sources we initially extract all the high voltage buses and connections between them, as well as the location of transmission and generation

Table 4 Outage Land power system statistics

Type	Number
400 KV buses	36
220 KV buses	64
132 KV buses	75
Total buses	**175**
400 KV to 400 KV branches	84
220 KV to 200 KV branches	115
132 KV to 132 KV branches	155
400 KV to 220 KV branches	29
400 KV to 132 KV branches	34
220 KV to 132 KV branches	4
Total branches	**421**
Generation stations	12
Maximum load forecast	15 GW
Installed generation capacity	17 GW

substations. Through this analysis we create a topological map of the system. We cross-validate the constructed topology through GIS services, by manually tracing high voltage lines and transmission towers throughout Outage Land. Fig. 2 shows an example of a generation plant, a transmission substation, and the incoming/outgoing transmission lines. Transmission towers are depicted as black dots and transmission lines as red lines. We undertake the laborious manual process of generating the topology of Outage Land's entire transmission network through GIS services and compare it with the topology extracted through other public resources, validating our findings. GIS mapping for the entire system required 40 man-hours. Although this process may be automated, we refrain from doing so because of the restrictive Terms and Conditions of commercial GIS software that do not allow automatic mining of satellite imagery.

Towards enabling contingency studies of Outage Land we again rely on OSINT techniques to extract operational characteristics of the power system. We fuse information from various sources to identify bus voltage levels, installed generation details, transmission line characteristics and load estimations. We again cross-validate the extracted information by comparing TSO reports, vendor success stories and news articles. The transmission network of Outage Land operates at three voltage levels: 132 KV, 220 KV and 400 KV. The components and connections for each voltage level are presented in Table 4.

6.2 Identifying critical operation points

The public information and the model of Outage Land is appropriately formatted to be used as input to MATPOWER, an open source MATLAB package, that solves power flow and optimal power flow problems [63]. The numerical testing we perform considers only the contingencies associated with tripped transmission lines,

since these are the targets identified in Section 2.1. For each $N-1$ contingency scenario (tripping of a power line) we compare the power flows in the resulting tripped network with the corresponding thermal limits of the transmission lines. If the contingency does not violate these thermal constraints, then the line is marked as "safe", otherwise it is characterized as "dangerous". Similarly, in the $N-2$ contingency case a pair of different lines (i, j) are tripped simultaneously, and a set of constraints are used to identify the events that lead to thermal limits violations.

We rely on the DC approximation for the $N-k$ contingency problem. The power system is described by the vectors of voltage angles θ_i, where $i = 1,..,n$ and n the number of buses in the system. In this scenario, the DC power flow equations have the form of $B\theta = p$, where B is the $n \times n$ nodal DC susceptance matrix and p is the vector of real power injections at the buses of the system. The matrix B can be represented as $B = MYM^T$, where M is the $n \times N$ connection matrix with $1s$ representing the beginning bus of the branch and -1 its end. Y is the diagonal $N \times N$ matrix of branch susceptances. Therefore, the vector of power flows can be described as $v = YM^T\theta = YM^TB^{-1}p$.

In addition to the $N-1$ contingencies we also want to identify the dangerous $N-2$ contingencies to increase the impact and probability of a successful attack. To that end, we use the pruning algorithm proposed in [60]. The algorithm excludes islanding conditions as they do not cause cascading failure propagation. The effect of each tripped line is described with a LODF matrix L which relates the change of flow in a monitored line i that follows after the tripping of line j with original flow v_j, i.e., the matrix element $L_{ji} = (\hat{v_j} - v_j)/v_i$, relates the change of the flow through line j from v_j (before outage) to $\hat{v_j}$ (after outage) with the flow v_i through line i before the outage. In order to find the relation between single and two line contingency LODFs, the LODF matrix becomes [60]: $L = YM^TB^{-1}\widetilde{M}(1-\widetilde{Y}\widetilde{M}^TB^{-1}\widetilde{M})^{-1}$, where $\widetilde{M}$ is the $n \times k$ submatrix of M corresponding to the outaged lines and similarly $\widetilde{Y}$ is the $k \times k$ outaged line submatrix of Y. This expression is applicable both to single and double line outage events. Direct comparison of these expressions allows us to relate the two LODF matrices. The double outage effect is:

$$\hat{v_l} - v_l = \frac{L_{li}(v_i + L_{ij}v_j)}{1 - L_{ji}L_{ij}} + \frac{L_{lj}(v_j + L_{ji}v_i)}{1 - L_{ji}L_{ij}} \tag{1}$$

In this relation we denote the outage lines by i, j and consider the change of the flow on some arbitrary line l. The contingency occurs whenever the absolute value of the flow at line l exceeds a critical value, i.e., $|\hat{v_l}| > v_l^{critical}$ that can for example be the thermal rating of a transmission line. We preprocess the model by converting it to a line-reduced network by aggregating radial branches into single nodes, a typical procedure for contingency analyses. We name the resulting nodes *transmission links*, as they are collections of transmission lines that connect the same edges (i.e., transmission substations).

From our analysis we identify 228 dangerous $N-1$ transmission link contingencies. They result from at least one violation of 72 transmission links in the 231-line reduced network, and are drawn as blue nodes in Fig. 3. All 228 dangerous $N-1$

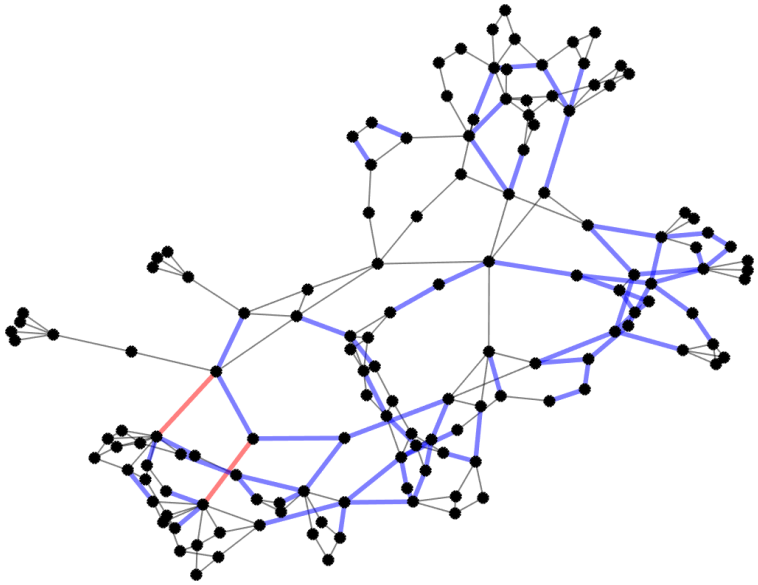

Fig. 3 Non-geographical network topology of Outage Land grid. Blue lines represent all the $N-1$ contingency transmission links, red lines represent the most critical set of $N-2$ contingency transmission links.

contingencies include more than one transmission lines to form their transmission link, meaning that more than one transmission lines need to be compromised to realize the contingency scenario they describe.

To investigate $N-2$ transmission link contingencies, we increase limits on the lines that caused $N-1$ contingencies, and rerun the $N-1$ analysis in search of $N-2$ contingencies. We identify 1174 such $N-2$ transmission link contingency scenarios in the reduced network, out of the possible 26565 combinations. In these results, some lines appear more frequently than others. To identify the most critical links, we rank the results based on their frequency of appearance. The set of most critical $N-2$ contingency links are presented as red nodes in Fig. 3.

The identified transmission links from the above study pose the greatest risk for a wide area power outage, hence they would constitute a natural target for a malicious adversary. The significance of the identified links is intuitively corroborated by their physical proximity to densely populated areas in Outage Land where end consumers are mostly situated. In theory, *any* of the above identified contingencies will result in a blackout, allowing an adversary to select which links to attack based on accessibility, whether the transmission lines are overhead or underground, or other criteria.

7 Discussion and related work

Ostensibly, this study seems to suggest that the cyber threat against power systems is under-estimated, as it serves as a proof-of-concept that attacks are enabled by the plethora of public sources of information. However, before tripping the alarm bells, we must take into consideration the resiliency and robustness of power grids around the world, as well as the readiness and experience of power engineers in handling blackouts and power outages. In the U.S. alone, 3879 blackouts were reported just in 2016, lasting an average of 48 minutes and affecting almost 18 million people across the country, causing an annual monetary loss of a staggering $150 billion [27]. All across the globe, blackouts are a usual occurrence typically caused by weather phenomena, accidents, animals, equipment failures etc. As a consequence, power system operators have experience in handling power outage scenarios. Furthermore, the power industry and government stakeholders have direct financial incentives in addition to societal obligations to direct efforts and funding in blackout prediction and prevention mechanisms, and ensuring shorter recovery times [23]. The objective investigated in this chapter, per the threat model, is causing a wide area power outage. This is not the same as a *prolonged* power outage. For example prolonged outages may require the destruction or incapacitation of critical not-in-stock equipment (e.g., large transformers and generators), something that typically requires physical tampering.

In general, public dissemination of information and transparency are beneficial for progress. With this work, our aim is not to constraint the release of information, classifying all power grid information as confidential. Rather, we aim to highlight the sensitivity of certain pieces of information that could enable malicious adversaries to launch cyberattacks against a power system, and assist in better understanding public information regarding power systems.

7.1 Why is sensitive information publicly available?

Throughout our study, we discovered a multitude of publicly available resources that contain sensitive details on power systems. Besides understanding what intelligence can be generated using such information, it is also important to understand *why* such information is part of the public domain. Regarding data that can be used for modeling a system, it is sometimes made publicly available by mistake, for example due to misconfiguration errors or improper access controls on public websites. In the case of identifying the location of power system assets through GIS, there is unintended leakage of information because of the nature of the satellite mapping techniques. Improper risk assessments from power system operators can also result in releasing sensitive information. Information can be leaked by third parties, such as power system devices vendors, when they release success stories that include operational details of real systems.

Alternatively, information can be intentionally released to the public. Given the strong dependence of nations on electric power supply, the power industry is often considered part of public utility infrastructure. Even in cases where power utilities are private corporations, they are regulated by public utilities commissions. For example, the U.S. Federal Energy Regulatory Commission (FERC) regulates all power utility companies. As public utilities, certain details regarding the operation of power systems must be transparent to citizens, in some cases mandated by law. For example, in some jurisdictions operational and procedural details must be provided in order to justify electricity pricing rates [18]. Furthermore, because of the threat of climate change, public interest groups and international organizations require publication of data to enable monitoring and regulating the environmental impact of electric utilities. Governments may also require the release of data to ensure transparency and fair competition between private companies [43], and data may be released towards promoting a more reliable and efficient power grid [53].

7.2 Prevention and mitigation strategies

In dealing with the cyber threat to power systems, efforts should be directed to efficient and effective prevention and mitigation strategies. These can start by following appropriate cybersecurity practices. Several best practices guides, such as the Guide to Industrial Control Systems Security by NIST [57] and the ISA99 Industrial Automation and Control Systems Security [34], offer practices that can thwart or impede attacks against power systems. Vendors of power system devices should harden the security of their products by revisiting their threat models, which might not consider malicious tampering as a serious threat. Given the direct impact of power system resiliency to the wider public, policymakers could accelerate action with regulatory policies that would incentivize power system operators and vendors to adopt good security practices.

From a technical perspective, field devices could be hardened at the different layers of the system, namely the hardware, firmware, software, network and operation layers [40]. When it comes to hardware, access through debug ports, such as JTAG and UART, should be disabled and the supply chain should be more strictly controlled to disallow adversaries to obtain physical copies of critical devices. Hardware support for cryptographic operations can enable the use of secure cryptographic primitives in power systems. For firmware, static images should be encrypted and firmware updates should be signed by the respective vendors. On the software layer, proper security mechanisms and risk assessments are necessary, along with the ability of secure and verifiable updates that do not require downtimes. Regarding the network layer, industrial devices deployed in power systems should never be directly connected to the internet, and field networks and business networks should be segmented to thwart attacks that rely on lateral movement. Industrial protocols with no security mechanisms should be redesigned or replaced with secure counterparts. On the operation layer, effective physical security mechanisms should be

employed, and the overall operation of the system could inform anomaly detection schemes that aim to act as early indicators of attacks by gathering system-wide data.

As regards to information that can expose critical operational characteristics and/or enable attack vectors against a power system, more careful consideration must be taken when deciding its classification and targeted audience. Power system shareholders can carry out periodic reviews of publicly available information, gaining visibility into the different sources of information that concern their systems. In addition, requirements and methods for proper handling and control of potentially critical information can be established. For example, techniques that anonymize or randomize the information before its release could simultaneously provide transparency while protecting critical and sensitive data.

7.3 Related work

In the power engineering academic community, there exist studies that utilize real data to model complex power systems. Several studies use the Polish grid model included in MATPOWER cases [63], which is based on data collected in 2000 and 2001 from the website of the Polish transmission system operator. However, information about the current state of the Polish grid is not available. In addition to the Polish grid, real data from the power grid in Great Britain (GB) in 2013 are available as a MATPOWER case. These data were assembled from National Grid public data and reports and used for optimal power flow studies [17]. Real U.S. power grid data obtained from the Platts GIS database are used in [22] to estimate the vulnerability of the U.S. power grid to geographically correlated failures. In contrast, in this chapter, we build a model of a real system from the ground up leveraging and fusing a variety of public resources, and validate the created model using GIS imagery.

8 Conclusions

Motivated by the observation that adversaries actively employ OSINT techniques for campaigns against the energy sector, we undertake a study of the amount and quality of intelligence that can be extracted using publicly available resources. We present a broad study on the sources of intelligence that can be leveraged to model a large power system, analyze the model and finally study how to exploit it. We experimentally evaluate the feasibility of an open source campaign by constructing and validating the model of a real system using only publicly available information, and analyzing the system to identify its critical points using contingency analysis. With this study we aim to provide insight into the threat cyberattacks based on publicly available resources pose to power systems. Our work can assist energy sector stakeholders and regulators take informed decisions, and more carefully handle information dissemination concerning sensitive characteristics of their power systems.

References

1. Bangladesh power cut plunges millions into darkness. http://reuters.com
2. Binwalk firmware analysis tool. https://github.com/ReFirmLabs
3. Blackout watch: Brazilian blackout 2009. http://pacw.org
4. Embedded device hacking. http://devttys0.com
5. European network of TSOs for electricity. http://entsoe.eu
6. How an entire nation became Russia's testlab for cyberwar. http://wired.com
7. India Northern Regional Load Despatch Centre. http://nrldc.in
8. Industrial Control Systems Cyber Emergency Response Team. http://ics-cert.us-cert.gov
9. Map of PMUs with synchrophasor data flows in North America. http://naspi.org
10. National Vulnerability Database. http://nvd.nist.gov
11. Open energy information. http://openei.org
12. Open power system data platform. http://open-power-system-data.org
13. Rebels tied to blackout across most of Pakistan. http://nytimes.com
14. Shodan search engine. http://shodan.io
15. U.K. Electricity Ten Year Statement 2016. http://nationalgrid.com
16. Ukraine's power outage was a cyber attack: Ukrenergo. http://reuters.com
17. Network data of real transmission networks. http://maths.ed.ac.uk (2013)
18. International energy statistics. http://eia.gov (2017)
19. Abraham, S., Efford, J.R.: Final report on the August 14, 2003 blackout in the U.S. and Canada. Tech. rep., Power System Outage Task Force (2004)
20. Alonso, F., Greenwell, C.: Underground vs. Overhead: Power line installation-cost comparison and mitigation. Electric Light and Power **22** (2016)
21. Bakshi, S.: Report of the enquiry committee on grid disturbance in Northern region on 30th July 2012 and in Northern, Eastern & North-Eastern region on 31st July 2012. Tech. rep., Indian Ministry of Power (2012)
22. Bernstein, A., Bienstock, D., Hay, D., Uzunoglu, M., Zussman, G.: Power grid vulnerability to geographically correlated failures: Analysis and control implications. In: INFOCOM. pp. 2634–2642. IEEE (2014)
23. Campbell, R.J.: Weather-related power outages and electric system resiliency. Tech. rep., Congressional Research Service (2012)
24. Davis, C.M., Overbye, T.J.: Multiple element contingency screening. Transactions on Power Systems **26**(3), 1294–1301 (2011)
25. Davis, C., Chmieliauskas, A., Nikolic, I.: Enipedia. Energy & Industry group, Faculty of Technology, Policy and Management, TU Delft (2015)
26. Durumeric, Z., Wustrow, E., Halderman, A.: Zmap: Fast internet-wide scanning and its security applications. In: Usenix Security (2013)
27. Eaton: Blackout tracker: United States annual report 2016 (2017)
28. Eles: Slovenia's transmission network: Annual report 2015. http://eles.si
29. Elia: Belgium electrical transmission network: Annual report 2016. http://elia.be
30. Fachkha, C., Bou-Harb, E., Keliris, A., Memon, N., Ahamad, M.: Internet-scale probing of CPS: Inference, characterization and orchestration analysis. Network and Distributed System Security Symposium (2017)
31. Gharavi, H., Ghafurian, R.: Smart grid: The electric energy system of the future. Proceedings of the IEEE (2011)
32. Grainger, J., Grainger, W., Stevenson, W.: Power system analysis. McGraw-Hill Education (1994)
33. Igure, V.M., Laughter, S.A., Williams, R.D.: Security issues in SCADA networks. Computers & Security **25**(7), 498–506 (2006)
34. International Society of Automation: ISA99: Industrial automation and control systems security (2015)
35. Kaplunovich, P., Turitsyn, K.: Fast and reliable screening of N-2 contingencies. Transactions on Power Systems **31**(6), 4243–4252 (2016)

36. Keliris, A., Maniatakos, M.: Remote field device fingerprinting using device-specific MODBUS information. In: International Midwest Symposium on Circuits and Systems. pp. 1–4. IEEE (2016)
37. Keliris, A., Maniatakos, M.: Demystifying advanced persistent threats for industrial control systems. Dynamic Systems & Control Magazine **5**(1) (2017)
38. Knake, R.: A cyberattack on the U.S. Power Grid. chap. Contingency Planning Memorandum No. 31. Council on Foreign Relations (2017)
39. Konstantinou, C., Maniatakos, M.: Impact of firmware modification attacks on power systems field devices. In: International Conference on Smart Grid Communications. pp. 283–288. IEEE (2015)
40. Konstantinou, C., Maniatakos, M.: Security analysis of smart grid. Communication, Control and Security Challenges for the Smart Grid **2**, 451 (2017)
41. Langner, R.: Stuxnet: Dissecting a cyberwarfare weapon. Security & Privacy **9**(3), 49–51 (2011)
42. Lee, R.M., Assante, M.J., Conway, T.: Analysis of the cyber attack on the Ukrainian power grid. Tech. rep., SANS Industrial Control Systems (2016)
43. Momoh, J., Mili, L.: Economic market design and planning for electric power systems, vol. 52. John Wiley & Sons (2009)
44. Morison, K., Wang, L., Kundur, P.: Power system security assessment. Power and Energy Magazine **2**(5), 30–39 (2004)
45. NERC: Disturbance Reports 1992-2009
46. NERC: FAC-011-2: system operating limits methodology for the operations horizon (2009)
47. OpenStreetMap: Power networks. http://openstreetmap.org
48. Orebaugh, A., Ramirez, G., Beale, J.: Wireshark & Ethereal network protocol analyzer toolkit. Syngress (2006)
49. Pajic, S.: Power system state estimation and contingency constrained optimal power flow: A numerically robust implementation (2007)
50. Project Group Turkey: Report on blackout in Turkey on 31st March 2015. Tech. rep., European Network of Transmission System Operators for Electricity (2015)
51. ProSoft Technology: Power success stories. http://prosoft-technology.com
52. Reimann, J., Rose, J.: Eclipse SCADA: The definite guide (2015)
53. Roland Berger: Study regarding grid infrastructure development: European strategy for raising public acceptance. Tech. rep., European Commission Tender No. ENER/B1/2013/371 (2014)
54. SCADA Innovations: Success stories. http://scadainnovations.com
55. Siemens: High voltage substation references. http://energy.siemens.com
56. Stott, B., Alsac, O., Alvarado, F.: Analytical and computational improvements in performance-index ranking algorithms for networks. International Journal of Electrical Power & Energy Systems **7**(3), 154–160 (1985)
57. Stouffer, K., Falco, J., Scarfone, K.: Guide to industrial control systems security. NIST special publication **800**(82), 16–16 (2011)
58. Symantec: Dragonfly: Western energy sector targeted by sophisticated attack group. http://symantec.com
59. Technavio: Global Smart Grid Transmission and Distribution Equipment Market 2016-2020. http://technavio.com
60. Turitsyn, K.S., Kaplunovich, P.A.: Fast algorithm for N-2 contingency problem. In: 46th Hawaii International Conference on System Sciences (HICSS). pp. 2161–2166. IEEE (2013)
61. U.S. DHS and FBI: US-CERT: Advanced Persistent Threat Activity Targeting Energy and Other Critical Infrastructure Sectors. http://us-cert.gov/ncas/alerts/TA18-074A
62. Zachariadis, T., Poullikkas, A.: The costs of power outages: A case study from Cyprus. Energy Policy **51**(Supplement C), 630 – 641 (2012)
63. Zimmerman, R.D., Murillo-Sánchez, C.E., Thomas, R.J.: MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. Transactions on power systems **26**(1), 12–19 (2011)

View publication stats