Journal of Electronics and Electrical Engineering
https://ojs.wiserpub.com/index.php/JEEE

Review

A Review of Protocol Analysis-Based Secure Traffic Monitoring and Anomalous Behavior Detection Technologies for Power Monitoring Systems

Shanshan Bai[1], Pengyu Wang[1], Tianle Gao[1], Zhaohui Hu[1], Guanling Wang[2], Kai Wang[2]*

[1] Guangdong Provincial Key Laboratory of Digital Grid Technology, Southern Power Grid Digital Grid Research Institute Co., Ltd., Guangzhou, 510663, China
[2] College of Electrical Engineering, Qingdao University, Qingdao, 266071, China
E-mail: wangkai@qdu.edu.cn

Received: 28 March 2026; **Revised:** 3 July 2026; **Accepted:** 9 July 2026; **Published:** 25 August 2026

Abstract: Focusing on the typical characteristics of power monitoring systems—including dedicated protocols, stable communication relationships, stringent real-time requirements, and security incidents with severe operational consequences—this paper presents a systematic review of protocol analysis-based secure traffic monitoring and anomalous behavior detection. The review begins with the network architecture, representative protocols such as International Electrotechnical Commission (IEC) 104, IEC 61850, and Distributed Network Protocol 3 (DNP3), and their associated security risks, thereby clarifying the foundational role of protocol-level visibility in power monitoring scenarios. It then surveys protocol analysis methods, including deep packet inspection, flow-level feature analysis, machine learning, and hybrid identification approaches, as well as anomaly detection strategies targeting connection behavior, operational status, and physical consistency. On this basis, it further summarizes common issues in publicly available datasets, evaluation metrics, and experimental design. Existing studies indicate that related technologies are moving from single rule matching toward the coordinated use of protocol semantic understanding, behavioral modeling, and multi-layer fusion analysis. Nevertheless, substantial gaps remain in power-specific data resources, cross-scenario generalization, stealthy anomaly recognition, and engineering interpretability. Finally, the paper discusses future research directions in light of the practical requirements of power monitoring systems.

Keywords: power monitoring system, protocol analysis, secure traffic monitoring, anomaly detection, industrial control network

1. Introduction

1.1 *Background and motivation*

With the continuous development of the new-type power system, dispatch automation, substation automation, and remote monitoring platforms have become increasingly dependent on communication networks. Power monitoring systems are not merely communication infrastructures but are also important information-support platforms for power-system situational awareness and operational decision-making. Their core functions include telemetry, status acquisition, disturbance monitoring, and state estimation. State estimation is widely regarded as one of the fundamental applications

Copyright ©2026 Kai Wang, et al.
DOI: https://doi.org/10.37256/jeee.5220269954
This is an open-access article distributed under a CC BY license
(Creative Commons Attribution 4.0 International License)
https://creativecommons.org/licenses/by/4.0/

supporting secure and stable grid operation because it provides a real-time assessment of network operating conditions and serves as the basis for contingency analysis, voltage stability assessment, and dispatch decision-making [1].

In modern Wide Area Measurement Systems (WAMS), Phasor Measurement Units (PMUs) are deployed at critical buses and substations to provide synchronized measurements of voltage and current phasors [2]. The collected synchrophasor data are transmitted through communication networks to Phasor Data Concentrators (PDCs), where data aggregation, alignment, and quality control are performed before being delivered to control centers and state-estimation applications [3]. Consequently, the correctness, timeliness, and integrity of communication traffic directly affect the observability of the grid and the reliability of monitoring functions.

Power monitoring systems have thus evolved into critical infrastructure for ensuring the secure and stable operation of the grid [4]. When communication networks are compromised by unauthorized access, malicious commands, data manipulation, or denial-of-service attacks, the resulting impact may propagate from communication layers to monitoring, state estimation, protection, and dispatch functions, thereby affecting overall grid security and operational stability [5]. At the same time, boundary expansion, access by heterogeneous devices, and the normalization of remote operation and maintenance have continually enlarged the attack surface. Compared with conventional Information Technology (IT) networks, these systems are characterized by strong protocol specialization [6], relatively fixed service chains, and strict real-time constraints. Once abnormal communication, unauthorized access, or malicious control occurs, its impact can directly propagate to monitoring, control, and dispatch functions.

In recent years, the research focus has gradually shifted from simple perimeter protection and static rule-based alarms to full-traffic protocol identification, behavior analysis, and anomaly detection [7]. The reason is straightforward: relying only on port numbers, packet lengths, or a small number of manually defined rules is increasingly insufficient for scenarios involving specialized protocols, concealed service interactions, and evolving attack patterns. Especially with the widespread deployment of Electrotechnical Commission (IEC) 104, IEC 61850, Distributed Network Protocol 3 (DNP3), and similar protocols, the focus of security monitoring is no longer limited to whether abnormal traffic exists [8]. It is moving toward protocol semantic understanding, characterization of connection relationships, recognition of operational-state deviation, and attribution of anomaly sources.

At present, however, several common challenges remain. High-quality power-specific datasets are scarce, and many reported results still rely on a small number of general industrial datasets [9]. Many methods still exhibit clear shortcomings in cross-scenario transfer, stealthy anomaly detection, and interpretability. When confronting slowly evolving anomalies, cross-layer correlated anomalies, and complex attack chains, a single model is often unable to balance accuracy, timeliness, and engineering usability. In view of this, this paper systematically reviews the literature along the main line of "protocol parsing-behavior understanding-state modeling-comprehensive evaluation" [10], with the aim of providing a clearer reference for subsequent method design and engineering deployment.

1.2 *Review methodology*

This review follows a structured literature-survey process. Relevant studies were collected from IEEE Xplore, Web of Science, Scopus, and ScienceDirect. The primary search keywords included "power monitoring system", "Supervisory Control And Data Acquisition (SCADA) security", "IEC 61850", "IEC 60870-5-104", "DNP3", "protocol analysis", "traffic monitoring" and "anomaly detection". The survey mainly focuses on studies published between 2015 and 2025.

Figure 1 illustrates the literature screening procedure used in this review. Relevant publications were initially sourced from major scientific databases. After removing duplicates, a title and abstract screening was conducted to exclude studies not directly related to power monitoring systems, industrial communication protocols, traffic monitoring, or anomaly detection. The remaining papers were then evaluated in full text. Only studies that made significant contributions in terms of methodology, experiments, or practical applications were retained for detailed analysis. This process enhances the transparency and reproducibility of the review methodology.

The inclusion criteria were: (1) peer-reviewed journal or conference publications; (2) studies involving protocol analysis, traffic monitoring, intrusion detection, or anomaly detection in power or industrial control environments; and (3) studies reporting technical methods, datasets, or evaluation results. Duplicate publications, non-English papers, and studies unrelated to operational monitoring environments were excluded. The selected literature was subsequently

categorized according to protocol analysis methods, connection-behavior analysis methods, operational-state anomaly detection methods, datasets, and evaluation approaches to facilitate systematic comparison and discussion.

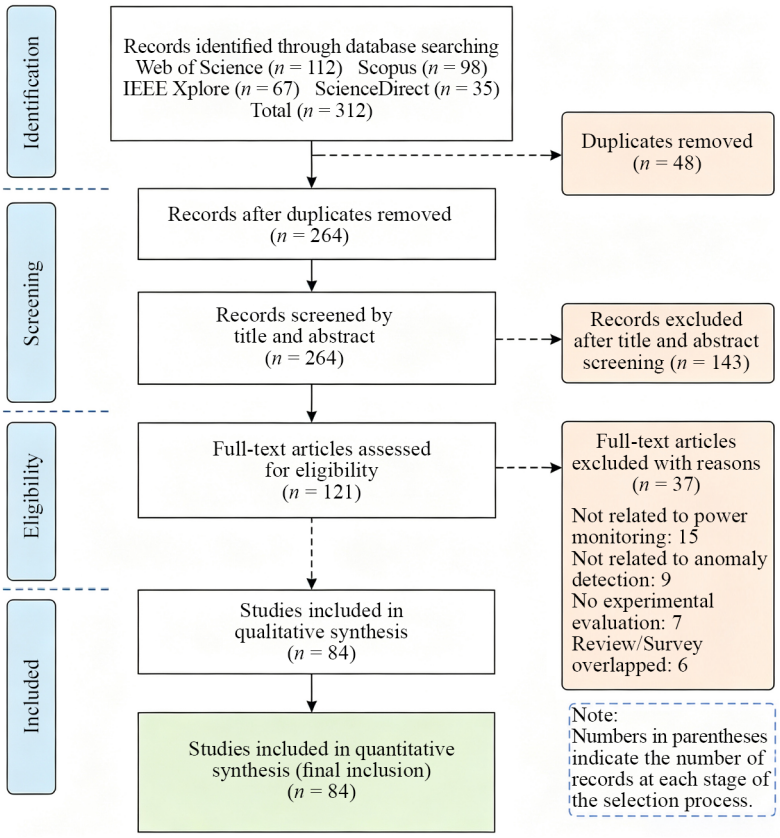


Figure 1. Literature screening process

2. Research foundations of secure traffic monitoring for power monitoring systems

Power monitoring systems carry critical services such as telemetry, telesignaling, telecontrol, remote adjustment, status monitoring, and alarm reporting. Their communication networks are both the operational backbone of these services and a major locus of security exposure. To establish a clear background for the later discussion of protocol identification, connection behavior analysis, and anomaly detection, this section examines four aspects: network architecture, representative protocols, security threats, and secure traffic monitoring tasks. The goal is to clarify why protocol-level full-traffic visibility is indispensable in power monitoring environments and how this requirement shapes subsequent method design.

2.1 *Network architecture of power monitoring systems and the basis of traffic formation*

From a structural perspective, power monitoring systems can usually be abstracted as a layered architecture composed of the control center, the communication network, substation or plant-side systems, and field terminal devices. The control center typically includes the human-machine interface, database servers, dispatch application servers, and simulation analysis modules, and is responsible for system-wide operational monitoring, data management, control decision-making, and command issuance. The substation or plant side includes station-level devices, bay-level intelligent electronic devices, and process-level measurement and execution units that undertake field-state acquisition, protection and control, status

interaction, and command execution. Krause et al. [11] summarized this structure as a typical grid operation network: the office network mainly handles general information processing, whereas the process control network connects control rooms, substations, and field devices and constitutes the core communication plane of power monitoring services.

As shown in Figure 2, the communication pattern of a power monitoring system is not the highly random and open "any-to-any" traffic model of the general Internet. Instead, it is a communication pattern with strict constraints, strong correlations, and tight temporal order, organized around dispatch control links, in-station automation links, and remote acquisition links. A relatively fixed master-slave interaction is usually maintained between the control center and substations; communication peers, periods, directions, and service contents are relatively stable. Inside a substation, devices interact in a more fine-grained real-time manner around protection, measurement and control, interlocking, and state synchronization [12]. This means that under normal operating conditions, power monitoring traffic is highly amenable to modeling: over long periods, the same device often communicates only with specific peers, and its interaction content and rhythm show clear regularity. Once abnormal communication peers, abnormal command sequences, abnormal communication frequency, or abnormal connection topologies appear, network behavior has often deviated from normal business logic. For this reason, power monitoring systems are naturally suited to secure monitoring based on full-traffic acquisition, protocol parsing, and behavior modeling, rather than depending solely on boundary alarms or simple five-tuple statistics. In addition to traditional SCADA communication chains, modern power monitoring systems increasingly incorporate WAMS architectures [13]. A typical WAMS consists of PMUs deployed at substations, communication networks for synchrophasor transmission, one or more PDCs for data concentration, and control-center applications responsible for visualization, state estimation, oscillation monitoring, and situational awareness. From a cybersecurity perspective, the communication links connecting PMUs, PDCs, and control centers constitute critical monitoring channels. Any abnormal traffic, packet loss, delay manipulation, replay behavior, or malicious data injection may affect the quality of synchronized measurements and subsequently influence monitoring and decision-support functions. Therefore, protocol-level visibility and anomaly detection are also important requirements in PMU-PDC communication environments.

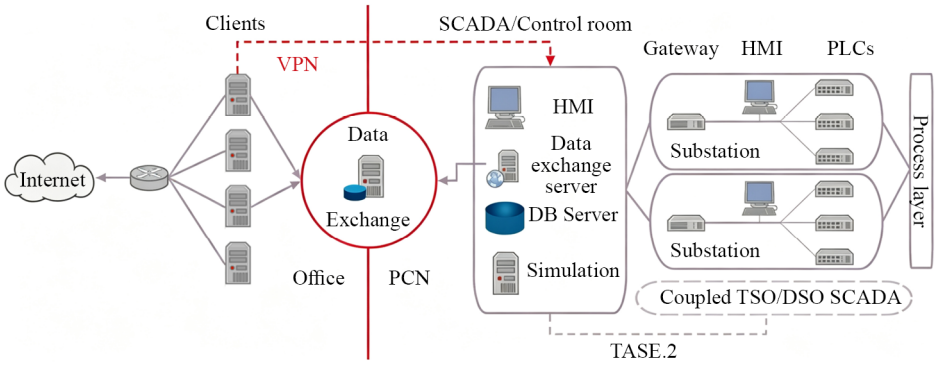


Figure 2. Typical relationship between the office network and the process control network in a grid operation network

Compared with enterprise networks, power monitoring systems also exhibit distinctive real-time and high-availability requirements. Krause et al. [11] emphasized that in critical infrastructure such as power systems, the Confidentiality, Integrity, and Availability (CIA) triad cannot be applied mechanically according to traditional IT logic; availability often has the highest priority, and no security mechanism should significantly impair the continuous operation of control and monitoring services. In other words, secure traffic monitoring in power monitoring systems must not only "see" anomalies, but also operate in a low-intrusion, low-latency, and highly stable manner. This requirement, in turn, determines that protocol analysis and anomaly detection methods must balance real-time performance, interpretability, and deployment cost.

2.2 *Typical communication protocols in power monitoring systems and their monitoring significance*

The reason why power monitoring systems emphasize traffic monitoring based on protocol analysis lies in their heavy dependence on specialized or semi-specialized industrial protocols. Besides SCADA-oriented protocols such as IEC 60870-5-104 and DNP3, synchrophasor communication in WAMS commonly relies on standards defined by IEEE C37.118 [14]. These standards specify the formats and transmission mechanisms for synchronized phasor measurements exchanged between PMUs and PDCs. Compared with conventional polling-based monitoring protocols, synchrophasor traffic exhibits higher reporting frequencies, stricter latency requirements, and stronger time-synchronization dependence. Consequently, monitoring the communication quality and protocol behavior of PMU-PDC traffic is an important component of secure traffic monitoring in modern power systems. Krause et al. [11] pointed out that in process control networks, the most representative wide-area monitoring protocols are DNP3 and IEC 60870-5-104. The former is widely used in North America and parts of Asia, while the latter is broadly adopted in Europe and other regions for SCADA communication. Khandekar et al. [15] further noted that IEC 60870-5-104 is a Transmission Control Protocol/Internet Protocol (TCP/IP)-based protocol that usually uses TCP port 2404 and is widely employed in SCADA systems in the energy sector. However, its native design lacks sufficient authentication and authorization mechanisms, making it vulnerable to unauthorized command injection and man-in-the-middle attacks.

For in-station communication, especially in digital substations, IEC 61850 is a more representative standard system. IEC 61850 substations typically consist of station, bay, and process levels; they support interoperability among Intelligent Electronic Devices (IEDs) from different vendors and realize equipment information import and export through standardized configuration languages. Ashraf et al. [16] similarly summarized the IEC 61850 communication architecture as comprising station, bay, and process levels: the station level mainly includes HMI and station control equipment, the bay level is centered on IEDs, and the process level corresponds to sensors, actuators, and CT/PT measurement devices. In other words, IEC 61850 not only defines communication protocols but also deeply embeds power service objects and in-station control semantics, which makes protocol fields themselves carry explicit service meaning.

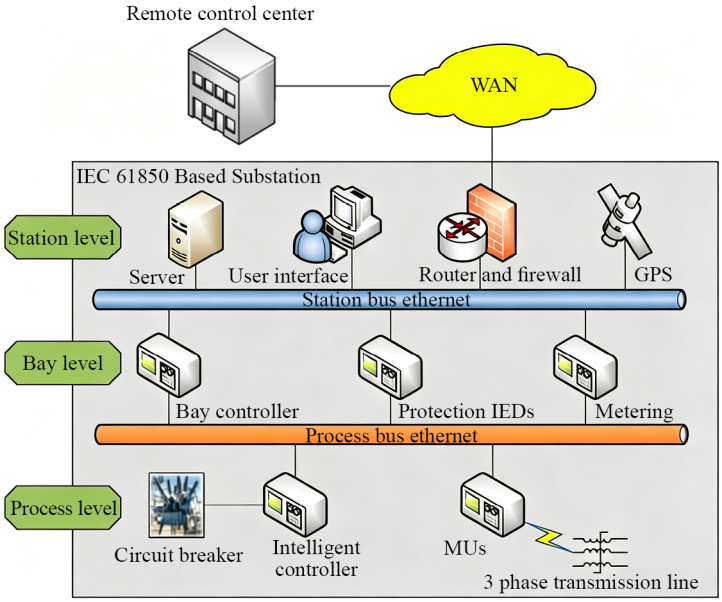


Figure 3. IEC 61850 substation communication architecture

As shown in Figure 3, the layered structure of IEC 61850 determines that different levels undertake different communication tasks, and it also means that traffic characteristics cannot be described adequately by simply applying conventional IT network methods. For secure traffic monitoring, this has two implications. First, it is necessary to

identify what protocol a packet belongs to. Second, it is also necessary to understand the role of that packet in the current service chain. If monitoring remains at the port or packet-length level, it is often difficult to distinguish normal control interaction from disguised malicious traffic, and even harder to judge whether a protocol message appears on an unreasonable device, at an unreasonable time, or in an unreasonable context. Therefore, secure monitoring in power systems often needs to proceed simultaneously from four dimensions: protocol fields, service types, message-mapping relationships, and interaction timing.

In IEC 61850 scenarios, distinguishing among transmission services is particularly important. Ashraf et al. [16] summarized that common IEC 61850 services include Manufacturing Message Specification (MMS), Generic Object Oriented Substation Event (GOOSE), Generic Substation State Event (GSSE), Sampled Measured Values (SMV), and time synchronization. Among them, GOOSE and sampled value messages are highly time-sensitive and are often mapped directly to lower layers to satisfy rapid transmission requirements, whereas MMS is more often used for status and supervisory information exchange. In other words, different message types differ substantially in timing constraints, communication patterns, and service consequences. If these services cannot be differentiated during monitoring, anomaly analysis aligned with service logic becomes difficult.

Table 1 makes clear that protocol analysis in power monitoring systems should not stop at identifying a protocol name; it also needs to capture the service semantics and timing constraints behind each message category. For example, if GOOSE messages appear at an abnormal location, abnormal time, or abnormal rate, the deviation may already indicate a violation of the intended substation control logic.

Table 1. Major IEC 61850 transmission services and timing requirements

Communication architecture	Service	Message type	Application type	Time requirement (Milli-second)	Communication mapping
Client-Server (SCADA)	ACS	2	Moderate speed	100	Ethernet TCP/IP
		5	File transfer	$\geq$ 1,000	
Publisher-subscriber	GOOSE, GSSE	1A	Trip	3-10	Ethernet
		1B	Others	20-100	
	SMV	4	Measurement data	3-10	
	TS	6	Synchronization	N/A	Ethernet UDP/IP

2.3 *Major security threats faced by power monitoring systems and their traffic manifestations*

The security threats faced by power monitoring systems include both conventional cyberattacks and attacks specifically targeting power-service processes and industrial protocol semantics. Once attackers gain access to a substation communication network, they may not only obtain system topology, measurement information, and maintenance records, but may also issue control commands, thereby causing broader operational risks. Krause et al. [11] likewise emphasized that as remote accessibility increases and legacy and modern systems coexist, grid communication networks face increasingly complex attack vectors, including denial-of-service, abuse of communication links, misuse of remote maintenance entry points, and targeted attacks against process-control protocols.

At the protocol level, both IEC 104 and IEC 61850 have typical security weaknesses. Khandekar et al. [15] noted that IEC 104 suffers from a lack of encryption, authentication, and authorization mechanisms, and is therefore vulnerable to traffic analysis, man-in-the-middle attacks, and unauthorized command injection. Ashraf et al. [16] systematically summarized network and message attacks in IEC 61850 environments, including malformed packet attacks, denial-of-service, Address Resolution Protocol (ARP) spoofing, man-in-the-middle attacks, configuration tampering, and attacks that modify GOOSE or SMV messages. These attacks do not always manifest as high-volume floods. Some are fine-

grained manipulations at the field level, some are covert deviations in connection relationships, and some alter in-station message propagation paths to affect control outcomes.

As shown in Figure 4, threats in power monitoring systems are not confined to a single layer. They often span the station, bay, and process levels and simultaneously act on network communication, protocol messages, and device configurations. This feature means that secure traffic monitoring cannot focus only on coarse-grained traffic anomalies. It must also consider three categories of manifestations: packet-level anomalies such as abnormal function codes, illegal field combinations, and tampered message contents; connection-level anomalies such as abnormal communication peers, changed access directions, and sudden session-frequency shifts; and behavior-level anomalies such as disrupted service timing, message chains inconsistent with normal control logic, and long-term deviations in asset communication profiles. In other words, monitoring that genuinely serves power monitoring systems must build a continuous mapping across the protocol, flow, and behavior layers.

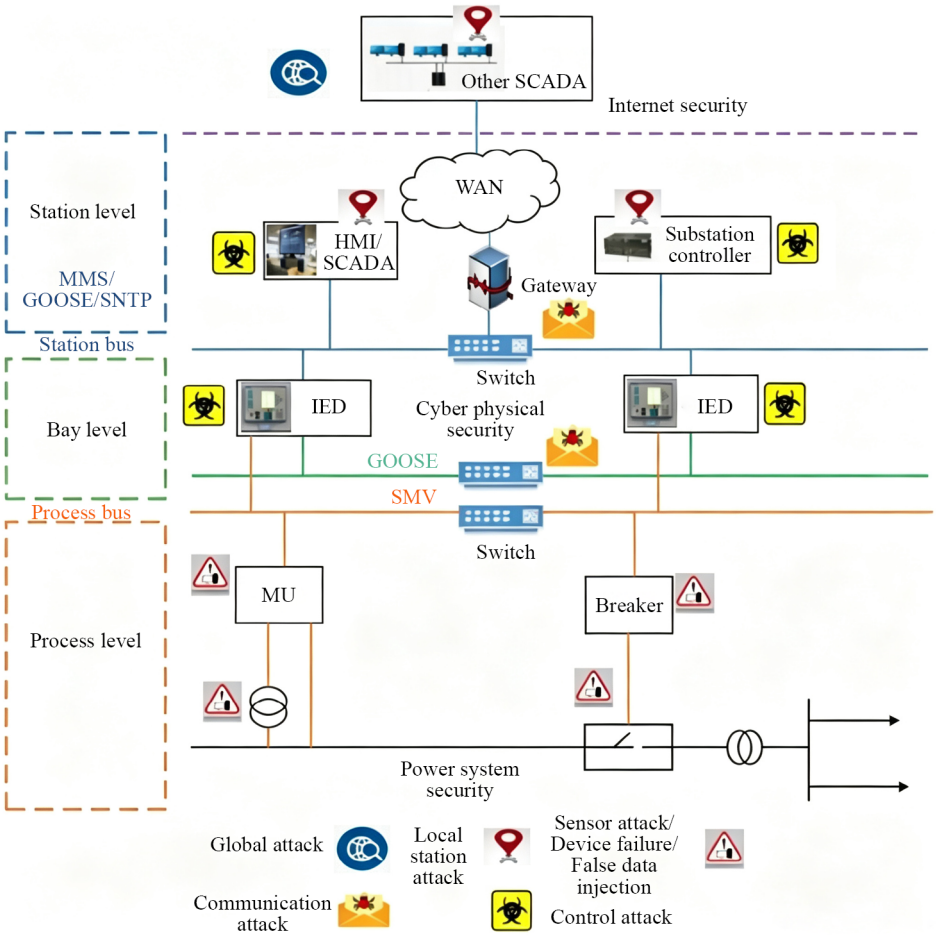


Figure 4. Security threats at different levels of IEC 61850

From the broader perspective of grid security, the consequences of attacks are no longer limited to local communication failures or single-station incidents. Several serious attacks and destructive events have occurred worldwide, showing that once the power grid is subjected to targeted cyberattacks, the outcome may include power outages, business paralysis, and even wider socioeconomic losses. For this review, the implication is clear: secure traffic monitoring in power monitoring systems is not simply a traffic-classification problem. It is a fundamental component of situational awareness, risk discovery, and emergency response for power-system cybersecurity.

2.4 *Task objectives and main research line of secure traffic monitoring*

Based on the network architecture, protocol systems, and threat characteristics discussed above, the tasks of secure traffic monitoring in power monitoring systems can be summarized at four levels. The first is protocol identification and semantic parsing, that is, identifying industrial protocols such as IEC 104, DNP3, and IEC 61850 from full traffic and recovering key service fields. The second is connection behavior analysis, used to characterize communication peers, directions, periodic rhythms, and access boundaries among devices. The third is operational-state modeling and anomaly detection, which focuses on persistent or stealthy deviations from normal patterns. The fourth is associating detection results with device roles, service impacts, and engineering risks so as to form interpretable security judgments.

These levels are not isolated from one another. If protocol identification is inaccurate, subsequent behavior modeling lacks a reliable foundation. If connection relationships are not characterized sufficiently, anomaly alarms are difficult to explain. Without long-term observation of system behavior, it is also difficult to assess the true risk of a small number of abnormal packets. The remaining sections therefore discuss protocol identification, connection behavior analysis, operational-state modeling, and experimental evaluation in sequence to keep the technical line coherent.

3. Key technologies for traffic identification based on protocol analysis

Protocol analysis occupies a foundational position in secure traffic monitoring for power monitoring systems because these networks are not dominated by open and random Internet application traffic. Instead, they operate around industrial control protocols with explicit service semantics and fixed interaction logic. In dispatch automation, substation automation, and in-station control networks, behavior analysis and anomaly detection lack reliable context unless protocols are first identified and their semantics understood.

From the perspective of research evolution, protocol identification has moved from shallow static matching to deep semantic recovery, from single-rule judgment to multi-feature fusion, and from deterministic methods to learning-based methods. Early approaches emphasized port numbers, message formats, and signature features. As encapsulation patterns multiplied, protocol variants increased, and encrypted traffic emerged, researchers gradually turned to flow-level characteristics, statistical patterns, and machine-learning models. Along this trajectory, the following subsections discuss Deep Packet Inspection (DPI), flow-level feature analysis, learning-based identification, and hybrid frameworks.

3.1 *Basic process and research role of protocol identification*

From an implementation perspective, protocol identification typically includes data acquisition, traffic preprocessing, protocol-feature extraction, feature screening, classification, and result mapping. The industrial-control protocol keyword feature extraction method proposed by Yu et al. [17] clearly illustrates this pipeline. It first extracts protocol payloads from raw network traffic and unifies packet lengths, then obtains candidate feature items through segmentation and redundant-data filtering, and finally mines representative protocol keyword features using association rules and trains a classifier for traffic identification. Their study emphasized that industrial-control protocols feature relatively fixed structures, explicit functions, and strong traffic periodicity, and thus protocol identification in Industrial Control System (ICS) should fully exploit these characteristics rather than mechanically copying methods designed for general Internet traffic.

As shown in Figure 5, the preprocessing stage itself already reflects a research logic distinct from ordinary network traffic classification. The focus is not generic application-type recognition, but recovering core fields that can stably characterize protocol type and functional role from industrial-protocol data. This is especially important in power monitoring systems. In IEC 104, DNP3, Modbus/TCP, IEC 61850 MMS/GOOSE/SMV, and similar scenarios, messages often contain not only protocol-category information but also service semantics related to measurement points, objects, function codes, message types, directionality, and timing constraints. If only coarse-grained classification is achieved, without identifying key fields or service types, the supporting effect of protocol identification on anomaly detection and behavior understanding will be greatly weakened.

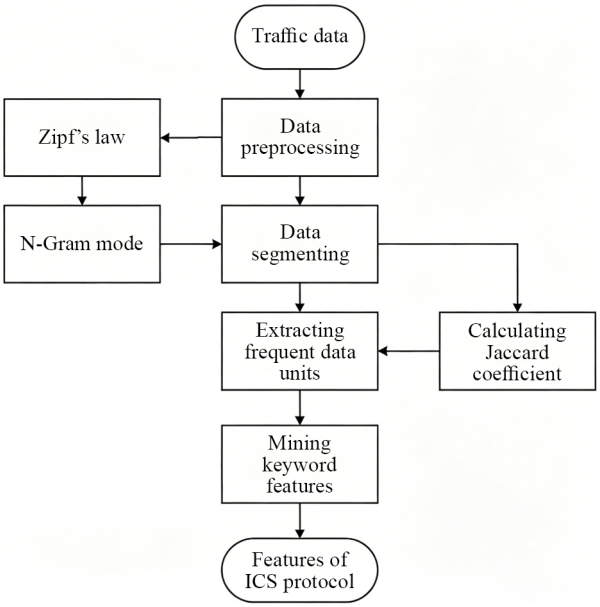


Figure 5. Industrial control protocol feature extraction process

From a review perspective, the role of protocol identification in power monitoring systems is reflected mainly in three aspects. First, it is the basis for full-traffic visibility. Without it, it is difficult to determine what services are actually running on the network. Second, it is a prerequisite for behavior analysis, because whether a connection behavior is abnormal often depends on protocol roles and interaction types. Third, it is the entry point for building anomaly-detection models oriented to service semantics, because many industrial-protocol attacks do not manifest as obvious traffic floods, but rather as field tampering, abnormal message ordering, or service-type mismatch.

3.2 *DPI-based protocol identification methods*

DPI is one of the most representative traditional approaches to protocol identification. Its basic idea is to parse packet payloads in depth, extract protocol signatures, and determine protocol categories by matching them against rule libraries, keyword libraries, or protocol grammar templates. Yu et al. [17] explicitly stated that payload-based classification is essentially DPI. In scenarios with known protocols, clear signatures, and unencrypted or unobfuscated messages, it can achieve high accuracy.

For power monitoring systems, DPI has long retained practical value for two reasons. First, many industrial protocols contain explicit function codes, object identifiers, service fields, and message formats, making rule-based and field-based identification feasible. Second, power monitoring traffic is relatively stable and service boundaries are clear, so deep parsing based on rules remains highly deployable in engineering practice. In substations or dispatch communication scenarios with known protocols and relatively fixed devices, DPI can often identify not only the protocol name but also fine-grained semantics such as control commands, status reporting, file transfer, and measurement services.

However, the strengths of DPI are also its limits. Yu et al. [17] noted that such methods depend heavily on prebuilt signature databases and perform poorly on encrypted traffic. As proprietary protocol variants, multilayer encapsulation, and vendor-specific extensions become more common, protocol identification relying solely on fixed rules faces high maintenance cost, poor transferability, and limited robustness. In future power monitoring networks, if more secure encapsulation, tunneling, or vendor-private extensions are deployed, static-signature DPI alone will become increasingly difficult to sustain. Therefore, DPI should be regarded as a high-precision parser for known protocols rather than the only method in the identification pipeline.

3.3 *DFI- and flow-level feature-based protocol identification methods*

DFI which analyzes individual packet payloads directly, flow-level protocol identification emphasizes extracting statistical, directional, temporal, and behavioral features from sessions or interaction sequences composed of multiple packets. It does not insist on fully restoring the semantics of every single message, but instead seeks to capture the behavioral contour of a protocol from continuous communication patterns. As a result, it is often more flexible when facing proprietary protocols, incomplete fields, or restricted payload access.

Holasova et al. [18] noted that protocol recognition and traffic classification in OT environments often need to analyze multiple granularities of data, including flows, frames, and other structured objects. With the development of machine learning and neural networks, many studies have moved beyond single-packet analysis and instead use continuous traffic or feature sequences for protocol identification. This observation is especially relevant to power monitoring systems, because their communications themselves are highly periodic and sequential. Polling-response, periodic IED reporting, and GOOSE event broadcasting all imply that protocol identification cannot be separated from temporal context and interaction structure.

In power monitoring systems, flow-level features usually involve four categories: basic statistical features such as packet length, packet count, direction ratio, and flow duration; temporal features such as periodicity, inter-arrival time, burstiness, and time-distribution pattern; structural features such as request-response modes, message sequences, and multi-message combinations; and semantic-mapping features that jointly model inter-message field associations, service roles, and business context. Compared with pure DPI, this approach is better suited to proprietary protocols, protocol variants, and interactions whose surface fields are similar but whose behavioral logic differs.

3.4 *Machine learning and deep learning for protocol identification*

As industrial protocol environments become more complex, more studies have attempted to use machine learning and deep learning to enhance the generalization of protocol identification. Yu et al. [17] observed that machine-learning methods typically rely on manually designed features and trained classifiers, whereas deep-learning methods attempt to reduce dependence on handcrafted features and learn representations directly from the data. Their comparison showed that deep models may achieve high accuracy in some tasks, but their training and inference costs are often substantial and may not always be optimal under the high real-time requirements of industrial control systems. The representative performance characteristics of these industrial control traffic classification methods are summarized in Table 2, including their feature-extraction strategies, training time, classification time, and reported accuracy.

Table 2. Performance comparison of industrial control traffic classification methods

Method	Feature extraction	Training time	Classification time	Accuracy
Machine Learning (ML)-based	Manual	0.31 s	Unknown	97.2%
		Unknown	Unknown	100%
		104.85 s	12.75 s	82.31%
		2.21 s	0.0505 s	99%
		Unknown	0.187 s	99%
DL-based	Automatic	100 s	Unknown	Unknown
		Unknown	Unknown	95%
		54.2 s	0.185 ms	98.98%
		61.31 s	2.36 s	99%
		121.9 min	Unknown	94%
ML-based	Automatic	0.34 s	0.264 s	99.99%

Table 3. Comparison of protocol recognition and traffic classification studies in IT and OT scenarios

Methods	Type	Year	Technique	Model	ML type	Protocols	Encoding	Real-time	Epochs	Layers	Accuracy [%]	Datasets
Protocol recognition	OT	2021	Convolutional Neural Network (CNN)	Attention Mechanism-based 1-Dimensional Convolutional Neural Network and Long Short-Term Memory (AM-1DCNN + LSTM)	Supervised	4	No	No	20	-	93.0	Own
		2023	Deep Neural Network (DNN)	Protocol Reverse Engineering for Industrial Control Protocols Based on Unsupervised Learning and Deep Neural Network Methods (PREIUD)	Unsupervised	1	No	No	-	-	-	Own
	IT	2011	Network Packet Inspection	Deterministic Finite-state Automaton	-	9	No	No	-	-	-	Own
		2012	Fingerprinting	-	-	4	No	Yes	-	-	95.0	Own
		2017	CNN + RNN	CNN + RNN-2a	Supervised	15	No	No	60-90	9	99.6	RedIRIS
		2020	CNN	PtrCNN	Supervised	4	No	Yes	20	8	96-100	Defense Advanced Research Projects Agency (DARPA)
		2020	CNN	-	Both	3	No	No	-	-	75.8-89.8	Own
		2021	Pattern matching algorithm	-	Supervised	4	No	No	-	-	93.8-100	DARPA
		2021	CNN	Improved Convolutional LSTM	Supervised	12	Yes	No	-	-	97.5	ISCX 2016
		2023	CNN	-	Supervised	8	Yes	No	-	-	98.2	ISCX VPN-nonVPN
Traffic classification	OT	2019	ML	Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes (NB)	Supervised	1	Yes	No	-	-	95.0	Own
		2019	Traffic Fingerprinting	CART	-	-	No	No	-	-	94.8	Secure Water Treatment (SWaT), SCADA Network Data Sets for Intrusion Detection Research
		2020	ML	KNN, SVM, DT, Naive Bayes with Gaussian (NBG), BKNN, Bagging Tree (BT), Random Forest (RF), AdaBoost, Gradient Boosting (GB)	Both	1	No	No	-	-	99.7	Own
		2022	DNN	-	Supervised	1	Yes	No	100	10	94.5	Own
		2022	ML	DT	Supervised	2	No	No	-	-	99.9	Own
		2022	RNN	-	Supervised	1	No	No	-	-	97.5	Own
	IT	2009	ML	C4.5, AdaBoost, NB, SVM, Repeated Incremental Pruning to Produce Error Reduction (RIPPER)	Supervised	8	Yes	No	-	-	98.4	DARPA, AMP, MAWI
		2017	CNN	1D-CNN	Supervised	12	Yes	No	40	7	99.5	ISCX VPN-nonVPN
		2018	CNN	CNN-LSTM	Supervised	9	Yes	No	30	8	91.0	ISCX VPN-nonVPN
		2021	Fuzzy Inference System	Fuzzy Inference System	-	6	Yes	Yes	-	-	90.9	ISCXVPN 2016

In engineering terms, the quality of a protocol-identification method cannot be judged only by classification accuracy. Training complexity, inference latency, and deployment constraints are equally important. This issue is especially critical in power monitoring systems, where protocol identification is not merely an offline research problem but often a continuously running online monitoring function. A model may be highly accurate, but if it is resource-intensive or too slow, its engineering value becomes limited. Comparative results reported by Yu et al. [17] suggest that lightweight classifiers combined with industrial-protocol feature extraction may be better positioned to balance high accuracy and low latency. This also illustrates that the method-selection logic for industrial protocol recognition is not identical to that for Internet application classification.

At a more general level, Holasova et al. [18] compared protocol-identification studies in IT and OT scenarios and found that OT-oriented research remains far less abundant, while publicly available datasets are also fewer. Consequently, many high-performing intelligent identification methods are still validated only on small-scale or self-built datasets. This implies that although protocol-identification research has become more intelligent, the fundamental constraints for power monitoring systems remain data availability and engineering transferability.

Taken together, Table 3 suggests that protocol identification for power monitoring systems should not simply pursue maximal model complexity. Greater practical value lies in methods that preserve protocol semantics, remain interpretable, satisfy real-time deployment constraints, and retain transferability across sites and operating conditions.

3.5 *Hybrid protocol identification methods and their engineering value*

Current studies increasingly adopt hybrid identification frameworks because a single method rarely satisfies the joint requirements of high accuracy, strong generalization, low latency, and high interpretability. Here, "hybrid" does not simply mean stacking multiple features, but integrating shallow network features, message-field features, flow-level statistics, and service-context features across different layers so as to improve robustness and engineering adaptability. This line of thinking fits the reality of power monitoring systems particularly well: protocols are relatively fixed, services are diverse, traffic is stable, and attacks may remain covert.

A representative related idea is the multilayer detection architecture reflected in ARIES. Although that work mainly focused on smart-grid intrusion detection rather than protocol recognition alone, it clearly demonstrated how multilayer traffic analysis can work in coordination. The system sets separate detection layers for network flows, Modbus/TCP packets, and operational data so as to identify attacks and anomalies at different levels. The study shows that in industrial and power environments, relying on a single data layer is often insufficient, whereas combining flow-level analysis, packet-level analysis, and higher-level operational semantics can significantly enhance both detection capability and interpretability.

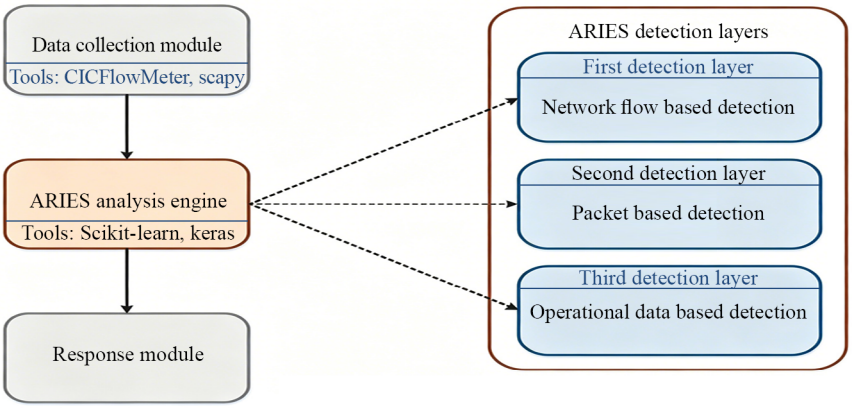


Figure 6. Multi-layer traffic analysis and detection architecture for smart grids

As illustrated in Figure 6, the value of multilayer analysis lies in the fact that protocol identification is no longer treated as an isolated point task, but as a structured input to a larger system for traffic understanding and anomaly sensing. In power monitoring systems, the ultimate value of protocol identification is not simply outputting a protocol label, but to support judgments about what kinds of services are communicating, whether the communication is reasonable, and whether protocol behavior deviates from the normal baseline. The engineering significance of hybrid methods therefore lies mainly in three aspects: they compensate for the weakness of single-method DPI on unknown and variant traffic; they use flow-level and temporal features to better capture stealthy behavior; and they improve result interpretability through multi-layer feature combination, thereby supporting subsequent connection behavior analysis and abnormal asset discovery.

4. Methods for network connection behavior analysis in power monitoring systems

Once protocol identification and traffic understanding are in place, the research focus naturally shifts to a higher-level question: how to identify anomaly clues from communication relationships among devices, interaction rhythms, and long-term operating patterns. Unlike methods that focus only on single packets or individual flows, connection behavior analysis asks who communicates with whom, at what frequency, and whether such relationships conform to established operational boundaries. In power monitoring scenarios, many attacks do not immediately appear as malformed packets [19]. They first manifest as changes in communication peers, shifts in communication paths, or disruptions in periodic rhythms. For that reason, connection behavior analysis has substantial engineering value.

4.1 *Basic connotation of network connection behavior analysis*

Power monitoring refers to the continuous acquisition, transmission, processing, and utilization of operational information associated with generation, transmission, distribution, and consumption facilities. Typical monitoring tasks include telemetry collection, equipment-status supervision, alarm management, event recording, synchrophasor acquisition, and support for state-estimation applications. The objective is to maintain observability of the physical power system and provide reliable information for operation, protection, and dispatch decisions. In power monitoring systems, "connection behavior" does not merely refer to session setup and teardown in the conventional networking sense. It also includes device-role relationships, communication directions, access frequencies, message organization patterns, and repeated temporal regularities. Kim et al. [20] pointed out that network communication in industrial control systems can be divided into two broad classes: communication triggered by human operations and event-driven automatic communication. In most ICS scenarios, however, periodic communication is dominant, which is one of the key features distinguishing ICS networks from general IT networks. Because supervision and control services are highly automated and stable, most normal traffic is predictable, and this predictability is exactly what makes connection behavior analysis feasible.

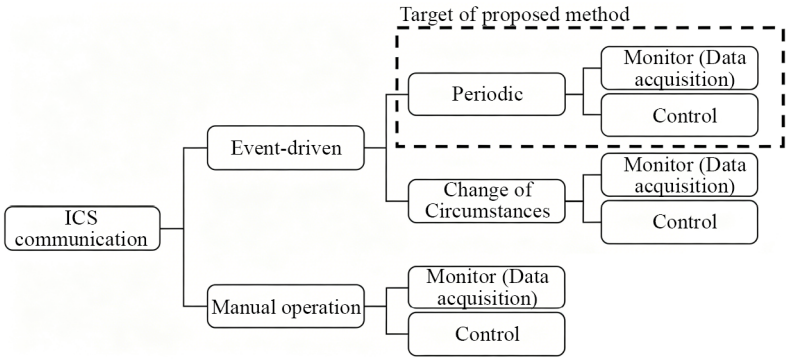


Figure 7. Classification of communication types in industrial control systems

As shown in Figure 7, normal communication in industrial control networks is not randomly scattered. It is jointly constrained by control logic, acquisition mechanisms, and automation processes. This characteristic is even more pronounced in power monitoring systems. Stable polling-response relationships usually exist between dispatch master stations and field terminals, long-standing reporting and control interaction chains exist between IEDs and higher-level station-control devices, and even non-periodic event-triggered messages generally follow identifiable service conditions and propagation paths. Accordingly, connection behavior analysis goes beyond conventional statistical feature extraction and instead emphasizes the establishment of a normal connection baseline: identifying which objects should communicate in a given environment, at what rhythm the communication should occur, and whether communication paths are structurally repeatable.

4.2 *Connection behavior analysis based on statistical features and temporal regularities*

In early studies of network connection behavior, a common approach was to characterize communication patterns using traffic statistics and temporal features such as packet counts, byte counts, duration, direction ratio, inter-arrival time, and period length. The main advantage of such methods lies in their low implementation cost and their ability to capture the "fixed peers plus fixed period" characteristics common in power monitoring systems. Kim et al. [20] pointed out that most communication in industrial control systems is periodic data acquisition, and that if stable periodic patterns can be extracted from long-term packet captures, normal behavior can be defined automatically and deviations can be recognized as anomalies.

In power monitoring systems, periodic communication is not just an incidental phenomenon; it is often a direct external manifestation of service logic. Periodic telemetry reporting, status synchronization, polling-based reading, heartbeat messages, and data refresh at fixed frequencies all imply that traffic presents relatively stable repetition on the time axis. For that reason, temporal regularity is not merely an auxiliary feature in connection behavior analysis, but often the key basis for determining whether communication remains normal.

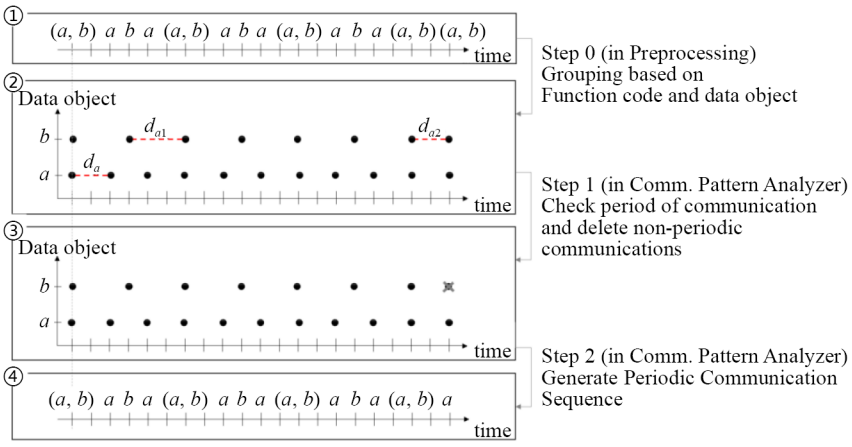


Figure 8. Four-step periodic analysis process in a communication pattern analyzer

As shown in Figure 8, communication-pattern analysis is not a matter of simply computing average inter-arrival times. In environments where multiple peers, function codes, and data objects coexist, it is necessary to separate the communications that exhibit stable periodicity step by step. The four-step analysis process given by Kim et al. [20] indicates that the key lies not in modeling all packets uniformly, but in first grouping them according to function code and data object, then validating periodicity within each group, and finally retaining the portions that represent normal operational rhythm.

4.3 *Graph-based behavior analysis*

When the monitoring target expands from individual connections to the overall communication relationship of the system, connection behavior analysis naturally develops into graph-based analysis. In such methods, devices, hosts, IEDs, service objects, or even flows themselves can be represented as nodes, while access, control, reporting, forwarding, or request-response relationships are modeled as edges. Traffic then becomes not a set of isolated temporal events but a structurally constrained relational network. Seo et al. [21] noted that graph-structure anomaly detection is centered on learning the normal connectivity patterns, centrality, paths, and subgraph structures of nodes and edges in a network so as to identify rarely occurring communication paths, lateral-movement behavior, or abnormal links.

In power monitoring systems, graph modeling is especially valuable because many anomalies do not arise from an illegal packet as such, but from communication relationships that should not exist under normal service logic. A field device that should communicate only with an upstream controller may suddenly begin communicating with a non-designated peer; a remote maintenance entry point may unexpectedly connect to a core IED; or a service may spread along paths that have never appeared before. These phenomena are better captured by communication graphs than by single-packet content.

As shown in Figure 9, once communication objects are abstracted as nodes and data-transfer relationships as directed edges, key relationships and potentially high-risk paths become much clearer. Liu et al. [22] used the number of connections as an important factor in critical-asset ranking and represented data-transfer directions by directed links. Although their study was oriented more toward asset criticality assessment, it provides a direct methodological inspiration for connection behavior analysis: connection density, directionality, and centrality are themselves important behavioral features, and anomalies involving highly connected or highly central assets may have impacts far beyond those of ordinary edge nodes.

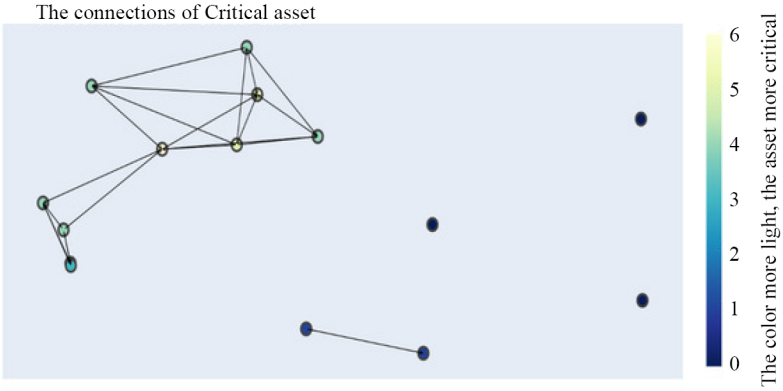


Figure 9. Illustration of connection relationships among critical assets

4.4 *Behavior analysis for asset profiling and critical node identification*

An important extension of connection behavior analysis in power monitoring systems is asset profiling and critical-node identification [22]. Rather than asking only whether an abnormal connection currently exists, asset profiling emphasizes long-term modeling of the behavioral characteristics of each device, server, IED, or communication node, including typical peers, common service types, direction distributions, periodicity, and network role. Under this view, the monitored object is no longer only a single flow, but whether a given asset continuously deviates from its normal behavioral profile.

Liu et al. [22] argued that with the increase of attacks against ICS, operators must understand the consequences of attacks on certain assets and identify which assets are more critical in the system as a whole. Their work not only discussed asset discovery, but also implemented a connection-based method for identifying critical assets and validated

its value under multiple attack scenarios. This idea is highly relevant to power monitoring systems: connection behavior analysis is not limited to abnormal-pattern detection, but can also support operational questions such as which assets deserve prioritized monitoring and which links should receive stronger protection.

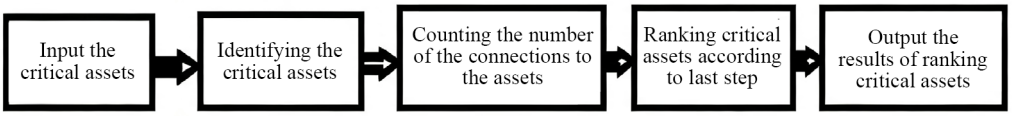


Figure 10. Critical asset analysis workflow based on asset discovery and connection relationships

As shown in Figure 10, asset profiling and critical-node identification generally involve multiple steps, including asset discovery, relation modeling, connection statistics, criticality ranking, and attack-impact analysis. This matters for a review because it shows that network connection behavior analysis has developed from a mere aid to detection into a foundational capability for situational awareness.

4.5 *Trends toward multi-layer feature fusion in connection behavior analysis*

With continued methodological development, connection behavior analysis in power monitoring systems is gradually evolving from single-feature analysis toward multi-layer feature fusion. Seo et al. [21] identified graph-structure anomaly detection as a dedicated method category in ICS/Cyber-Physical System (CPS) and summarized the main research objects, learning paradigms, model types, attack types, and datasets in that area. Their review indicates that graph-based methods are no longer confined to static relationship graphs. They increasingly incorporate dynamic adjacency, spatiotemporal features, and heterogeneous graph structures to represent complex relationships among hosts, flows, services, and time windows.

Table 4. Graph-structure-based network anomaly detection methods in ICS/CPS

Data characteristics utilized	D, C, P	Learning method	Model type	Attack type	Dataset
Graph-based	D	Sup + Fed	APPNP Graph Convolution + 1D-CNN	Distributed Denial-of-Service (DDoS)	CIC Modbus Dataset 2023, Edge-IIoTset
	D	Uns	Graph Neural Network (GNN)-based Hetero-SAGEConv	DDoS, Replay, Reconnaissance, Injection	CIC Modbus 2023
	D	Uns	Bloom + Hypergraph Kohonen	Injection, Denial-of-Service (DoS), Reconnaissance	Gas Pipeline, SWaT
	D, C, P	Uns	Ens	Intrusion, Error, Failure	ADFANet, AndMal17, CICIDS2017/2018, etc.
	D	Sup	NN	DoS, Bruteforce	Custom-built dataset

Note: D = Detection; C = Classification; P = Prediction; Sup = Supervised learning; Uns = Unsupervised learning

The trend reflected in Table 4 is clear: connection behavior analysis is no longer satisfied with manually constructed low-dimensional statistics. It increasingly depends on graph representation learning, heterogeneous relationship modeling, and spatiotemporal feature fusion to characterize the actual behavior of industrial networks. This development is highly consistent with the needs of power monitoring systems, where communication relationships are inherently heterogeneous and hierarchical. If hosts, flows, protocol services, and time windows can be jointly represented within a unified graph

model, connection behavior analysis becomes not only more capable of finding anomalies but also more capable of explaining them.

It is important to note, however, that multi-layer feature fusion does not render traditional statistical and periodicity-based methods obsolete. On the contrary, in strongly constrained service environments such as power monitoring systems, simple features such as periodicity, directionality, and fixed communication relationships often provide the strongest interpretability. A more promising direction is therefore to combine these explicit service- and physics-related features with graph representations, deep models, and asset profiling to form a comprehensive framework that is both interpretable and capable of recognizing complex anomalies [23].

5. Operational-state modeling and anomalous behavior detection in power monitoring systems

If protocol identification answers what traffic is present and connection behavior analysis answers whether those flows conform to established interaction relationships, operational-state modeling asks whether the system as a whole remains on a normal operating trajectory. This stage goes beyond packet content and connection topology. It learns how normal states evolve over time and uses that knowledge to recognize persistent, stealthy, or cross-layer anomalies. Figure 11 summarizes the main technical lineages of current ICS anomaly-detection methods and provides the conceptual framework for this section [24].

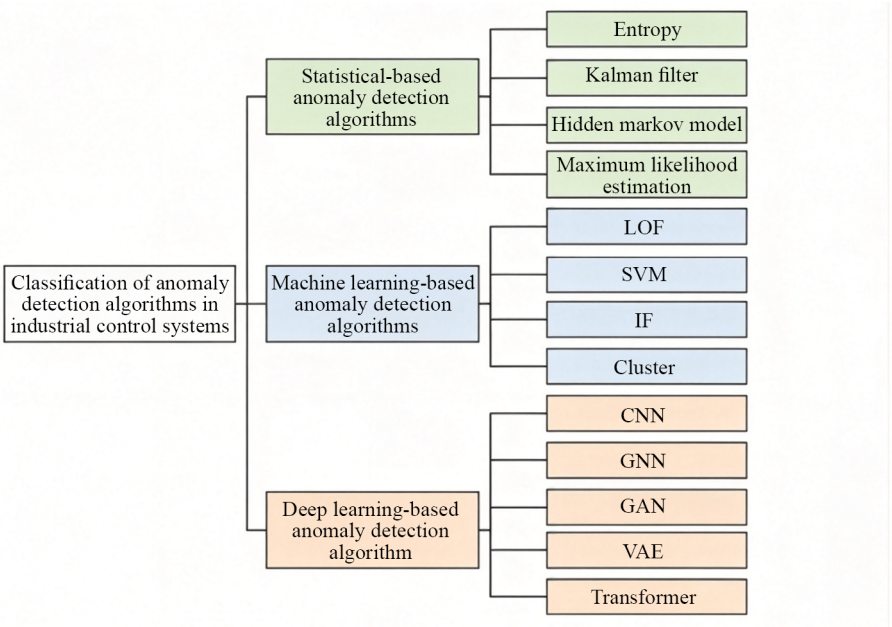


Figure 11. Main categories of anomaly detection methods for ICS

As shown in Figure 11, anomaly-detection methods for power monitoring systems are not isolated from one another. They all revolve around the central problem of how to characterize the normal operating state. Traditional methods explicitly define normality through rules, statistical distributions, or state-transition models. Machine-learning methods distinguish normal and abnormal behavior through boundaries, distances, or decision surfaces in feature space. Deep-learning methods exploit neural representation learning to infer normal operating patterns from multivariate time series, network traffic, or graph structures. For power monitoring systems, these categories each have different strengths: rule-based and statistical methods emphasize interpretability and engineering priors; machine-learning methods are suitable

when structured features are informative; and deep-learning and graph models are more appropriate for complex coupling, long-range dependency, and stealthy anomaly scenarios [21].

5.1 *Rule-based and statistical anomaly detection*

In industrial control and power monitoring environments, one of the earliest classes of anomaly-detection methods was built on rules, thresholds, and statistical modeling. Such methods assume that under normal operation the system follows relatively stable statistical distributions, boundaries, or ranges of variation. Once observations exceed a threshold, deviate substantially from a probability model, or exhibit abrupt statistical shifts, an anomaly is declared. Song et al. [25] noted that traditional anomaly-detection methods often depend on mathematical models constructed by domain experts to reflect normal system behavior. Representative methods include entropy-based techniques, Kalman filtering, hidden Markov models, and maximum-likelihood estimation. These methods detect anomalies by measuring the degree of deviation from normal behavior, but they usually depend heavily on expert knowledge and often assume that system states are observable or explicitly modelable, which is not always easy in real industrial systems.

Table 5. Statistical- and entropy-oriented anomaly detection methods based on network traffic in ICS/CPS

Data characteristics utilized	D, C, P	Learning method	Model type	Attack type	Dataset
Statistical based	D	Sup	GCNN, RS, RT, XGBoost	Port Scanning, Botnet/Malware, DDoS, C&C, Malware Delivery	IoT-23
	D	Sup	LSTM, XGBoost	DoS/DDoS, Brute Force	NF-BoT-IoT
	D	Sup, Uns	DAE, DFFNN	DoS/DDoS, APT, Spyware	NSL-KDD, UNSW-NB15
	D	Sup, Uns, Semi	Decision Tree (DT), Random Forest (RF), K-NN, SVM, etc.	DDoS	Infineon Factory Packet Capture (PCAP) + DDoSDB
	D	Uns	DAGMM	DoS/DDoS, Flooding Attack, Malicious Operation	ICS-CPS Testbed (Zhejiang Univ.), CIC-IDS2017
	D, C	Sup	AE, Principal Component Analysis (PCA), DT, DNN	Replay, DoS, False Data Injection, Command Injection, Flow Manipulation	SWaT, Gas Pipeline
	D	Sup	DPL-FSAD	DoS/DDoS, Probe, R2L/U2R, Fuzzing, Malware, Information Theft, IEC-61850 Protocol Attacks	UNSW-NB15, NSL-KDD, TON_IoT, ERENO IEC-61850
	D	Uns	ST-OCBLS	Unknown/Zero-Day, Protocol Abuse, Port Misuse, DoS, Probe	NSL-KDD, UNSW-NB15
	D	Semi	Lightweight 1D-CNN	Replay, DoS, False Data Injection (FDI), Command Injection, Stealthy Multi-Point Attack, Sensor Spoofing, Adversarial Evasion	SWaT, Battle of the Attack Detection Algorithms (BATADAL), Water Distribution (WADI)
	D	Sup, Uns	RF, SVM, Multilayer Perceptron (MLP), AE, K-Means	MITM, DoS	Cyber-Security Modbus ICS Dataset
	D, C	Semi	ADESSA	DoS, Probe, R2L, U2R	NSL-KDD, SWaT

Note: D = Detection; C = Classification; P = Prediction; Sup = Supervised learning; Uns = Unsupervised learning

For power monitoring systems, rule-based and statistical methods offer two prominent advantages. First, they are usually directly associated with service logic, physical boundaries, or control principles, and therefore their results are relatively easy to interpret in highly trusted monitoring environments. Second, they often entail low computational complexity and modest online-processing overhead, making them well suited to the real-time requirements of monitoring systems. Especially in dispatch master stations, substation automation systems, and protection devices, anomaly discovery

based on fixed logical relationships, boundary conditions, or simple statistical changes remains an important first line of defense.

At the same time, these methods have clear limitations. Stealthy attacks and slowly evolving anomalies often do not trigger obvious threshold violations at the outset. When operating conditions vary frequently, when multiple operating modes coexist, or when many variables are coupled in complex ways, single-variable thresholds and simple statistics are insufficient to describe the normal state completely. Rules and statistical models are also costly to maintain across scenarios. Therefore, while they remain important building blocks, they are rarely sufficient on their own in complex environments.

An important implication of Table 5 is that statistical and entropy-oriented methods remain well matched to communication-centric attacks, especially DoS, DDoS, scanning, probing, and traffic-distribution distortions that materially alter session counts, packet-length distributions, or port-usage ratios. Their limitations become more evident, however, when anomalies are deeply entangled with process evolution, control semantics, or long-horizon temporal dependencies.

5.2 *Hidden Markov Model (HMM)-based operational-state modeling*

Among state-modeling approaches, Hidden Markov Models (HMMs) are especially representative. The central idea is that the true operating state of the system is not directly observable, but can be inferred from observation sequences through hidden states and their transition probabilities. In power monitoring systems, where operation is often stage-based, ordered, and time-correlated, HMMs are particularly suitable for describing how normal behavior evolves across stable states. Once the observed sequence deviates significantly from the learned normal model in transition probability, observation probability, or overall sequence likelihood, the behavior can be flagged as anomalous. Song et al. [25] highlighted the value of HMMs in this regard: they can integrate anomaly cues from different detectors into a sequential classification process, thereby helping distinguish ordinary faults from cyberattacks.

As shown in Figure 12, HMM-based methods elevate detection from instant-by-instant judgment to sequence-level modeling. Rather than asking only whether a single sample is abnormal, they ask whether an entire observation sequence conforms to normal state evolution. This is highly relevant to power monitoring systems because many stealthy attacks work by gradually altering communication rhythms, message order, control-feedback relationships, or device operating patterns over an extended period. Compared with simple thresholding, HMMs are better at capturing time-dependent relationships and therefore perform well in identifying slowly evolving anomalies, order-disruption anomalies, and stage-wise attacks.

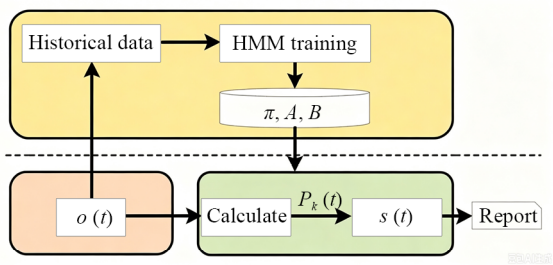


Figure 12. HMM-based anomaly classification mode

HMMs are also flexible in what they model. The observation layer can consist of traffic statistics, protocol events, device states, or even outputs of several preliminary detectors. Song et al. [25] explicitly adopted such a strategy: network-state and device-state detectors first generated anomaly observations, and an HMM then classified the sequence-level pattern to distinguish attacks from faults. Even so, HMMs require predefinition of state numbers and observation structures, and traditional HMMs may struggle when state spaces become very large or when nonlinear coupling is pronounced. They thus remain a classic but not exclusive tool for operational-state anomaly detection [26].

5.3 *Machine learning-based anomaly detection*

As monitoring data become richer and anomaly patterns more complex, machine-learning methods have been increasingly used for anomaly detection. Compared with rule-based or purely statistical methods, machine-learning approaches usually do not rely on explicit handcrafted thresholds [27]. Instead, they learn decision boundaries, distance structures, or distributions in feature space to separate normal from abnormal behavior. In power monitoring systems, such methods are especially suitable for structured measurement features, traffic statistics, and protocol-behavior features [28], and they often perform well when feature separability is reasonable and sample sizes are moderate [29].

A representative process-data-driven example is the measurement-based framework proposed by Mokhtari et al. [30] Rather than monitoring packets alone, their Measurement Intrusion Detection System (MIDS) framework uses system-wide measurement data collected across the ICS hierarchy and compares models such as KNN, decision trees, and random forests for anomaly detection. Their key point is that supervisory-layer measurement data can form a second defense line against stealthy attacks, complementing network intrusion detection rather than replacing it.

As shown in Figure 13, the general workflow of machine-learning anomaly detection includes data collection, feature selection, model training, and classification. In engineering terms, its value lies in mapping diverse observations from sensors, controllers, SCADA systems, and monitoring links into a common feature space and then using a classifier to discriminate normal from abnormal operation. Compared with HMMs, which emphasize temporal state transitions, machine-learning methods rely more heavily on the representation of the current sample or time window, and are therefore efficient when the boundary between normal and abnormal classes is relatively clear.

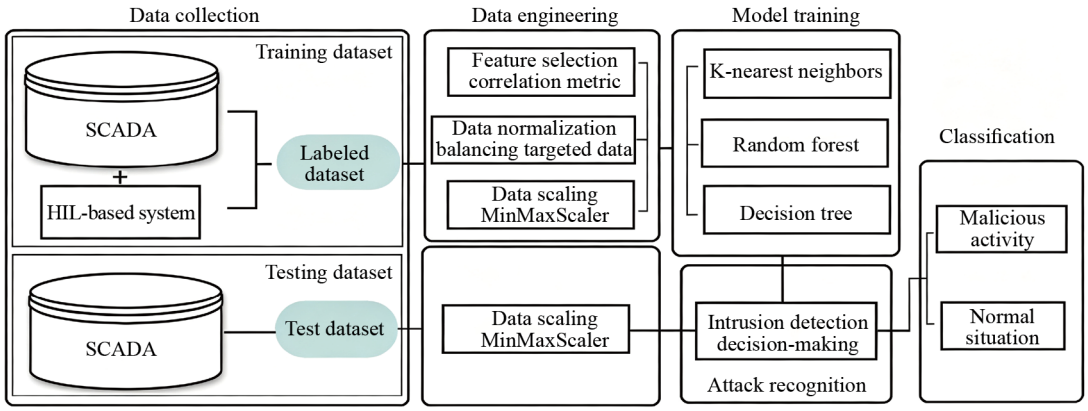


Figure 13. ICS anomaly detection workflow based on measurement data

Table 6. Performance comparison of machine-learning models using measurement data

	KNN	Decision Tree Classifier (DTC)	RF
Precision	0.9732	0.9937	0.9976
Recall	0.9729	0.9937	0.9976
F1-score	0.9729	0.9937	0.9976
Accuracy	0.9729	0.9937	0.9976
Area Under the Curve (AUC)	0.9729	0.9937	0.9976
Fitting time [s]	173	5.8	2.21
Prediction time [s]	104	0.0283	0.0505

Table 6 highlights a key engineering consideration: machine-learning-based anomaly detection should be assessed not only by detection quality but also by computational cost. Even when a model performs best on a benchmark dataset,

deployment value is limited if training overhead, inference latency, or hardware demand conflicts with online monitoring requirements.

5.4 *Deep learning-based anomaly detection*

When the detection target expands from low-dimensional features to high-dimensional multivariate time series, complex traffic patterns, and long dependency chains, deep learning begins to show stronger advantages. Song et al. [25] observed that deep learning has become a major research hotspot in ICS anomaly detection and can be broadly divided into prediction-based models and reconstruction-based models [31]. Prediction-based models, such as CNNs, LSTMs, GNNs, and Transformers [32], detect anomalies by forecasting future states or next-step values and examining prediction residuals [33]. Reconstruction-based models, such as autoencoders and Generative Adversarial Networks (GANs), learn compact representations of normal data and identify anomalies by reconstruction error [34]. This distinction is useful for power monitoring systems because it clearly separates two different conceptions of anomaly: whether normal dynamics can still be predicted correctly, and whether the current state can still be represented by a model of normal behavior.

From the perspective of process data, the main strength of deep learning lies in modeling temporal dependence and multivariate coupling. Seo et al. [21] further divided process-data-driven anomaly detection into prediction-residual methods, reconstruction-error methods, and sensor-correlation methods, and summarized many representative studies in each category. Prediction-residual methods are well suited to attacks that perturb physical processes without necessarily altering network statistics, such as false data injection, malicious parameter manipulation, and gradual sensor drift [35]. Reconstruction-error methods are effective for injection attacks, sensor or actuator faults, and zero-day anomalies [36]. Correlation-based methods target scenarios in which individual variables still appear normal but the dependency structure among multiple variables has been disrupted.

Table 7. Prediction-residual-based anomaly detection methods for process data in ICS/CPS

Data characteristics utilized	D, C, P	Learning method	Model type	Attack type	Dataset
Statistical based	D	Uns	MLP (Distributed Linear Deep Learning)	FDI, Replay, Command Injection, DoS	SWaT, WADI
	D	Uns	MLP + LSTM	DoS, Malicious Command, Script Injection, Parameter Injection	SWaT
	D	Uns	1D-CNN + Multi-head Self-Attention	FDI, Sensor Spoofing, Stealthy Attack	SWaT/WADI
	D	Uns	HMM	DoS, Noise Attack, Protocol Violation, Buffer Attack, Sensor-Actuator Physical Attack	Avionics Testbed, Consumer Robot Testbed-iRobot Create 2
	D	Uns	LSTM	Malicious Code Execution, Coordinated Stealthy Attack, Replay, Command Injection	Custom-built dataset
	D, P	Uns	Functional Mock-up Interface (FMI) + Functional Mock-up Reference Architecture for Control (FMRAC) + Adaptive Sliding Control (ASC)	Ramp Attack (Data Integrity Attack)	Offshore Wind Farm Benchmark
	D	Uns	Temporal State-Transition-based Process-aware IDS	Process-Oriented Attack	SWaT
	D	Semi, Uns	TCN + Trans	Global Anomaly, Contextual Anomaly, Seasonal Anomaly, Trend Anomaly	SWaT, WADI

Note: D = Detection; C = Classification; P = Prediction; Sup = Supervised learning; Uns = Unsupervised learning

Table 8. Reconstruction-error-based anomaly detection methods for process data in ICS/CPS

Data characteristics utilized	D, C, P	Learning method	Model type	Attack type	Dataset
Reconstruction-error-based	D	Uns	MM + AFA	Injection Attack	FCC Fractionator Simulation Dataset, BATADAL
	D	Uns	CNN + LSTM AE	Step Injection, Slope Injection, DoS, Injection	HIL-based Augmented ICS (HAI)
	D	Uns	NN-Oneclass	Sensor, Actuator Fault, FDI, Command Manipulation	SWaT
	D	Uns	VAE-LSTM	Sensor/Actuator Fault, False Data Injection, Command Manipulation	SWaT
	D	Uns	BiLSTM	Operational Anomaly, Contextual Sequence Deviation, Process Feedback Error	HAI
	D	Uns	KNN, Linear SVM, AE	Physical intrusion devices on CAN fieldbus	Custom-built dataset
	D	Uns	PCA, One-Class Support Vector Machine (OCSVM), Local Outlier Factor (LOF), kNN, Isolation Forest (IF)	FDI	Custom-built dataset
	D	Sup + Uns	AE + IF + XGBoost + RF + LSTM	Injection, Tampering, DoS, Reconnaissance	SWaT, Wind Turbine SCADA

Note: D = Detection; C = Classification; P = Prediction; Sup = Supervised learning; Uns = Unsupervised learning

Table 9. Sensor-correlation-based anomaly detection methods for process data in ICS/CPS

Data characteristics utilized	D, C, P	Learning method	Model type	Attack type	Dataset
Sensor-correlation-based	D, C	Sup	Deep RNN-based Attention, Deep LSTM, Deep Bi-LSTM	Injection, Spoofing Attack, Noise Attack	SWaT, WADI, GHL
	D	Uns	CNN, RNN, LSTM, GRU	Command Injection, Replay, DoS, Physical Process Manipulation	SWaT/WADI
	D	Sup	ARF, HAT	DoS, Spoofing, Command Injection, Physical Fault, Sabotage, Insider Attack	aNormalies/WDT/HAI
	D	Sup	GA-mADAM-LSTM	Injection, Tampering, DoS, Reconnaissance	SWaT/WADI
	D	Uns	GCN + LSTM VAE	Injection, Tampering, DoS, Multi-Stage	SWaT, BATADAL
	D, P	Sup + Uns	Super Learner Ensemble + Isolation Forest + BBN	Injection, Tampering, Sabotage, DoS	aNormalies
	D	Weak-Sup	C-ary Tree MIL Framework	Fault, Tampering, DoS, Overflow, Sensor Anomaly	SWaT/WADI/AIOPS/GHL
	D	Sup + Uns	Trans, RF	FDI, Setpoint Manipulation, Sensor Fault, Controller Parameter Attack	HAI
	D	Sup	SVM, KNN, MLP, DT, RF, Gaussian Naïve Bayes	Memory, Parameter Tampering	Custom-built dataset
	D	Sup	RF, KNN, SVM, NN	FDI	Gas Pipeline Dataset
	D	Sup	Extra Trees, AdaBoost	FDI	ICS Cyber Attack Power System (Triple-Class), IEEE 14-Bus FDI Dataset, IEEE 57-Bus FDI Dataset
	D	Uns	OCSVM	FDI, Command Injection	GRFICSv2
	D	Sup	OCSVM	FDI, Replay, Command Injection	Custom-built dataset
	D	Sup	DNN	FDIA	IEEE 118-bus Simulation
	D	Sup	KAN	DDoS, Reconnaissance, Replay, MitM Injection	SWaT, WADI, ICS-Flow
	D	Sup	CNN + Attn	FDI, Replay, Spoofing, Control Injection	SWaT, WADI

Note: D = Detection; C = Classification; P = Prediction; Sup = Supervised learning; Uns = Unsupervised learning

Taken together, the trends summarized in Tables 7-9 indicate that deep-learning-based anomaly detection is no longer concerned only with isolated time series. It increasingly emphasizes modeling the intrinsic coupling within multivariate systems. This is especially important for power monitoring systems, where anomalies often do not involve only one measurement point; rather, they disrupt coordination among multiple measurements, control variables, and communication links. Prediction-based methods capture whether the system evolves as expected, reconstruction-based methods capture whether the current state still lies on the manifold of normal behavior, and correlation-based methods capture whether the physical or control dependencies among variables remain intact.

At the same time, deep learning is not without drawbacks [37]. It usually requires large amounts of training data, whereas real anomalous samples in power monitoring systems are scarce. Although detection performance may be high, interpretability is often limited, making it harder for operators to trace anomaly sources. Complex models may also be expensive to train and deploy on edge devices or in real-time monitoring settings [38]. Therefore, in power monitoring systems, deep learning is often more appropriate as a higher-level analysis component or as part of a layered framework that also includes rules, statistics, and lightweight machine learning.

5.5 *Graph models and operational-consistency-based anomaly detection*

As the importance of multivariate dependency and topology in industrial monitoring systems becomes more apparent, graph models have become an increasingly important direction in operational-state modeling. Compared with methods based solely on time series, graph models can simultaneously represent spatial associations and structural dependencies among variables, devices, or subsystems, making them better suited to anomalies characterized by correlation collapse, topological inconsistency, or cross-node propagation. Seo et al. [21] pointed out that process-data-driven correlation-based anomaly detection increasingly uses graph neural networks to model spatiotemporal dependencies among sensors and actuators, and that operational-consistency methods explicitly incorporate physical laws, control logic, and reachability constraints into anomaly decisions [39].

This direction is particularly important for power monitoring systems because power systems are strongly constrained by physical laws and topology. Even small deviations at a single measurement point may break voltage, frequency, power-balance, or flow-consistency relationships. Operational-consistency methods therefore focus not merely on statistical deviation, but on whether observed behavior remains logically and physically possible. Such methods are especially valuable for stealthy false-data injection, command tampering, and deception of control chains.

Table 10. Operational-consistency- and physical-constraint-based anomaly detection methods in ICS/CPS

Data characteristics utilized	D, C, P	Learning method	Model type	Attack type	Dataset
Leveraging operational-consistency	D	Uns	Bayesian Network + FSM	FDI, Command Injection, Control Tampering, Sensor Spoofing	SWaT
	D, P	Sup	Rule-based Function Simulator, DT	FDI, Command Injection, Control Tampering, Replay	SWaT
	D	Uns	PM-SEN, PM-ACT, Extended State Reconstruction (ESR)	FDI, Command Injection, Replay, Stealthy	SWaT, WADI
	D	Sup	XGBoost	Value Manipulation, Replay, Fuzzy, Zero-Day	Seat Leon 2018 OBD-II Dataset, KIA SOUL Dataset
	D	Sup	Rule-based SRI + DNN	Adversarial Attack, Zero-day, FDI, DoS, Spoofing, Replay	SWaT
	D	Sup	CNN + LSTM + DNN	FDI, Coordinated, Stealthy, Ramp, Random	IEEE 37-bus Simulation

Note: D = Detection; C = Classification; P = Prediction; Sup = Supervised learning; Uns = Unsupervised learning

As indicated by Table 10, operational-consistency methods and general deep-learning methods are better viewed as complementary rather than competing. The former improve trustworthiness and interpretability by introducing explicit rules, control logic, and physical constraints, while the latter improve the ability to fit complex patterns through representation learning. For power monitoring systems, this combination has substantial practical significance because detection must be not only accurate, but also explainable, actionable, and associated with low false-alarm burden [40].

5.6 *Multi-layer fusion detection frameworks and section summary*

In summary, operational-state modeling and anomaly detection in power monitoring systems are moving from isolated point detection toward multi-layer fusion frameworks. Representative approaches extract anomaly evidence from network state, device state, and process state simultaneously [41], and then perform joint reasoning through temporal or probabilistic models to reduce missed detections and false alarms under any single perspective. The framework shown in Figure 14 exemplifies this trend: front-end modules perform anomaly perception at different layers, while the back end carries out integrated correlation and type discrimination. The methods with the greatest future potential are therefore not isolated models, but collaborative detection systems that jointly consider protocol semantics, connection relationships, operational states, and engineering interpretability.

As illustrated in Figure 14, multilayer fusion architectures are highly consistent with engineering requirements in power monitoring systems. Front-end rule-and statistics-based modules support rapid warning, HMM and machine-learning components perform intermediate state discrimination, and deep learning or graph-based modules are better suited to handling complex multivariate coupling and covert anomalies.

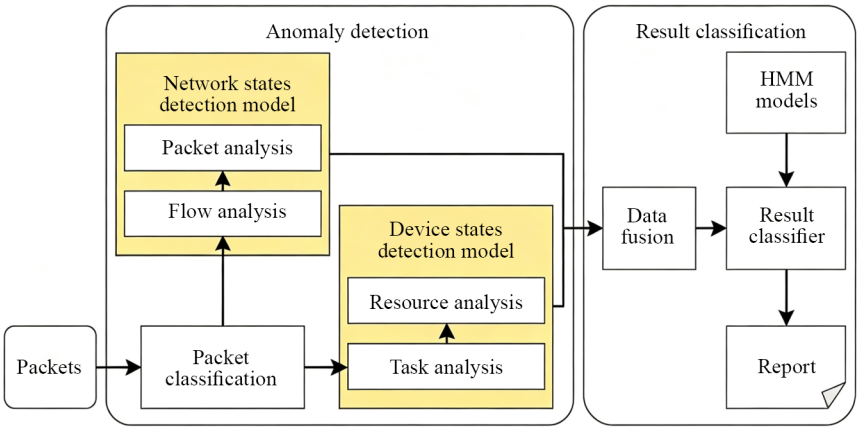


Figure 14. Multi-layer anomaly detection framework integrating network states, device states, and HMM classification

Table 11. A Comparative summary of representative anomaly detection methods for power monitoring

Category	Typical data	Accuracy	Interpretability	Cost	Scenarios
Statistical	Traffic statistics	Medium	High	Low	DoS, Scanning
HMM	Sequential events	Medium-High	Medium-High	Medium	Temporal anomalies
ML	Feature vectors	High	Medium	Low-Medium	Structured monitoring data
Deep learning	Multivariate time series	Very High	Low	High	Complex stealthy attacks
Graph-based	Topology & relationships	Very High	Medium	Very High	Dependency-aware detection
Operational-consistency	Process variables & constraints	High	Very High	Medium	FDI and process attacks

Table 11 provides a qualitative comparison of representative anomaly detection approaches from multiple practical perspectives, including data requirements, detection capability, interpretability, deployment cost, and application suitability. Rather than focusing solely on performance metrics, the comparison highlights the trade-offs that influence method selection in real-world power monitoring environments.

Statistical and rule-based approaches generally exhibit high interpretability and low deployment costs, making them suitable for detecting obvious anomalies such as scanning activities and denial-of-service attacks. However, their detection capability is often limited when facing complex or stealthy attack behaviors.

Machine learning methods improve detection accuracy by leveraging feature-based pattern recognition and are particularly effective for structured monitoring datasets. Deep learning approaches further enhance the ability to identify nonlinear and high-dimensional attack patterns but typically require substantial computational resources and provide limited explainability.

Graph-based methods offer significant advantages in modeling communication dependencies and network topology relationships, enabling the detection of sophisticated multi-stage attacks. Nevertheless, their deployment complexity and computational overhead remain relatively high.

Operational-consistency-based approaches provide strong interpretability because anomaly decisions can be directly linked to physical processes and operational constraints. These methods are especially effective for detecting false data injection attacks and process-level manipulations that may not be observable through network traffic alone. Overall, no single detection technique can simultaneously achieve optimal performance across all evaluation dimensions. The comparison suggests that future anomaly detection frameworks should combine multiple analytical perspectives to balance accuracy, interpretability, computational efficiency, and operational applicability.

More broadly, the methods discussed in this section can be regarded as a deeper extension of the protocol-identification and connection-analysis methods in the preceding sections. Those earlier sections mainly answered what is being transmitted and how it is transmitted, whereas this section addresses whether the system remains in a normal operating state and how anomalies manifest in state evolution. This line of reasoning also sets the stage for Section 6, which turns to datasets, evaluation metrics, and experimental design-elements that ultimately determine whether the effectiveness of these methods can be demonstrated convincingly.

5.7 *Discussion and research challenges*

The literature reviewed in this section reveals a clear evolution in anomaly detection for power monitoring systems. Early studies primarily relied on rule-based reasoning, statistical analysis, and manually designed thresholds, whereas recent research increasingly adopts machine learning, deep learning, graph neural networks, and operational-consistency-based approaches. Although these methods have significantly improved detection capability, several fundamental challenges remain unresolved.

First, a persistent gap exists between network-centric anomaly detection and process-aware anomaly detection. Many studies focus exclusively on either communication traffic or process measurements. However, real attacks against power monitoring systems often propagate across multiple layers, simultaneously affecting protocol behavior, communication relationships, and physical processes. As a result, detection models built on a single data source may fail to capture the complete attack lifecycle. Future research should therefore emphasize cross-layer modeling frameworks that jointly integrate protocol semantics, connection behavior, device states, and process variables.

Second, the increasing complexity of learning models raises concerns regarding explainability and operational trustworthiness. Deep neural networks and graph-based models often achieve excellent detection performance, yet their decision-making processes remain difficult for operators to interpret. In critical infrastructures such as power systems, anomaly alarms must be traceable and explainable because operational decisions may directly affect grid reliability. Consequently, combining data-driven learning with explicit operational constraints, physical laws, and engineering knowledge is likely to become an important research direction.

Third, the scarcity of publicly available power-specific datasets continues to hinder objective evaluation and fair comparison of detection methods. Most existing studies rely on datasets originally developed for general industrial control systems, such as SWaT, WADI, and HAI. Although these datasets provide valuable benchmarks, they cannot

fully represent the protocol characteristics, dispatching workflows, and operational constraints of actual power monitoring environments. The construction of realistic and standardized benchmark datasets therefore remains a key prerequisite for future progress.

Finally, deployment requirements impose additional constraints on practical applications. Power monitoring systems often require continuous monitoring, low latency, and high reliability. While advanced deep-learning and graph-based models may provide superior accuracy, their computational complexity and resource consumption can limit real-time deployment. Future anomaly detection systems will therefore need to balance detection performance, interpretability, robustness, and computational efficiency simultaneously.

Overall, the development trend observed throughout this review suggests that future anomaly detection frameworks will move beyond isolated algorithms toward integrated monitoring architectures. By combining protocol identification, connection-behavior analysis, operational-state modeling, and physical-consistency verification, such frameworks can provide a more comprehensive understanding of system security conditions and improve resilience against increasingly sophisticated cyber-physical attacks.

6. Typical datasets, evaluation metrics, and experimental design

For secure traffic monitoring in power monitoring systems, datasets, evaluation metrics, and experimental design are not peripheral issues. They determine whether reported conclusions are credible, whether results are comparable, and whether models have engineering transfer value. The protocol-identification, behavior-analysis, and state-detection methods discussed above ultimately have to answer practical questions: what data were used for validation, what metrics were used for evaluation, and how were the experiments designed? This section therefore summarizes typical datasets, evaluation metrics, and experimental design practices, while also highlighting common weaknesses repeatedly observed in the literature.

6.1 *Typical datasets and their applicable scenarios*

From the standpoint of data type, datasets commonly used in industrial control and power-monitoring anomaly detection can be divided into three broad categories [42]. The first consists of network-traffic datasets that describe protocol messages, traffic statistics, and communication relationships and are suitable for attacks such as reconnaissance, denial-of-service, man-in-the-middle, and protocol abuse. The second consists of process or operational datasets, including sensor, actuator, and control variables, which are more appropriate for false-data injection, command tampering, gradual drift, and physical-process anomalies. The third consists of hybrid or simulated datasets that jointly reflect network-layer and physical-layer attacks or generate rare scenarios in simulation environments [43]. Seo et al. [21] explicitly noted that network data are more suitable for "cyber-side" attacks, process data for "physical-side" anomalies, and hybrid data for cross-layer consistency analysis and multi-stage attack experiments.

Kim et al. [44] provided a concise summary of common public operational datasets, including ICS Cyber Attack Datasets, BATADAL, WADI, SWaT, and HAI. Although these datasets are not specifically tailored to power monitoring systems, they are widely used because they contain clear process backgrounds, temporal dependencies, and attack injection mechanisms. They are therefore still valuable for validating methods at the model-development stage. Nevertheless, the communication and control logic of water-treatment systems and general industrial testbeds cannot fully substitute for the protocol semantics and dispatch processes of power monitoring systems. If a study explicitly targets power scenarios, it should therefore use or supplement power-specific data whenever possible.

Although the datasets listed in Table 12 are not all specific to power monitoring systems, they remain widely used because they provide explicit process context, temporal dependence, and attack-injection mechanisms. For power-oriented studies, they are especially useful for validating model sensitivity to multivariate time-series anomalies, control-oriented attacks, and slowly evolving disturbances, while their limitations also remind researchers to supplement them with domain-specific power datasets whenever possible.

For power-system-oriented scenarios, some more targeted datasets have been reported. The BUT-IEC104-I dataset collected from a real smart-grid environment was described in a study on anomaly detection in IEC 104-based smart grids. It follows the communication characteristics of IEC 60870-5-104, contains 16 feature fields, and covers attacks such as DoS, injection, connection loss, and rogue-device behavior. Other studies have used open-source simulated power-system datasets. Other studies have used open-source simulated power-system datasets, one of which contains 37 event scenarios, 128 features, and 3,760 samples. Such datasets are useful for early-stage method comparison, but because simulated environments still differ from real monitoring networks, they should not be the sole basis for engineering conclusions.

Table 12. Representative operational datasets for industrial control systems

Name	Proposed	Attack points	Year
ICS cyber attack datasets	Tommy morris	32	2015
BATADAL	iTrust	14	2017
WADI	iTrust	15	2019
SWaT datasets	iTrust	41	2015
HAI datasets	NSR	50	2020

In recent years, hybrid and multi-domain datasets have become increasingly important. Seo et al. [21] pointed out that hybrid data combine network traffic and process data on a unified timeline so that communication-layer anomalies and physical-process anomalies can be analyzed together. Representative hybrid datasets include SWaT, WDT, and ICS-Flow. Such datasets better match the cross-layer propagation characteristics of real attacks and are particularly suitable for the protocol analysis-behavior analysis-state modeling line emphasized in this review.

Table 13. Summary of anomaly detection studies based on hybrid data fusion in ICS/CPS

Data characteristics utilized	D, C, P	Learning method	Model type	Attack type	Dataset
Data fusion-based	D	Uns	LSTM + GAN	MITM, DoS, Scan, Pump Failure, Sensor Breakdown, Leak	WDT
	D	Uns	LSP-DFA	DoS, Replay, Command Injection, Long-duration Attack, Masquerade Transition Attack, System Recovery Attack	Custom-built dataset
	D	Semi	CDF	FDI, DoS/DDoS, Replay, Command Injection, Spoofing	NSL-KDD, ICS, etc.
	D	Uns	GRU + AE	Stealthy, Non-stealthy, Amplification, Replay, FDI	Custom-built dataset
	D	Sup + Uns	CNN + DBSCAN + MPC Control	Cyber-Attack, Sensor Anomaly, Defect Detection	Custom-built dataset
	D, P	Sup	HMM	FDI, DoS, MITM, Buffer Overflow, Backdoor, Spoofing	Custom-built dataset
	D	Sup/Uns/Semi	RF, Extra Trees, GB, MLP, AE, LOF	DoS, FDI, Replay, Probe, Injection, Reconnaissance	Water Storage Tank, New Gas Pipeline, Power System, WADI, BATADAL, Tennessee Eastman
	D, P	Semi/Prob	Bayesian Network + Multi-modal Fusion + ATT&CK Mapping	Malware Injection, Lateral Movement, Privilege Escalation, Data Exfiltration	Power Terminal Network Simulation
	D	Sup + RL	Trans + LSTM AE + CANet + RF + SVM	DoS, MITM, Scan, Physic Fault	SCADA, WDT

Note: D = Detection; C = Classification; P = Prediction; Sup = Supervised learning; Uns = Unsupervised learning

Table 13 indicates that future dataset construction for power monitoring anomaly detection should move beyond isolated protocol traffic or isolated process variables. Greater long-term value lies in datasets that jointly cover protocol fields, network connections, device states, and physical-process variables on a synchronized time axis.

6.2 *Evaluation metrics for anomaly detection and their boundaries of applicability*

With regard to evaluation metrics, the most commonly used measures in industrial control and power-monitoring anomaly detection remain traditional classification metrics such as Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic-Area Under the Curve (ROC-AUC), and Precision-Recall-Area Under the Curve (PR-AUC). Other widely used measures, such as Kappa and G-mean, further underscore the field's reliance on classification-oriented performance assessment.

However, Kim et al. [44] specifically argued that time-series anomaly detection in ICS does not always fit well with conventional point-wise metrics. Industrial anomalies often have duration, attack windows, and recovery phases [44]. If performance is evaluated only at the point level, methods that cover long abnormal intervals while missing short anomalies or the onset of anomalies may still receive deceptively high scores. Moreover, after an attack ends, industrial systems may not return to normal immediately but may pass through a stabilization phase, which makes simple point-wise Precision and Recall insufficient to reflect actual detection quality.

As illustrated in Figure 15, anomaly-detection metrics can be roughly divided into point-wise and interval-wise categories. Point-wise metrics are derived from confusion matrices and suit scenarios where anomalous points are clearly labeled and appear in isolation. Interval-wise metrics emphasize how well an entire anomalous segment is covered, and are better suited to persistent attacks, slowly evolving anomalies, and industrial time series with recovery phases. For this reason, when the research target is long-term operational state or the evolution of anomalous intervals, interval-level metrics-or at least explicit reporting of window definitions, alarm delays, and anomaly-coverage ratios-should accompany conventional classification scores.

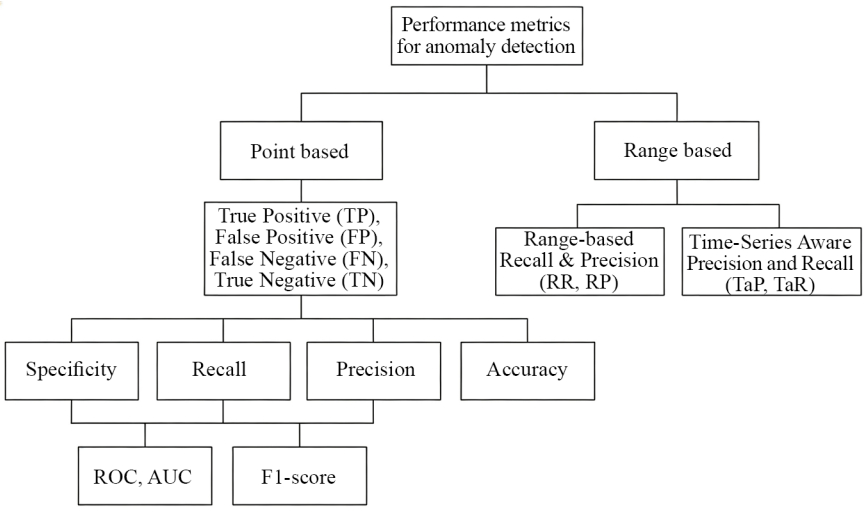


Figure 15. Performance metric system for industrial-control time-series anomaly detection

The point of Table 14 is to remind researchers that, in continuous process-oriented environments such as power monitoring systems, evaluation metrics should reflect engineering significance rather than only produce favorable scalar scores. A detector that reacts slightly late may be penalized heavily by point-wise metrics even when it successfully covers the operationally relevant attack interval.

Table 14. Example of performance misjudgment caused by point-wise evaluation

Method	Precision	Recall
1	0.67	0.50
2	1.00	0.625

6.3 *Key steps in experimental design and data preprocessing*

Compared with ordinary supervised-learning tasks, anomaly-detection experiments in industrial control and power monitoring systems require greater respect for the temporal structure of actual operation. The MADICS methodology explicitly proposed five steps for ICS anomaly detection: dataset preprocessing, feature filtering, feature extraction, anomaly-detection method selection, and validation [45]. A key point in that work is that ICS data are highly repetitive and temporally coherent, so data splitting and model training should preserve chronological order rather than arbitrarily shuffling samples.

As shown in Figure 16, a rigorous anomaly-detection experiment should not focus only on model selection. Sources of bias need to be controlled starting from the data layer. Training, validation, and testing sets should be split in time order to avoid leakage of future information. Categorical and continuous variables should be processed differently; for example, pumps and valves may require one-hot encoding, while continuous variables should be normalized or standardized. Features with zero variance, severe distributional inconsistency, or obvious instability should be filtered out to reduce noise.

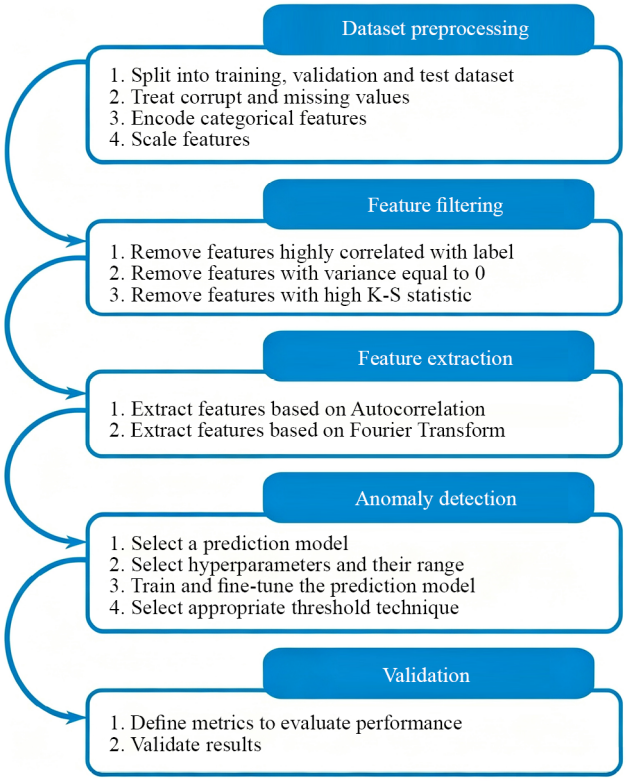


Figure 16. General experimental workflow for anomaly detection in industrial control systems

Another frequently overlooked but crucial issue is how to handle system warm-up and post-attack recovery periods. MADICS showed that at the beginning of data acquisition, ICS variables often go through an initialization process, and different variables require different times to reach stable values. If this warm-up period is directly included in the normal training set, the model may incorrectly learn unstable transitional behavior as normal. Similarly, data collected immediately after attacks may still reflect the process of returning to stability. Ignoring these stages can introduce label noise and impair model generalization. For power monitoring systems, analogous transitional phases may arise during station-control startup, communication-link establishment, periodic-task synchronization, or recovery after faults.

As shown in Figure 17, ICS variables do not necessarily enter a stable operating region immediately after data acquisition begins; a nontrivial initialization interval may be present. The same phenomenon often appears in power monitoring systems during control-center startup, communication-link establishment, and cyclic-task initialization, so warm-up segments should be isolated from the effective evaluation interval whenever possible.

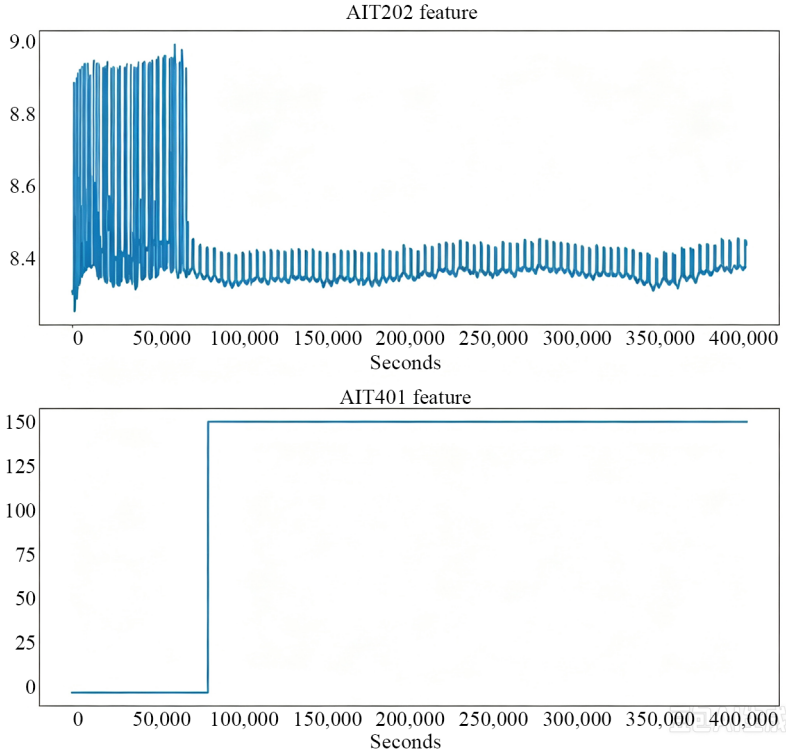


Figure 17. Example of feature stabilization during the warm-up stage of ICS data

6.4 *Common problems in current experimental research and recommendations*

Taken together, existing studies reveal four recurrent problems. First, many methods depend too strongly on a small number of public benchmarks such as SWaT, WADI, and HAI, which may lead to overfitting to a few scenarios and poor cross-system generalization. Second, a large number of papers still report mainly point-wise metrics, paying insufficient attention to persistent anomalies, alarm latency, and false-alarm duration. Third, details of preprocessing, windowing, and label construction are often underreported, making results hard to reproduce. Fourth, many models emphasize offline accuracy while discussing online deployment cost, latency, and alarm-handling burden only briefly or not at all.

For anomaly detection aimed at power monitoring systems, experimental design should therefore be more cautious. Data selection should ideally cover protocol traffic, connection behavior, and process variables simultaneously, and should prioritize datasets that match power-service logic. Evaluation should combine point-wise and interval-wise metrics and

should separately report detection latency, false-alarm persistence, and cross-scenario generalization whenever possible. Experimental procedures should disclose preprocessing, splitting, and labeling rules so that comparisons are transparent and reproducible. In short, experiments should not be treated as a peripheral appendix to a paper, but as a core component supporting the credibility of its conclusions.

6.5 *Comparative analysis of existing methods*

Existing secure traffic monitoring approaches can generally be divided into rule-based methods, traditional machine-learning methods, deep-learning methods, and hybrid approaches.

Rule-based methods usually provide high interpretability and low deployment complexity but exhibit limited adaptability to emerging attack patterns. Traditional machine-learning methods improve detection flexibility through statistical feature learning; however, their effectiveness depends heavily on feature engineering and data quality. Deep-learning methods are capable of automatically extracting complex traffic characteristics and often achieve superior detection performance, yet they generally require larger datasets and exhibit reduced interpretability. Hybrid approaches attempt to combine protocol semantics, behavioral features, and learning models, thereby improving robustness and detection coverage, although they may introduce additional computational and engineering complexity.

Overall, no single category currently achieves an optimal balance among detection accuracy, interpretability, real-time performance, deployment cost, and cross-scenario generalization.

7. Conclusion and future prospects

Overall, secure traffic monitoring in power monitoring systems has progressed from early rule matching and point-wise alarming toward a multi-layer methodological system in which protocol analysis, connection-behavior modeling, and operational-state detection are advanced in parallel. Yet several barriers still stand in the way of large-scale engineering deployment. High-quality power-specific data resources remain limited; cross-scenario generalization is still inadequate; stealthy anomalies and compound attack chains remain difficult to identify; and a stable balance between model interpretability and online deployability has not yet been achieved.

Several important challenges remain unresolved. First, publicly available power-specific datasets remain extremely limited, restricting the reproducibility and comparability of research results. Second, increasing adoption of encrypted communication mechanisms may reduce protocol visibility and complicate traffic analysis. Third, many existing methods are developed and evaluated under isolated protocol environments and therefore exhibit limited capability for cross-protocol and cross-scenario deployment. Fourth, the interpretability of AI-based anomaly detection remains insufficient for practical deployment in critical infrastructures where operational decisions require transparent reasoning. Fifth, digital-twin technologies provide new opportunities for security evaluation and anomaly verification, yet their integration with traffic-monitoring frameworks remains at an early stage. Finally, edge computing and distributed intelligence are becoming increasingly important because future monitoring systems must simultaneously satisfy requirements for real-time response, scalability, and deployment efficiency. Future work should focus less on pushing the accuracy of individual models in isolation and more on building an overall monitoring capability truly oriented to power-system scenarios. On the one hand, greater effort is needed to support the coordinated collection and sharing of protocol semantics, network traffic, connection relationships, device states, and process variables, thereby creating a data foundation closer to actual service processes. On the other hand, data-driven methods should be more tightly integrated with power-service logic, control laws, and physical constraints so that detection performance, interpretability, robustness, and transferability can improve together. Ultimately, the engineering value of the methods reviewed here will depend on whether they can provide stable support for security monitoring and risk handling in complex operational environments.

This paper has systematically reviewed the research foundations of secure traffic monitoring in power monitoring systems, key technologies for protocol analysis, methods for network connection behavior analysis, operational-state modeling and anomaly detection methods, and representative datasets and evaluation systems. The overall picture from the literature is clear: protocol analysis is the basis for security visibility, behavior analysis is the key to understanding

communication relationships, state modeling is the core of deeper situational awareness, and standardized data and evaluation systems determine the credibility of reported conclusions.

In addition, the manuscript was carefully reviewed to ensure consistency in terminology, abbreviations, figures, table descriptions, and the formatting of references. These improvements enhance readability and facilitate a clearer understanding of the literature reviewed.

Author contributions

Conceptualization, S. B. and K.W.; methodology, S.B. and P.W.; software, T.G.; validation, P.W., T.G. and Z.H.; formal analysis, S.B. and P.W.; investigation, S.B., P.W., T.G. and Z.H.; resources, K.W. and G.W.; data curation, S.B. and T.G.; writing-original draft preparation, S.B.; writing-review and editing, P.W., Z.H., G.W. and K.W.; visualization, T.G. and Z.H.; supervision, K.W.; project administration, K.W.; funding acquisition, K.W. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data availability statement

The present work is a literature review, and no new datasets have been generated in this study. All data analysed are from publicly available published articles, which are duly cited within the text.

Acknowledgements

The authors would like to thank the Guangdong Provincial Key Laboratory of Digital Grid Technology for providing technical support and experimental environment, as well as the anonymous reviewers for their valuable comments and suggestions.

Conflict of interest

The authors declare no conflicts of interest.

References

[1] Y.-S. Cho and Y.-H. Choi, "Methodology for implementing the state estimation in renewable energy management systems," *Energies*, vol. 14, no. 8, p. 2301, 2021.

[2] S. P. Singh and S. P. Singh, "Optimal cost wide area measurement system incorporating communication infrastructure," *IET Generation, Transmission & Distribution*, vol. 11, pp. 2814-2821, 2017.

[3] A. F. Tandjaoui, "Pmu and wireless sensor placement in WAMS: A joint resolution," *Journal of Systems Architecture*, vol. 175, p. 103758, 2026.

[4] B. Said, N. Deivanayagampillai, K. Jacob, and V. Gobinath, "Fuzzy hidden markov model using aggregation operators," *Journal of Computational and Cognitive Engineering*, vol. 4, pp. 1-7, 2023.

[5] T. C. Xygkis, J. Löfberg, and G. N. Korres, "Investigation of optimal phasor measurement selection for distribution system state estimation under various uncertainties," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1-16, 2025.

[6] O. Almomani, A. Saaidah, A. Adel, A. Alzaqebah, M. Almaiah, and Q. Shambour, "Enhance URL defacement attack detection using particle swarm optimization and machine learning," *Journal of Computational and Cognitive Engineering*, vol. 4, pp. 296-308, 2025.

[7] Y. Hu, A. Yang, H. Li, Y. Sun, and L. Sun, "A survey of intrusion detection on industrial control systems," *International Journal of Distributed Sensor Networks*, vol. 14, no. 8, pp. 1-14, 2018.

[8] M. Shehab and L. Alzabin, "Evaluating the effectiveness of stealth protocols and proxying in hiding VPN usage," *Journal of Computational and Cognitive Engineering*, vol. 4, pp. 186-194, 2024.

[9] A. Andreu, "Operational technology security-a data perspective," *Network Security*, vol. 2020, no. 1, pp. 8-13, 2020.

[10] M. F. Guato Burgos, J. Morato, and F. P. Vizcaino Imacaña, "A review of smart grid anomaly detection approaches pertaining to artificial intelligence," *Applied Sciences*, vol. 14, no. 3, p. 1194, 2024.

[11] T. Krause, R. Ernst, B. Klaer, I. Hacker, and M. Henze, "Cybersecurity in power grids: Challenges and opportunities," *Sensors*, vol. 21, no. 18, p. 6225, 2021.

[12] T. C. Xygkis and G. N. Korres, "Optimized measurement allocation for power distribution systems using mixed integer SDP," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 11, pp. 2967-2976, 2017.

[13] N. P. Theodorakatos, R. Babu, and A. P. Moschoudis, "The branch-and-bound algorithm in optimizing mathematical programming models to achieve power grid observability," *Axioms*, vol. 12, no. 11, p. 1040, 2023.

[14] N. P. Theodorakatos, R. Babu, C. A. Theodoridis, and A. P. Moschoudis, "Mathematical models for the single-channel and multi-channel PMU allocation problem and their solution algorithms," *Algorithms*, vol. 17, no. 5, p. 191, 2024.

[15] S. Khandekar, R. Murarishetti, C. Z. Y. Caven, N. G. F. Eric, and J. Zhou, "Network-based real-time detection of data manipulation attacks in industrial control systems," *International Journal of Information Security*, vol. 25, no. 1, p. 12, 2025.

[16] S. Ashraf, M. H. Shawon, H. M. Khalid, and S. M. Muyeen, "Denial-of-service attack on IEC 61850-based substation automation system: A crucial cyber threat towards smart substation pathways," *Sensors*, vol. 21, no. 19, p. 6415, 2021.

[17] C. Yu, Z. Zhang, and M. Gao, "An ICS traffic classification based on industrial control protocol keyword feature extraction algorithm," *Electronics*, vol. 12, no. 21, p. 11193, 2022.

[18] E. Holasova, R. Fujdiak, and J. Misurec, "Comparative analysis of classification methods and suitable datasets for protocol recognition in operational technologies," *Algorithms*, vol. 17, no. 5, p. 208, 2024.

[19] N. P. Theodorakatos, R. Babu, and M. D. Lytras, "Avoiding the maratos effect in non-convex optimization through piecewise convexity: A case study in optimal PMU placement problem," *Algorithms*, vol. 19, no. 1, p. 11, 2026.

[20] S. Kim, W. Jo, H. Kim, S. Choi, D.-I. Jung, H. Choi, and T. Shon, "Two-phase industrial control system anomaly detection using communication patterns and deep learning," *Electronics*, vol. 13, no. 8, p. 1520, 2024.

[21] J. K. Seo, J. Lee, B. Kim, W. Shim, and J. T. Seo, "AI-based anomaly detection in industrial control and cyber‑physical systems: A data-type-oriented systematic review," *Electronics*, vol. 15, no. 1, p. 20, 2025.

[22] C. Liu, Y. Alrowaili, N. Saxena, and C. Konstantinou, "Cyber risks to critical smart grid assets of industrial control systems," *Energies*, vol. 14, no. 17, p. 5501, 2021.

[23] L. Xu, K. Shang, X. Zhang, C. Zheng, and L. Pan, "Multi-scale feature fusion-based real-time anomaly detection in industrial control systems," *Electronics*, vol. 14, no. 8, p. 1645, 2025.

[24] A. Houkan, A. K. Sahoo, S. P. Gochhayat, and P. K. Sahoo, "Artificial intelligence approach to intrusion detection in industrial control systems with real world dataset generation and model evaluation," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 307, 2025.

[25] Y. Song, B. Bu, and L. Zhu, "A novel intrusion detection model using a fusion of network and device states for communication-based train control systems," *Electronics*, vol. 9, no. 1, p. 181, 2020.

[26] D. Myers, S. Suriadi, K. Radke, and E. Foo, "Anomaly detection for industrial control systems using process mining," *Computers & Security*, vol. 78, pp. 103-125, 2018.

[27] J. Cho and S. Gong, "Dynamic data abstraction-based anomaly detection for industrial control systems," *Electronics*, vol. 13, no. 1, p. 158, 2024.

[28] X. Zhao, L. Zhang, Y. Cao, K. Jin, and Y. Hou, "Anomaly detection approach in industrial control systems based on measurement data," *Information*, vol. 13, no. 10, p. 450, 2022.

[29] T. K. Das, S. Adepu, and J. Zhou, "Anomaly detection in industrial control systems using logical analysis of data," *Computers & Security*, vol. 96, p. 101935, 2020.

[30] S. Mokhtari, A. Abbaspour, K. K. Yen, and A. Sargolzaei, "A machine learning approach for anomaly detection in industrial control systems based on measurement data," *Electronics*, vol. 10, no. 4, p. 407, 2021.

[31] M. M. Aslam, A. Tufail, and M. N. Irshad, "Survey of deep learning approaches for securing industrial control systems: A comparative analysis," *Cyber Security and Applications*, vol. 3, p. 100096, 2025.

[32] F. Wang, Y. Liu, Z. Qin, and F. Dong, "A transformer-enhanced LSTM framework for robust malicious traffic detection in industrial control systems," *Knowledge-Based Systems*, vol. 321, p. 113725, 2025.

[33] C. Han and G. Gim, "Time-series-based anomaly detection in industrial control systems using generative adversarial networks," *Processes*, vol. 13, no. 9, p. 2885, 2025.

[34] J. Lu, X. Yang, Y. Liu, H. Zuo, F. Zhou, T. Yu, D. Liu, T. Deng, and L. Luo, "Transformer-autoencoder-based unsupervised temporal anomaly detection for network traffic with dual prediction and reconstruction," *Applied Sciences*, vol. 16, no. 4, p. 2143, 2026.

[35] J. Liu, Y. Sha, W. Zhang, Y. Yan, and X. Liu, "Anomaly detection method for industrial control system operation data based on time-frequency fusion feature attention encoding," *Sensors*, vol. 24, no. 18, p. 6131, 2024.

[36] R. Shrestha, M. Mohammadi, S. Sinaei, A. Salcines, D. Pampliega, R. Clemente, A. L. Sanz, E. Nowroozi, and A. Lindgren, "Anomaly detection based on LSTM and autoencoders using federated learning in smart electric grid," *Journal of Parallel and Distributed Computing*, vol. 193, p. 104951, 2024.

[37] W.-H. Choi and J. Kim, "Unsupervised learning approach for anomaly detection in industrial control systems," *Energies*, vol. 7, no. 2, p. 18, 2024.

[38] C. Wang, H. Liu, C. Li, Y. Sun, W. Wang, and B. Wang, "Robust intrusion detection for industrial control systems using improved autoencoder and bayesian gaussian mixture model," *Mathematics*, vol. 11, no. 9, p. 2048, 2023.

[39] Q. Zhu, Y. Ding, J. Jiang, and S.-H. Yang, "Anomaly detection using invariant rules in industrial control systems," *Control Engineering Practice*, vol. 154, p. 106164, 2025.

[40] D. T. Ha, N. X. Hoang, N. V. Hoang, N. H. Du, T. T. Huong, and K. P. Tran, "Explainable anomaly detection for industrial control system cybersecurity," *IFAC-PapersOnLine*, vol. 55, no. 10, pp. 1183-1188, 2022.

[41] M. M. Aslam, L. C. De Silva, R. A. A. H. M. Apong, and A. Tufail, "An optimized anomaly detection framework in industrial control systems through grey wolf optimizer and autoencoder integration," *Scientific Reports*, vol. 15, no. 1, p. 27579, 2025.

[42] M. Conti, D. Donadel, and F. Turrin, "A survey on industrial control system testbeds and datasets for security research," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2248-2294, 2021.

[43] M. Boeding, M. Hempel, H. Sharif, J. Lopez, and K. Perumalla, "A flexible OT testbed for evaluating on-device implementations of IEC-61850 GOOSE," *International Journal of Critical Infrastructure Protection*, vol. 42, p. 100618, 2023.

[44] G.-Y. Kim, S.-M. Lim, and I.-C. Euom, "A study on performance metrics for anomaly detection based on industrial control system operation data," *Electronics*, vol. 11, no. 8, p. 1213, 2022.

[45] A. L. Perales Gómez, L. Fernández Maimó, A. Huertas Celdrán, and F. J. García Clemente, "MADICS: A methodology for anomaly detection in industrial control systems," *Symmetry*, vol. 12, no. 10, p. 1583, 2020.