This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3131220, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier <doi>

Energy theft in smart grids: A survey on data-driven attack strategies and detection methods

AHLAM ALTHOBAITI[1], (Student Member, IEEE), ANISH JINDAL[2], (Member, IEEE), ANGELOS K. MARNERIDES[3], (Member, IEEE), UTZ ROEDIG[4], (Member, IEEE)

[1]School of Computing & Communications, Lancaster University, Lancaster, UK (e-mail: a.althobaiti@lancaster.ac.uk)
[2]School of Computer Science & Electronic Engineering, University of Essex, Colchester, UK (e-mail: anishjindal90@gmail.com)
[3]School of Computing Science, University of Glasgow, Scotland, UK (e-mail: angelos.marnerides@glasgow.ac.uk)
[4]School of Computer Science & Information Technology, University College Cork, Cork, Ireland (e-mail: u.roedig@cs.ucc.ie)

Corresponding author: Ahlam Althobaiti (e-mail: a.althobaiti@lancaster.ac.uk)

ABSTRACT The convergence of legacy power system components with advanced networking and communication facilities have led towards the development of smart grids. Smart grids are envisioned to be the next generation innovative power systems, guaranteeing resilience, reliability and sustainability and to facilitate energy production, distribution and management. Nonetheless, the development of such systems entails challenges covering a broad spectrum ranging from operational management up to data-driven power accounting and network security. Given the highly distributed properties of the modern grid, energy theft can now be observed at various transmission and distribution levels. Apart from the financial gain for a malicious actor, energy theft can also affect critical grid processes with a direct impact on its overall resilience and safety. This survey reviews recent energy theft strategies as well as detection methods from a data-driven perspective. By considering various operational and functional layers within modern smart grids we critically assess how energy theft can be formulated. Moreover, we provide an overview of the grid demand, supply and control chain with a focus on energy theft and associated security flaws that currently exist in the smart grid ecosystem. Different attack detection models for theft detection in the smart grid are categorized. Lastly, we discuss various open issues in the scope of data-driven energy theft detection methods and provide future directions to carry out research in this field.

INDEX TERMS Energy theft, Data-driven methods, Smart grid, Cybersecurity

Nomenclature

Abbreviation

$T\&D$	Transmission and distribution
ACC	Accuracy
AMI	Advanced metering infrastructure
ANN	Artificial neural network
AUC	Area under the curve
BMS	Building management system
CNN	Convolutional neural network
DC	Direct current
DL	Deep learning
DR	Demand response
DRES	Distributed renewable energy sources
DSO	Distribution system operator
EMS	Energy management system
FDI	False data injection
FIT	Feed in tariff
FPR	False positive rate
GBM	Gradient boosting machines
GPS	Global positioning system
GRU	Gated recurrent unit
HAN	Home area network
HEMS	Home energy management system
HV	High voltage
ICS	Industrial control system
IEA	International energy agency
IED	Intelligent electronic device
KNN	K-nearest neighbors
LAN	Local area network
LOF	Local outlier factor
LSTM	Long short term memory
LV	Low voltage

MITM	Man in the middle
MLP	Multi layer perceptron
MV	Medium voltage
NAN	Neighbourhood area network
OPF	Optimum path forest
PCA	Principal component analysis
PDC	Phasor data concentrator
PMU	Phasor measurement unit
RNN	Recurrent neural network
RTU	Remote terminal unit
SCADA	Supervisory control and data acquisition
SVM	Support vector machine
TPR	True positive rate
TSO	Transmission control operator
WAMS	Wide area measurement system
WAN	Wide area network

Notation

α	Theft coefficient on generation data
β	Theft coefficient on supply data
γ	Theft coefficient on demand data
G	T&D grid
M	Number of energy distribution buses
N	Number of total nodes
P	Number of prosumer nodes
Q	Number of consumer nodes
S	Number of grid supply nodes
Ec	Demand node energy consumption
Er	Prosumer node energy generation
Es	Energy supply by T&D control nodes
NTL	Cumulative non-technical energy loss
TL	Technical energy loss

I. INTRODUCTION

CYBER-PHYSICAL attacks on power grids aiming explicitly at energy theft are the most prominent and they have been reported to cause significant financial as well as functional losses to energy utility companies at a global scale. Hence, energy theft attacks cause major concerns to both providers and consumers. Irrespective of whether such attacks are executed by a single consumer or at a large scale, losses incurred to providers due to energy theft are undesirable and enormous. As reported in [1], energy theft causes utility companies to loose more than £19 billion yearly on a global basis.

Several studies carried out in 2019 point that almost 80% of 2000 UK residents were not aware that energy theft is directly affecting them [2]. The reported studies also reveal that due to energy theft, £20 are added yearly on average to a household bill. Thus, millions of clients pay for energy that they have never used and, most importantly, did not steal. In general, each year in the UK alone, energy worth £400 million is stolen leading to inflated customer bills [3].

In order to secure such energy and revenue losses, utility companies typically conduct physical inspections in the locations where energy theft is due to intensify [4]. Nonetheless, such conventional energy theft detection tracking is time-consuming, inaccurate, costly, and labour-intensive [5]. Therefore, to deploy more effective theft countermeasures, providers need to make use of the present electricity market driven by the need to collect and analyze data. The facilitation of data-driven operation drives utility providers to embed smart metering equipment in various levels of the electricity flow within smart grids [6]. The entire life cycle of gathering energy data runs through smart grid infrastructures which are categorized into electricity generation, transmission and distribution (T&D), and end-user infrastructure. This data collection infrastructure leads to the emergence of an advanced line of detection method driven by measurement-based data providing opportunities to address energy theft. Data-driven detection is able to reduce the risk of lateral attacks leading to energy theft and recognize anomalous system behaviours arising from such events. Thus, reduce revenue losses for service utilities.

Although, a variety of data-driven detection methods have been developed, malicious actors continue to discover innovative strategies in an attempt to perpetrate energy thefts across smart-grid infrastructures [7]. In this regard, the smart grid data measurements and monitoring infrastructure can pave the way for more approaches to fabricate next generation data-driven theft attacks, thus increasing relative energy and financial losses. McLaughlin *et al.* [8] and Jiang *et al.* [9] review these data-driven theft attacks from the perspective of power-system communication-layer architectures, based on adversary strategies targeting the integrity of the power system by manipulating power demand data. However, these surveys were not focused on energy theft and do not consider recent advances in modern smart grids, as the nature of vulnerabilities and threats related to energy theft are constantly changing due to the increasing intersection of power grids with Internet-enabled cyber-physical systems [10].

Motivated by these observations, we investigate and survey the advances in energy theft from different perspectives within the smart grid ecosystem revolving around energy data manipulation from all the three functions of demand, supply, and generation. A variety of vulnerabilities enable adversaries to exploit grid infrastructure components, communication networks and managements systems with the intention of gaining monetary benefit. Hence, in this survey we provide an overview of different types of energy theft attacks in smart grids. We audit the latest research on data-driven attacks enabling energy theft and outline key findings. Moreover, we also discuss the existing data-driven energy theft detection schemes and summarise outstanding challenges. This work serves as the first stop for general audiences as well as domain specialists looking for information and guidelines regarding energy theft in present-day smart grid systems and markets. We explicitly contribute in the wider research community for modern energy grids by providing:

1) The first survey paper covering the largest spectrum of data-driven attack strategies available in the literature used for carrying out energy theft in the modern electricity market.

2) A novel energy theft categorization model from the different smart-grid data flow perspectives.
3) A critical assessment of lessons learned from the application of various data-driven approaches presently used for detecting energy theft.
4) Recommendations for future research directions with respect to the design of data-driven energy theft detection schemes as tailored with an extensive analysis of open issues.

The remainder of this paper is organized as follows: section II focuses on the key infrastructures consisting the modern power grids such as to relate attack vectors associated with energy theft. Section III provides a comprehensive analysis on data-driven energy theft attacks. In section IV, we categorize and discuss data-driven algorithms used in energy theft detection systems. Section V presents the existing gaps in research for data-driven energy theft detection discussing open issues and recommends future research directions. Finally, in Section VI, we conclude and summarise this paper.

II. SMART GRID COMPONENTS

Energy theft may span over multiple logical or physical entities and can be instrumented via numerous attack vectors affecting one or more of the systems consisting the modern smart grid. Within this work, the various properties of energy theft are discussed in terms of the intrinsic characteristics of each of these infrastructures. Therefore this section is dedicated at presenting an overview of the infrastructure of the smart grid with its core components.

One of the main goals within the modern smart grid is to ensure the optimal operation of the electricity supply chain[1]. As shown in Fig. 1, the end-to-end energy supply chain is decomposed into three distinct phases; i) generation , ii) transmission and distribution (i.e. T&D) and, iii) end-user consumption. All three phases are directly dependent on explicit technologies, administrative domains and networked power system infrastructures. Each of these entities pose unique vulnerabilities that can enable energy theft [11], [12].

The energy generation phase is achieved within large, centralised power stations that nowadays are interfaced with power generation DRES deployments and are commonly owned by the national transmission energy network controlled by one or a set of transmission control operators (TSOs). Each TSO is engaged through a competitive energy trading market scheme with a number of distribution system operators (DSOs) in order to supply them with electricity to be distributed to end-consumers [2]. DSOs may also have a direct interface and own DRES deployments or they frequently have an energy trading contract with end-consumers or third-party DRES owners that contribute directly in the energy generation phase.

In general, any control and management (sub)systems alongside the electro-mechanical set of power systems enabling data and energy flows spanning the energy supply chain are underpinned by diverse and ubiquitous data communication technologies. Fig. 1, indicatively illustrates a variety of potential networking technologies and deployment setups that could be employed in modern smart grids. Similarly with the energy trading market, the business model behind the ownership of these deployments depends on a number of aspects related to country-level legislation and policies [13] and it is out of the interest within this paper.

A. ENERGY GENERATION

1) Centralised generation

Energy generation systems can be categorised to operate either in a centralised or a decentralised fashion. Centralised generation systems produce large-scale electricity at power stations, utilising fossil fuels and nuclear plants or renewable resources such as hydroelectric power plants, wind and solar farms. These centralised systems are usually placed in remote areas that are distant from the end users. and are linked to distributed stations owned by a given DSO via a network of HV transmission lines operated by a TSO [14]. The DSO stations are responsible for transmitting electricity through the medium and low voltage grids to multiple end-users [15].

2) Distributed Renewable Energy Sources (DRES)

DRES have evolved to act as an integral element of the electricity generation infrastructure aiding the needs of the backbone grid in terms of critical ancillary services (e.g., frequency regulation, reactive power) enabling grid stabilisation, diversifying energy trading and most importantly matching the peak during overloaded periods [16]–[18]. Moreover, DRES deployments are currently considered as the most suitable components for contributing towards the reduction of global carbon emissions [17]. According to the international energy agency (IEA), DRES deployments have contributed to 40% of the total primary energy supply globally in 2020 [19].

Energy generation billing and trading for DRES is currently achieved via two distinct systems; i) net metering and, ii) feed in tariffs (FIT). Net metering operates with a single meter and employs a model where prosumers use their own DRES-based generated power on-site and any surplus is considered as a future credit on their billing issued by their DSO. On the other hand, FIT operates based on two smart meters residing at the prosumer end dealing with the capturing of energy generation and consumption rates independently. By contrast to net metering, FIT decouples the monitoring process and facilitates a simpler data processing framework for energy trading as well as billing, thus it was extensively adopted in a number of developed countries such as the UK, Canada, Japan, China, and Australia [20].

Despite of the various benefits offered by DRES deployments, their direct dependency on natural resources (e.g., wind, solar radiation) that are in some cases unpredictable

[1] This paper is focused on energy delivered by electricity networks and not gas.

[2] In the USA a TSO may be referred to as an independent system operator (ISO) and a DSO as a regional transmission operator (RTO).

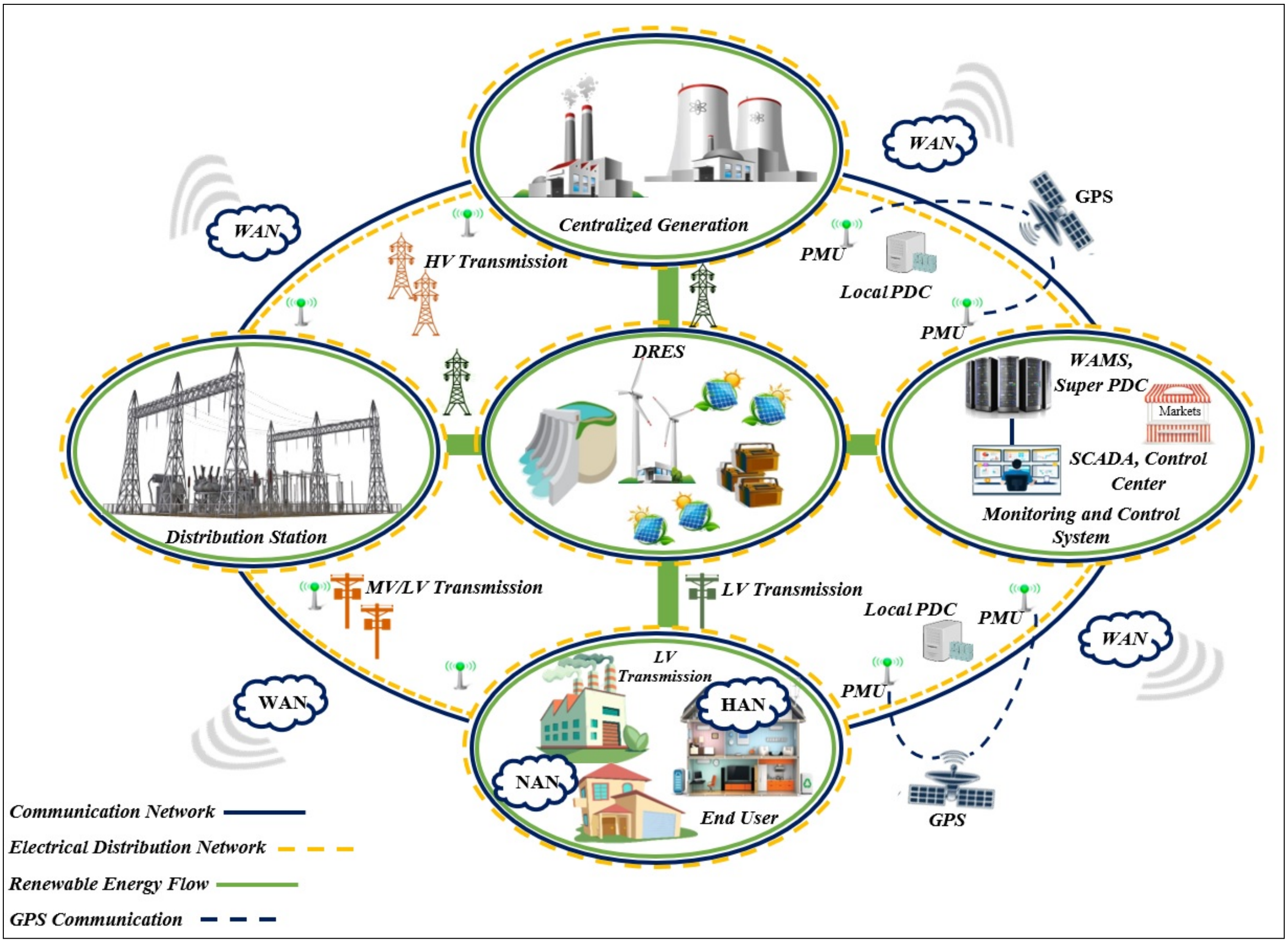

FIGURE 1: Phases and components of the energy supply chain in the smart grid.

to fully forecast may cause challenges and higher complexity within the overall grid optimisation process. Thus increasing risks related to aspects of reliability, resilience as well as safety [21], [22]. Nonetheless, a number of approaches have been proposed to confront complexity constraints through ramp strategies [17]. In parallel, the integration of DRES involves diverse types of data communication and system-on-chip technologies that are commonly manufactured with minimal security [23], [24]. Hence, enlarging the spectrum of cyber attacks that could be initiated such as to support potential energy theft acts [25].

B. ENERGY TRANSMISSION & DISTRIBUTION (T&D)

1) T&D energy flow

The T&D infrastructure is responsible for enabling the transmission of power and further distribution of electricity to the consumers. As depicted in Fig. 1, T&D infrastructures may be categorised into the low voltage (LV), high voltage (HV) and medium voltage (MV) power networks. Throughout the years, the topology for these power networks has evolved from an ordinary radial structure to interconnected or consistent networks, which has guaranteed higher reliability, operational economy, and best equipment use [26]. Primarily, the electricity produced by the centralized electricity generation systems is transported to different distribution stations over HV transmission lines, which is then supplied to the end-users through the widespread transmission lines of MV and LV networks. In parallel, modern T&D infrastructures also distribute energy generated at DRES deployments through MV-LV substations [26].

2) T&D data communication

The data communication network underpinning the operations of T&D infrastructures commonly consists of two types of networked deployments that interact with the end-consumer home area network (HAN). As demonstrated in Fig. 2, end-to-end data communication between the T&D infrastructure and a HAN is achieved via a wide area network (WAN) interacting with a set of neighbourhood area networks (NANs).

A WAN typically represents the aggregation of NANs and it is mapped at the scale of a city-wide network considering data flows related to energy distributed by multiple micro-

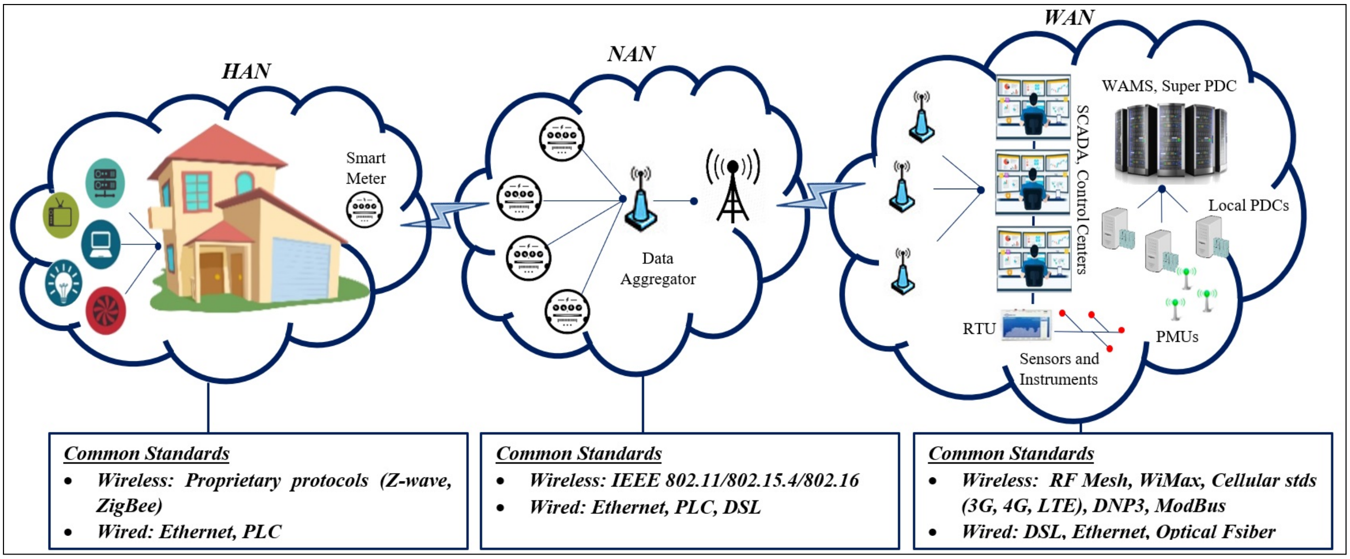
HAN
Smart Meter
NAN
Data Aggregator
WAN
WAMS, Super PDC
SCADA, Control Centers
Local PDCs
RTU
PMUs
Sensors and Instruments
Common Standards
• Wireless: Proprietary protocols (Z-wave, ZigBee)
• Wired: Ethernet, PLC
Common Standards
• Wireless: IEEE 802.11/802.15.4/802.16
• Wired: Ethernet, PLC, DSL
Common Standards
• Wireless: RF Mesh, WiMax, Cellular stds (3G, 4G, LTE), DNP3, ModBus
• Wired: DSL, Ethernet, Optical Fsiber

FIGURE 2: Exemplar Smart Grid network architecture highlighting some of the main data communication standards.

grids where each micro-grid is linked with a particular NAN. In real deployments, the structure of a WAN is quite diverse since it may consist of multiple networking technologies with varying physical, logical and software components dealing with network control and management [27], [28]. On the other hand, NANs can be considered as a subset of a WAN since they support smaller geographical regions and they act as proxies of a given WAN for functions related to connectivity and data aggregation of HANs with the main WAN. In general, a WAN or a set of WANs alongside related NANs and HANs are not necessarily always owned by corresponding TSOs or DSOs as they could be managed and maintained by third-party network providers (e.g., Internet Service Providers) or community entities (e.g., municipality).

3) Data acquisition & management

The actual interface of data communication with data-driven control and management of the processes explicit to reliable and resilience distribution of energy is achieved via network-enabled cyber-physical systems such as supervisory control and data acquisition (SCADA) systems. These systems are nowadays the most frequently used systems within modern T&D infrastructures. SCADA systems provide native integration of data communication technologies and system components such as remote terminal units (RTUs) and intelligent electronic devices (IEDs) [29], [30]. The data communication reliability offered by SCADA systems enables TSO/DSO control centres to develop close to real-time state estimation algorithms in order to optimise the grid's performance and increase situation-awareness [31], [32].

A relatively recent alternative approach to SCADA are wide area measurement systems (WAMS) [33]. WAMS are embedded with new data acquisition technologies facilitating synchronised measurements between remote T&D deployments (e.g, micro-grids, substations) and facilitate the basis for monitoring, operation and control [34]. In practise, WAMS may be decomposed by a set of distributed Phasor measurement units (PMUs) and phasor data concentrators (PDCs) that sample data related to the waveform and the analog voltage of remote sites through a global positioning system (GPS) clock [33], [35].

C. END-USER INFRASTRUCTURE

1) Advanced Metering Infrastructure (AMI)

AMIs are considered one of the fundamental components within the smart functionalities of the modern energy grid. The operation of such infrastructures achieves end-to-end metering in order to support the billing and trading processes between an end-consumer or prosumer and a DSO/TSO. A core innovation behind AMIs lies with the integration of smart meters within residential households or business buildings. In most developed and many of the developing countries, smart meters have replaced the traditional mechanical and analogue meters and they enable a variety of services. Apart from the real-time logging of measurements related to end-user energy consumption (i.e. demand data), smart meters also assess other features such as voltage levels as well as real-time monitoring [36].

As already mentioned, data captured by smart meters contribute to the overall demand response (DR) model and they are transmitted through low-powered communication and automation protocols (e.g., ZigBee, Z-Wave) in synergy with upper layer application protocols (e.g., HTTP/HTTPS) supported by their corresponding HAN. Fig. 2 provides an exemplar illustration in which smart meter measurements are locally aggregated within a HAN and are further distributed to the corresponding T&D infrastructure through an adjacent NAN interacting with a WAN. The sampling rate for measurements gathered by individual smart meters falls with a pre-defined schedule agreed between the end-consumer or prosumer with its corresponding DSO. Normally, measurements are agreed to be sent in $5, 15, 30$, or 60 minute intervals [36]–[39].

2) Energy Management System (EMS)

The adequate management and reactive control of energy usage and production in end-user deployments is achieved through the installation of EMS instances. Such instances may be directly interfacing with a given DSO or through proxy third-party stakeholders maintaining and supporting large-scale EMS deployments. From the end-user perspective, there is a variety of EMS types coming with specific functionalities such as home energy management systems (HEMSs) and building management systems (BMSs)[3]. In parallel, EMS can also be present at a larger scale deployed either at a centralised or a distributed topology aggregating measurements for the T&D insfrastructure [40]–[42]. Nonetheless, the main role of an EMS instance at the end-user infrastructure is to optimise energy consumption for an individual or a set of individuals through controlling the various appliances residing within a given building or household [43]. Hence, EMS software instances are usually composed of a controller instructed by advanced energy optimisation algorithmic components coupled with rule-based control functions orchestrating the operations of appliances [6], [44].

D. GRID EFFECTIVENESS PILLARS

The effectiveness of the grid in all levels depends on the performance of both quantitative as well as qualitative indicators. For instance, the reliable operation of the energy grid directly affects the well-being and safety of consumers whereas well-being is not a fully quantifiable parameter and, in parallel, grid reliability depends on quantifiable performance metrics (e.g., demand-supply rate) [45]. Moreover, cyber-physical challenges, such as attacks enabling energy theft may affect directly grid optimisation processes, thus impacting grid reliability with a cascading impact over user safety since some power system machinery could be affected and malfunctioning [46], [47]. The latter example has a number of parameters that are not necessarily quantifiable (e.g., grid security level, safety impact on consumers/prosumers), hence a holistic correlation scheme between the aforementioned pillars is an extremely challenging task.

As evidenced in Fig. 3 this work relates grid effectiveness with the three broad domains of reliability, resilience and safety that we refer to as pillars. We exploit definitions developed throughout the years and summarise the definitions of the three inter-related pillars in order to structurally assess the energy theft impact in the overall grid effectiveness [47]–[49]:

1) Grid Reliability: preservation of continuous energy supply to end consumers .
2) Grid Resilience: preservation of continuous energy supply to end users with an acceptable level of energy quality while under stress or faults.
3) User Safety: ensure that an individual or a group of individuals utilising or maintaining the grid and its services are not physically affected.

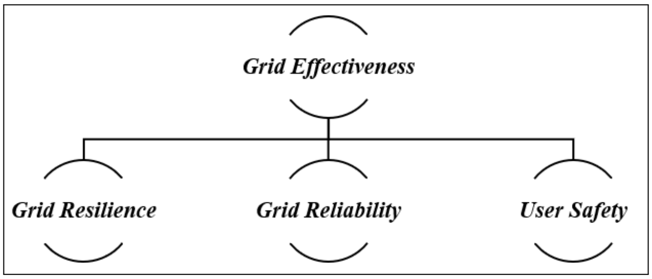

FIGURE 3: Grid effectiveness pillars.

This survey acknowledges that the highlighted pillars are considered widely as independent research domains themselves. Hence, deeper investigation on the structure and properties of these pillars is out of the context within this paper.

III. ENERGY THEFT

Energy theft can be broadly defined as the case where individuals do not pay their electricity bills or they are paying less than they should due to their meter being tampered or bypassed. Attack vectors underpinning energy theft span numerous vulnerability domains due to the emergence of a plethora of smart grid applications (e.g., energy trading) that rely on inherently vulnerable networked environments as a result of the convergence of diverse legacy power systems with Internet technologies (e.g., ICS deployments, metering). In general, energy theft can be instrumented through a variety of techniques exploiting both physical as well as data or communication-oriented properties of the current grid [4], [50], [51]. Hence, the adequate categorisation of energy theft types is a highly challenging task.

In order to address the aforementioned challenge and appropriately structure the focus within this work, we identify two distinct classes of energy theft:

1) Data-agnostic energy theft: the act of physical tampering of power components through techniques such as obstruction and bypass of electro-mechanical meters, cable hooking as well as modification of meter circuitry.
2) Data-driven energy theft: the act of manipulating and altering communication and/or consumption-related data generated and/or logged at any networked metering (e.g., smart-meter), management (e.g., SCADA system) and control device (e.g., PLC) as well as billing software (e.g., utility mobile apps) aiming at reporting false consumption information to the power distribution authority (e.g., a DSO).

Both classes target either of the bidirectional energy or data flow between different grid aggregation points (e.g.,

[3] Discussion of EMS variations is out of scope for this paper.

T&D, end-user, generation) and they have seen a considerable level of attention from the research community as well as the society in general [12], [20]. Moreover, both types have shown to be applicable in all three levels of aggregation within a smart grid. Hence, energy theft can be deployed in the power generation infrastructure, the T&D network as well as the end-consumer level.

This work argues that the main concept of a given data-driven theft attack can be abstracted by a discrete function in which inter-dependent variables are tailored based on the targeted infrastructures composing a complete smart grid deployment. Hence, the function may vary depending on the variable-specific adjustments conducted by a malicious actor based on the intrinsic properties of a given smart grid (sub)infrastructure (e.g., communication, power). Commonly, malicious actors attempt to target a set of diverse vulnerabilities of both system and network components from all three infrastructures described herein. Evidently, data-driven energy theft in all three infrastructures has considerably increased due to the data-oriented functioning of the business layer as envisaged in the current smart grid reference architectures (e.g., SGAM [52]).

A. TECHNICAL IMPACT OF ENERGY THEFT

Energy theft is underpinned by a large spectrum of cyber-physical attacks that span a number of organisation (e.g., physical security) as well as technology-oriented vulnerabilities (e.g., legacy ICS security). The current ubiquity offered by the bidirectional flow of energy and data in the current smart grid, alongside the highly distributed nature of various components (e.g., DRES) enable the composition of energy theft-related attempts.

In the year 2012, a German renewable power utility was targeted with the denial of service (DoS) attack when thousands of requests were sent to its server to block its operation [53]. This attack knocked off the Internet connection of the utility for five consecutive days. Such a scenario can serve as an opportunity for the malicious entities to orchestrate an energy theft without being detected causing major losses to the utility. More recently, there have been unconfirmed attempts on national grid infrastructure of the United States and United Kingdom wherein the potential hackers tried to break into the utility's network to disrupt their services [54], [55]. These attacks, if successful, have the potential to affect grid effectiveness by hampering the business model of the grid and can also lead to the infrastructure failure. Following such events, the US and UK security services issued warnings to the providers to raise the cyber-security standards in order to mitigate such attacks which have increased manifolds.

Cyber attacks on power grids were, and still are in the majority, instrumented with the intent to manipulate data flows of the various grid resources and/or services. A number of data-driven cyber attacks (e.g., as in [56], [57]) caused catastrophic faults on system components leading vital grid optimisation processes to malfunction. Thus, causing a decay on the overall grid effectiveness. For instance, the infamous cyber-attack in Ukraine's T&D infrastructure resulted in power outages that have affected around $225,000$ consumers for several hours [56]. Another similar attack in 2019 targeted the major electricity supplier in South Africa's Johannesburg which caused major disruption in the electricity supply for some resident areas leaving these without electricity [58]. The attackers used ransomware to encrypt the files and computer systems of the utility, which affected the ability of the customers to buy pre-paid electrical energy and later hindered with the response towards localised blackouts.

All of these attacks target for grid/service failure which in turn hampers the grid infrastructure by causing temporary or permanent damage to the grid assets. These attacks primarily exploited the open and existing network vulnerabilities to target the electrical infrastructure in the power sector. The scale of these attacks will only increase with time (more so where all the entities are connected over the Internet), however, using the data-driven techniques, these attacks can be mitigated to a great extent.

B. ENERGY THEFT MODEL

Energy theft in the context of the smart grid can be abstracted using various generalised approaches. We indicate ways in which data-driven energy theft can be modeled from the perspective of manipulating generation, supply and demand data respectively. The proposed approaches rely on the notation denoted in Table 1.

TABLE 1: Energy theft model notation.

Ec	Demand node energy consumption
Er	Prosumer node energy generation
Es	Energy supply by T&D control nodes
NTL	Cumulative non-technical energy loss
TL	Technical energy loss
G	T&D grid
S	Number of grid supply nodes
M	Number of energy distribution buses
N	Number of total nodes
P	Number of prosumer nodes
Q	Number of consumer nodes
α	Theft coefficient on generation data
β	Theft coefficient on supply data
γ	Theft coefficient on demand data

As depicted in Fig. 4, we consider a grid G in a NAN to be defined by a set of N connected nodes and M connecting energy buses. A node is indicated as a prosumer node if it has a local DRES; otherwise, the node is indicated as a demand node.

Let $TL_j(t)$ denote technical energy losses caused by wires and equipment resistance under the normal, theft-free condition in the j^{th} bus, where $j \in M$. We also consider the cumulative non-technical energy loss, NTL in G expressed as:

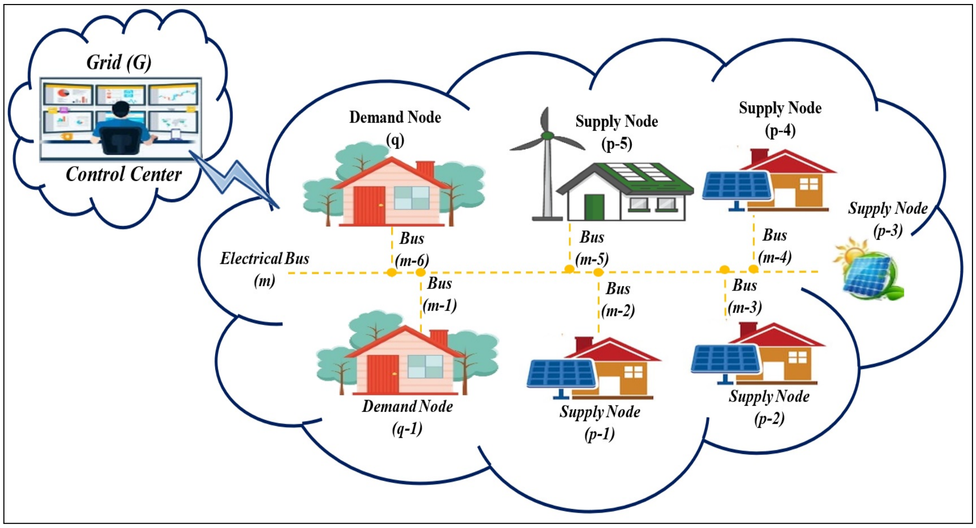

FIGURE 4: Energy grid model consisting of supply and demand nodes.

$$NTL(t) = \sum_{i=1}^{S} Es_i(t) + \sum_{k=1}^{P} Er_k(t) - \left(\sum_{h=1}^{Q} Ec_h(t) + \sum_{j=1}^{M} TL_j(t) \right) \tag{1}$$

The first two terms in equation 1 denote the total energy supplied via $i \in S$ T&D supply nodes in G and the total energy generated by $k \in P$ prosumer nodes where $P \subset S$. The latter two terms are effectively subtracted by the former two since they refer to total energy consumed by $h \in Q$ consumer nodes where $Q \subset N$ and the aggregation of technical losses caused by energy transmission over $j \in M$ buses respectively. The range of values for the final term in equation 1 is normally between 5% and 8% of the transmitted energy from the T&D infrastructure [4], [59].

1) Generation data-oriented theft

We consider that various data manipulation attacks may be conducted on DRES generation data [60] and [10] and two-metering end-user deployments [20] on the prosumer site. Alongside the inability to accurately predict weather fluctuations affecting energy generation, we abstract the total electrical energy injected to the power grid by the $k \in P$ supply nodes during an energy theft attempt to be:

$$Er(t) = \sum_{k=1}^{P} \alpha_k Er_k(t) \tag{2}$$

where $\alpha_k(t)$ is the theft coefficient for each supply node and two outcomes for this coefficient are possible being:

$$\begin{cases} \alpha_k(t) > 1, & malicious\ prosumers \\ \alpha_k(t) = 1, & honest\ prosumers \end{cases}$$

Each supply node $k \in P$ has a theft coefficient α at time t. In the legitimate case where no attack is present, the theft coefficient $\alpha_k(t)$ equals 1; meaning that there are no discrepancies in the DRES generation measurement at node k, since $Er_k(t) = \alpha_k Er_k(t)$. However, in the generation data-oriented theft scenarios, the DRES generation measurements entailed within $Er_k(t)$ are scaled by an attacker based on an arbitrarily selected percentage, represented by $\alpha_k(t)$. For instance, the attacker in such a scenario may report 200% of the actual measurements when $\alpha_k(t) = 2$. Hence, we abstract malicious prosumers that report falsified metering for their DRES generation process. Consequently, the non-technical energy loss, NTL, will be greater than that for the normal case (i.e. equation 1); since $\sum_{k=1}^{P} \alpha_k Er_k(t) > \sum_{k=1}^{P} Er_i(t)$.

2) Supply data-oriented theft

Let assume the generalised direct current (DC) model described in [61], [62] such as the energy supply in our grid G by S supply nodes to be defined as:

$$\sum_{i=1}^{S} Es_i(t) = \mathbf{J} \left(\sum_{j=1}^{M} \theta_j(t) \right) + \sum_{i=1}^{S} e_i(t) \tag{3}$$

where $\mathbf{J} \left(\sum_{j=1}^{M} \theta_j(t) \right)$ are the state variables composed of the voltages phase angles within a Jacobian matrix $\mathbf{J}$ and $\sum_{i=1}^{S} e_i(t)$ is the measurement error from supply nodes assumed to adhere to Gaussian noise e.

In data-driven energy theft, malicious actors normally manipulate a subset of measurement data to alter metering. Hence, the aggregation of energy supply Es from all supply nodes can be defined as:

$$\sum_{i=1}^{S} Es_i(t) = \mathbf{J} \left(\sum_{j=1}^{M} \theta_j(t) \right) + \sum_{i=1}^{S} e_i(t) + \beta_i(t) \tag{4}$$

where $\beta_i(t)$ is a vector representing maliciously injected data within the legitimate measurements captured by a given T&D control center. Essentially, $\beta_i(t)$ can be mapped as a False Data Injection (FDI) attack instrumented at various levels (e.g., communication protocol, metering protocol etc.).

3) Demand data-oriented theft

Consumers or prosumers are also capable to lie on their demand data by utilising FDI techniques to cause under or over-reporting of energy consumption [23], [63]–[66]. We denote as $\gamma_i(t)$ to be the theft coefficient of node i at time t and O to be the set of consumers or prosumers providing falsified demand request data, where $O = P \cap Q$. Considering a demand data-oriented theft the non-technical loss NTL can be represented as [67]:

$$NTL = \sum_{i=1}^{O} \gamma_i Ec_i(t) \tag{5}$$

In this case, the NTL should be greater than that for the normal case; since $\sum_{i=1}^{O} \gamma_i Ec_i(t) < \sum_{i=1}^{Q} Ec_i(t)$. Hence, the two possibilities for $\gamma_i(t)$ would be:

$$\begin{cases} 0 \leq \gamma_i(t) < 1, & malicious\ consumer/prosumer \\ \gamma_i(t) = 1, & honest\ consumer/prosumer \end{cases}$$

In more detail, each consumer/prosumer $i \in O$ has a theft coefficient γ at time t. In the legitimate case assuming no attack enabling energy theft, there are no discrepancies in the demand measurements denoted by $Ec_i(t)$, since the relative theft coefficients $\gamma_i(t) = 1$ and $Ec_i(t) = \gamma_i(t)Ec_i(t)$. However, in the demand data-oriented theft, the attacker manipulates the demand measurement signal Ec_i at time t by enforcing an arbitrarily selected percentage entailed within $\gamma_i(t)$. Therefore, the attacker under reports demand measurements and just reports a small portion of measurements on a regular basis. For instance, an attacker could potentially report 50% of the actual demand data, when $\gamma = 0.5$.

C. ENERGY THEFT STRATEGIES

Data-driven energy theft is orchestrated either through targeted or random methods [17], [68]. Targeted energy theft refers to instances where a malicious actor has full awareness of the vulnerability spectrum for a given system consisting of a set of nodes (e.g., DRES deployment) and purposely injects data such as to compromise its operation. Random methods usually refer to scenarios where a malicious actor disturbs the operation of individual nodes (e.g., a single DRES) by randomly flooding the application protocol dealing with metering data or by injecting corrupted measurement values while a node communicates with a centralised monitoring component (e.g., a SCADA system). In general, energy theft triggered by random methods is detected with higher precision [69].

Both targeted or random methods for energy theft may be triggered by a number of cyber-physical attack techniques. The most common technique employed in the context of energy theft is the combination of man in the middle (MITM) with false data injection (FDI) [70]. These attempts refer to cases where an individual with malicious intent intercepts and redirects communication traffic between a smart meter and an energy monitoring entity (e.g., SCADA instance in a NAN) to its own hardware. Traffic is redirected to the malicious actor such as to modify legitimate measurements and further inject falsifying metering information and retransmit it to the monitoring component in order to affect the energy billing process. Regardless of the attack scenario underpinning energy theft, there are always some necessary steps to be undertaken by a malicious actor. Fig. 5 briefly provides some core steps that are frequently practised.

We highlight four steps that in many cases are used concurrently in a given attack; i) reconnaissance, ii) scanning, iii) exploitation, and iv) access. Hence, there exists a number of variations of how the aforementioned synergistic use of MITM and FDI can be instrumented [57], [70], [71]. For instance, malicious actors could intercept general traffic at specific data recording entities (e.g., microgrid backend server) that they were aware of due to either scanning or reconnaissance such as to jeopardise the final data writing process with crafted, falsified measurements [8]. Other examples, include a combination of physical tampering of meters at various power grid levels (e.g., T&D, end-user smart-meters) where an attacker could identify through simple social engineering and bypassing of authentication protocols through ANSI optical ports with software such as Terminator that enables access [20]. In parallel, sophisticated MITM and FDI techniques may also consider the overall topology of a given grid deployment [72] in order to bypass any detection mechanisms whereas other utilise adversarial machine learning in order to game optimisation, scheduling and control processes within the EMS [17], [23]. The aforementioned technique is relatively new and exploits the deficiencies of automated management functions by manipulating and crafting falsified training data to machine learning-based processes that profile several measurements (e.g., ramp rate, power factor, reactive power) [23].

Given the diversity of the cyber-physical attack vectors enabling energy theft [12], [24], this work organises the various attack strategies based on their instrumentation and further impact in Table 2. As depicted, there has been a large volume in literature identifying, studying and further demonstrating that such attacks can be initiated at various aggregation levels by utilising different types of resources (e.g., SCADA, PV panels). Interestingly different types of attacks affect explicit grid efficiency pillars that we introduced in Section II.

IV. DATA-DRIVEN DETECTION METHODS

As briefly discussed in Section III, energy theft can be *data-agnostic* and resulted purely from physical tampering of various grid components, or *data-driven* via manipulating, destroying or corrupting software processes with the goal to modify any data related to energy demand, generation or consumption. Throughout the years, both the industry and

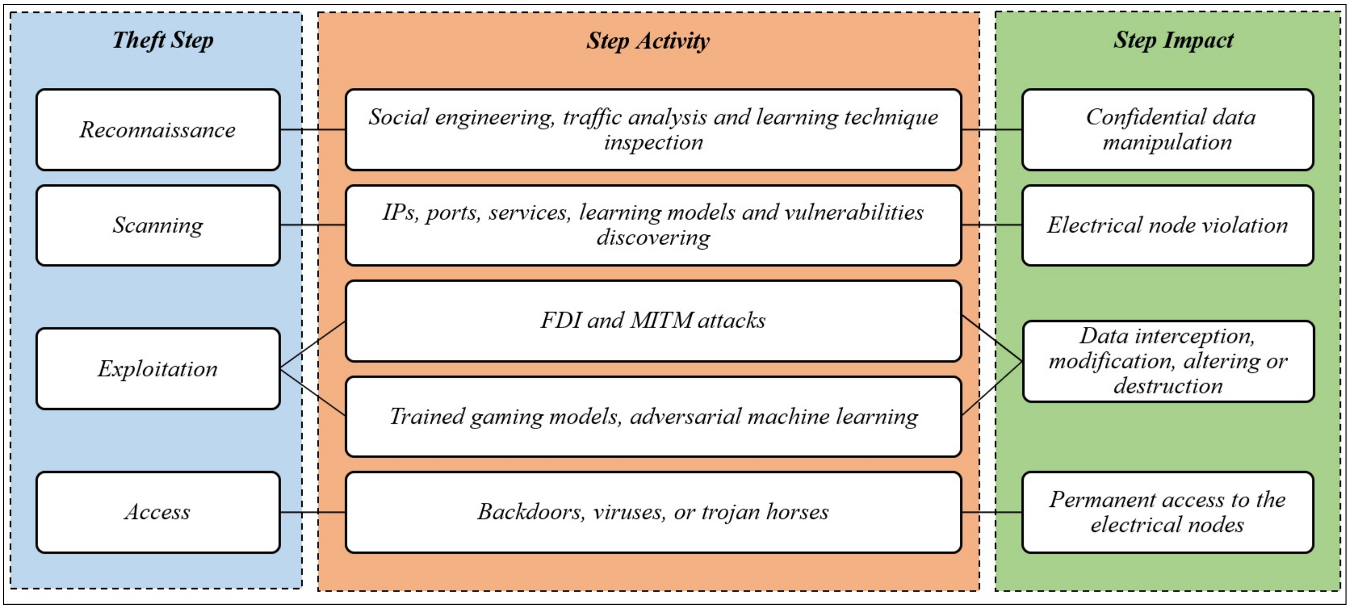
Theft Step
Step Activity
Step Impact
Reconnaissance
Social engineering, traffic analysis and learning technique inspection
Confidential data manipulation
Scanning
IPs, ports, services, learning models and vulnerabilities discovering
Electrical node violation
Exploitation
FDI and MITM attacks
Trained gaming models, adversarial machine learning
Data interception, modification, altering or destruction
Access
Backdoors, viruses, or trojan horses
Permanent access to the electrical nodes

FIGURE 5: Steps and associated activities in cyber-physical attacks enabling energy theft.

the research community have developed and employed techniques in aiming to detect any energy theft-related activities. In general, energy theft detection methods are structured under two main categories; i) hardware-based detection and ii) data-driven detection. Since the focus of this work is on the data-driven aspects of energy theft, only the latter category is discussed in this section.

Generally, the data-driven energy theft detection is achieved through the algorithmic solution composition focusing on deviations of data related to aspects such as metering and billing. Hence, such detection schemes place a strong emphasis on analysing data patterns through a variety of statistical tools and the majority utilises machine learning techniques. As depicted in Fig. 6, this work stratifies and discusses data-driven energy theft detection with respect to three main categories; i) classification-based, ii) regression-based and, iii) clustering-based detection.

Given the diversity of theft scenarios and associated attack vectors over different data aggregation levels on the smart grid infrastructure, detection methods have been employed either at a centralised or a distributed fashion. Table 3 provides a comprehensive summary of methods introduced in past literature over the last decade. Evidently, the majority of methods consider a combinatorial use of algorithmic techniques in order to address specific challenges ranging from data pre-processing and filtering up to statistical correlation analysis. Furthermore, some formulations are broadly used (e.g., artificial neural networks -ANNs and support vector machines - SVMs) over different types of attacks operating under diverse data types gathered at various smart grid data aggregation components.

Complementary, Table 4 illustrates the experimental approach underpinning the methods summarised in Table 3 and further provides their outcomes. As depicted, each method was employed over energy theft use cases involving a number of nodes within the actual grid and utilised specific statistical features. In summary, we identify a range of raw as well as post-processing features that are utilised within the listed methods. Thus, there exist techniques involving one or more of basic statistical features (e.g., mean, min/max), frequency and temporal domain features (e.g., signal periodicity frequency components), scaling on independently distributed raw data, clustering or probability-based similarity metrics as well as locality (e.g, geolocation coordinates), auxiliary (e.g., number of energy appliances) and environmental features (e.g., temperature, humidity).

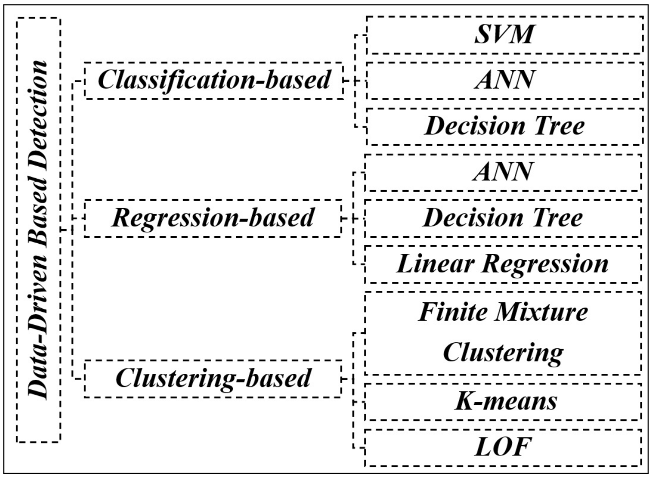

FIGURE 6: Data-driven energy theft detection categories.

TABLE 2: Overview of the data-driven energy theft attacks.

Ref.	Strategies	Infrastructure	Resource	Attack Effect	Remarks
[60], [73]	Generation meter manipulation	Generation	PVs	Grid reliability	Introduces physical attack functions applied to inject energy into PV power systems.
[20]	Generation meter manipulation	Generation	PVs	Grid reliability	Introduces attack functions applied to manipulate the reported energy generation profile of PV power systems.
[10]	Generation meter manipulation	Generation	PVs Wind turbines	Grid reliability	Introduces stealthy adversary model initiated by generation meters managed by prosumers.
[25]	Generation meter manipulation	Generation	PVs Wind turbines	Grid reliability	Assumes attacks manipulating the average of net generation while the detection mechanism is perceptible.
[31]	Monitoring and control systems manipulation	T&D	PMU	Grid resilience	Summarizes different methods applied to commit data-driven theft against the grid measurements through WAMS manipulation.
[74]	Monitoring and control systems manipulation	T&D	PMU PDC	Grid resilience	Assumes attackers compromise one or more of the PMUs, PDCs, communication links or/and routers.
[75]	Monitoring and control systems manipulation	T&D	SCADA	Grid resilience	Makes various assumptions about the attacks in the context of the current security mechanisms in SCADA networks.
[76]	Monitoring and control systems manipulation	T&D	PMU	Grid resilience	Assumes the attacker has access to only the PMU measurements at buses where the PMU has been compromised.
[77]	Monitoring and control systems manipulation	T&D	SCADA PMU	Grid resilience	Assumes the attacker only compromises a single state variable. The attacker alters all the measurements to project the desired changed state variable.
[72]	Monitoring and control systems manipulation	T&D	SCADA	Grid resilience	Assumes the attacker can access several SCADA's sensors to compromise several measurements.
[78]	Monitoring and control systems manipulation	T&D	SCADA	Grid resilience	Introduces a more realistic attack where the attackers have only inaccurate and incomplete information because of their restricted access to the grid.
[79]	Consumption meter manipulation	End-user	Smart meter	User safety	Introduces data-driven attacks enabling time-variant modifications on load profiles of the end users.
[67]	Consumption meter manipulation	End-user	Smart meter	User safety	Generates and labels real-time attack patterns for use with supervised detection algorithms.
[63]	Consumption meter manipulation	End-user	Smart meter	User safety	Models the energy loss resulting from meter manipulating, meter malfunctioning, and y illegal bypassing.
[64]	Consumption data manipulation	End-user	Smart meter	User safety	Introduces theft attacks based on the manipulation of the smart meter, AMI, appliances load profiles, and withdrawing heavy appliances from the actual measurements.
[65]	Consumption meter manipulation	End-user	Smart meter	User safety	Presents theft attack assuming the customer has DRES.
[66]	Consumption meter manipulation	End-user	Smart meter	User safety	Introduces a theft attack designed by a fraudulent employee who fabricates the consumption measurements based on the past readings, instead of reading the actual measurements from the smart meter.

1) Classification-based detection

Messinis *et al.* [80] proposed a classification system to detect energy theft conducted at the end-user infrastructure. The introduced solution was assessed over simulations replaying the Irish Smart Energy Trail dataset and its operation relied on the synergistic use of an SVM classifier, a power optimization scheme and a voltage sensitivity analysis component. In practise, the SVM classifier was producing a weight function based on the annual active energy consumption for a consumer that was expressed as the probability for committed fraud. The proposed system achieved a high accuracy of 99.4%. However, this system required the utilization of additional features such as voltage and active energy data to detect theft. The problem with utilizing such sensitive measures is that it can expose customer data to privacy violations. Moreover, features associated to real-time ancillary services (e.g., active/reactive energy require adequate signal smoothing techniques for complete conversion over the time-frequency domain; an element missing from this piece of work as it is not encapsulated within SVM formulations or the proposed pre-processing stage. Thus, we argue that such methods may not be generic enough.

A synergistic use of SVMs and decision trees for theft detection in end-user infrastructure was proposed by Jindal *et al.* [4]. Decision-tree formulation operates on various features, including the numbers of heavy appliances and persons, to generate the predicted consumption of each consumer. Then, an SVM-based classifier is used to detect malicious consumers. Results show that the proposed method can be implemented in real-time scenarios as the false positive rate is significantly reduced to 5.12%. A similar combination was adopted by Althobaiti *et al.* in [10] to detect malicious prosumers in generation infrastructure. To rigorously analyze DRES generation data, an XGBoost and SVM were combined for the proposed method. An XGBoost algorithm was run on freely available weather data and used to calculate the energy generated by the DRES, and SVM was used for measurement classification. The results show improved accuracy of 98% for theft detection. However, applying the proposed methods to a large-scale theft detection process remains limited due to the computational complexity resulting from the synergistic use of multiple data-driven algorithms in such detection methods.

Variations of the conventional SVM formulation in syn-

TABLE 3: Overview of the data-driven energy theft detection methods.

Ref.	Technique	Nature		Attack Infrastructure	Attack Type	Data Type
		Centred	Distributed			
[80]	SVM, Voltage Sensitivity Analysis, Breakout Detection Package	√		End-user	Demand data manipulation	Consumption
[4]	Decision Tree, SVM		√	T&D End-user	Attacks caused NTL	Consumption
[10]	XGBoost, SVM		√	Generation	Generation data manipulation	PV and wind turbine measurements
[62]	SVM, Density based anomaly detection, PCA		√	T&D	SCADA data manipulation	Network measurements
[1]	Convolutional ANN, Paillier Algorithm, SVM, Random Forest, Logistic Regression	√		End-user	Demand data manipulation	Consumption
[81]	Wide & Deep Convolutional ANN, Three-Sigma Rule, Random Forest, Convolutional ANN, SVM, Logistic Regression	√		End-user	Demand data manipulation	Consumption
[82]	DL, Generalized Linear Modeling, Random Forest, GBM	√		T&D	SCADA data manipulation	Network measurements
[20]	Deep Feed Forward ANN, Deep Recurrent ANN, Deep Convolutional Recurrent ANN, SVM, ARIMA		√	Generation	Generation data manipulation	PV measurements
[83]	OPF, SVM, Bayesian Classifier, Logistic Regression	√		T&D End-user	Demand data manipulation (Direct tapping)	Consumption
[84]	Random Forest, Logistic Regression, SVM, K-means	√		T&D End-user	Demand data manipulation (Direct tapping)	Consumption
[85]	SVM, KNN, Random Forest, Logistic Regression	√		T&D End-user	Demand data manipulation (Direct tapping)	Consumption
[7]	Logistic Regression, KNN, Fourier Transform, Random Forest		√	T&D End-user	Demand data manipulation (Direct tapping)	Consumption
[86]	XGBoost, K-means, KNN, SVM, Logistic Regression	√		Generation T&D End-user	Attacks caused NTL	Consumption
[67]	XGBoost, CatBoost, LightGBM	√		End-user	Demand data manipulation	Consumption
[87]	MLP, RNN, LSTM, GRU, Simple Moving Average		√	T&D End-user	Demand data manipulation (Direct tapping)	Home appliances data
[65]	Linear regression, SVR, ANN, Radial Basis Function Network	√		End-user	Demand data manipulation	Consumption
[88]	Decision Tree	√		T&D End-user	Demand data manipulation (Direct tapping)	Consumption
[89]	Random Forest, Decision Tree		√	T&D End-user	Demand data manipulation (Direct tapping)	Consumption
[90]	SVM, K-means	√		End-user	Demand data manipulation	Consumption
[91]	Finite Mixture Clustering, Genetic Programming, ANN, Random Forest, SVM, KNN, GBM	√		End-user	Demand data manipulation	Consumption
[92]	Local Outlier Factor, KNN, Maximal Information Coefficient, Clustering by Fast Search and Find of Density Peaks	√		End-user	Demand data manipulation	Consumption

ergy with principal component analysis (PCA) was also the basis behind the work of Esmalifalak *et al.* in [62]. The evaluation of SVM-based formulations was based on labelling load data that were simulated as stochastic processes such as to comply with pragmatic power system behaviour in the T&D system infrastructure. PCA was initially employed in order to reduce the high dimensionality of the simulated measurements and they were firstly labelled within the training process of a supervised SVM formulation. Subsequently, newly generated measurements were tested over the supervised model and the identification of outliers implying theft detection was feasible with 95% accuracy. However, due to the dependence of the proposed scheme on PCA, there exists a high likelihood of a trade-off between the loss of important information included in the simulated measurements and the dimensionality reduction process.

Recent developments in the area of deep learning (DL) enabled the composition of adequate energy theft detection schemes. Yao *et al.* in [1] demonstrated a novel synergy of convolutional neural networks (CNN) and the Paillier cryptosystem in order to maintain user privacy but also detect energy theft. Under a similar mindset, a modified wide and deep CNN was proposed in [81] in which the wide component of the customised CNN deals with global consumption features whereas the deep CNN component was more focused on profiling the consumer's consumption periodicity such as to detect deviations implying energy theft at end-user level. The superiority of DL-based energy theft detectors was also illustrated at the work in [82] where a number of traditional and ensemble classifiers such as random forests, and gradient boosting machines (GBM) were compared with a CNN-based classifier using T&D infrastructure measurements. Similarly, the work by Ismail *et al.* in [20] demonstrates the applicability of a DL-based detection solution based on measurements that are captured at DRES deployments. However, such theft detection methods entail enormous computational costs due to the large amount of data required to effectively train fully supervised DL-based detectors.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3131220, IEEE Access

TABLE 4: Experimental approaches of surveyed studies on data-driven energy theft detection.

Ref.	Number of Nodes ($\approx$)	Features	Evaluation Metrics (Best algorithm) (%)	Experimental Evaluation		Percent of Attacked Samples ($\approx$) (%)
				Simulation	Testbed	
[80]	$5K$	Statistical, Auxiliary, Scaling, Frequency	ACC = 99.4, FPR = 0 TPR = 98.9, AUC = 99.9	√	√	50
[4]	$1k$	Scaling, Auxiliary, Environmental, Temporal	ACC = 92.5, FPR = 5.12	√		20
[10]	300	Environmental	ACC = 98, AUC = 99.6		√	—
[62]	$1k$	Similarity, Auxiliary	F-score = 95	√		—
[1]	$42k$	Similarity	ACC = 92.67		√	—
[81]	$42k$	Statistical, Scaling	AUC = 96.86		√	9
[82]	$100k$	Auxiliary	ACC = 97.7, F-score = 98.78 AUC = 98.53	√		—
[20]	71	Auxiliary	ACC = 99.3, FPR = 0.22 F-score = 99.55	√		—
[83]	$42k$	Statistical, Auxiliary	ACC = 83, F-score = 80.9		√	—
[84]	$3.5M$	Similarity, Temporal, Locality, Auxiliary	AUC = 75.03		√	$10 - 90$
[85]	$700k$	Locality, Auxiliary	AUC = 62.8		√	$1 - 90$
[7]	425	Statistical, Frequency, Scaling	ACC = 98.37, FPR = 0 F-score = 87.50		√	16
[86]	$57k$	Statistical, Similarity, Auxiliary	AUC = 91		√	$5.38 - 8.37$
[67]	$5k$	Statistical	FPR = 4, TPR = 97		√	50
[87]	1	Auxiliary	ACC = 99.96		√	—
[65]	980	Auxiliary	–		√	–
[88]	$5k$	Temporal	–		√	–
[89]	1	Auxiliary, Environmental, Temporal	ACC = 95.78, AUC = 100		√	–
[90]	$5K$	Auxiliary, Similarity	FPR = 0.1, TPR = 94		√	–
[91]	$4k$	Statistical, Similarity	ACC = 99, AUC = 99.8		√	–
[92]	$3.5k$	Statistical, Similarity	AUC = 91.84		√	12

Several studies have also provided insightful comparisons of various classification-based energy theft detection schemes and insights on the performance of particular statistical features. For instance, the work by Fernandes *et al.* in [83] introduces the use of a customised optimum path forest (OPF)-based detection scheme for attacks that target explicitly energy theft. In evaluations of industrial and end-user consumption data the proposed scheme outperformed conventional classifiers such as SVM and Bayesian classifiers with respect to detection accuracy. However, with respect to log loss function, SVM achieved the best value, outperforming the customised OPF-based detection scheme.

Meira *et al.* in [84], examine a diverse set of spatiotemporal and exogenous features based on four criteria, namely, auxiliary, similarity, locality and temporal. The performance of the selected features was investigated through the classification processes of customised SVM, logistic regression and random forest formulations. It was clearly revealed that features derived only from consumption measurements (such as similarity features) are adequate for the accurate detection of energy theft attacks. However, such a detection study entails computational processes on further features from historical consumption measurements, which limits the application of this method in large-scale detection scenarios.

In parallel, the study by Glauner *et al.* in [85] demonstrates that the classification process under various algorithms (e.g., SVMs) reveals that features related to aggregated neighbourhood consumption alongside locality parameters outperformed individual meter time series distributions. However, we argue that energy theft detection based on the utilization of features related to neighbourhood consumption and locality parameters may not be generic enough, due to the fact that the consumption patterns of those who belong to the same geographical domain differ from one another.

The assessment of features pointing to energy theft in synergy with classification performance were also one of the main focus areas in the studies conducted in [7], [86] and [67]. Through the application and comparison of classification-based ensemble methods (e.g., XGBoost, CatBoost, LightGBM) with conventional classifiers (e.g., ANNs, SVMs) over simulated attack scenarios it was revealed that ensemble methods contribute significantly towards computationally-efficient and more accurate theft detection. However, ensemble-based detection methods pose some instability since a slight variation in the training data would unavoidably entail substantial restructuring of the main tree-based detection model. Thus, imposing higher computational costs. Nonetheless, the work by Ashrafuzzaman *et al.* in [82] demonstrates the superiority of deep learning-based theft detection schemes over any ensemble-based approaches compared, where the detection accuracy based on the deep learning technique was 97.7%.

Despite the relatively high accuracy performance and reliability of classification-based techniques, the aforementioned detection methods require labelled data from malicious and energy theft-free behaviours. Obtaining such data is either challenging in a real scenario or, even if they exist, they do not cover all possible theft-attack behaviours [93]. Theft-free data can be collected from historical grid measurements, however, malicious data (i.e., theft samples) covering the spectrum of theft behaviours for a particular node hardly exist. In such cases, the performance of the detection method is limited due to malicious sample unavailability. These methods may remain unsuccessful in detecting more advanced and stealthy attacks that are not available in training data, which directly affects the overall detection performance [94].

2) Regression-based detection

M. Li *et al.* [87] proposed a modular energy theft detection system consisting of a three-stage decision making process achieving 99.96% on theft detection accuracy. The first stage relies on a multi-model power consumption prediction system based on Multi Layer Perceptron (MLP) ANN, Long Short Term Memory (LSTM) ANN, Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU). The second stage deals with monitoring a moving average whereas the third stage employs a customer's historical measurements to determine occasional maximum energy consumption in order to make a final decision on a theft attack. Although interesting results are achieved, the proposed method is undynamic for any future changes in consumption patterns, since the main focus of such a system is the utilization of historical consumption measurements in the detection process.

The behavioral profile of normal energy consumption was assessed by Cody *et al.* [88] in order to detect deviations implying energy theft. The conducted experiments revealed that consumption values can be predicted using decision tree learning and they can be categorised into normal or fraudulent based on the threshold root mean squared error value. Any value exceeding this threshold indicates a possible energy theft attack. However, the prediction formulation proposed in this study can be improved through the utilization of further comprehensive features, such as numbers of appliances and providing the prediction model with additional details to determine consumers' energy consumption patterns.

Complementary work in [89] achieves regression based on random forests to predict the expected energy consumption over the US-wide consumption profiles for 2014. Through the use of various performance metrics (e.g., prediction accuracy, classification error rate) forecasting through random forests achieved 95.78% of prediction accuracy and outperformed a decision tree-based approach that reached 91.6% accuracy. Thus, providing a quite effective energy theft prediction scheme. However, such a scheme cannot be considered as generic since energy consumption is usually characterized by invariable variance or non-stationary behaviour. Therefore, the fundamental principles underpinning random forests model could become inappropriate for identifying short-term irregularities in energy consumption.

A data-driven regression model was proposed by Y, Gao *et al.* [65] for energy theft detection. Instead of using unreliable topology information and parameters from secondary network, this method was based on modified linear regression algorithm. It uses only the voltage data and consumer's consumption data making it more feasible to adopt. Finally, the training data from real world smart meter was used to validate proposed method and results illustrate effective identification of cases related to energy theft. However, customers' data may be vulnerable to privacy breaches due to the dependence upon voltage measurements.

Overall, despite the applicability of the aforementioned methods to identify advanced energy-theft attacks, regression-based methods regularly demonstrate longer detection times than other detection categories. In such cases, regression techniques are principally employed in the first stage of theft-detection methods and require additional procedures to reach a final decision during the detection process. This in turn is a time-consuming task and limits the applicability of such methods in a real-time energy trading scenario, where the time required to detect theft activities is influential in preventing any losses.

3) Clustering-based detection

A clustering-based theft detector utilising consumption patterns was also proposed by Jokar *et al.* [90]. In order to improve classification accuracy, the number of clusters in the examined dataset was filtered through Silhouette plots and subsequently clusters were hierarchically labelled across various consumption profiles. The resulted outcomes of this approach demonstrate that even with low measurement sampling intervals, the algorithm is scalable and achieves a detection rate of 94%. However, the proposed technique required the installation of transformer meters, which increased the monetary cost of such systems.

An alternative approach based on genetic algorithms and finite mixture modeling for composing clusters of consumption in order to identify customer segmentation and potential outliers was presented by Razavi *et al.* in [91]. In fact, the proposed method outperforms a number of classification-based approaches such as k-nearest neighbours (KNN), ANN and SVM by 99.8% in the area under the curve for theft detection. However, such a detection system cannot be applied in a real-time scenario, since the results achieved indicate that there is an increase in the relative to physical inspection.

An outlier-based detector of three modules was presented by Peng *et al.* in [92]. The proposed method applied local outlier factor (LOF) and the KNN algorithm as the basis to detect theft at the end-user infrastructure. Firstly, consumption profiles were analysed with k-means and subsequently outlier candidates were selected based on the deviation of each consumer from the relative cluster centers. Finally, the anomaly ranking of the selected candidates was calculated using the LOF algorithm. Although the proposed detector achieves reasonably high detection accuracy of 91.84%, it still fails to detect linear theft, where an attacker manipulates the consumption profile to reduce it at a constant rate.

Despite the fact that clustering-based methods can be used in scenarios of scarcity, minimal or zero availability of malicious intent, these methods will normally produce an end result with a high false-positive rate. To construct a clustering-based model, no assumptions of labelled data from malicious and theft-free behaviours are made. As a result, the detection model can identify any abnormal patterns as malicious behaviours [93]. In general, abnormalities may occur due to non-malicious activities (e.g., smart-meter misconfigurations), leading to an increase of false-positive rates resulted by clustering-based theft detection mechanisms.

4) Comprehensive analysis

Undoubtedly, malicious actors continue to target a diverse set of vulnerabilities present over various system, network and algorithmic components serving the (sub)infrastructures composing a smart grid deployment. Hence, attackers intend to launch energy theft attacks through a variety of techniques that target the evasion from current detection schemes. Evidently, data-driven methods for detecting energy theft distilled by learning, profiling and detecting abnormalities are considered as a means to adaptively engage with new attack vectors.

In general, data-driven energy theft detection schemes leverage three conceptual and data-driven procedures; (i) data-processing and model-selection stages covering aspects of data sanitisation and feature selection, (ii) model-training procedure which varies across classification, clustering and regression detection methods and (iii) decision-making procedure which includes applying a model trained on new data such as to pinpoint anomalies that could relate with malicious activity.

Given the "ad-hoc" employment of most of the detection methods presented herein over specific use cases, we argue that there is no universal data-driven methodology covering all aggregation levels in a given smart grid deployment. In general, the aforementioned three levels, categorized into energy generation, T&D, and end-user infrastructures have different probabilities for the deployment of theft and different vulnerabilities exploited by malicious actors. Such factors should be taken into consideration when a method is designed to detect energy-theft attacks.

However, the utilization of a hybrid data-driven model has proven to be more robust than adopting a single model in detecting attack vectors underpinning energy theft. Such hybrid methods are considered to make combinatorial use of two or more data-driven models. In such methods, the entire theft detection method leverages the analytic process of each candidate model to achieve a specific action. All achieved actions are subsequently integrated into one detection system in order to complement each other and mitigate the limitations of the others.

Furthermore, the utilization of data from multiple and diverse sources can create a more reliable method for detecting energy-theft attacks over smart-grid infrastructures. Detection methods utilizing a single data source are constrained to build a candidate model fitting specific data measurements, thus its suitability is not generic. Moreover, the candidate model is sensitive to the samples it was trained with, which may potentially have been manipulated to falsify the detection method to cope with new adversarial objectives. However, by acquiring the data from various sources which have less likelihood to be accessible to adversaries can significantly increase the reliability and performance of the detection method.

The adoption of data-driven methods that utilise multiple and diverse data feeds would unavoidably invoke trade-offs spanning across performance, privacy and computational complexity. For instance, data-driven theft detection at the end-user infrastructure method would require a privacy-aware data processing and aggregation scheme. Hence, in order to detect theft in DRES infrastructure, the detection method should not rely on data that are not available to utility providers such as EMS measurements. Such measurements are usually maintained by the DRES owner and not accessible to any third party. Thus, there could be some limitations in terms of the granularity of the anomaly detection process employed by the theft detection scheme. On the other hand, energy theft-detection process in the T&D infrastructure inherently requires the utilization of high volume of network and system log measurements. Therefore, an anticipated high computational cost would be implied and thus limit the real-time capabilities of a given theft detection scheme.

V. PRESENT GAPS AND FUTURE DIRECTIONS

Despite the various solutions proposed in terms of energy theft detection, there exist various gaps and open issues thus requiring further attention within future research directions. Within this section, we highlight and discuss some of the challenges and we further summarize potential future research directions. As depicted in in Fig. 7 we decompose the gaps spectrum into (i) measurement-driven , (ii) machine learning and (iii) security-related challenges.

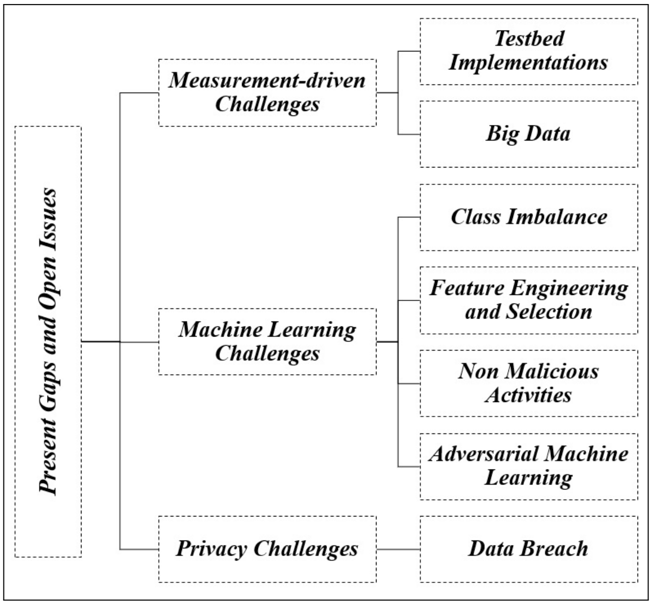
Present Gaps and Open Issues
Measurement-driven Challenges
Testbed Implementations
Big Data
Machine Learning Challenges
Class Imbalance
Feature Engineering and Selection
Non Malicious Activities
Adversarial Machine Learning
Privacy Challenges
Data Breach

FIGURE 7: Present gaps in energy theft detection.

A. MEASUREMENT-DRIVEN CHALLENGES

a: Testbed scenarios and datasets

Diverse energy-related data sets, different network infrastructures, and multi-faceted energy theft-related attacks are studied in most of the presented works as discussed in this study.

However, there is a notable lack of commonly available (and applied) prototype implementation on realistic large-scale testbed as well as datasets such as to enable extensive experimental verification nor experimental reproducibility tailored for energy theft detection [93]. Most testbeds and their corresponding datasets are principally designed in an ad-hoc fashion for specific projects limiting the generalisation of findings [95]. Therefore, we argue that it is of crucial importance to build benchmark testbeds and properly designed platforms such as to test connections and security features of a system and maintain alignment with the pragmatic and rapidly emerging design requirements of current and future smart grid deployments.

b: Measurements & Big data

Volume, velocity, and variety are the traditional traits and they naturally challenge any analysis domain within the smartgrid ecosystem [96]. Hence, the adequate comprehension and optimisation of these diverse traits during data collection, processing and analysis over particular smartgrid scenarios such as energy theft detection is of vital importance. For instance, there are 27 million consumers that consume domestic electricity in the UK alone. These consumers have more than 100 million data points that are collected either quarterly or half-yearly. These points are used by the energy suppliers to store, record and use in the billing system and identifying abnormal conditions that could relate to specific energy theft-related attacks. However, with smart metering, to collect the data from these many data points, at a thirty-minute sampling rate, will require a substantial amount of resources. For example, at least 4500 to 9000 times more of the present data size will be required to be processed by the energy suppliers, and therefore this leads to a significant augmentation in data size [97]. Thus, there is a strong requirement for efficiently coupling the measurement requirements for granular energy monitoring with optimised storage as well as data processing solutions.

B. MACHINE LEARNING CHALLENGES

a: Class imbalance

Class imbalance problem is a traditional problem existing for supervised or semi-supervised learning having direct implications on energy theft detection. In particular, this problem occurs when one of the classes (in a multi-class problem) has significantly more number of samples than the other classes, thus the training model is biased leading the testing phase to classify events towards the majority class label [98]. Hence, in the case of learning for theft instances in which are by far less than legitimate instances, the class imbalance problem would result on a classifier to incorrectly label malicious instances to the majority of normal behaviour. It is therefore important to establish adequate ground truth datasets with correct scaling factors through the training phase of learning processes by assigning correct weight parameters to malicious samples. Thus, addressing the limitations from the class imbalance problem [99]. Nonetheless, the composition of concrete ground truth labels for theft instances is also a topic aligned with the needs of optimised feature engineering and selection as we discuss next.

b: Feature engineering and selection

Feature engineering accompanied by efficient feature selection is a powerful foundation for addressing the aforementioned class imbalance problem as well as tailoring a learning procedure to identify energy theft instances. Evidently, it is common in many energy theft detection processes to operate over insufficient or incomplete feature vectors and experience class imbalance as well as model over-fitting (i.e., learn the only specific pattern in a given dataset), thus affecting significantly detection accuracy. Therefore, designing and engineering new features can improve the performance of machine learning detection methods [99], [100].

c: Non-malicious abnormal activities

A classical problem within anomaly detection is the distinction of classes between anomalous events. Energy theft-related attacks could relate to statistical abnormalities and have similar properties as anomalous events that are caused by legitimate intent (e.g., smart meter misconfiguration). A great challenge is to compose adequate classification and clustering schemes that are able to pinpoint the differences between malicious and legitimate processes and further highlight the specific properties entailed within an energy theft incident. There can be many reasons that the ambiguities in electrical node output patterns may occur. These can happen owing to several altered causes such as new device installation (for example, a new DRES) or changing in the electricity usage habit of the residential end users [93]. This, in turn, increases the overall inspection cost [4] as once the model classifies an energy theft attack, physical inspection is essential for final verification and that is a costly procedure [90]. It can, therefore, be argued that there is a requirement for more research in the improvement of the proposed detection methods in terms of reconsigning the theft detection activities and reducing the false positive alarms [99].

d: Adversarial machine learning

As already described, it is feasible for an adversary to manipulate end-user data or game the algorithmic learning procedure in a targeted manner. These particular types of attacks are called adversarial machine learning attacks which are carried out for the purpose of theft detection. For example, carrying out an attack where input data is made to look like normal electrical data, i.e., crafting an attack that seems normal to the machine learning algorithm or changing the weights of the trained ML model. These scenarios can maximize the predicted loss or falsify trained models to new adversarial objectives [23], [101], [102]. Moreover, handcrafted rule-based attacks are more sophisticated (than automated attacks) and proposes different challenges, and therefore a generalized detection model will not provide promising results [23]. Thus, more studies are required to investigate the capabilities and the limitations of existing machine learning detection algorithms with respect to adversarial machine learning.

C. PRIVACY CHALLENGES

a: Data breach

Most of the energy theft attack detection schemes utilize (some of) the private information of consumers/prosumers, such as smart meter readings and user load/generation profiles. While this information can help to detect the theft attacks to a certain extent, it should still be kept in mind that disclosing such private data may raise concerns about the user's safety and breach his/her privacy. These data breach threats can occur in different stages of the theft detection process, including data collection, transmission and storage. Such sensitive breached information might be purchased by interested third parties such as marketing companies which can use this data to sell their products to possible customers. Apart from this, if criminals get their hands on this sensitive data, the daily routine of a household can be analyzed from electricity usage/generation pattern to carry out crimes. Therefore, detecting energy theft attacks while maintaining privacy of information is a challenging task, but there is a notable lacking of intelligent privacy-preserving detection schemes in the works of the energy theft [103].

VI. FUTURE RESEARCH DIRECTIONS

1) Measurement-driven solutions

a: Testbed simulation, emulation and hardware

Future works should consider the measurement-driven challenges that affect energy theft detection frameworks. The energy theft activities should be ratified by experimental environments and for this to happen, there is a strong need to include testbed software simulation, emulation and hardware for carrying out energy theft analysis. For instance, a cloud-based environment can be created to store smart grid data which can be used in these testbeds to conduct energy theft analysis [30]. With simulation software and emulation hardware, a quick verification of new concepts can be achieved efficiently which can then be easily transferred to power system industry and for more extensive public use. Moreover, these testbeds create interesting educational platforms to understudies which would spur the research interests to conduct multi-user experimental facilities for several smart grid applications [95].

b: Big data schemes

To collect, store, and process monitoring data various diverse data sources in smart grid results to the big data challenges as discussed earlier. To cater to these challenges the two important future directions include the creation on big data analysis platforms and reducing the complexity of such data. For the former, cloud computing technology has been used to create big data platforms by the many industries since this technology is scalable, self-organizing, and adaptive. Therefore, platforms such as Hadoop, Cassandra, and Hive in conjunction with cloud computing can be used by utility providers for smart grid big data analysis [104]. For the latter (to reduce the data complexity), different techniques such as dimensionality reduction, distributed optimization algorithms, and active learning can be useful to analyze big data efficiently [105]. Different studies reported that the computational process of the summarized and produced data rather than the original data stream can result in an acceptable relative error [106]. Therefore, these dimensionality reduction techniques are useful for reducing the communication cost, computing complexity, and storage resource utilization for smart grid big data analysis [107].

2) Machine learning solutions

a: Class Imbalance

Class imbalances happen when there are less samples in one of the target classes for machine learning algorithms or a close similarity in the number of samples in considered classes. To enhance the learning results associated with imbalanced data classes (and improve on their bias), three primary methods can be utilized: data-level, algorithm-level and hybrid techniques [108]. In the data-level techniques, the concentration is on the modification of training set to allow more balanced distributions for oversampling (more minority groups' samples) and undersampling (fewer majority groups' samples). The algorithm-level techniques modify the learners that already exist to eliminate their bias for majority groups. However, good insight is required into the modified learning algorithm and real discovery of reason for skewed mining distributions. Some popular algorithmic techniques include cost-sensitive approach (to insert different penalties for every group of samples) and one-class learning (concentrating on the specific target groups). The hybrid techniques use the combination of methods as mentioned above, by reducing their weaknesses and making use of their strengths [108].

b: Feature engineering and selection schemes

We argue that future research directions could place stronger focus on particularly exploring algorithmic and system-wide principles to facilitate automated feature engineering and selection methods. The feature engineering process can extend the original detection model's feature vector by adding new features that are calculated based on other input features. These engineered features may be the differences, averages, or other statistical transformations of the original feature vector, helping in better understanding of the interactions amongst these features. This process is similar to the statistical transformations performed by human analysts for constructing an engineered feature formulas. The task of feature engineering and selection is mainly a time-consuming task and each model type will respond in different manner to different engineered feature types [109]. However, in general, the selected and engineered featured would help in achieving the maximum probability of success for the ML algorithms to detect energy theft [110]. Typically for feature engineering and selection, many methods can be used such as mathematical functions, deep feature synthesis components, expansion reduction, evolution-centric, multi layer neural networks and hyper parameter optimization [109], [111].

c: False positive rate-reduction schemes

A meta-learning scheme can be helpful to reduce the false positive rates resulting from non-malicious activities in the process of energy-theft attack detection. Meta-learning can be defined as a learning process involving the collection of knowledge from past experience in order to use it in future learning [112]. Meta-learning is required by the theft-attack detection system to combine various classifiers (by taking note of their behaviours) and adopting an integration rule to reduce false positives. In the literature, the main meta-learning techniques include stacking, bagging, voting and boosting. In the voting approach, each classifier has one vote, and the classification that has the highest votes determines the final prediction. In stacking learning, the process adopts a layered architecture wherein each layer has one or more classification techniques. A layer's projection is applied to extend the original vector of the feature with the closest instance. The bagging approach creates a combination of classifiers through the manipulation of training samples in a base classifier. It selects one base classifier and invokes it many times using several training samples. Boosting, in contrast to bagging learning, generates various basic classifiers through a procedure in which examples of data sets receive new weights in sequence [113].

d: Adversarial machine learning schemes

With respect to adversarial machine learning, a binary classifier-based intrusion detection system trained on available device behaviour logs is imperative [23]. This system can attempt to tag approaching instances as either malicious or benign, using features which are generated in real-time from streams of energy data. Through gradual training instances expansion and feature generation refinement, this system can produce a confidence score that can be utilized to set recall/precision. This will allow having low maintenance overheads and fewer false alerts as compared to a manual system. The underlying intrusion detection system can employ a broader range of features including outgoing data from the control algorithm [23]. As also discussed in [23], [24] malicious behaviours can be detected using other associated features such as network properties (e.g. packet size, packet arrival time) and communication security (e.g. certificate fingerprints, negotiated cyber suite).

3) Privacy preserving schemes

Privacy-preserving schemes can be used in two ways to detect energy theft attacks; one, focusing on protecting the identities of users, and the other, emphasising protecting the data of users [114]. For the first aspect, pseudonym, anonymization, and virtual ring have been used. Pseudonym is considered to be a common user identity protection approach. The registration process for a pseudonym often involves many data protection methods, such as ring signature and zero-knowledge proof [114]. Anonymizing smart-grid data is one of the methods approved by the National Institute of Standards and Technology [115]. The main goal of anonymization is to enable smart grids' nodes to communicate in an anonymous manner with various smart-grid service providers by using different pseudonyms. Another common method for user-identity preservation is a virtual ring, where a ring signature is used to validate the identity of users, without knowing their actual identity, by a control centre [116]. On the other hand, for the second aspect, emphasising protecting users' data, many methods can be used, such as data aggregation or authentication methods. Data aggregation is a well-known scheme which is used to protect the data of smart-grid users. It generally includes data obfuscation algorithms and homomorphic encryption [103], [114]. Authentication methods are efficient countermeasures for privacy-related attacks and are usually based on key public infrastructure [117].

VII. CONCLUSION

Smart power grids aim towards resilient, reliable and sustainable operation of legacy power systems and also the integration of smart business models for the optimised use of energy by consumers. Nonetheless, their complex system architecture in which diverse and heterogeneous infrastructures interconnect, facilitates the basis for a number of attacks that enable energy theft. Energy theft attacks affect critical grid processes and facilitate financial gain for malicious actors. To present the overall overview of such actors and their energy theft activities, we conduct a through study of data-driven energy theft attack and detection techniques in this paper for smart grid systems. In this regard, we firstly present the smart grid components in the energy supply chain with a focus on their data communication along with the pillars to access grid effectiveness. The impact of energy theft in the smart grid is then discussed by critically assessing how energy theft can be formulated by manipulating demand, supply, and generation data. The data-driven energy theft attack examples are then discussed along with their enabling activities. Furthermore, we categorize extensive studies addressing the data-driven aspect of energy theft detection and summarizing the experimental approaches for such studies. Lastly, we highlight various open issues and challenges still persisting in the area of energy theft detection. We summarise and further indicate future research directions for data-driven energy theft.

ACKNOWLEDGMENT

This work has received funding from the EU H2020 EASY-RES project (grant agreement no. 764090) and also sponsored by Taif University and the Royal Embassy of Saudi Arabia Cultural Bureau.

REFERENCES

[1] D. Yao, M. Wen, X. Liang, Z. Fu, K. Zhang, and B. Yang, "Energy theft detection with energy privacy preservation in the smart grid," IEEE Internet of Things Journal, 2019.

[2] S. E. Safe, “Report energy theft | electricity theft, gas theft in the uk,” 2018, accessed on: March 2020. [Online]. Available: https://www.stayenergysafe.co.uk/

[3] S. Yorukoglu, F. Nasibov, M. Mungan, and M. Bagriyanik, “The effect of the types of network topologies on nontechnical losses in secondary electricity distribution systems,” IEEE Transactions on Industry Applications, vol. 52, no. 5, pp. 3631–3643, 2016.

[4] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, “Decision tree and svm-based data analytics for theft detection in smart grid,” IEEE Transactions on Industrial Informatics, vol. 12, no. 3, pp. 1005–1016, 2016.

[5] A. Aldegheishem, M. Anwar, N. Javaid, N. Alrajeh, M. Shafiq, and H. Ahmed, “Towards sustainable energy efficiency with intelligent electricity theft detection in smart grids emphasising enhanced neural networks,” IEEE Access, vol. 9, pp. 25 036–25 061, 2021.

[6] A. K. Marnerides, P. Smith, A. Schaeffer-Filho, and A. Mauthe, “Power consumption profiling using energy time-frequency distributions in smart grids,” IEEE Communications Letters, vol. 19, no. 1, pp. 46–49, 2014.

[7] Z. Aydin and V. C. Gungor, “A novel feature design and stacking approach for non-technical electricity loss detection,” in 2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia). IEEE, 2018, pp. 867–872.

[8] S. McLaughlin, D. Podkuiko, and P. McDaniel, “Energy theft in the advanced metering infrastructure,” in International Workshop on Critical Information Infrastructures Security. Springer, 2009, pp. 176–187.

[9] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, “Energy-theft detection issues for advanced metering infrastructure in smart grid,” Tsinghua Science and Technology, vol. 19, no. 2, pp. 105–120, 2014.

[10] A. Althobaiti, A. Jindal, and A. K. Marnerides, “Data-driven energy theft detection in modern power grids,” in Proceedings of the Twelfth ACM International Conference on Future Energy Systems, 2021, pp. 39–48.

[11] A. O. Otuoze, M. W. Mustafa, and R. M. Larik, “Smart grids security challenges: Classification by sources of threats,” Journal of Electrical Systems and Information Technology, vol. 5, no. 3, pp. 468–483, 2018.

[12] A. Jindal, A. Schaeffer-Filho, A. K. Marnerides, P. Smith, A. Mauthe, and L. Granville, “Tackling energy theft in smart grids through data-driven analysis,” in 2020 International Conference on Computing, Networking and Communications (ICNC), 2020, pp. 410–414.

[13] M. J. Burke and J. C. Stephens, “Political power and renewable energy futures: A critical review,” Energy Research & Social Science, vol. 35, pp. 78 – 93, 2018, energy and the Future.

[14] EPA, “Centralized generation of electricity and its impacts on the environment,” 2018, accessed on: March 2020. [Online]. Available: https://www.epa.gov/energy/centralized-generation-electricity-and-its-impacts-environment

[15] S.-C. Yip, W.-N. Tan, C. Tan, M.-T. Gan, and K. Wong, “An anomaly detection framework for identifying energy theft and defective meters in smart grids,” International Journal of Electrical Power & Energy Systems, vol. 101, pp. 189–203, 2018.

[16] C. Li and B. Shen, “Accelerating renewable energy electrification and rural economic development with an innovative business model: A case study in china,” Energy Policy, vol. 127, pp. 280–286, 2019.

[17] D. M. Shilay, K. G. Lorey, T. Weiz, T. Lovetty, and Y. Cheng, “Catching anomalous distributed photovoltaics: An edge-based multi-modal anomaly detection,” arXiv preprint arXiv:1709.08830, 2017.

[18] A. Banshwar, N. K. Sharma, Y. R. Sood, and R. Shrivastava, “Renewable energy sources as a new participant in ancillary service markets,” Energy strategy reviews, vol. 18, pp. 106–120, 2017.

[19] IEA, “Global energy review 2020,” 2020, last accessed 05-May-2020. [Online]. Available: https://www.iea.org/reports/global-energy-review-2020?utm_campaign=IEA\%20newsletters&utm_source=SendGrid&utm_medium=Email

[20] Mahmoud, M. Ismail, M. Shahin, Naidu, and E. Serpedin, “Deep learning detection of electricity theft cyber-attacks in renewable distributed generation,” Transactions on Smart Grid, pp. 73–102, 2020.

[21] B. Zhao, X. Wang, D. Lin, M. M. Calvin, J. C. Morgan, R. Qin, and C. Wang, “Energy management of multiple microgrids based on a system of systems architecture,” IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 6410–6421, 2018.

[22] X. Han, K. Heussen, O. Gehrke, H. W. Bindner, and B. Kroposki, “Taxonomy for evaluation of distributed control strategies for distributed energy resources,” IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 5185–5195, 2018.

[23] M. C. Bor, A. K. Marnerides, A. Molineux, S. Wattam, and U. Roedig, “Adversarial machine learning in smart energy systems,” in Proceedings of the Tenth ACM International Conference on Future Energy Systems, ser. e-Energy ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 413–415. [Online]. Available: https://doi.org/10.1145/3307772.3330171

[24] A. Jindal, A. K. Marnerides, A. Scott, and D. Hutchison, “Identifying security challenges in renewable energy systems: A wind turbine case study,” in Proceedings of the Tenth ACM International Conference on Future Energy Systems, ser. e-Energy ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 370–372. [Online]. Available: https://doi.org/10.1145/3307772.3330154

[25] V. B. Krishna, C. A. Gunter, and W. H. Sanders, “Evaluating detectors on optimal attack vectors that enable electricity theft and der fraud,” IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 4, pp. 790–805, 2018.

[26] C. Wang, J. Wu, J. Ekanayake, and N. Jenkins, Smart electricity distribution networks. CRC Press, 2017.

[27] E. U. Ogbodo, D. Dorrell, and A. M. Abu-Mahfouz, “Cognitive radio based sensor network in smart grid: Architectures, applications and communication technologies,” IEEE Access, vol. 5, pp. 19 084–19 098, 2017.

[28] Y. Zhang, T. Huang, and E. F. Bompard, “Big data analytics in smart grids: a review,” Energy Informatics, vol. 1, no. 1, p. 8, 2018.

[29] T. Liu, J. Tian, Y. Gui, Y. Liu, and P. Liu, “Sedea: State estimation-based dynamic encryption and authentication in smart grid,” IEEE Access, vol. 5, pp. 15 682–15 693, 2017.

[30] S. Tan, D. De, W.-Z. Song, J. Yang, and S. K. Das, “Survey of security advances in smart grid: A data driven approach,” IEEE Communications Surveys & Tutorials, vol. 19, no. 1, pp. 397–422, 2017.

[31] A. Sundararajan, K. Tanwir, A. Moghadasi, and A. I. Sarwat, “Survey on synchrophasor data quality and cybersecurity challenges, and evaluation of their interdependencies,” Journal of Modern Power Systems and Clean Energy, vol. 7, no. 3, pp. 449–467, 2019.

[32] J. C. do Prado, W. Qiao, L. Qu, and J. R. Agüero, “The next-generation retail electricity market in the context of distributed energy resources: Vision and integrating framework,” Energies, vol. 12, no. 3, p. 491, 2019.

[33] S. RB and D. GM, “A survey of wide area measurment technology in electrical networks,” in 2015 International Conference on Computing Communication Control and Automation. IEEE, 2015, pp. 521–526.

[34] M. Rezaee and M. H. Y. Moghaddam, “Sdn-based quality of service networking for wide area measurement system,” IEEE Transactions on Industrial Informatics, 2019.

[35] D.-A. Tian and G. Sansavini, “Impact of degraded communication on interdependent power systems: the application of grid splitting,” Electronics, vol. 5, no. 3, p. 49, 2016.

[36] S. N. Lighari, B. B. Jensen, A. A. Shaikh et al., “Attacks and their defenses for advanced metering infrastructure,” in 2014 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). IEEE, 2014, pp. 148–151.

[37] R. Mohammad, “Ami smart meter big data analytics for time series of electricity consumption,” in 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, 2018, pp. 1771–1776.

[38] A. Ghosal and M. Conti, “Key management systems for smart grid advanced metering infrastructure: A survey,” IEEE Communications Surveys & Tutorials, 2019.

[39] A. Ikpehai, B. Adebisi, and K. Rabie, “Broadband plc for clustered advanced metering infrastructure (ami) architecture,” Energies, vol. 9, no. 7, p. 569, 2016.

[40] D. Arcos-Aviles, J. Pascual, L. Marroyo, P. Sanchis, and F. Guinjoan, “Fuzzy logic-based energy management system design for residential grid-connected microgrids,” IEEE Transactions on Smart Grid, vol. 9, no. 2, pp. 530–543, 2016.

[41] B. V. Solanki, K. Bhattacharya, and C. A. Canizares, “A sustainable energy management system for isolated microgrids,” IEEE Transactions on Sustainable Energy, vol. 8, no. 4, pp. 1507–1517, 2017.

[42] G. K. Venayagamoorthy, R. K. Sharma, P. K. Gautam, and A. Ahmadi, “Dynamic energy management system for a smart microgrid,” IEEE transactions on neural networks and learning systems, vol. 27, no. 8, pp. 1643–1656, 2016.

[43] A. Jindal, B. Bhambu, M. Singh, N. Kumar, and S. Naik, “A heuristic-based appliance scheduling scheme for smart homes,” IEEE Transactions on Industrial Informatics, vol. 16, no. 5, pp. 3242–3255, 2020.

[44] G. M. U. Din, A. U. Mauthe, and A. K. Marnerides, “Appliance-level short-term load forecasting using deep neural networks,” in 2018 International Conference on Computing, Networking and Communications (ICNC). IEEE, 2018, pp. 53–57.

[45] N. Shokoya and A. Raji, “Electricity theft: A reason to deploy smart grid in south africa,” in 2019 International Conference on the Domestic Use of Energy (DUE). IEEE, 2019, pp. 96–101.

[46] S. A. Salinas and P. Li, “Privacy-preserving energy theft detection in microgrids: A state estimation approach,” IEEE Transactions on Power Systems, vol. 31, no. 2, pp. 883–894, 2016.

[47] R. Czechowski and A. M. Kosek, “The most frequent energy theft techniques and hazards in present power energy consumption,” in 2016 Joint Workshop on Cyber-Physical Security and Resilience in Smart Grids (CPSR-SG). IEEE, 2016, pp. 1–7.

[48] C. Kawann, “Reliability of the us electric system–recent trends and current issues,” 2002.

[49] F. H. Jufri, V. Widiputra, and J. Jung, “State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies,” Applied Energy, vol. 239, pp. 1049–1065, 2019.

[50] T. J. Bihl and S. Hajjar, “Electricity theft concerns within advanced energy technologies,” in 2017 IEEE National Aerospace and Electronics Conference (NAECON). IEEE, 2017, pp. 271–278.

[51] T. B. Smith, “Electricity theft: a comparative analysis,” Energy policy, vol. 32, no. 18, pp. 2067–2076, 2004.

[52] S. G. C. CEN-CENELEC-ETSI, “Group.(2012),” Smart Grid Reference Architecture, pp. 1–107, 2012.

[53] A. Neslen, “European renewable power grid rocked by cyber-attack,” 2012. [Online]. Available: https://www.euractiv.com/section/energy/news/european-renewable-power-grid-rocked-by-cyber-attack/

[54] B. Sobczak, “’Cyber event’ disrupted U.S. grid networks — DOE,” 2019. [Online]. Available: https://www.eenews.net/stories/1060242741

[55] S. Pfeifer, N. Fildes, and A. Ram, “Energy sector on alert for cyber attacks on uk power network,” 2018. [Online]. Available: https://www.ft.com/content/d2b2aaec-4252-11e8-93cf-67ac3a6482fd

[56] L.-Y. Lu, H. J. Liu, H. Zhu, and C.-C. Chu, “Intrusion detection in distributed frequency control of isolated microgrids,” IEEE Transactions on Smart Grid, vol. 10, no. 6, pp. 6502–6515, 2019.

[57] C. Peng, H. Sun, M. Yang, and Y.-L. Wang, “A survey on security communication and control for smart grids under malicious cyber attacks,” IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 49, no. 8, pp. 1554–1569, 2019.

[58] B. News, “Ransomware hits johannesburg electricity supply,” 2019. [Online]. Available: https://www.bbc.co.uk/news/technology-49125853

[59] E. EIA, “International energy outlook 2016,” 2016, accessed on: March 2020. [Online]. Available: https://www.eia.gov/outlooks/ieo/pdf/electricity.pdf

[60] X. Yuan, M. Shi, and Z. Sun, “Research status of electricity-stealing identification technology for distributed pv,” in 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT). IEEE, 2015, pp. 2031–2034.

[61] Y. Liu, P. Ning, and M. K. Reiter, “False data injection attacks against state estimation in electric power grids,” ACM Transactions on Information and System Security (TISSEC), vol. 14, no. 1, p. 13, 2011.

[62] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, “Detecting stealthy false data injection using machine learning in smart grid,” IEEE Systems Journal, vol. 11, no. 3, pp. 1644–1652, 2017.

[63] J. Y. Kim, Y. M. Hwang, Y. G. Sun, I. Sim, D. I. Kim, and X. Wang, “Detection for non-technical loss by smart energy theft with intermediate monitor meter in smart grid,” IEEE Access, vol. 7, pp. 129 043–129 053, 2019.

[64] S. K. Singh, R. Bose, and A. Joshi, “Energy theft detection for ami using principal component analysis based reconstructed data,” IET Cyber-Physical Systems: Theory & Applications, vol. 4, no. 2, pp. 179–185, 2019.

[65] Y. Gao, B. Foggo, and N. Yu, “A physically inspired data-driven model for electricity theft detection with smart meter data,” IEEE Transactions on Industrial Informatics, 2019.

[66] S. Sharma and A. Majumdar, “Unsupervised detection of non-technical losses via recursive transform learning,” IEEE Transactions on Power Delivery, 2020.

[67] R. Punmiya and S. Choe, “Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing,” IEEE Transactions on Smart Grid, vol. 10, no. 2, pp. 2326–2329, 2019.

[68] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, “Sparse malicious false data injection attacks and defense mechanisms in smart grids,” IEEE Transactions on Industrial Informatics, vol. 11, no. 5, pp. 1–12, 2015.

[69] K. G. Lore, D. M. Shila, and L. Ren, “Detecting data integrity attacks on correlated solar farms using multi-layer data driven algorithm,” in 2018 IEEE Conference on Communications and Network Security (CNS). IEEE, 2018, pp. 1–9.

[70] Z. El Mrabet, N. Kaabouch, H. El Ghazi, and H. El Ghazi, “Cyber-security in smart grid: Survey and challenges,” Computers & Electrical Engineering, vol. 67, pp. 469–482, 2018.

[71] P. Engebretson, The basics of hacking and penetration testing: ethical hacking and penetration testing made easy. Elsevier, 2013.

[72] L. Xie, Y. Mo, and B. Sinopoli, “False data injection attacks in electricity markets,” in 2010 First IEEE International Conference on Smart Grid Communications. IEEE, 2010, pp. 226–231.

[73] X. Yuan, M.-g. Shi, and Z. Sun, “Research of electricity stealing identification method for distributed pv based on the least squares approach,” in 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT). IEEE, 2015, pp. 2471–2474.

[74] S. Pal, B. Sikdar, and J. H. Chow, “Classification and detection of pmu data manipulation attacks using transmission line parameters,” IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 5057–5066, 2017.

[75] A. Ashok, M. Govindarasu, and V. Ajjarapu, “Online detection of stealthy false data injection attacks in power system state estimation,” IEEE Transactions on Smart Grid, vol. 9, no. 3, pp. 1636–1646, 2018.

[76] S. Basumallik, S. Eftekharnejad, N. Davis, and B. K. Johnson, “Impact of false data injection attacks on pmu-based state estimation,” in 2017 North American Power Symposium (NAPS). IEEE, 2017, pp. 1–6.

[77] S. K. Singh, K. Khanna, R. Bose, B. K. Panigrahi, and A. Joshi, “Joint-transformation-based detection of false data injection attacks in smart grid,” IEEE Transactions on Industrial Informatics, vol. 14, no. 1, pp. 89–97, 2017.

[78] A. Tajer, “False data injection attacks in electricity markets by limited adversaries: stochastic robustness,” IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 128–138, 2017.

[79] K. Zheng, Q. Chen, Y. Wang, C. Kang, and Q. Xia, “A novel combined data-driven approach for electricity theft detection,” IEEE Transactions on Industrial Informatics, vol. 15, no. 3, pp. 1809–1819, 2018.

[80] G. M. Messinis, A. E. Rigas, and N. D. Hatziargyriou, “A hybrid method for non-technical loss detection in smart distribution grids,” IEEE Transactions on Smart Grid, vol. 10, no. 6, pp. 6080–6091, 2019.

[81] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, “Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids,” IEEE Transactions on Industrial Informatics, vol. 14, no. 4, pp. 1606–1615, 2018.

[82] M. Ashrafuzzaman, Y. Chakhchoukh, A. A. Jillepalli, P. T. Tosic, D. C. de Leon, F. T. Sheldon, and B. K. Johnson, “Detecting stealthy false data injection attacks in power grids using deep learning,” in 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC). IEEE, 2018, pp. 219–225.

[83] S. E. Fernandes, D. R. Pereira, C. C. Ramos, A. N. Souza, D. S. Gastaldello, and J. P. Papa, “A probabilistic optimum-path forest classifier for non-technical losses detection,” IEEE Transactions on Smart Grid, 2018.

[84] J. A. Meira, P. Glauner, R. State, P. Valtchev, L. Dolberg, F. Bettinger, and D. Duarte, “Distilling provider-independent data for general detection of non-technical losses,” in 2017 IEEE Power and Energy Conference at Illinois (PECI). IEEE, 2017, pp. 1–5.

[85] P. Glauner, J. A. Meira, L. Dolberg, R. State, F. Bettinger, and Y. Rangoni, “Neighborhood features help detecting non-technical losses in big data sets,” in 2016 IEEE/ACM 3rd International Conference on Big Data Computing Applications and Technologies (BDCAT). IEEE, 2017, pp. 253–261.

[86] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, “Detection of non-technical losses using smart meter data and supervised learning,” IEEE Transactions on Smart Grid, 2018.

[87] W. Li, T. Logenthiran, V.-T. Phan, and W. L. Woo, “A novel smart energy theft system (sets) for iot based smart home,” IEEE Internet of Things Journal, 2019.

[88] C. Cody, V. Ford, and A. Siraj, "Decision tree learning for fraud detection in consumer energy consumption," in 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). IEEE, 2015, pp. 1175–1179.

[89] S. Nallathambi, "Prediction of electricity consumption based on dt and rf: An application on usa country power consumption," in 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE). IEEE, 2017, pp. 1–7.

[90] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity theft detection in ami using customers' consumption patterns," IEEE Transactions on Smart Grid, vol. 7, no. 1, pp. 216–226, 2016.

[91] R. Razavi, A. Gharipour, M. Fleury, and I. J. Akpan, "A practical feature-engineering framework for electricity theft detection in smart grids," Applied Energy, vol. 238, pp. 481–494, 2019.

[92] Y. Peng, Y. Yang, Y. Xu, Y. Xue, R. Song, J. Kang, and H. Zhao, "Electricity theft detection in ami based on clustering and local outlier factor," IEEE Access, vol. 9, pp. 107 250–107 259, 2021.

[93] G. M. Messinis and N. D. Hatziargyriou, "Review of non-technical loss detection methods," Electric Power Systems Research, vol. 158, pp. 250–266, 2018.

[94] L. A. P. Júnior, C. C. O. Ramos, D. Rodrigues, D. R. Pereira, A. N. de Souza, K. A. P. da Costa, and J. P. Papa, "Unsupervised non-technical losses identification through optimum-path forest," Electric Power Systems Research, vol. 140, pp. 413–423, 2016.

[95] M. H. Cintuglu, O. A. Mohammed, K. Akkaya, and A. S. Uluagac, "A survey on smart grid cyber-physical system testbeds," IEEE Communications Surveys & Tutorials, vol. 19, no. 1, pp. 446–464, 2016.

[96] J. Hu and A. V. Vasilakos, "Energy big data analytics and security: challenges and opportunities," IEEE Transactions on Smart Grid, vol. 7, no. 5, pp. 2423–2436, 2016.

[97] T. Wilcox, N. Jin, P. Flach, and J. Thumim, "A big data platform for smart meter data analytics," Computers in Industry, vol. 105, pp. 250–259, 2019.

[98] S. K. Gunturi and D. Sarkar, "Ensemble machine learning models for the detection of energy theft," Electric Power Systems Research, vol. 192, p. 106904, 2021.

[99] A. Maamar and K. Benahmed, "Machine learning techniques for energy theft detection in ami," in Proceedings of the 2018 International Conference on Software Engineering and Information Management. ACM, 2018, pp. 57–62.

[100] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "The challenge of non-technical loss detection using artificial intelligence: A survey," arXiv preprint arXiv:1606.00626, 2016.

[101] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, and J. D. Tygar, "Adversarial machine learning," in Proceedings of the 4th ACM workshop on Security and artificial intelligence. ACM, 2011, pp. 43–58.

[102] Y. Chen, Y. Tan, and D. Deka, "Is machine learning in power systems vulnerable?" in 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm). IEEE, 2018, pp. 1–6.

[103] S. Salinas, M. Li, and P. Li, "Privacy-preserving energy theft detection in smart grids," in 2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON). IEEE, 2012, pp. 605–613.

[104] B. P. Bhattarai, S. Paudyal, Y. Luo, M. Mohanpurkar, K. Cheung, R. Tonkoski, R. Hovsapian, K. S. Myers, R. Zhang, P. Zhao et al., "Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions," IET Smart Grid, vol. 2, no. 2, pp. 141–154, 2019.

[105] M. Li, K. Zhang, J. Liu, H. Gong, and Z. Zhang, "Blockchain-based anomaly detection of electricity consumption in smart grids," Pattern Recognition Letters, vol. 138, pp. 476–482, 2020.

[106] A. Jindal, N. Kumar, and M. Singh, "A unified framework for big data acquisition, storage, and analytics for demand response management in smart cities," Future Generation Computer Systems, vol. 108, pp. 921–934, 2020.

[107] P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis, "Big data analytics for dynamic energy management in smart grids," Big Data Research, vol. 2, no. 3, pp. 94–101, 2015.

[108] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," Progress in Artificial Intelligence, vol. 5, no. 4, pp. 221–232, 2016.

[109] J. Heaton, "An empirical analysis of feature engineering for predictive modeling," in SoutheastCon 2016. IEEE, 2016, pp. 1–6.

[110] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," ACM Computing Surveys (CSUR), vol. 50, no. 6, p. 94, 2018.

[111] U. Khurana, H. Samulowitz, and D. Turaga, "Feature engineering for predictive modeling using reinforcement learning," in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[112] J. Li and M. Hu, "Continuous model adaptation using online meta-learning for smart grid application," IEEE Transactions on Neural Networks and Learning Systems, 2020.

[113] I. Possebon, A. da Silva, L. Granville, A. Schaeffer-Filho, and A. Marnerides, "Improved network traffic classification using ensemble learning," IEEE Symposium on Computers and Communications (ISCC) 2019, 2019.

[114] Z. Guan, G. Si, X. Zhang, L. Wu, N. Guizani, X. Du, and Y. Ma, "Privacy-preserving and efficient aggregation based on blockchain for power grid communications in smart communities," IEEE Communications Magazine, vol. 56, no. 7, pp. 82–88, 2018.

[115] S. Afrin and S. Mishra, "An anonymized authentication framework for smart metering data privacy," in 2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). IEEE, 2016, pp. 1–5.

[116] T. Alladi, V. Chamola, J. J. Rodrigues, and S. A. Kozlov, "Blockchain in smart grids: A review on different use cases," Sensors, vol. 19, no. 22, p. 4862, 2019.

[117] W.-L. Chin, Y.-H. Lin, and H.-H. Chen, "A framework of machine-to-machine authentication in smart grid: a two-layer approach," IEEE Communications Magazine, vol. 54, no. 12, pp. 102–107, 2016.

AHLAM ALTHOBAITI (Graduate Student Member, IEEE) received a master's degree in Computer Science from the Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia in 2016. She is currently pursuing a PhD degree at the School of Computing and Communications (SCC), Lancaster University, Lancaster, UK. She holds the post of lecturer in the Computer Science Department, College of Computers and Information Technology, Taif University, Saudi Arabia. Her research interests include data analytics, artificial intelligence, cyber-physical systems and industrial system cybersecurity with special focuses on anomaly detection for smart energy systems, cybersecurity for Distributed Renewable Energy Sources (DRES) and energy theft in modern power grids. She has peer-reviewed published articles on network protocol design and energy modelling and the security of DRES. She received a First-Class Honours Award from Taif University in 2010 and a scholarship for a PhD degree.

ANISH JINDAL (Member, IEEE) has been working as a Lecturer in the School of Computer Science and Electronic Engineering (CSEE), University of Essex since Mar 2020. Prior to this, he worked as a senior research associate in the School of Computing Communications, Lancaster University, UK from Oct. 2018 to Mar. 2020. He completed his PhD, M.E. and B. Tech. degrees in computer science engineering in 2018, 2014, and 2012, respectively. He is the recipient of the Outstanding PhD Dissertation Award, 2019 from the IEEE Technical Committee on Scalable Computing (TCSC) and was conferred with the IEEE Communication Society's Outstanding Young Researcher Award for the Europe, Middle East and Africa (EMEA) Region, 2019. He has served as General co-chair, TPC co-chair, TPC member, Publicity chair and Session chair of various reputed conferences and workshops including IEEE ICC, IEEE WoWMoM and IEEE INFOCOM and IEEE GLOBECOM. He is also the guest editor of various journals including Software: Practice and Experience (Wiley) and Computers (MDPI). His research interests are in the areas of smart cities, data analytics, artificial intelligence, cyber-physical systems, wireless networks and security. He is a member of the IEEE and actively involved with various working groups and committees of IEEE and ACM related to smart grid, energy informatics and smart cities.

ANGELOS K. MARNERIDES (Member, IEEE) is Senior Lecturer (eq. tenured Associate Professor) of Computer Science in the School of Computing Science at the University of Glasgow, UK and leads the Glasgow Cyber Defence Group (GCDG). His research revolves around applied and data-driven security and resilience for Internet-enabled cyber physical systems, the Internet at scale and programmable networks. His research has received significant funding from industry (e.g. Fujitsu, BAE, Raytheon) and governmental bodies (e.g. EU, IUK, EPSRC) and he has been invited to serve as a grant reviewer for national (e.g., EPSRC) and international bodies (e.g., EU, Israeli Innovation Authority). He has been a member of the IEEE and the ACM since 2007 and served as a Technical Programme Committee (TPC) member, TPC track and workshop co-chair and organiser for several top-tier IEEE conferences (e.g. IEEE ICC, IEEE GLOBECOM) leading to receiving IEEE ComSoc contribution awards in 2016 and 2018. He obtained his PhD in Computer Science (2011) from Lancaster University and held lectureships, postdoctoral and visiting researcher positions at Lancaster University (UK), Carnegie Mellon University (USA), University of Porto (Portugal) and University College London (UK).

UTZ ROEDIG (Member, IEEE) was awarded a PhD degree in computer science from Darmstadt University of Technology, Germany in 2002. Between 2002 and 2006, he was a postdoctoral researcher in the Department of Computer Science at University College Cork, Ireland. he currently serves as a Full Professor of Computer Science at University College Cork (UCC), Ireland. Before moving to Cork he was a Professor at Lancaster University, UK, where he led the Academic Centre of Excellence in Cyber Security Research (ACE-CSR). He frequently serves as a TPC member at international conferences such as DCOSS, EWSN and IPSN, and he is a grant reviewer for international funding bodies such as the EPSRC (UK), ESF (EU) and FWO (Belgium). Aspects of his research include computer networks and security and he has published over 150 peer-reviewed papers in this field. Within this area, his work focus is on distributed embedded systems; this domain may also be described as Wireless Sensor Networks (WSN), Cyber Physical Systems (CPS) or the Internet of Things (IoT).

• • •