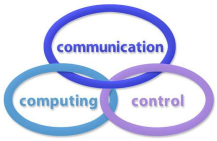
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 21, Issue: 5, Month: October, Year: 2026
Article Number: 7401, https://doi.org/10.15837/ijccc.2026.5.7401



CCC Publications

Attack-Chain Temporal Heterogeneity and Cross-Domain Detection Difficulty in IoT Security Datasets

Xiaoyu Zhao, Lei Bu, Xing Yang,Shufang He

Xiaoyu ZHAO
Geely University of China
Chengdu, Sichuan, 610000, China
zxy2570332345@gmail.com

Lei BU*
Zhejiang Dahua Technology Co., Ltd.
Chengdu, Sichuan, 610000, China
*Corresponding author:18382268541@163.com

Xing YANG
Geely University of China,
Chengdu, Sichuan, 610000, China
18980091368@163.com

Shufang HE
Geely University of China
Chengdu, Sichuan, 610000, China
heshufang@guc.edu.cn

Abstract

Intrusion detection models often show substantial performance discrepancies across network security datasets, even under identical feature representations and learning paradigms. Models that perform well on edge-oriented datasets frequently degrade when applied to large-scale and heterogeneous datasets such as ToN-IoT. Existing studies mainly attribute this issue to model limitations, while data-level temporal and structural characteristics remain underexplored. This paper proposes a data-centric attack-chain framework to analyze intrinsic detection difficulty across datasets. Using a unified attack chain construction protocol, we conduct a comparative analysis of Edge-IIoTset and ToN-IoT, focusing on attack chain length distributions, inter-event temporal intervals, stage transition dynamics, and structural complexity. Robustness is evaluated under multiple chain segmentation thresholds. Results show that ToN-IoT exhibits long-tailed chain distributions, sparse temporal organization, and degraded attack-type label transitions, which weaken stable temporal dependencies and hinder cross-dataset generalization. These findings provide a data-level explanation for performance degradation and highlight the importance of temporal attack chain analysis in intrusion detection evaluation.

Keywords: Intrusion Detection; Attack Chain Analysis; Temporal Heterogeneity; Dataset Bias; Data-Centric Security Analysis.

1 Introduction

With the expansion of Internet of Things (IoT) infrastructures and large-scale networked systems, modern network environments have become increasingly distributed, heterogeneous, and dynamic [1]. Correspondingly, cyberattacks have evolved from isolated events into complex, multi-stage processes unfolding over extended time periods. As a core component of network defense architectures, intrusion detection systems (IDS) are responsible for identifying malicious activities from massive volumes of network traffic. In recent years, machine learning and deep learning techniques have been widely adopted in IDS research and have demonstrated promising detection performance across multiple public benchmark datasets [2]. However, cross-dataset evaluations consistently reveal substantial performance inconsistencies, even when intrusion detection models are trained and tested under identical feature representations, model architectures, and training protocols. In particular, models that achieve near-perfect performance on edge-oriented datasets such as Edge-IIoTset often suffer significant performance degradation when directly applied to larger-scale and structurally more complex datasets, including ToN-IoT. Motivating empirical evidence for this phenomenon is illustrated in Figure 1.

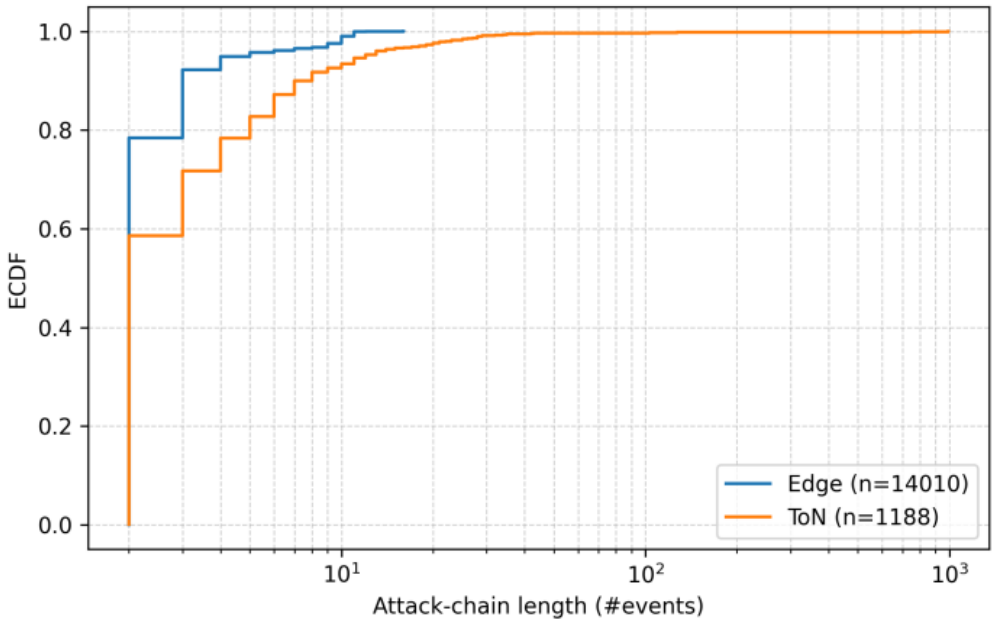


Figure 1: Motivating Evidence

Existing studies commonly attribute the aforementioned performance gap to factors such as domain shift, class imbalance, or insufficient model generalization capability. However, these explanations largely remain within a model-centric analytical paradigm and fall short of addressing a more fundamental question: why the data itself becomes difficult to learn. We argue that cross-dataset detection difficulty is, to a large extent, rooted in differences in the temporal and structural organization of attack behaviors. Specifically, different datasets may exhibit distinct attack-chain compositions, temporal distributions, and stage evolution patterns. These factors directly affect a model's ability to capture attack semantics and temporal dependencies. To this end, this work adopts a data-centric explanatory perspective on attack-chain analysis and proposes an analytical framework for characterizing and explaining the temporal heterogeneity of attack chains in network security datasets. Rather than improving detection models, the proposed framework aims to explain how dataset-level temporal organization constrains learnability. The overall workflow of the proposed framework is illustrated in Figure 2.

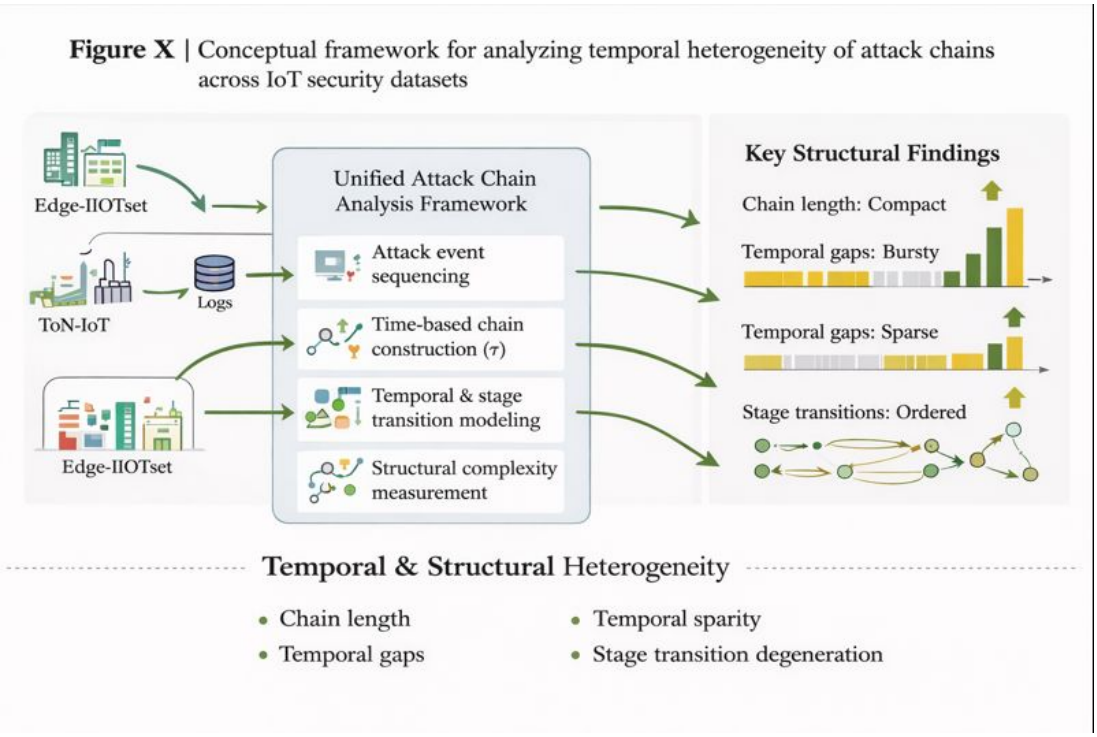


Figure 2: Conceptual framework for analyzing temporal heterogeneity of attack chains across loT security datasets

Under a unified and reproducible attack-chain construction protocol, the proposed framework consistently models attack events and conducts a systematic analysis along multiple dimensions, including attack-chain length, inter-event temporal intervals, attack-type label transitions, and structural complexity. In addition, a lightweight and stable learning behavior probe is introduced to reveal how different datasets respond to an identical learning process under consistent training conditions. The main contributions of this work are summarized as follows:

(1) We propose a unified and reproducible framework for attack-chain construction and analysis, with a focus on data-centric characterization of temporal organization.

(2) We systematically reveal significant heterogeneity in the temporal organization of attack chains across different IoT security datasets, highlighting its role as a key data-level factor affecting learnability.

(3) We provide a data-level explanation for the higher intrinsic detection difficulty of ToN-IoT, by identifying temporal sparsity, long-tailed structures, and degraded stage transition patterns as fundamental constraints on learnability.

(4) We design and implement a unified learning behavior probe, which isolates data-induced effects under controlled settings and supports explanatory analysis of cross-dataset generalization.

2 Related Work

2.1 Intrusion Detection and Cross-Dataset Generalization in IoT and Industrial Networks

With the continuous expansion of Internet of Things (IoT) and Industrial Internet of Things (IIoT) systems, network environments have become increasingly heterogeneous, characterized by complex communication patterns and diverse attack behaviors. Traditional intrusion detection approaches that rely on static rules or shallow feature representations have gradually exposed limitations in generalization under such environments. In recent years, machine learning and deep learning techniques have been extensively adopted for intrusion detection tasks and have achieved strong detection performance on multiple public benchmark datasets [3]. To support data-driven security research, the

community has developed a variety of publicly available datasets targeting IoT/IIoT scenarios, including ToN-IoT [**?**], Edge-IIoTset [6], Bot-IoT [7], IoT-23 [8], and CICIoT2023 [9]. These datasets differ substantially in terms of data collection environments, attack types, and labeling strategies, and have been widely used to evaluate the effectiveness of intrusion detection models across diverse application scenarios [10]. However, numerous studies have shown that models achieving superior performance on a single dataset often suffer significant performance degradation in cross-dataset evaluations. This phenomenon has been repeatedly observed in comparative studies involving classic benchmark datasets [11] and has been shown to be prevalent in IoT/IIoT contexts [12]. Recent work further indicates that variations in data collection environments, attack behaviors, and data organization strategies across different IoT/IoMT datasets can directly influence intrusion detection outcomes, highlighting the necessity of accounting for dataset-specific characteristics when analyzing detection performance [13].

2.2 Attack-Chain Modeling and Temporal Organization Characteristics

Attack chains and the Cyber Kill Chain have been widely employed to characterize the multi-stage evolution of complex cyberattacks and play an important role in threat analysis and defensive decision-making. Existing studies commonly model and analyze multi-stage attack behaviors using rule-based event correlation, attack graph models, or stage partitioning strategies that rely on prior knowledge [14]. Some works further adopt statistical analyses to characterize the overall distribution of attack traffic or events, as exemplified by analytical studies on datasets such as IoT-23 [15]. However, most attack-chain–related research primarily focuses on logical relationships between attack stages, path reachability, or attack pattern identification, while relatively little attention has been paid to the systematic analysis of how attack events are temporally organized at the dataset level. Characteristics such as temporal continuity of attack events, inter-event interval distributions, long-tailed attack-chain structures, and degraded stage transition patterns are often implicitly assumed to be comparable across different datasets. This assumption, however, remains insufficiently validated under cross-dataset settings [16]. The lack of explicit consideration of temporal organization properties limits the ability of existing attack-chain models to explain performance discrepancies observed in cross-dataset intrusion detection. Although sessionization or event aggregation mechanisms have been introduced in intrusion detection research [17], their primary objective is still to improve detection performance rather than to systematically analyze the learnability differences of attack-chain temporal structures themselves.

2.3 Dataset Bias and Data-Centric Security Analysis

In recent years, machine learning research has gradually shifted from a paradigm centered on model architectures and training strategies toward a data-centric analytical perspective, which emphasizes the decisive role of data distributions, structural complexity, and data collection processes in determining model performance [18]. In the field of network security and intrusion detection, existing studies have investigated issues such as dataset bias, class imbalance, and distribution drift, and have proposed domain adaptation, transfer learning, and federated learning techniques to mitigate performance degradation under cross-domain settings [19].

In addition, to address concept drift and long-term distributional changes, some studies have introduced memory replay or online updating mechanisms to enhance model robustness in dynamic environments [20]. Although these approaches improve detection performance under cross-domain or cross-temporal conditions to some extent, their primary focus remains on model-level optimization or adaptation [19]. Relatively little attention has been paid to explaining learnability differences of attack behaviors across datasets from the perspective of data structure itself, which limits a deeper understanding of the mechanisms underlying cross-dataset intrusion detection failures.

2.4 Positioning and Research Paradigm of This Work

In summary, although extensive research has been conducted on IoT/IIoT intrusion detection models, attack-chain modeling techniques, and cross-domain adaptation strategies, there remains a

lack of in-depth and reproducible studies on how temporal and structural differences in attack chains systematically affect cross-dataset detection difficulty. Most existing work addresses cross-dataset performance degradation from the perspectives of model architecture design or training strategy optimization, while the temporal and structural organization of attack behaviors at the dataset level has received comparatively limited attention.

In contrast to prior studies, this work conducts a systematic comparative analysis of the temporal organization and structural dynamics of attack chains in the Edge-IIoTset and ToN-IoT datasets under a unified attack-chain construction protocol and controlled learning settings. Rather than proposing a new detection model, we adopt a data-centric analytical perspective and treat learning models as controlled learning behavior probes to reveal the intrinsic impact of attack-chain temporal and structural properties on detection learnability.

Unlike traditional attack-chain or session-based studies, which primarily aim to improve detection performance or describe attack behavior patterns, this work introduces a fundamentally different analytical paradigm.

Our framework is not designed to enhance model performance, but to explain learnability from a data-centric perspective. By fixing model architectures and training protocols, we isolate dataset-level temporal and structural properties as independent variables, enabling a causal interpretation of cross-dataset generalization gaps.

This paradigm shifts the focus from "how to design better models" to "how data organization constrains learning," thereby extending attack-chain analysis from descriptive modeling to explanatory analysis.

To further clarify the positioning of this study within the intrusion detection research landscape, Figure 3 provides a conceptual characterization of existing IDS research paradigms along two dimensions: model-centric versus data-centric and detection performance versus structural interpretability.

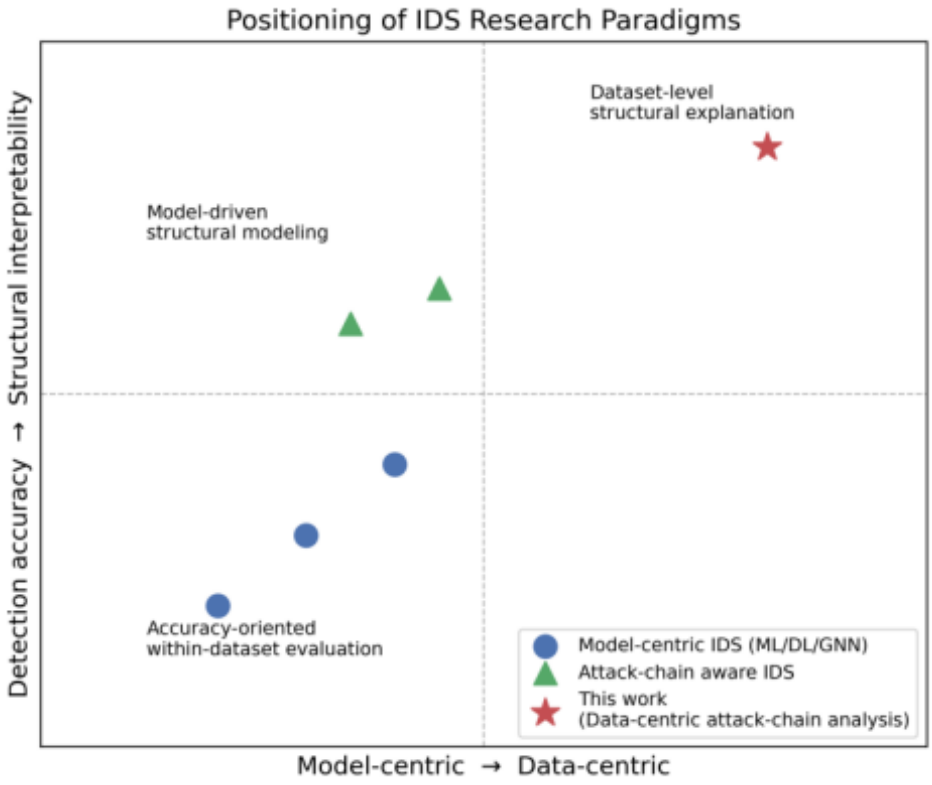


Figure 3: Positioning of intrusion detection research paradigms: model-centric vs data-centric and performance vs interpretability.

As illustrated in the figure, existing studies predominantly focus on model-centric performance optimization or attack-aware modeling strategies, whereas this work concentrates on the temporal

and structural organization of attack chains at the dataset level, and explains the fundamental causes of cross-dataset detection performance degradation through structured analysis.

3 Methodology

To systematically characterize differences in the temporal and structural organization of attack behaviors in network security datasets, this work proposes a data-centric attack-chain temporal heterogeneity analysis framework. Unlike research paradigms centered on detection performance optimization, the proposed methodology does not aim to improve model accuracy. Instead, by fixing learning conditions, it examines whether attack behaviors exhibit stable and learnable temporal structures at the data level, thereby revealing the structural origins of intrinsic detection difficulty across different datasets.

The overall analysis workflow is illustrated in Figure 4 and consists of three main stages:(1) data preprocessing and attack event modeling;(2) unified attack-chain construction;(3) analysis of attack-chain temporal and structural heterogeneity.The core idea of this workflow is to compare attack-chain characteristics across different datasets under fully consistent modeling and analytical rules, so that observed performance discrepancies can be attributed, as much as possible, to data structure itself rather than to model design or training strategies.

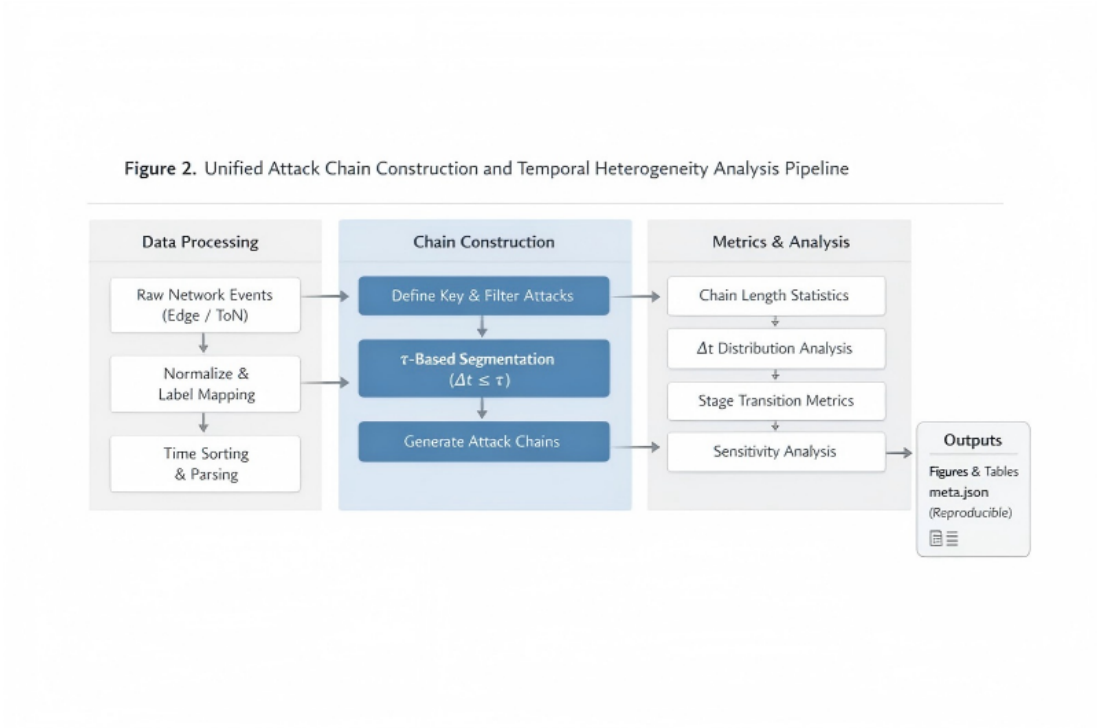


Figure 4: Unified attack-chain construction and temporal heterogeneity analysis framework

3.1 Problem Formulation and Analysis Objectives

In network security datasets, attack behaviors are typically recorded in the form of discrete events. Each event is associated with a timestamp, communication features, and labels indicating the attack type. Learning-based intrusion detection models generally make an implicit assumption that attack behaviors exhibit a certain degree of temporal continuity and structural stability along the time axis, enabling statistical learning methods to capture attack patterns.This assumption implies that when temporal continuity is violated, the stability of learned patterns may be compromised, leading to degraded detection and generalization performance.

Instead of focusing on detection capability, this study examines whether differences in temporal and structural organization of attack events disrupt the estimation of stable temporal dependencies,

thereby degrading detection and generalization performance. Let the set of attack events be defined as:

$$\mathcal{E} = \{e_i = (t_i, x_i, y_i)\}_{i=1}^{N}$$

where t_i denotes the timestamp of the i-th event, x_i represents its network traffic feature description, and y_i indicates the corresponding attack type label(used as the stage indicator in this study). An attack chain is then defined as an ordered sequence of events that satisfies temporal continuity constraints:

$$C = \{e_k, e_{k+1}, \ldots, e_{k+m}\}$$

The temporal continuity constraint between adjacent events is defined as:

$$\Delta t_i = t_{i+1} - t_i \leq \tau$$

where τ denotes the chain segmentation threshold. It should be emphasized that this work does not assume any specific value of τ to be optimal. Instead, the robustness of the observed differences is examined through subsequent sensitivity analysis.

3.2 Unified Attack-Chain Construction Protocol

To ensure the comparability of analytical results across different datasets, a unified attack-chain construction protocol that does not depend on dataset-specific characteristics is adopted throughout all experiments. Specifically, attack events are grouped within a consistent communication scope defined by source–destination pairs combined with protocol information (i.e., flow-level aggregation). This design minimizes the impact of subjective modeling choices on the results. In cases where identifiers are missing or ambiguous, events are filtered or assigned based on the closest available communication attributes to avoid incorrect chain merging.

Subsequently, events are grouped incrementally according to the temporal continuity principle: when the time interval between adjacent attack events exceeds the threshold τ, the current attack chain is segmented into an independent instance. The same procedure is applied across all datasets without dataset-specific adjustments.

The unified attack-chain construction protocol follows the principles below:Only events labeled as attacks are considered;A fixed temporal threshold τ is used for chain segmentation;Original attack type labels are retained as node attributes and directly used as stage indicators in subsequent transition analysis.

By adopting a unified attack-chain construction protocol, this work minimizes the potential impact of subjective modeling choices on the analytical results, ensuring that the observed differences primarily reflect intrinsic disparities in the temporal organization of attack behaviors across datasets.

It should be noted that the proposed attack-chain construction approach is conceptually aligned with the widely adopted sessionization modeling paradigm in intrusion detection research, which aggregates temporally adjacent attack events into semantically coherent attack instances to capture behavioral continuity and contextual structure.

Table 1 summarizes the key parameters and configuration settings used throughout the attack-chain construction and analysis process, thereby ensuring that cross-dataset comparisons are conducted under controlled and reproducible conditions.

3.3 Modeling Temporal Organization Characteristics of Attack Chains

The temporal organization of attack chains determines the length and continuity of temporal context available to learning models and constitutes a critical data-level factor influencing the learnability of intrusion detection. To characterize differences in the temporal organization of attack behaviors across network security datasets, this work models attack-chain temporal properties from two perspectives: the distribution of attack-chain lengths and the inter-event time intervals between adjacent attack events.

Table 1: Parameters for attack-chain construction and analysis

Parameter	Value / Range	Description
τ	$\{1, 2, 5, 10, 30\}$ s	Time threshold for attack-chain segmentation
ϵ	1×10^{-6}	Numerical constant to avoid log(0)
Feature Dim	42	Unified feature space after feature alignment
Model	Fixed MLP probe (3-layer, 256 hidden units, ReLU, dropout=0.2)	Controlled learning probe
Training Protocol	Fixed	Identical optimizer and training epochs across datasets

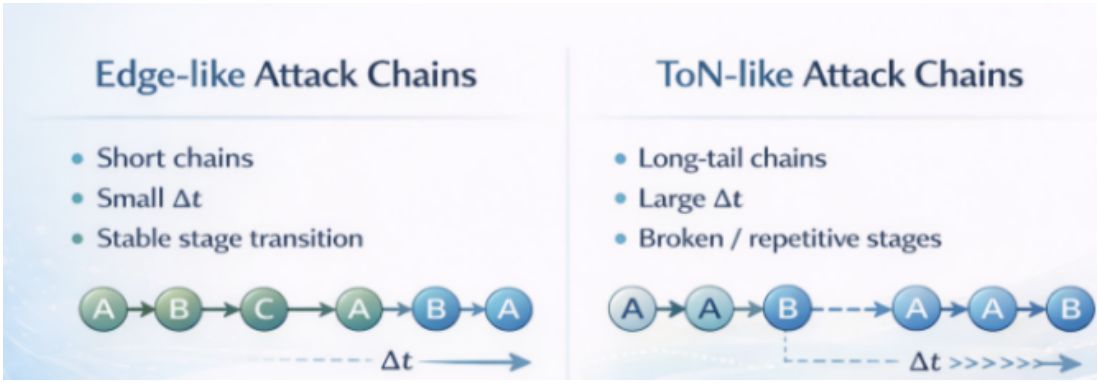


Figure 5: Conceptual illustration of temporal organization patterns in attack chains

Figure 5 provides a conceptual comparison between Edge-like and ToN-like temporal organization patterns of attack chains. The former typically exhibits compact and temporally continuous chains, whereas the latter is characterized by long-tailed, sparse, and irregular temporal structures. This intuitive contrast motivates the subsequent quantitative modeling and statistical analysis.

3.3.1 Modeling Attack-Chain Length

The length of an attack chain is defined as the number of attack events contained in a single chain:

$$L_C = |C| = m + 1$$

Attack-chain length reflects the ability of attack behaviors to form continuous temporal context along the time axis. The length of an attack chain is defined as the number of events it contains. It reflects the extent to which attack behaviors form continuous temporal context.However, excessively long-tailed or highly irregular chain distributions may introduce noise and reduce the consistency of temporal patterns, thereby negatively affecting learnability.

To characterize the long-tailed properties of different datasets, this work analyzes the attack-chain length distribution and its empirical cumulative distribution function (ECDF):

$$F_L(l) = \mathbb{P}(L \leq l) = \frac{1}{M} \sum_{j=1}^{M} \mathbf{1}(L_j \leq l)$$

3.3.2 Modeling Inter-Event Temporal Intervals

The inter-event time interval Δt_i describes the temporal density of adjacent attack events and serves as a key indicator of attack continuity. Larger inter-event intervals indicate increased temporal sparsity, which weakens the estimation of continuous dependencies.

To capture temporal differences spanning multiple orders of magnitude, inter-event intervals are modeled on a logarithmic scale:

$$g_i = \log_{10}(\Delta t_i + \epsilon)$$

where ϵ is a small constant introduced to avoid log(0).

3.3.3 Modeling Attack attack-type label transitions

In this study, attack stages are directly represented by the original attack type labels provided by each dataset.Beyond temporal scale, the evolutionary patterns of attack behaviors across different stages constitute an important structural assumption for learning models. To characterize the dynamic evolution of attack stages, an attack stage transition matrix is constructed as:

$$P_{ij} = \mathbb{P}(S_{t+1} = j \mid S_t = i)$$

This matrix describes the stage-level evolution of attack behaviors, and its structural stability directly affects a model's ability to learn attack paths.

3.4 Structural Complexity Metrics and Learnability Indicators

To further characterize the structural organization of attack chains, this work introduces multiple structural complexity metrics, which are treated as indirect indicators of attack behavior learnability.

Branching Factor: measures the number of alternative branches available in attack-type label transitions.

$$B = \frac{1}{|S|} \sum_{i \in S} |\{j : P_{ij} > 0\}|$$

Stage Entropy: reflects the uncertainty of stage evolution.

$$H = -\sum_{i \in S} \pi_i \sum_{j \in S} P_{ij} \log(P_{ij} + \epsilon)$$

where π_i denotes the stationary probability of stage i.

Repetition Ratio: measures the degree of adjacent stage repetition and is used to characterize long but low-information-density degenerate chains.

$$R = \frac{1}{L-1} \sum_{u=1}^{L-1} \mathbf{1}(S_u = S_{u+1})$$

These metrics serve as indicators of structural regularity and learnability of attack behaviors.Specifically, Stage entropy measures the uncertainty of stage transitions, where higher values indicate less predictable patterns.The repetition ratio captures structural redundancy, where excessive repetition may reduce effective information density and limit the learnability of attack patterns.

3.5 Sensitivity Analysis and Robustness Verification of the Chain Segmentation Threshold

Given that attack-chain construction relies on the temporal threshold parameter τ, a systematic sensitivity analysis is conducted over the following set of threshold values: $[\tau \in \{1, 2, 5, 10, 30\}\, s]$.

These values are selected to cover a range of temporal scales commonly observed in IoT environments, from short bursty communication patterns to more temporally sparse and intermittent attack behaviors.For each value of τ, attack-chain statistics and the corresponding stage transition matrices are generated, and key metrics are summarized across different thresholds to form a robustness evaluation table. To quantify whether relative differences vary with respect to τ, a metric difference function is further defined as:

$$\Delta_k(\tau) = m_k^{\mathrm{ToN}}(\tau) - m_k^{\mathrm{Edge}}(\tau)$$

where $m_k(\tau)$ denotes the value of the k-th statistical or structural metric (e.g., median chain length, inter-event interval quantiles, entropy) under threshold τ. If the sign and magnitude of $\Delta_k(\tau)$ remain consistent across different values of τ, the observed inter-dataset difference pattern is considered robust.

3.6 Datasets and Experimental Configuration

This study selects Edge-IIoTset, ToN-IoT, and UNSW-NB15 as representative network security datasets for analysis. UNSW-NB15 is included as a supplementary dataset to further validate the generality of the findings. All datasets contain timestamp fields suitable for temporal ordering as well as labeled attack information. To ensure fairness and reproducibility, all data are processed using a unified preprocessing pipeline, including timestamp normalization, feature field alignment, connection-state harmonization, and binary attack/normal label generation. The original attack-type labels are preserved after preprocessing and are directly used in the subsequent structural analysis.All statistical analyses are conducted under identical experimental environments and configuration settings.

3.7 Learning Behavior Probe and Analysis Pipeline

To establish a systematic link between learning-based detection behavior and attack-chain temporal structure, we construct a unified learning–analysis pipeline. The purpose of this pipeline is not to develop new detection models, but to use learning models as controlled probes under fixed training conditions, in order to reveal dataset-dependent differences in learnability.

By fixing feature representations, model architectures, and training protocols across datasets, the pipeline minimizes the influence of model design and training variability. Consequently, observed performance differences can be primarily attributed to intrinsic temporal and structural properties of attack behaviors.The overall pipeline, illustrated in Figure 6, consists of five stages:

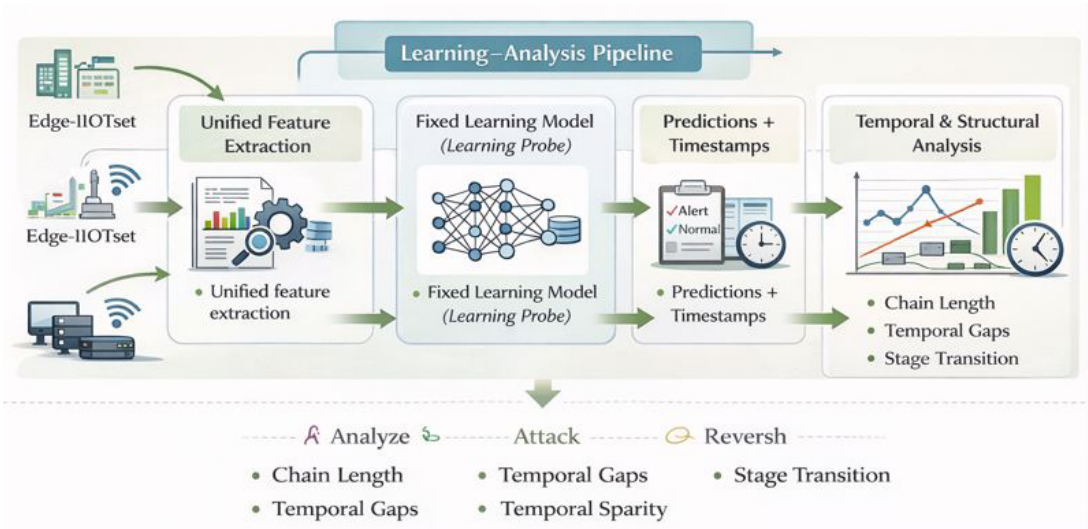


Figure 6: Learning-Analysis Pipeline for Data-Centric Attack Chain Analysis

(1) Feature Representation and Data Preparation

Raw data from different datasets are first aligned into a unified feature space by retaining only semantically consistent features. Numerical features are normalized, and missing values and outliers are handled using consistent preprocessing strategies. This ensures that all datasets are represented under comparable input distributions.

(2) Model Training

Learning models are trained under a fixed architecture and unified training protocol. The goal is not to optimize detection performance, but to use models as controlled probes that respond to dataset-specific temporal and structural characteristics. Identical configurations are used across datasets to ensure comparability.

(3) Prediction Generation

Trained models generate predictions for each event, which are recorded together with timestamps and attack type labels, forming the basis for subsequent analysis.

(4) Prediction-Based Attack-Chain Reconstruction

Predicted events are reorganized into temporally ordered attack chains following the unified construction protocol. This step transforms independent predictions into structured sequences with explicit temporal context.

(5) Temporal and Structural Analysis

The reconstructed attack chains are analyzed using the temporal and structural metrics defined in Sections 3.3–3.5. This enables systematic comparison of attack-chain organization across datasets. It also reveals its impact on learnability.Overall, the proposed pipeline provides a unified and reproducible framework for linking learning behavior with data-level temporal and structural characteristics, enabling a data-centric explanation of cross-dataset detection performance differences.

3.8 Probe Model Configuration

To ensure reproducibility, the learning probe used in this study is implemented as a fixed multilayer perceptron (MLP) under a binary classification setting, where all attack events are grouped into one positive class and normal events are treated as the negative class. Accordingly, Accuracy, Precision, Recall, F1-score, and AUC are all computed under the binary setting.

The probe model consists of three fully connected layers with hidden dimension 256. Hidden layers use ReLU activation, and dropout with a rate of 0.2 is applied after each hidden layer. The output layer produces a two-dimensional score, followed by softmax for binary classification.

Training is performed using the AdamW optimizer with a learning rate of $1e^{-3}$, batch size 1024, and 60 training epochs. The loss function is cross-entropy loss. Early stopping is not applied. To control randomness, all runs use a fixed random seed of 42 for model initialization, data shuffling, and training.

No additional class rebalancing strategy (e.g., reweighting or resampling) is applied. The purpose of this probe is not to maximize predictive performance, but to provide a stable and controlled learning process for comparing dataset-induced differences in learnability.

4 Results and Analysis

This section presents a series of progressively structured experiments that start from observable detection performance phenomena and gradually uncover the intrinsic differences in the temporal and structural organization of attack chains across network security datasets. By doing so, we provide a data-centric explanation for the degradation of cross-dataset detection performance from a structural perspective. It should be emphasized that the objective of the experiments in this section is not to evaluate or compare the optimal performance of different models, but rather to verify the prevalence of performance degradation under controlled learning conditions and to further analyze its structural causes at the data level.

4.1 Cross-Dataset Detection Performance Analysis Based on Learning Models

Before conducting an in-depth analysis of the temporal and structural characteristics of attack chains, we first examine detection performance to verify whether stable and significant performance discrepancies objectively exist across different network security datasets. To this end, representative learning-based intrusion detection models are evaluated under various dataset configurations using a unified feature space and a consistent training protocol.

The purpose of this subsection is not to establish a performance benchmark, but to confirm—through reproducible experimental results—the existence of cross-dataset performance degradation under controlled learning conditions, thereby providing direct motivation for the subsequent attack-chain-based structural analysis [9].

4.1.1 Experimental Setup and Data Partitioning

To ensure the comparability of experimental results, multiple representative training–testing configurations are designed and evaluated under identical feature representations and a unified training

pipeline, aiming to systematically characterize how different network security datasets influence model learning behavior. Specifically, the experimental settings include the following scenarios:

In-domain evaluation:

Training and testing are conducted on the same dataset, including Edge → Edge and ToN → ToN, representing standard detection scenarios in edge environments and large-scale complex network environments, respectively;

Direct cross-dataset evaluation:

Models are trained on one dataset and directly tested on another (e.g., Edge → ToN and ToN → Edge) to assess cross-dataset generalization capability without any adaptation [21];

Cross-dataset adaptation:

Models are first pre-trained on a source dataset and then fine-tuned using a limited number of samples from the target dataset, with evaluations performed on both source and target test sets to analyze the effect of transfer learning on cross-dataset detection performance [22];

Reverse validation:

The roles of source and target datasets are swapped for the above configurations to verify whether the observed performance discrepancies are direction-independent and structurally consistent.

Together, these experimental configurations cover a wide range of learning scenarios, from in-domain detection and direct cross-dataset evaluation to adaptation and reverse validation, enabling a systematic comparison of dataset-dependent learning behavior under identical learning conditions.

4.1.2 Detection Performance Comparison Results

To characterize learning behavior differences under both in-domain and cross-dataset conditions, Table 2 reports the detection performance of the probe model across multiple settings, including in-domain evaluations (Edge → Edge, ToN → ToN, and UNSW → UNSW), cross-domain transfers (Edge → ToN and ToN → Edge), and an adaptation scenario.The evaluation metrics include Accuracy, Precision, Recall, F1-score, and AUC.

Table 2: Detection performance under different dataset settings

Category	Train → Test	Accuracy	Precision	Recall	F1-score	AUC
In-domain	Edge → Edge	0.9991	0.9975	0.9999	0.9987	0.9998
In-domain	ToN → ToN	0.9512	0.9452	0.9580	0.9516	0.9896
In-domain	UNSW → UNSW	0.9599	0.9642	0.9560	0.9601	0.9932
Cross-domain	Edge → ToN	0.5029	0.4039	0.8006	0.5369	0.5940
Cross-domain	ToN → Edge	0.6325	0.6012	0.6648	0.6318	0.7124
Adaptation	Edge → ToN (fine-tuned)	0.7563	0.7128	0.7435	0.7278	0.8126

It can be observed that the model achieves stable performance under both in-domain settings. In contrast, a significant performance degradation occurs in the direct cross-dataset evaluation from Edge to ToN, indicating that cross-dataset generalization failure remains an objective and reproducible phenomenon even when the input representation and training protocol are strictly controlled [23].The adaptation results further show that partial performance recovery can be achieved through limited fine-tuning, although a substantial gap compared to in-domain performance still remains.

It should be noted that the experimental results presented in this study are based on a single representative learning model. This model is intentionally employed as a controlled learning behavior probe rather than as a performance-optimized detector. Under this setting, the observed performance degradation reflects the response of the learning process to dataset organization characteristics, rather than being attributed to model-specific architectural properties.

The results on the UNSW-NB15 dataset further validate the robustness of the findings, showing that the model achieves consistently high learnability under the same controlled setting. This indicates that the observed performance degradation in cross-dataset scenarios is not caused by model instability, but is closely related to dataset-specific temporal and structural characteristics.

4.1.3 Summary and Motivation for Further Analysis

In summary, when a learning model is trained on Edge-IIoTset and directly applied to the ToN-IoT dataset, consistent and significant detection performance degradation is always observed under controlled learning conditions. Specifically, although the model achieves stable and relatively high detection performance in both in-domain settings (Edge → Edge and ToN → ToN), its accuracy, AUC, and other key metrics exhibit a sharp decline in the direct Edge → ToN cross-dataset evaluation, with some performance indicators approaching the level of random discrimination.

Notably, this performance degradation occurs despite identical feature representations, model architectures, and training procedures, indicating that the phenomenon cannot be attributed to differences in model configuration or training strategies. Instead, it is more likely rooted in intrinsic differences in how attack behaviors are organized across network security datasets.

The stable occurrence of cross-dataset performance degradation under controlled learning conditions directly suggests that model generalization failure has a clear data-level origin rather than arising from incidental training instability or model-specific factors. This observation provides direct motivation for the subsequent analysis in this study.

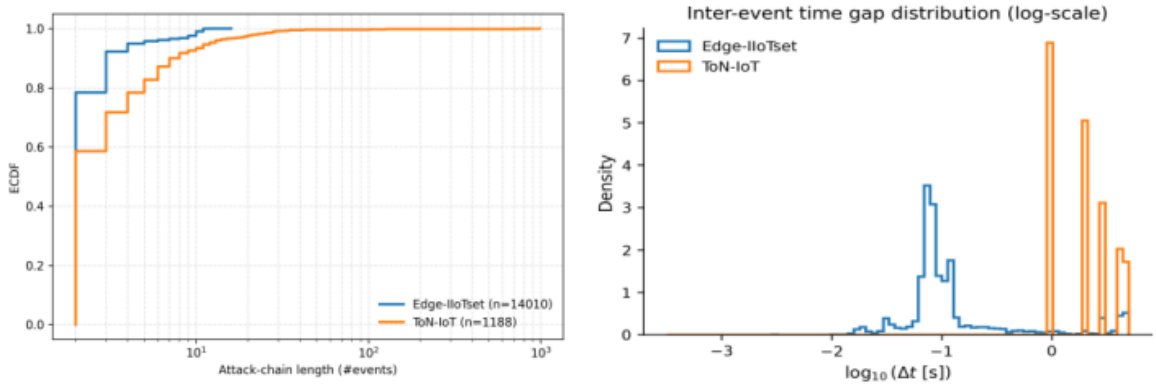
Based on these findings, the remainder of this paper shifts focus away from model-level explanations and instead examines differences in the temporal organization and structural complexity of attack chains. In particular, we investigate how intrinsic attack behavior organization in different network security datasets affects learnability, thereby revealing the structural sources of cross-dataset detection difficulty.

These observations further indicate that cross-dataset performance degradation is unlikely to be caused by insufficient model capacity or training instability. Instead, they motivate the hypothesis that fundamental differences in the intrinsic temporal organization of attack behaviors across datasets may violate the implicit learning assumptions of sequence-based detectors. This observation also suggests that different model classes may exhibit varying sensitivity to such temporal irregularities, which will be further discussed in Section 5.

4.2 Overall Temporal Structural Characteristics of Attack Chains

Long and temporally continuous attack chains provide stable sequential context for learning temporal dependencies. As illustrated in Figure 7(a), Edge-IIoTset consistently exhibits such structures under different conditions, whereas ToN-IoT fails to form comparable temporal contexts even when the time threshold τ is relaxed.

Under a unified attack chain construction protocol, we systematically compare the temporal structural characteristics of attack chains in Edge-IIoTset and ToN-IoT across different time threshold settings τ. The analysis focuses on attack chain length distributions and inter-event temporal interval characteristics, with the results shown in Figure 7.



(a) attack chain length ECDF (b) log-scale inter-event interval distribution

Figure 7: Temporal organization of attack chains across datasets

(1) Attack Chain Length Characteristics

Under all τ settings, Edge-IIoTset consistently exhibits a higher average attack chain length, with the emergence of extremely long chains as τ increases. This indicates strong temporal continuity among attack events, enabling dense, session-like aggregation over relatively wide time windows. Such temporal organization facilitates the preservation of multi-stage attack behaviors within a single attack chain, thereby providing learning models with stable and coherent sequential context.

In contrast, attack chains in ToN-IoT are substantially shorter overall, and their growth with increasing τ remains relatively limited. This observation suggests a weak temporal aggregation capability of attack events in ToN-IoT. Even when the chain segmentation constraint is relaxed, sustained and complete attack session structures are difficult to form. (2) Inter-Event Temporal Interval Characteristics Across all τ settings, the average inter-event time interval Δt in ToN-IoT is consistently larger than that in Edge-IIoTset and exhibits a monotonic increasing trend as τ increases. This indicates that attack behaviors in ToN-IoT are temporally sparser and more irregular, with events commonly separated by long idle periods and pronounced temporal discontinuities.

Taken together, these results reveal clear differences in the temporal organization of attack behaviors across datasets. In Edge-IIoTset, attack events tend to form continuous and compact temporal contexts, whereas the highly sparse and fragmented temporal structure in ToN-IoT weakens effective temporal continuity. As a result, multi-stage attacks are less likely to establish stable sequential dependencies within limited time windows, thereby reducing the ability of learning models to reliably capture attack patterns from a data-structural perspective.

To facilitate quantitative comparison of these temporal-structural differences, Table 3 summarizes key temporal and structural statistics of attack chains in Edge-IIoTset and ToN-IoT under $\tau = 5$ s.

Table 3: Statistical Comparison of Temporal and Structural Characteristics of Attack Chains ($\tau = 5$ s, with 95% bootstrap confidence intervals)

Dataset	#Chains	Median Chain Length	Mean Gap (s)	Median Gap (s)	Branching Factor	Stage Entropy	Repetition Ratio
Edge-IIoTset	820	1.00 (1.00–1.00)	0.0766 (0.01–0.15)	0.0027 (0.00–0.04)	1.2085 (1.18–1.24)	1.7777 (1.40–1.87)	0.1048 (0.09–0.12)
ToN-IoT	24087	1.00(1.00-1.00)	1.6307 (1.32–2.00)	1.0000 (1.00–2.00)	1.0665 (1.06-1.07)	1.5397 (1.36–1.62)	0.1257 (0.12–0.13)

Although both datasets exhibit a median chain length of 1, indicating that most attack chains consist of isolated events, clear differences arise in temporal gaps and structural metrics. Edge-IIoTset shows extremely small inter-event gaps, reflecting strong temporal continuity, whereas ToN-IoT exhibits much larger gaps, indicating sparse and fragmented temporal organization.

Moreover, the bootstrap confidence intervals are well separated across datasets, suggesting that these differences are not driven by chain-count imbalance but reflect intrinsic differences in attack behavior organization. The statistics in Table 3 provide data-level evidence of systematic temporal differences between the datasets and establish the basis for subsequent analysis of attack stage structures.

4.3 Analysis of Attack attack-type label transitions and Structural Dynamics

To further examine the evolutionary patterns of attack behaviors across different stages, this study constructs stage transition matrices constructed directly from the original attack type labels (treated as stage indicators) and conducts a comparative analysis of the dynamic stage structures in Edge-IIoTset and ToN-IoT. The results are illustrated in Figure 8.

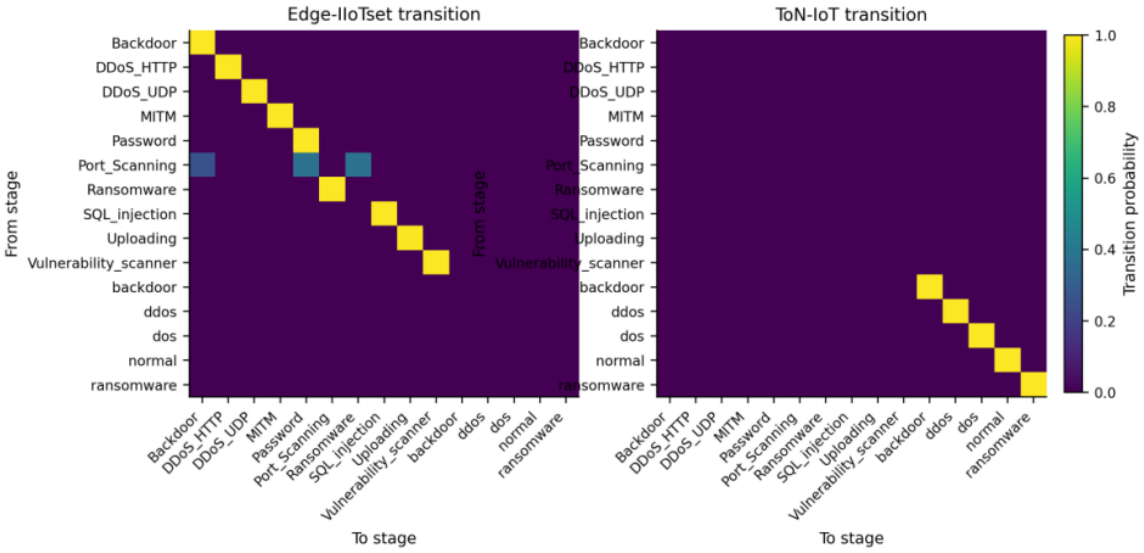


Figure 8: Attack Type Transition Matrices for Edge-IIoTset and ToN-IoT

In Edge-IIoTset, the attack stage transition structure exhibits relatively clear and stable sequential relationships, with transition probabilities concentrated along a small number of dominant paths. This structural property indicates strong regularity in attack behavior at the stage level, which facilitates learning models in capturing stage evolution patterns and forming stable discriminative representations.

In contrast, the stage transition matrix of ToN-IoT shows pronounced degenerative characteristics: transition probabilities among different stages tend to be more uniformly distributed, and clear dominant evolution paths are largely absent. This implies that the discriminative power of stage information is substantially reduced at the structural level, thereby weakening the effectiveness of attack type labels in supporting the model's decision-making process.

4.4 Structural Complexity and Heterogeneity Metrics

Building upon the analyses of chain length and attack-type label transitions, this study further provides a quantitative characterization of attack chains from the perspective of structural complexity. Figure 9 summarizes key indicators such as branching factor, stage entropy, and repetition pattern frequency of attack chains across different datasets.

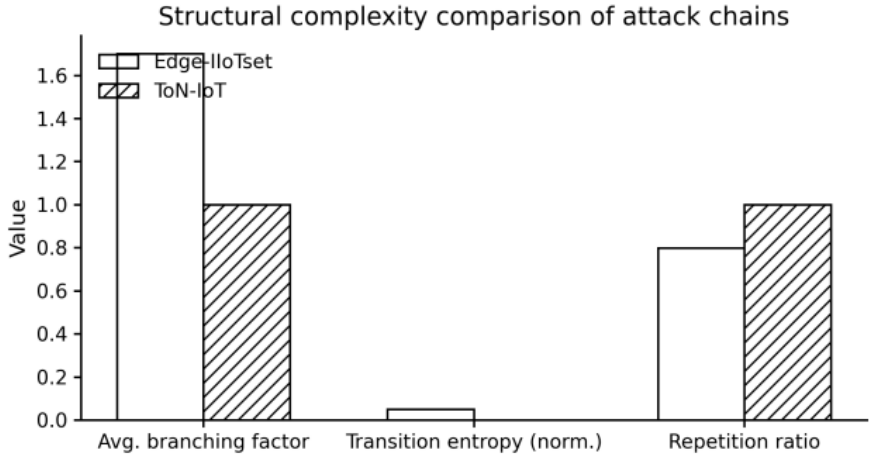


Figure 9: Structural Complexity Comparison of Attack Chains

The results show that ToN-IoT exhibits significantly higher branching factors and stage entropy

than Edge-IIoTset, indicating greater uncertainty in stage selection and attack evolution paths. Such high-entropy structures lead to more dispersed potential evolution trajectories of attack chains, thereby increasing the difficulty for learning models to capture stable and consistent structural patterns.

At the same time, a large number of repetitive or weakly discriminative stage sequences are observed in ToN-IoT, causing attack chains to exhibit a degenerative pattern characterized by increased length but reduced information density. This phenomenon suggests that greater structural complexity does not translate into richer discriminative information; instead, it further weakens the model's ability to exploit effective structural features.

Similar conclusions have been reported in related studies, which indicate that high temporal uncertainty in attack behavior and increased complexity of attack path structures substantially raise the difficulty of learning stable temporal and structural dependencies in intrusion detection tasks [24].

4.5 Sensitivity Analysis of the Chain Segmentation Threshold

To verify whether the above temporal and structural differences are influenced by the choice of attack chain construction parameters, we conducted a systematic sensitivity analysis on the chain segmentation threshold τ. Experiments were repeated under different settings of $\tau \in \{1, 2, 5, 10, 30\}$ seconds, and the corresponding results are summarized in Table 4.

Table 4: Sensitivity Analysis of Attack-Chain Statistics under Different τ (with 95% bootstrap confidence intervals for representative metrics)

τ (s)	Edge #Chains	ToN #Chains	Edge Median Gap (95% CI)	ToN Median Gap (95% CI)	Edge Stage Entropy (95% CI)	ToN Stage Entropy (95% CI)	Edge Repetition	ToN Repetition
1	1775	48996	0.0026 (0.00–0.09)	0.00 (0.00–1.00)	1.78 (1.06–1.79)	1.54 (1.47–1.58)	0.30	0.06
2	1188	39151	0.0027 (0.00–0.11)	1.00 (1.00–1.00)	1.78 (0.79–1.76)	1.54 (1.43–1.60)	0.07	0.09
5	820	24087	0.0027 (0.00–0.04)	1.00 (1.00–2.00)	1.78 (0.40–1.77)	1.54 (1.36–1.62)	0.10	0.13
10	444	14756	0.0027 (0.00–0.05)	2.00 (1.00–2.00)	1.78 (0.39–1.77)	1.54 (1.27–1.61)	0.14	0.14
30	27	2203	0.0027 (0.00–0.05)	2.00 (1.00–7.00)	1.78 (0.32–1.78)	1.54 (0.97–1.66)	0.36	0.24

Although different values of τ affect the absolute number of attack chains and their lengths, the relative differences between Edge-IIoTset and ToN-IoT remain highly consistent across all τ settings.ToN-IoT consistently exhibits larger inter-event gaps and lower stage-structure concentration, whereas Edge-IIoTset shows more compact temporal aggregation and more stable structural patterns. These differences are supported by the bootstrap confidence intervals in Table 4, which remain well separated across datasets despite large differences in chain counts and this indicates that the observed temporal heterogeneity is not driven by dataset size imbalance or parameter selection, but reflects intrinsic differences in attack-chain organization. Consistent effect size patterns across τ further confirm the robustness of these findings.Alternative grouping strategies (e.g., host-level aggregation) yield consistent relative patterns, despite variations in absolute statistics.

4.6 Summary of Results and Mechanism-Level Discussion

Based on the above experimental results, the following conclusions can be drawn at the mechanism level:

(1)Different network security datasets exhibit significant differences in the temporal and structural organization of attack chains, and such differences directly affect the ability of learning models to stably model attack behaviors;

(2)The temporal sparsity, degenerated attack-type label transitions, and high-entropy yet low-information-density structural characteristics observed in ToN-IoT jointly disrupt the learning of continuous temporal dependencies and dominant evolution paths from multiple perspectives;

(3)The observed temporal heterogeneity of attack chains remains robust across different chain segmentation thresholds, validating the reliability and stability of the analysis conclusions.

Overall, this chapter provides a data-structural explanation for why detection models that perform well on Edge-IIoTset struggle to generalize directly to large-scale network security datasets such as ToN-IoT. These findings further highlight the necessity of incorporating a data-centric, attack-chain temporal structure analysis perspective into the evaluation and design of intrusion detection systems, rather than relying solely on increased model complexity or enhanced feature engineering.

5 Discussion

This study conducts a systematic investigation into the intrinsic mechanisms underlying intrusion detection performance discrepancies across network security datasets, from the perspective of temporal and structural organization of attack chains. Unlike prior studies that predominantly attribute detection performance degradation to model architectures or training strategies, this work performs a controlled comparative analysis of Edge-IIoTset and ToN-IoT under a unified learning paradigm and identical feature configurations. The results demonstrate that even under fully controlled learning conditions, significant and stable performance differences persist, and their root cause lies in the intrinsic heterogeneity of attack-chain temporal organization and structural complexity.

From a mechanistic perspective, the temporal organization of attack chains directly determines the quality of sequential context available to learning models. In Edge-IIoTset, attack behaviors typically exhibit short and stable inter-event intervals, strong temporal continuity, and clear attack-type evolution patterns. These characteristics align well with the implicit assumptions of most sequence-based and structured learning models regarding stable dependencies, thereby facilitating the learning of consistent and reusable attack patterns. In contrast, attack chains in ToN-IoT commonly present long-tailed length distributions, sparse and irregular temporal gaps, and degenerated stage transition structures. Such properties introduce pronounced temporal discontinuities, which jointly undermine the model's ability to effectively capture attack context across multiple dimensions.

These findings provide a data-structural explanation for a long-observed but insufficiently understood phenomenon: detection models that perform well on edge-oriented datasets often fail to generalize directly to larger-scale and structurally more complex network security datasets. The inclusion of UNSW-NB15 further demonstrates that this pattern is not limited to the two primary datasets, but persists across datasets with different scales and collection settings.Furthermore, the results indicate that when attack chains exhibit high temporal and structural heterogeneity, improvements based solely on increased model complexity or optimized training strategies are often insufficient to overcome the fundamental learning bottlenecks imposed by data organization.

In addition, the identified temporal and structural heterogeneity may affect different model classes in distinct ways. Sequence-based models (e.g., RNNs or Transformers), which rely on temporal continuity and contextual dependencies, are more sensitive to irregular temporal gaps and disrupted attack-type label transitions, whereas non-sequence-based models may be less affected due to their reliance on local feature distributions.

This suggests that the impact of dataset-level temporal organization varies across model types, and that improving generalization in complex datasets may require model designs that explicitly account for temporal heterogeneity. This observation is derived from data-level analysis and has not been explicitly validated through comparative experiments.

Therefore, the key contribution of this work lies not in proposing a new detection model, but in revealing that dataset-level temporal organization acts as a fundamental determinant of learnability, thereby shifting intrusion detection research from a model-centric optimization paradigm toward a data-centric explanatory paradigm.

Despite the effectiveness of the proposed analysis framework in revealing intrinsic dataset differences, several limitations remain. First, attack-chain construction primarily relies on temporal continuity assumptions. Although multi-threshold sensitivity analyses confirm the robustness of the main conclusions, incorporating more sophisticated event correlation or causal modeling mechanisms may further enhance the fidelity of attack-chain representation. Second, this study focuses mainly on temporal and label-level transition structure of attack behavior, while deeper modeling of traffic semantic features remains an important direction for future extension.

From an engineering and practical perspective, these findings offer direct implications for the design and deployment of intrusion detection systems. In network environments where attack chains are temporally compact and well-organized, existing learning-based models can generally achieve stable detection performance. However, in scenarios characterized by highly sparse and disordered attack behavior, effective detection may require the integration of multi-scale temporal modeling, structural preprocessing, or data-level diagnostic mechanisms to mitigate the adverse effects of attack-chain temporal heterogeneity.

More specifically, these findings suggest that (1) dataset construction should preserve temporal continuity and realistic attack progression patterns; (2) preprocessing strategies such as temporal segmentation or adaptive time-windowing may help alleviate temporal sparsity; and (3) model design should explicitly incorporate temporal-aware mechanisms, such as sequence modeling or hybrid architectures that combine temporal and local feature representations.

6 Conclusion

This study systematically investigates the intrinsic causes of cross-dataset intrusion detection performance degradation from a data-structural perspective, with a particular focus on how the temporal and structural organization of attack chains affects detection learnability. The results indicate that the performance degradation observed on the ToN-IoT dataset is not merely attributable to increased attack-chain length or higher structural complexity. Instead, it primarily stems from fundamental differences in the temporal organization of attack behaviors.

Compared with Edge-IIoTset, ToN-IoT consistently exhibits larger inter-event time gaps and stronger temporal sparsity and irregularity across different time-threshold settings. Such temporal characteristics hinder multi-stage attacks from forming stable contextual representations within limited time windows, thereby weakening the model's ability to effectively learn and exploit sequential dependencies.

Further analysis demonstrates that even under strictly controlled conditions—where the feature space, learning paradigm, and training protocol are fully unified—detection models still struggle to achieve stable generalization in cross-dataset scenarios. This observation suggests that the intrinsic temporal structure of attack chains constitutes a critical data-level factor constraining detection performance, independent of model architecture or training strategy.

By adopting a data-centric analytical framework, this work provides a new perspective for understanding the long-standing challenges of cross-domain generalization in intrusion detection. The findings further imply that future research should incorporate systematic diagnostics of attack-chain temporal organization prior to model evaluation and method design, in order to support more reliable and interpretable conclusions regarding cross-dataset detection performance.

Funding

This work was supported by the 2025 Sichuan Provincial Education Digitalization Research Project (No. 2025LXKTPS393) and the 2025 Research Project on Computer Basic Education Teaching of the Association of Fundamental Computing Education in Chinese Universities (No. 2025-AFCEC-592).

References

[1] Montoya, G.A.; Lozano-Garzón, C.; Paternina-Arboleda, C.; Donoso, Y. (2025). A Mathematical Optimization Approach for Prioritized Services in IoT Networks for Energy-constrained Smart Cities, *International Journal of Computers Communications & Control*, 20(1), 6912, 2025.

[2] Hamdi, M.; Bouhamed, H.; Badreddine, F.; Alkanhel, R. (2024). Deep recurrent neural networks distributed on a Hadoop/Spark cluster for fall detection, *International Journal of Computers Communications & Control*, 19(3), 6428, 2024.

[3] Rahman, M. M.; Al Shakil, S.; Mustakim, M. R. (2024). A survey on intrusion detection systems in IoT networks, *Computer Science Review*, 45, 100082, 2024.

[4] Moustafa, N.; Ahmed, M.; Ahmed, S. (2020). Data analytics-enabled intrusion detection: Evaluations of ToN_IoT Linux datasets, *arXiv preprint*, 2020.

[5] Alsaedi, A.; Moustafa, N.; Tari, Z.; Mahmood, A.; Anwar, A. (2020). TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems, *IEEE Access*, 8, 179285–179300, 2020.

[6] Ferrag, M. A.; Friha, O.; Hamouda, D.; Maglaras, L.; Janicke, H. (2022). Edge-IIoTset: A comprehensive realistic cybersecurity dataset for IoT and IIoT applications, *IEEE Access*, 10, 3165809, 2022.

[7] Koroniotis, N.; Moustafa, N.; Sitnikova, E.; Turnbull, B. (2019). Bot-IoT dataset: A realistic botnet dataset in IoT for network forensic analytics, *Future Generation Computer Systems*, 99, 300–317, 2019.

[8] Ghani, H.; Salekzamankhani, S.; Virdee, B. (2020). IoT-23: A labeled dataset with malicious and benign IoT network traffic, *Zenodo*, 2020.

[9] Thakkar, A.; Lohiya, R. (2020). A review of the advancement in intrusion detection datasets, *Procedia Computer Science*, 167, 1890–1899, 2020.

[10] Sharafaldin, I.; Lashkari, A. H.; Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization, In *Proceedings of ICISSP 2018*, 108–116, 2018.

[11] Moustafa, N.; Slay, J. (2016). The evaluation of network anomaly detection systems: UNSW-NB15 dataset analysis, *Information Security Journal*, 25(1–3), 18–31, 2016.

[12] Al Nuaimi, T.; Al Zaabi, S.; Alyilieli, M.; AlMaskari, M.; Alblooshi, S.; Alhabsi, F.; Yusof, M. F.; Al Badawi, A. (2023). Comparative evaluation of intrusion detection systems on Edge-IIoT datasets, *Information Security Journal*, 32(2), 200–215, 2023.

[13] Doménech, J.; León, O. (2025). Evaluating and enhancing intrusion detection systems in IoMT, *Internet of Things*, 20, 101631, 2025.

[14] Stellios, I.; Kotzanikolaou, P. (2021). Assessing IoT-enabled cyber-physical attack paths, *Computers & Security*, 109, 102316, 2021.

[15] Ghadami, R. (2025). An intrusion detection system in IoT with deep learning, *Scientific Reports*, 15, 22074, 2025.

[16] Parlanti, T. S.; Catania, C. A. (2025). Temporal analysis of NetFlow datasets for intrusion detection, *arXiv preprint*, 2025.

[17] Yu, Y.; Long, J.; Cai, Z. (2017). Session-based network intrusion detection using deep learning, In *SecureComm 2017*, 200–212, 2017.

[18] Ng, A. Y.; et al. (2023). Data-centric artificial intelligence: A survey, *ACM Computing Surveys*, 56(7), 1–42, 2023.

[19] Wu, J.; Wang, Y. (2025). TriHID: Domain adaptation-based IoT intrusion detection, *Expert Systems with Applications*, 226, 129543, 2025.

[20] Andresini, G.; Pendlebury, F.; Pierazzi, F.; Loglisci, C. (2021). INSOMNIA: Concept-drift robustness in intrusion detection, In *ACM CCS 2021*, 1234–1248, 2021.

[21] Zhang, H.; Zhang, Z.; Huang, H.; Yang, H. (2025). Wasserstein distance guided transformer for intrusion detection, *Computers & Security*, 119, 104562, 2025.

[22] Sun, B.; Saenko, K. (2016). Deep CORAL: Correlation alignment for domain adaptation, In *ECCV 2016*, 443–450, 2016.

[23] Nasreen, F. A. H.; Khraisat, A. (2025). Adaptive memory replay for intrusion detection, *Computer Networks*, 239, 111712, 2025.

[24] Parlanti, T. S.; Catania, C. A. (2025). Temporal analysis framework for intrusion detection systems, *arXiv preprint*, 2025.

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Zhao, X.; Bu, L.; Yang, X.; He, S. (2026). Attack-Chain Temporal Heterogeneity and Cross-Domain Detection Difficulty in IoT Security Datasets, *International Journal of Computers Communications & Control*, 21(5), 7401, 2026.
https://doi.org/10.15837/ijccc.2026.5.7401