scientific reports

OPEN

Research on encrypted malicious traffic detection in power information interaction: application of the electricity multi-granularity flow representation learning approach

Zhifu Wu[1], Xianfu Zhou[2], Xindai Lu[3], Liqiang Yang[2]✉, Siqi Shen[3] & Dong Yan[2]

With the rapid digital transformation of power systems, encrypted communication technologies are increasingly adopted to enhance data privacy and security. However, this trend also creates potential covert channels for malicious traffic, making the detection of encrypted malicious traffic a critical challenge. Current detection methods often struggle to capture both fine-grained semantic interactions during the TLS handshake and global temporal patterns in traffic behavior, particularly in domain-specific contexts like power systems. This paper proposes the Electricity Multi-Granularity Flow Representation Learning (E-MGFlow) approach to address these issues. E-MGFlow integrates field-level and packet-level granularity analyses, leveraging a multi-head attention mechanism and bidirectional LSTM to effectively capture local semantic details and global traffic dynamics. The method is further optimized for power systems by incorporating device state information and bidirectional communication features. Experimental results on the DataCon dataset and a power information interaction dataset demonstrate that E-MGFlow significantly improves detection performance, achieving 93.64% precision and 93.76% recall with a low false positive rate of 6.52%. The approach offers substantial practical value for securing power system networks against sophisticated cyber threats, ensuring timely detection and defense against potential attacks.

Keywords Crypto malicious traffic detection, Multi-granularity representation learning, Power systems, Network security, Information interaction

Power systems, as critical infrastructures, face growing cybersecurity threats due to increasing digitization and connectivity[1]. The adoption of encrypted communication protocols, such as TLS/SSL, enhances data security but also creates opportunities for malicious actors to conceal their activities within legitimate traffic. In the power domain, detecting encrypted malicious traffic presents unique challenges that distinguish it from generic scenarios: (1) Obscured Traffic Analysis: Encryption hides payload details, rendering traditional detection methods reliant on content inspection ineffective. (2) Complex Traffic Patterns: Power systems exhibit a mix of periodic control signals and event-driven communications, complicating the identification of anomalies. (3) Stringent Real-Time Requirements: The need for uninterrupted service demands low-latency detection, yet many existing approaches are too resource-intensive for real-time deployment. (4) Heterogeneous Environment: The diversity of devices, protocols, and legacy systems in power networks hinders the development of unified detection strategies. Conventional detection methodologies, which rely on message content analysis, are rendered ineffective against encrypted traffic, thereby rendering network security detection a significant challenge within the encrypted traffic domain[2].

In the contemporary era, the proliferation of network traffic encryption technologies has necessitated the development of advanced detection methods for encrypted traffic. The academic community has responded by

[1]State Grid Corporation of China, Beijing 100032, China. [2]Huzhou Power Supply Company, State Grid Zhejiang Electric Power Co., Ltd., Huzhou 313099, China. [3]State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310007, China. ✉email: liqiangyang123@163.com

proposing a variety of detection methodologies, which can be broadly categorized into three distinct groups: decryption analysis-based methods[3], traffic statistical feature-based methods[4], and machine learning and deep learning-based methods[5]. (1) The earliest encrypted traffic detection relies on the decryption operation, which obtains the plaintext by decrypting the encrypted traffic, and then performs regular malicious traffic detection. Zhou et al.[6] decrypt SSL traffic and detect malicious content. Although this method has certain advantages in terms of accuracy, it has significant limitations in practical applications. First, the decryption process may compromise user privacy and disrupt the end-to-end security architecture. Additionally, it demands significant computational resources and time, making this approach unsuitable for large-scale real-time detection scenarios. (2) In order to avoid direct decryption, many researchers have turned to relying on statistical features of traffic for detection. Aladaileh et al.[7] designed a detection system based on entropy comparison by analyzing statistical features in encrypted traffic, such as traffic entropy, traffic packet size, and time interval. Although traffic statistical feature-based methods perform well in some scenarios, they often fail to capture the deep behavioral characteristics of malicious traffic. This limitation arises from their over-reliance on traffic features, which makes them ineffective against diverse encryption protocols, complex traffic patterns, or advanced attack tactics. (3) With the development of machine learning and deep learning techniques, researchers have begun to try to utilize these techniques to improve the accuracy of encrypted malicious traffic detection. Early machine learning methods, such as Support Vector Machine (SVM)[8] and Random Forest (RF)[9] based models, demonstrated good performance in the binary classification task of encrypted traffic. Liu et al.[10] proposed an incremental learning algorithm based on TLS, IP and DNS flow features. The algorithm is able to continuously update the feature database to cope with the ever-changing encrypted malicious traffic. However, these methods still depend on manual feature engineering. This reliance makes it challenging to adapt to diverse traffic patterns and results in poor performance when handling unseen samples or unknown attacks.

To overcome these problems, academics are gradually turning to automatic feature extraction models based on deep learning. Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models are widely used in the field of network traffic analysis. Zhou et al.[11] proposed a CNN-based crypto-malicious traffic detection method to analyze the traffic data by transforming it into grayscale images and achieved a high accuracy rate. Similarly, Li et al.[12] used an LSTM model to capture the temporal features of the traffic during cryptographic handshake, and demonstrated better generalization ability in detecting malicious traffic. Although deep learning approaches reduce the reliance on manual feature extraction to some extent, existing models primarily focus on global traffic behavior. They often overlook fine-grained semantic interactions during the cryptographic handshake. This may lead to degradation of detection performance under complex attack patterns.

Existing cryptographic malicious traffic detection methods still face a number of challenges when dealing with diverse and complex attack methods. First, decryption-based methods have significant problems in privacy protection and real-time performance, making it difficult to be applied on a large scale. Second, traffic statistical feature-based detection methods exhibit insufficient generalization capabilities when dealing with complex encryption protocols or dynamic traffic. Finally, although deep learning methods provide new ideas for detection, they lack effective modeling of local behaviors of traffic (e.g., key field interactions during handshaking), making it difficult to adequately capture multi-level traffic semantic information. To solve these problems, this paper proposes a cryptographic malicious flow detection method based on E-MGFlow. E-MGFlow integrates field-level and packet-level flow behaviors by modeling semantic representations at multiple granularities. It leverages the multi-head attention mechanism and the bidirectional long short-term memory network (BiLSTM) to capture both local and global interaction features, significantly enhancing detection accuracy and efficiency.

In this paper, we propose the Electricity Multi-Granularity Flow Representation Learning (E-MGFlow) method to tackle the challenge of detecting encrypted malicious traffic in power systems. Our approach offers the following key contributions:

(1) Multi-Granularity Representation Learning: Unlike most existing methods that rely on a single granularity, E-MGFlow integrates both field-level and packet-level analyses. This dual-granularity approach captures fine-grained semantic interactions during the TLS handshake using a multi-head attention mechanism and global temporal patterns in traffic behavior via bidirectional LSTM, enabling a richer representation of encrypted traffic tailored to power systems. This addresses a key limitation of prior work, where single-granularity methods often miss critical contextual details.

(2) Domain-Specific Optimization: While many existing approaches are generic and not optimized for specific domains, E-MGFlow incorporates power system-specific features, such as device state and power load data. This customization enhances the model's applicability and performance in real-world power system scenarios, filling a gap in current methods that typically overlook such domain-relevant information.

(3) Innovative Feature Fusion: E-MGFlow employs a linear layer to fuse local and global representations, surpassing basic stacking or concatenation techniques commonly used in prior work. This fusion method allows the model to learn weighted combinations of local and global features, capturing complex interactions that simple concatenation might miss, thus significantly improving detection accuracy for malicious traffic.

Related work

This section introduces the foundational techniques that form the basis of the Electricity Multi-Granularity Flow Representation Learning (E-MGFlow) algorithm: Long Short-Term Memory (LSTM) networks and the Multi-Head Attention mechanism. These established methods are critical for understanding the E-MGFlow framework and are presented here to provide the necessary context for the subsequent sections.

Long short-term memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) specifically designed to capture long-term dependencies in sequential data[13]. Unlike traditional RNNs, which often struggle with vanishing or exploding gradients, LSTMs incorporate a gating mechanism that allows them to retain or discard information selectively over time. This makes them particularly effective for modeling time-series data, such as network traffic, where temporal patterns are crucial.

An LSTM unit consists of three primary gates—input, forget, and output gates—that regulate the flow of information. The mathematical formulation of an LSTM unit at time step t is as follows. LSTM introduces a gating mechanism on the basis of recurrent neural network to better capture the timing information[14]. The formula of LSTM is as follows:

(1) Input gate

$$i_t = \sigma\left(W_i P_t + U_i h_{t-1} + b_i\right) \tag{1}$$

Where, i_t: the activation value of the input gate, which controls how much the current input affects the state of the unit.

W_i: input weight matrix, connecting the input data.

P_t: feature vector of the current packet.

U_i: hidden state weight matrix, connecting the hidden state of the previous moment.

h_{t-1}: the hidden state of the previous moment.

b_i: bias term of the input gate.

σ : sigmoid activation function.

(2) Forget gate

$$f_t = \sigma\left(W_f P_t + U_f h_{t-1} + b_f\right) \tag{2}$$

Where, f_t: the activation value of the forgetting gate, which controls how much the state of the cell at the previous moment affects the state of the current cell.

W_f, U_f, b_f: the same as the input gate, the weight and bias of the oblivion gate, respectively.

(3) Candidate states

$$\widetilde{c}_t = \tanh(W_c P_t + U_c h_{t-1}) \tag{3}$$

Where, $\widetilde{c}_t$: the candidate cell state at the current moment, representing the new information that will be added to the cell state.

W_c, U_c: weight matrices, for the current package and the previous hidden state, respectively.

(4) Unit state

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t \tag{4}$$

Where, c_t: the state of the cell at the current moment, indicating the state after control by the input gate and the forget gate.

c_{t-1}: the unit state at the previous moment.

(5) Output gate

$$o_t = \sigma\left(W_o P_t + U_o h_{t-1} + b_o\right) \tag{5}$$

Where, o_t: the activation value of the output gate, which controls how much the current cell state affects the hidden state.

(6) Hidden state

$$h_t = o_t \odot \tanh\left(c_t\right) \tag{6}$$

Where, h_t: the hidden state at the current moment, indicating the characteristic representation of the power flow at this moment.

LSTMs have been widely adopted in applications such as natural language processing, time-series forecasting, and network traffic analysis due to their ability to model complex sequential dependencies.

Multi-head attention mechanism

The Multi-Head Attention mechanism is a key component of transformer architectures, designed to capture relationships between different elements in a sequence by computing attention scores across multiple subspaces[15]. This allows the model to focus on various aspects of the input simultaneously, enhancing its ability to model complex interactions.

The attention function is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{7}$$

Where Q, K, and V represent the query, key, and value matrices, respectively. The Multi-Head Attention mechanism computes attention across multiple heads in parallel:

$$\text{MultiHead}(Q, K, V) = \text{Concat}\left(Z_1, Z_2, \ldots, Z_{\text{num\_heads}}\right) W^o \tag{8}$$

Where, where each head is:

$$Z_i = \text{Attention}(Q_i, K_i, V_i) \tag{9}$$

This mechanism has proven effective in various tasks, including machine translation, text classification, and, more recently, in analyzing sequential data such as network traffic.

Recent advances in encrypted data processing

Recent developments in the field of encrypted data processing have introduced a variety of innovative techniques aimed at enhancing privacy and security across diverse applications. For example, efficient privacy-preserving spatial range queries for outsourced encrypted data enable secure geospatial analysis without compromising user privacy[16]. This approach allows users to query encrypted spatial data stored on untrusted servers, ensuring confidentiality while supporting applications such as location-based services. Similarly, verifiable multi-keyword search schemes for encrypted cloud data have been proposed to facilitate secure and efficient data retrieval in cloud environments[17,18]. These methods guarantee the accuracy and verifiability of search results, addressing key challenges in encrypted data management. Both techniques underscore the increasing focus on privacy-preserving mechanisms in data processing, tackling issues related to secure querying and retrieval over encrypted datasets. However, while these approaches primarily concentrate on data access and retrieval, our work targets a distinct challenge: the detection of encrypted malicious traffic. The E-MGFlow method integrates multi-granularity representation learning to identify malicious activities within encrypted traffic, specifically tailored for power systems where security and real-time detection are critical. This differentiation establishes E-MGFlow as a novel contribution to cybersecurity, bridging the domains of privacy-preserving data processing and real-time threat detection.

E-MGFlow algorithm

This paper proposes the E-MGFlow, a multi-granularity flow detection framework based on multi-granularity representation learning, for analyzing encrypted flows in electric power information interactions. Figure 1 illustrates the overall architecture of the proposed Electricity Multi-Granularity Flow Representation Learning (E-MGFlow) framework. This figure visually demonstrates how the E-MGFlow framework integrates field-level and packet-level analyses to capture both fine-grained semantic interactions (e.g., during the TLS handshake) and global temporal patterns in the traffic.

The overall architecture is divided into three main components: (1) General framework (Fig. 1a) provides a high-level overview of the E-MGFlow framework. The input to the framework is the encrypted traffic data, which is processed through two parallel paths: field-level granularity analysis and packet-level granularity analysis. (2) Local behavior modeling based on word vectors (Fig. 1b) zooms in on the field-level granularity analysis. The encrypted traffic is first divided into sessions, and within each session, the handshake messages (e.g., Client Hello, Server Hello) are extracted. Key fields from these messages (e.g., encryption components, certificate information) are selected and converted into word vectors. (3) Global behavior modeling based on spatio-temporal features (Fig. 1c) focuses on the packet-level granularity analysis. The encrypted traffic is divided into packets, and the top 30 packets are selected to represent the global behavior of the traffic. Each packet is characterized by features such as packet size, time interval, transmission direction, power load information, and device state.

To provide a comprehensive understanding of the E-MGFlow algorithm, Fig. 2 presents a detailed flowchart of the entire process. The E-MGFlow algorithm processes encrypted traffic data to classify it as benign or malicious through a multi-step approach. It starts by receiving encrypted traffic data as input, which is then segmented into individual sessions for separate analysis. Within each session, handshake messages—such as Client Hello and Server Hello—are extracted to examine the initial setup of secure connections. For field-level granularity analysis, the algorithm selects key fields from these handshake messages, converts them into word vectors, and applies Multi-head Attention to capture semantic interactions among the fields, enabling simultaneous focus on various aspects of the handshake data. Concurrently, packet-level granularity analysis is conducted by dividing the traffic into packets, selecting the top 30 based on significance, and characterizing each with features like size, time interval, transmission direction, power load, and device state. A Bidirectional Long Short-Term Memory (BiLSTM) network then processes this packet sequence to capture temporal dependencies, analyzing the data in both forward and backward directions for contextual understanding over time. The representations from both field-level and packet-level analyses are subsequently fused to integrate insights from these dual granularities. Finally, a linear layer followed by a Softmax function classifies the traffic as either benign or malicious based on the combined features.

Spatio-temporal based global behavior modeling

In order to explore the semantic information of traffic global communication in power information interaction, the traffic is divided by packet granularity. According to statistics, the vast majority of sample packets in the dataset are less than 30 in number, so the top 30 packets of traffic are selected as representatives of global traffic. They are characterized as follows:

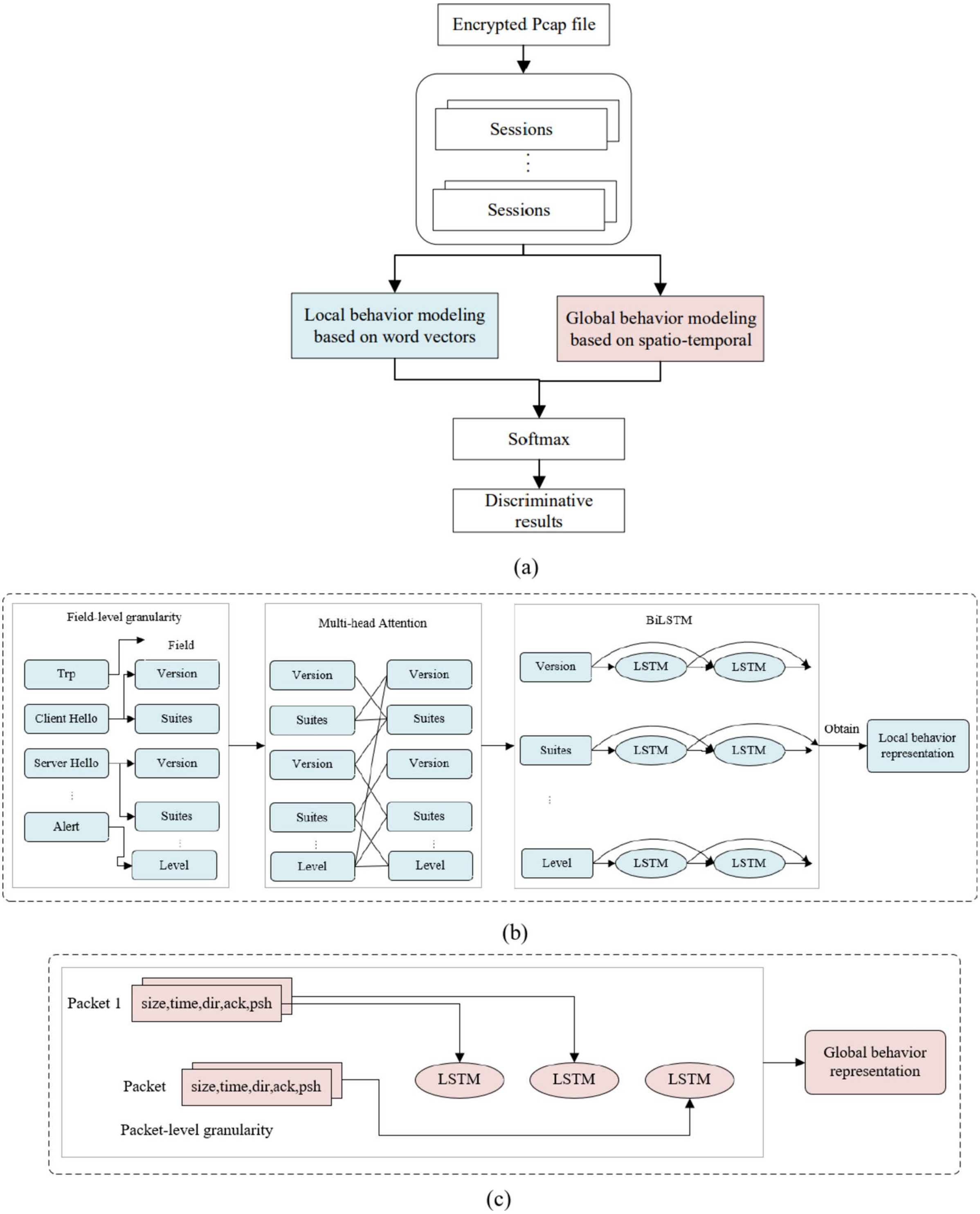

Fig. 1. E-MGFlow based encrypted malicious traffic detection framework. (**a**) General framework. (**b**) Local behavior modeling based on word vectors. (**c**) Global behavior modeling based on spatio-temporal.

$$S_{\text{behavior}} = \{P_1, P_2, \ldots, P_{30}\} \tag{10}$$

Where $P_t = \{'size, time, dir, load, state''\}$.In the t-th packet (P_t), it contains the following information:

size: load size, indicating the amount of data (e.g., real-time power consumption) transmitted by the power message.

time: time interval, indicates the time interval (in milliseconds) since the last packet.

dir: Transmission direction, indicates the direction of data transmission (e.g. uplink or downlink).

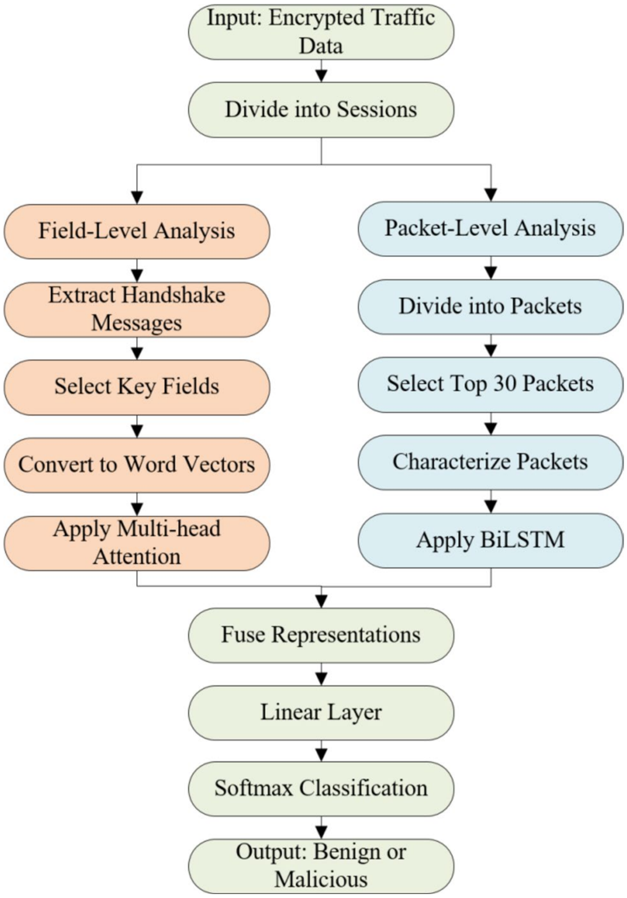

Fig. 2. Flowchart of the E-MGFlow algorithm.

load: Power load information, indicates the power usage at this point in time.

state: device state information, indicating the current operating state of the power device (e.g., normal, faulty, etc.).

An LSTM model captures the temporal dependencies in this sequence:

$$\text{behavior} = \text{LSTM}(P_1, P_2, \ldots, P_{30}) \tag{11}$$

The above modeling method can effectively capture the timing characteristics of the power flow and its state changes, which facilitates the subsequent detection of malicious flows.

Local behavior modeling based on word vectors

In the process of electric power communication, the communicating parties will perform handshake interaction before data transmission to ensure the security and reliability of the data. In this paper, this process is defined as local behavior. The local behavior mainly includes two stages of TCP three times handshake and TLS encrypted handshake between power devices.

When analyzing local behaviors, it may not be obvious to observe the maliciousness of traffic in one phase alone. However, if the interaction behavior of that phase with other phases is mined, it can enhance the detection of malicious traffic, as shown in Fig. 3. In power message interactions, malware may implement attacks by tampering with certificate or status information during the handshake.

This paper proposes three key interaction behavior patterns:

(1) Interaction between different message fields. In malicious samples, the Client Hello may list several supported encryption components, while the Server Hello might ultimately select encryption component 0xc02f, representing a relationship between alternatives. The client side may detect that the server's certificate has expired, triggering an alert such as "Certificate Expired."

(2) Interaction between the fields in the message. In the message Certificate, subject and issuer are the certificate holder and issuer respectively. In malicious samples, there are often low-credit organizations or themselves issuing certificates.

(3) Interaction between messages. In the message Server Key Exchange, changing the key passed in the message Server Hello is a behavioral interaction resulting from a certain type of message responding to a specific message.

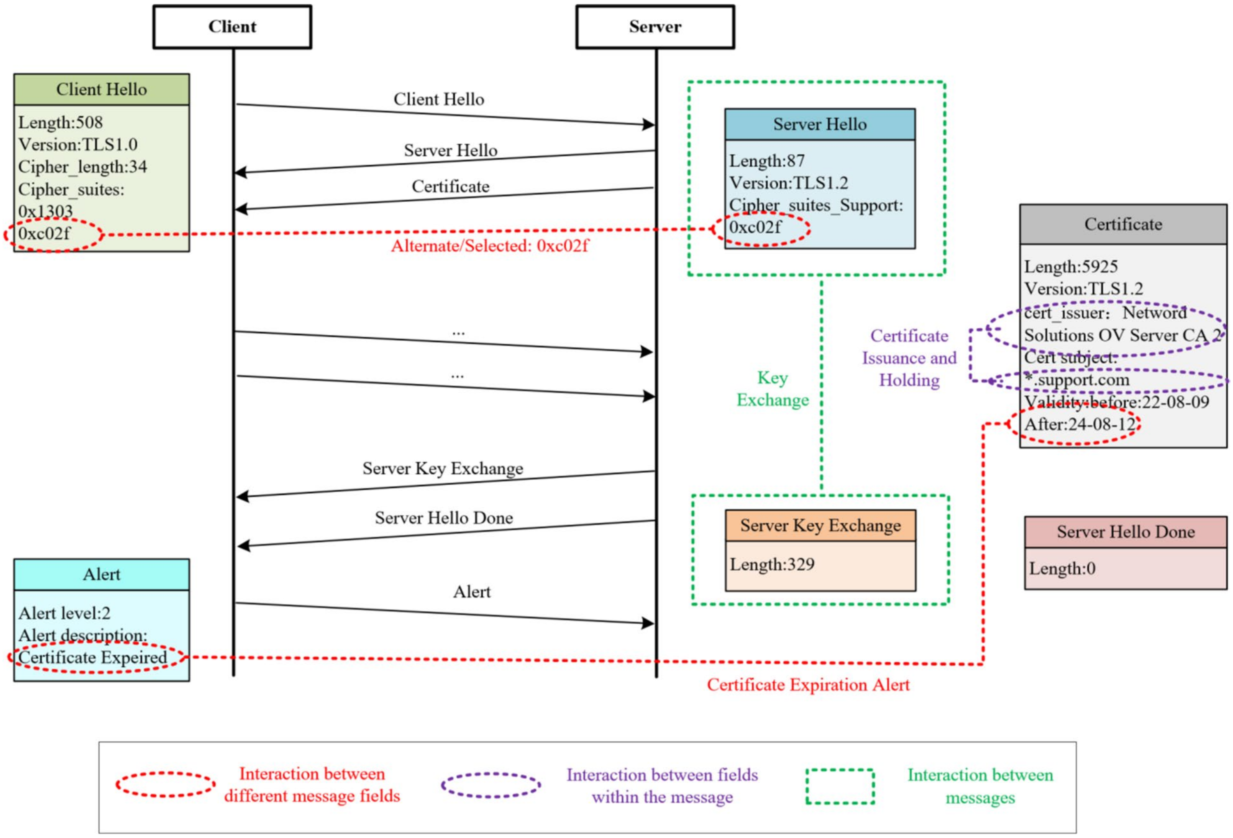

Fig. 3. Interaction during handshake for malicious samples.

In order to learn the semantics of the above three interactions in a session, in this paper, key fields (words) in a session are extracted and formed into word sequences (sentences), which are then used in a natural language processing model for representation learning.

The word sequences extracted from the session have the following characteristics in the representation learning process: (a) Different sessions contain different numbers of messages. Different types of messages contain different numbers of key fields (i.e., the length of field sequences corresponding to each session may be different). (b) Different key fields are affected differently by contextual fields. (c) There are differences in the order of occurrence of different fields on the semantic characterization of the fields.

Thus, in this paper, the session is divided into messages, which are further divided into different fields. The Multi-head Attention mechanism is then employed to learn the semantics of the field context. The session is divided into messages as follows:

$$S_{\text{handshake}} = \{R_1, R_2, \dots, R_n\} \tag{12}$$

Where R represents the i-th message.

In the local behavioral modeling process, the session includes TCP three-way handshake and TLS encryption phase handshake messages. The types of encrypted messages in the handshake process include Handshake, Change Cipher Spec, Application Data, and Alert, representing different phases of the handshake process. Since specific data transmission is beyond the scope of this paper, the Application Data message is discarded.

For each record, based on field granularity, the representation is as follows:

$$R_i = \{w_1, w_2, \dots, w_{m_i}\} \tag{13}$$

Where w is the j-th key field in the record, and each message contains multiple fields.

During the field selection process, filtering is based on the following criteria:

(1) Randomly generated fields: Some fields (e.g., keys) are merely parameters in the key exchange process and do not contribute to identifying malicious behavior; thus, these fields need to be removed. If a message (e.g., Server Key Exchange) only has a specific interaction, the fields are left empty.

(2) Ability to distinguish between normal and malicious fields: Certain fields (e.g., encryption components, certificate issuance times) exhibit distribution differences between normal and malicious traffic. Malware authors tend to focus more on content encryption rather than the choice of encryption algorithms, often opting for older encryption components.

Based on the aforementioned selection criteria, several key fields are selected for each message. For the vector representation of each key field, the corresponding byte value of the field itself is used bb. Additionally, due to varying lengths of different fields, the vector length of the input model needs to be unified. Given that most key field lengths are less than 4 based on statistics, the key field lengths are standardized to 4. For fields shorter than 4, the high bits are padded with 0s. For fields longer than 4, every 4 bytes form a word vector. Fields shorter than 4 bytes are also padded to 4 bytes. The initial representation of each field is

$$w_{\mathrm{raw}_j} = [b_1, b_2, b_3, b_4] \tag{14}$$

On this basis, to differentiate the semantic differences of fields in different positions, this paper selects the message type $type_r$ and handshake type $type_h$ of the TLS protocol as prefixes to form a vector representation of the key fields:

$$w_j = [\mathrm{prefix}; w_{\mathrm{raw}_j}] \tag{15}$$

Where

$$\mathrm{prefix} = \{type_{\mathrm{record}}, type_{\mathrm{handshake}}\} \tag{16}$$

Bidirectional LSTM for temporal dependencies

By splicing multiple words, a complete field sequence is formed. To mine key information, this paper adopts Multi-head Attention to calculate the weights of the key fields, enabling the model to fully capture the interactions among fields during the communication process.

Through the aforementioned process, the contextual semantics between fields are extracted. To mine the semantics of the communication handshake from all fields, bidirectional BiLSTM is utilized to extract the back-and-forth timing information between key fields, obtaining the communication handshake semantics of the flow:

$$h_{\mathrm{handshake}} = \mathrm{BiLSTM}(w_1, w_2, \ldots, w_m) + \mathrm{BiLSTM}(w_m, \ldots, w_2, w_1) \tag{17}$$

After dividing the traffic according to packet-level granularity and field-level granularity, the spatio-temporal-based global behavioral representation S_{behavior} and the word vector-based local behavioral representation $h_{\mathrm{handshake}}$ are obtained. The semantic representation vectors from both granularities are then fused, and after passing through a linear layer (Linear), the final discriminative result (output) is produced by the Softmax function, as shown below:

$$session = [S_{\mathrm{behavior}}; h_{\mathrm{handshake}}] \tag{18}$$

Where, $session$ represents the comprehensive semantic vector obtained by concatenating the global behavior representation S_{behavior} and the local behavior representation $h_{\mathrm{handshake}}$. The session vector is then processed through a linear layer followed by a Softmax function to produce the final classification result:

$$output = \mathrm{Softmax}\left(\mathrm{Linear}(\ session\ )\right) \tag{19}$$

Where, $output$ denotes the probability distribution over each class, which is used to predict whether the traffic is malicious. Linear represents the linear transformation layer, used to fuse the features of global and local behaviors. With the aforementioned modeling approach, the local and global characteristics of power flow can be effectively captured. Then, the predicted labels are calculated based on the output:

$$\hat{y} = \mathrm{argmax}\left(output\right) \tag{20}$$

During the training process, the loss function must be calculated to optimize the model parameters. The loss function is calculated using the formula:

$$\mathcal{L} = -\alpha \sum\nolimits_{k=1}^{C} y_k \log\left(\mathrm{output}_k\right) \tag{21}$$

Where, $\mathcal{L}$ indicates the loss value of the model, k is the class index iterating over all classes C, y_k is the one-hot encoded true label, and output_k is the predicted probability for class k. The weighting factor α is used to address sample imbalance. The loss function guides the direction of model training, improving model accuracy by minimizing the loss value.

Experimental results

Data set and experimental setup

In our binary classification experiments, we utilized the DataCon2020 dataset, which comprises 90,842 malicious traffic samples and 30,235 normal traffic samples, totaling 121,077 samples[19]. For this study, we employed a subset of this dataset, consisting of 7,000 training samples (3,500 benign and 3,500 malicious) and 5,000 testing samples (2,500 benign and 2,500 malicious). This subset, constituting approximately 10% of the original dataset, was selected based on the following criteria and practical considerations: First, the computational requirements for training the E-MGFlow model, which integrates multi-head attention mechanisms and bidirectional LSTM

Dataset type	Benign PCAP file	Malicious PCAP file
Training set	3500	3500
Test set	2500	2500

Table 1. Sample distribution statistics.

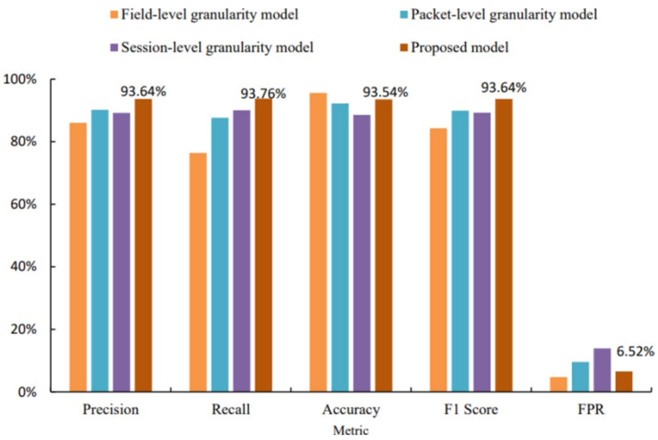

Fig. 4. Prediction results of each granularity model.

networks, necessitated a reduction in dataset size. Processing the entire dataset would demand substantial computational resources and time, exceeding the constraints of this research. The chosen subset enabled efficient experimentation while still allowing for a thorough evaluation of the model's performance. Second, the original DataCon2020 dataset presents a class imbalance, with malicious samples outnumbering benign ones by approximately 3:1. To address this imbalance and prevent bias toward the majority class, we constructed a balanced subset with an equal number of benign and malicious samples. This balance facilitates the model's ability to learn discriminative features without overfitting to the more prevalent class. Third, the subset was carefully curated to preserve the diversity of traffic patterns, encryption protocols, and attack types found in the full dataset. This ensures that the subset remains representative of the broader data distribution, thereby supporting the generalizability of the experimental results to real-world scenarios, particularly those relevant to power systems. These considerations collectively ensure that the reduced dataset size does not compromise the validity of our findings.

These datasets are widely used for encrypted traffic detection tasks, providing diverse and representative traffic patterns. The power information interaction dataset, derived from actual power system communication flows, includes benign traffic (e.g., device status reporting and power load information) and malicious traffic (e.g., DoS attacks and fake device identity attacks). This dataset captures the unique characteristics of power system communication, such as device state changes and load variations, enabling the validation of E-MGFlow in real-world power system scenarios.

Parameter Selection: The parameters for the E-MGFlow model were selected based on a combination of empirical tuning and domain-specific considerations. For instance, the number of attention heads in the multi-head attention mechanism was set to 8, as this value has been shown to effectively capture diverse semantic interactions in similar sequence modeling tasks. The learning rate was set to 0.001, which is a standard choice for training deep learning models to ensure stable convergence. Additionally, the batch size was chosen to be 64, balancing computational efficiency with the need to capture sufficient variability in the training data.

To comprehensively assess the predictive effectiveness of the model, various metrics are considered, including Accuracy, Recall, Precision, F1-score, and False Positive Rate (FPR). These metrics provide comprehensive quantitative criteria for evaluating a model's ability to recognize malicious traffic within the context of power information interactions.

Experimental results for DataCon2020 and DataCon2021 datasets

Binary classification experiments (DataCon2020 dataset)

For binary classification experiments on encrypted malicious traffic, this study employs the open-source dataset provided by DataCon2020. The dataset is divided into training and testing sets, as depicted in Table 1.

As observed from Table 1, the training set comprises 7000 PCAP files, with 3500 benign and 3500 malicious PCAP files each. The test set includes 5000 PCAP files, with 2500 benign and 2500 malicious PCAP files each.

In this paper, four feature sets of varying granularities (field level, packet level, session level, and fusion level) are extracted from the training data, and the model is trained. The fused results are obtained by summing and averaging the predicted probabilities of the four base classifiers. The experimental results are illustrated in Fig. 4.

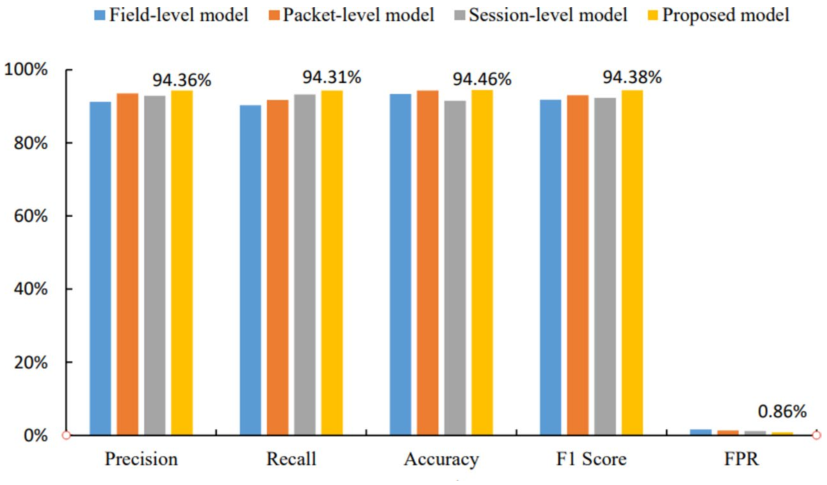

Fig. 5. 12 classification experiment results.

Task type	Accuracy	Recall	Precision	F1 score	FPR
100 classification	95.60%	96.19%	95.77%	95.64%	1.35%

Table 2. Performance of encrypted traffic 100 classification.

The prediction results in Fig. 4 indicate that the proposed model can synthesize the advantages of each model, enhancing the prediction precision and recall rate to 93.64% and 93.76%, respectively. This is particularly crucial for power systems, where the timely detection of malicious activities can prevent catastrophic failures, such as grid destabilization or widespread power outages. The model's low FPR of 6.52% further ensures that benign traffic is not misclassified, reducing the risk of unnecessary disruptions in power system operations. The fusion of field-level and packet-level features enhances the model's ability to capture both local and global traffic behaviors, leading to improved detection performance. These results indicate that the proposed model can effectively distinguish between benign and malicious traffic.

Multi-classification experiments (DataCon2021 dataset)

To ascertain the effectiveness of the method proposed in this paper on multi-classification tasks, this section utilizes the open-source dataset provided by DataCon2021 for experimental analysis[20]. The specific multi-classification tasks include a 12-classification task for encrypting malware and agent software traffic and a 100-classification task for encrypting agent traffic.

The dataset for the 12-classification task is based on the Part 1 section of DataCon2021 and includes encrypted traffic generated by malware, with malicious traffic considered as the 12th category, corresponding to label 11. The objective of the task is to differentiate between encrypted malicious traffic and agent software traffic and to further subdivide encrypted agent traffic by identifying the specific agents generating the traffic.

Three distinct granularity features are employed in this task: field-level granularity, packet-level granularity, and session-level granularity. The specific acquisition process of these three granularity features is identical to the encrypted malicious traffic detection method that integrates multi-granularity features, although the classification models selected differ. Experiments reveal that the three classifiers, obtained by training these three granularity features separately and then employing soft voting, can achieve high prediction accuracy, precision, recall, and F1 values. The probability (i.e., FPR) of misclassifying agent software traffic as malware traffic is low, as demonstrated in Fig. 5.

As evident from Fig. 5, the model in this paper performs well across all indices, achieving 94.36% precision and 94.31% recall, with the FPR reduced to 0.86%. This demonstrates the proposed model's ability to handle complex traffic patterns and differentiate between various types of encrypted traffic.

The dataset for the 100-classification task is derived from the Part 2 section of DataCon2021. The objective of this classification task is to detect which website the packets in the PCAP file were generated by visiting. In this task, packet-level features are used for training. It is found experimentally that classifiers (e.g., Random Forest), obtained by training with packet-level features, can achieve high prediction accuracy, recall, precision, and F1 values, as shown in Table 2.

The experimental results demonstrate that the proposed model, combined with multi-granularity features, performs exceptionally well in multi-classification tasks. It effectively identifies the type of encrypted traffic in network information interactions, significantly improving detection accuracy.

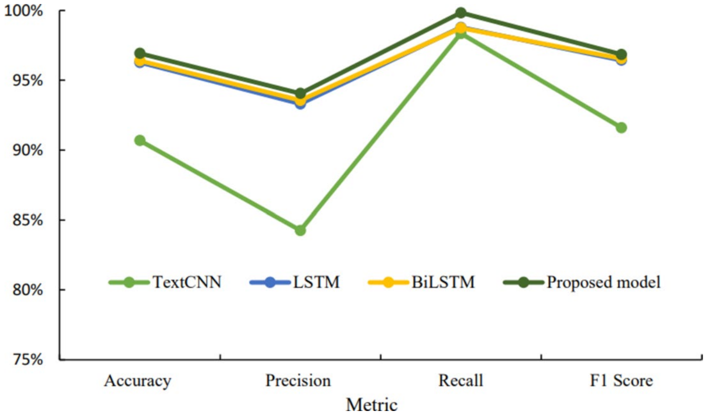

Fig. 6. Ablation comparison of different models.

Dataset type	Benign PCAP file	Malicious PCAP file
Training set	4000	4000
Test set	3000	3000

Table 3. Sample distribution of the electricity dataset.

Ablation study on local behavior modeling (DataCon dataset)

To investigate the impact of each enhancement on model performance, this paper designs ablation experiments to compare different models used as the base model for local behavioral features. The purpose of the ablation experiments is to assess the performance variation of each model in network information traffic detection. The specific models include those from literature[21] (TextCNN), literature[12] (LSTM), literature[22] (BiLSTM), and the BiLSTM combined with the Multi-head Attention mechanism, which is the model proposed in this paper. The baseline models—TextCNN, LSTM, and BiLSTM—were selected for the following reasons: (1) TextCNN: This model excels at capturing spatial features from traffic data, making it a suitable baseline for evaluating E-MGFlow's field-level analysis capabilities. (2) LSTM: As a widely used model for temporal sequence modeling, LSTM helps assess the importance of packet-level temporal dependencies in our approach. (3) BiLSTM: By processing sequences bidirectionally, BiLSTM provides a robust benchmark for evaluating E-MGFlow's ability to leverage both past and future contexts in traffic flows.These models were chosen to represent key aspects of traffic analysis (spatial, temporal, and bidirectional temporal), enabling a comprehensive evaluation of E-MGFlow's multi-granularity design.

The rationale for selecting these models is rooted in their ability to represent different aspects of network traffic analysis: spatial patterns (TextCNN), temporal dependencies (LSTM and BiLSTM), and fine-grained semantic interactions (BiLSTM with Multi-head Attention). These models are representative of the state-of-the-art in encrypted traffic detection and provide a comprehensive basis for evaluating the performance of the proposed E-MGFlow approach. The experimental results are depicted in Fig. 6.

It can been seen from Fig. 6 that the E-MGFlow model, which combines BiLSTM with Multi-head Attention, achieves an accuracy of 88.1%, precision of 76.5%, recall of 99.7%, and F1-score of 87.6%. In comparison, the baseline models (TextCNN, LSTM, and BiLSTM) achieve lower performance, with accuracy ranging from 62.9 to 85.9%, precision from 37.0 to 74.3%, recall from 93.3 to 95.9%, and F1-score from 66.6 to 86.1%. This demonstrates that the integration of Multi-head Attention significantly enhances the model's ability to capture fine-grained semantic interactions, leading to improved detection performance. The ablation experiments reveal that the fusion model proposed in this paper can substantially enhance the performance of malicious traffic detection within the context of network information interaction. Compared to other models, it holds significant advantages in terms of four metrics.

Experimental results on the power dataset

This section presents the experimental evaluation of the E-MGFlow model for detecting malicious traffic in power information interaction scenarios. The experiments were conducted using a power-domain dataset, which comprises real-world communication flows from power systems. The dataset includes benign traffic, such as device status reports and power load information, and malicious traffic, such as Denial-of-Service (DoS) attacks and fake device identity attacks. For training, the dataset contains 4000 benign and 4000 malicious PCAP files, while the test set includes 3000 benign and 3000 malicious PCAP files. The detailed distribution of the power dataset is presented in Table 3.

Model	Precision (%)	Recall (%)	F1-score (%)	FPR(%)	Response time (ms)	Detection latency (ms)
E-MGFlow	93.64	93.76	93.55	6.52	90	140
GCN-ETA	89.12	90.23	89.45	7.89	120	180
Incremental learning	88.45	89.67	88.89	8.12	110	170
Transformer-based model	90.34	91.12	90.56	7.45	100	160

Table 4. Comparative analysis of E-MGFlow and recent encrypted traffic detection models.

Model	Precision (%)	Recall (%)	F1-score (%)	FPR (%)	Response time (ms)	Detection latency (ms)
E-MGFlow	93.64	93.76	93.55	6.52	90	140
Literature [26]	94.30	93.22	96.74	7.84	116	175

Table 5. Performance comparison of E-MGFlow with state-of-the-art method.

Comparative analysis with recent models

To assess the performance of E-MGFlow, it was compared against several recent models relevant to encrypted traffic detection: GCN-ETA[23], an incremental learning algorithm[24], and a transformer-based model[25]. These models were chosen based on their demonstrated adaptability to diverse network environments.

The evaluation metrics include precision, recall, F1-score, false positive rate (FPR), and real-time performance, with results summarized in Table 4. E-MGFlow achieved a precision of 93.64%, surpassing GCN-ETA (89.12%), the incremental learning model (88.45%), and the transformer-based model (90.34%). This suggests that E-MGFlow is more effective at correctly identifying malicious traffic while minimizing misclassifications of benign traffic. In terms of recall, E-MGFlow recorded 93.76%, compared to GCN-ETA (90.23%), the incremental learning model (89.67%), and the transformer-based model (91.12%), indicating its superior ability to detect all instances of malicious traffic.

The F1-score, which balances precision and recall, was 93.55% for E-MGFlow, outperforming GCN-ETA (89.45%), the incremental learning model (88.89%), and the transformer-based model (90.56%). Additionally, E-MGFlow exhibited the lowest FPR at 6.52%, compared to GCN-ETA (7.89%), the incremental learning model (8.12%), and the transformer-based model (7.45%). A low FPR is particularly important in power systems to avoid unnecessary disruptions. For real-time applicability, E-MGFlow demonstrated a response time of 90 ms and a detection latency of 140 ms, making it viable for operational deployment. These findings confirm that E-MGFlow provides robust performance improvements over the compared models in the context of power system traffic detection.

Comparison with state-of-the-art method on power dataset

To further validate E-MGFlow's effectiveness, it was compared with a state-of-the-art method from reference[26], an improved lightweight quantum convolutional neural network. Although reference[26] reported high accuracy (94.30%) and F1-score (96.74%) on the DataCon2020 dataset, its performance on power-domain traffic required evaluation. Thus, both E-MGFlow and the method from reference[26] were tested on the power dataset described earlier, with results presented in Table 5.

E-MGFlow achieved a precision of 93.64%, slightly higher than the 92.50% of reference[26], indicating a modest improvement in correctly identifying malicious traffic. The recall for E-MGFlow was 93.76%, compared to 91.22% for reference[26], suggesting better detection coverage. The F1-score of E-MGFlow (93.55%) also exceeded that of reference[26] (91.85%), reflecting a stronger balance between precision and recall.

Regarding the false positive rate, E-MGFlow recorded 6.52%, lower than the 7.84% of reference[26], which is advantageous for reducing false alarms in power systems. In terms of real-time performance, E-MGFlow showed a response time of 90 ms and detection latency of 140 ms, outperforming reference[26] with 116 ms and 175 ms, respectively. These metrics highlight E-MGFlow's suitability for time-sensitive applications.

The superior performance of E-MGFlow on the power dataset can likely be attributed to its incorporation of domain-specific features, such as device state and power load information, which enhance its adaptability to power system traffic patterns.

Ablation study on multi-granularity fusion in E-MGFlow (power dataset)

To thoroughly evaluate the contributions of each component in E-MGFlow, we performed ablation experiments comparing our full model against several variants, including a standalone Multi-head Attention model. This additional comparison specifically addresses the need to verify whether E-MGFlow's multi-granularity approach outperforms a model relying solely on multi-head attention.

The evaluated models are as follows. (1) Multi-head Attention-only Model: This baseline employs the Multi-head Attention mechanism to process field-level features extracted from handshake messages, without incorporating packet-level temporal features or BiLSTM. (2) BiLSTM-only Model: This variant uses BiLSTM to capture temporal dependencies from packet-level features, excluding field-level analysis. (3) E-MGFlow (Full Model): Our proposed method, integrating field-level and packet-level analyses via Multi-head Attention and

Model	Precision (%)	Recall (%)	F1-score (%)	FPR (%)
Multi-head Attention-only	85.2	86.7	85.9	8.3
BiLSTM-only	87.5	88.9	88.2	7.6
E-MGFlow (full model)	93.6	93.8	93.7	6.5

Table 6. Ablation study results on the power dataset.

BiLSTM, followed by feature fusion. The performance results on the power dataset are presented in Table 6 below.

As shown in Table 6, the Multi-head Attention-only model achieves a precision of 85.2%, recall of 86.7%, and F1-score of 85.9%, with a false positive rate (FPR) of 8.3%. While this model effectively captures semantic relationships within field-level features, it lacks the temporal context provided by packet-level analysis. The BiLSTM-only model performs better, with a precision of 87.5%, recall of 88.9%, and F1-score of 88.2%, due to its focus on temporal patterns, though it misses fine-grained field-level interactions.

In contrast, E-MGFlow, combining both granularity levels, achieves superior performance with a precision of 93.6%, recall of 93.8%, and F1-score of 93.7%, alongside the lowest FPR of 6.5%. These results confirm that the integration of field-level and packet-level analyses in E-MGFlow significantly enhances detection performance, outperforming the standalone Multi-head Attention model and validating the effectiveness of our approach for identifying encrypted malicious traffic in power systems.

Conclusion

The Electricity Multi-Granularity Flow Representation Learning (E-MGFlow) approach presented in this paper adeptly tackles the challenge of detecting encrypted malicious traffic within power systems. Through the utilization of multi-granularity representation learning, E-MGFlow captures temporal features and local behavioral nuances in power information interactions, thereby enhancing the precision of malicious traffic detection. The experimental outcomes illustrate that E-MGFlow markedly surpasses traditional detection methods across a spectrum of datasets, presenting substantial practical value for network security management in power systems. It is poised to swiftly recognize and counter potential cyber threats, safeguarding the reliable operation of power systems. Future research will concentrate on further refining the E-MGFlow algorithm to bolster its adaptability and generalization capabilities within varied network environments. Moreover, investigations will delve into the applicability of the E-MGFlow method to traffic detection tasks in other sectors, such as finance or medical information systems, to evaluate its efficacy in diverse contexts.

Data availability

The data that support the findings of this study are available from the corresponding author Liqiang Yang upon reasonable request.

Received: 27 January 2025; Accepted: 14 May 2025
Published online: 22 May 2025

References

1. Saeed, S., Altamimi, S. A., Alkayyal, N. A., Alshehri, E. & Alabbad, D. A. Digital transformation and cybersecurity challenges for businesses resilience: issues and recommendations. *Sensors* **23** (15), 6666 (2023).
2. Saleem, J., Islam, R. & Islam, M. Z. Darknet traffic analysis: A systematic literature review. *IEEE Access*. **12**, 42423–42452 (2024).
3. Alzaabi, F. R. & Mehmood, A. A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. *IEEE Access*. **12**, 30907–30927 (2024).
4. Xu, S. J., Kong, K. C., Jin, X. B. & Geng, G. G. Unveiling traffic paths: explainable path signature feature-based encrypted traffic classification. *Comput. Secur.* **150**, 104283 (2025).
5. Zang, X. et al. Encrypted malicious traffic detection based on natural Language processing and deep learning. *Comput. Netw.* **250**, 110598 (2024).
6. Elmaghraby, R. T., Aziem, N. M. A., Sobh, M. A. & Bahaa-Eldin, A. M. Encrypted network traffic classification based on machine learning. *Ain Shams Eng. J.* **15** (2), 102361 (2024).
7. Aladaileh, M. A. et al. Effectiveness of an entropy-based approach for detecting low-and high-rate DDoS attacks against the SDN controller: experimental analysis. *Appl. Sci.* **13** (2), 775 (2023).
8. Ferriyan, A., Thamrin, A. H., Takeda, K. & Murai, J. Encrypted malicious traffic detection based on Word2Vec. *Electronics* **11** (5), 679 (2022).
9. hang, X. et al. Deep-forest-based encrypted malicious traffic detection. *Electronics* **11** (7), 977 (2022).
10. Liu, J., Wang, J., Yan, T., Qi, F. & Chen, G. Unknown traffic recognition based on Multi-Feature fusion and incremental learning. *Appl. Sci.* **13** (13), 7649 (2023).
11. Zhou, Y. et al. Identification of encrypted and malicious network traffic based on one-dimensional convolutional neural network. *J. Cloud Comput.* **12** (1), 53 (2023).
12. Li, Y. et al. One-class LSTM network for anomalous network traffic detection. *Appl. Sci.* **12** (10), 5051 (2022).
13. Malashin, I., Tynchenko, V., Gantimurov, A., Nelyub, V. & Borodulin, A. Applications of long Short-Term memory (LSTM) networks in polymeric sciences. *Rev. Polym.* **16** (18), 2607 (2024).
14. Coelho, C., Costa, M. F. P. & Ferrás, L. L. Enhancing continuous time series modelling with a latent ODE-LSTM approach. *Appl. Math. Comput.* **475**, 128727 (2024).
15. Chen, J. et al. GCN-MHSA: A novel malicious traffic detection method based on graph convolutional neural network and multi-head self-attention mechanism. *Comput. Secur.* **147**, 104083 (2024).
16. Miao, Y. et al. Efficient privacy-preserving Spatial range query over outsourced encrypted data. *IEEE Trans. Inf. Forensics Secur.* **18**, 3921–3933 (2023).

17. Li, Y., Xu, C., Xu, L., Mei, L. & Zhu, Y. Verifiable searchable encryption scheme with flexible access control in the cloud. *J. Parallel Distrib. Comput.* **197**, 105025 (2025).
18. Zamani, M., Safkhani, M. & Daneshpour, N. SCF-VPEKS: secure channel free verifiable public key encryption with keyword search. *Wirel. Netw.*, 1–18. (2025).
19. Dang, Z. et al. Semi-Supervised Learning for Anomaly Traffic Detection via Bidirectional Normalizing Flows. *arXiv* (2024).
20. Wang, H., Zhang, Y., Guo, Z., Li, T. & Fan, L. Format preserving encryption of sensitive data in database. In *International Conference on Cyber Security, Artificial Intelligence, and Digital Economy (CSAIDE 2023)*. SPIE, 206–212 (2023)
21. Liu, J., Yan, Z., Chen, S., Sun, X. & Luo, B. Channel attention TextCNN with feature word extraction for Chinese sentiment analysis. *ACM Trans. Asian Low-Resour. Lang. Inform. Process.* **22** (4), 1–23 (2023).
22. Singla, P., Duhan, M. & Saroha, S. An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. *Earth Sci. Inf.* **15** (1), 291–306 (2022).
23. Chen, J. et al. TLS-MHSA: an efficient detection model for encrypted malicious traffic based on Multi-Head Self-Attention mechanism. *ACM Trans. Priv. Secur.* **26** (4), 1–21 (2023).
24. Bovenzi, G. et al. Benchmarking class incremental learning in deep learning traffic classification. *IEEE Trans. Netw. Serv. Manage.* **21** (1), 51–69 (2023).
25. Zheng, W., Zhong, J., Zhang, Q. & Zhao, G. MTT: an efficient model for encrypted network traffic classification using multi-task transformer. *Appl. Intell.* **52** (9), 10741–10756 (2022).
26. Xiong, Q. et al. A modified lightweight quantum convolutional neural network for malicious code detection. *Quantum Sci. Technol.* **10** (1), 015007 (2024).

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Zhifu Wu, Xianfu Zhou; data collection: Xianfu Lu, Liqiang Yang; analysis and interpretation of results: Zhifu Wu, Liqiang Yang, Siqi Shen; draft manuscript preparation: Liqiang Yang, Dong Yan. All authors reviewed the results and approved the final version of the manuscript.

Funding

This work is supported by the science and technology project of State Grid Corporation of China: “Defense System Design and Key Technologies Research on Cyber Security of New Power System” (Grand No. 5700–202358388 A-2-3-XG).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.