PAIEL: Protocol-Aware and Context-Integrated Protocol Explanation Using LLMs for SOCs

Takeshi Kaneko[*], Hiroyuki Okada[*], Rashi Sharma[†], Tatsumi Oba[*], Naoto Yanai[*]
[*]Panasonic Holdings Corporation
kaneko.takeshi001@jp.panasonic.com, okada.hiroyuki001@jp.panasonic.com,
oba.tatsumi@jp.panasonic.com, yanai.naoto@jp.panasonic.com
[†]Panasonic R&D Center Singapore
rashi.sharma@sg.panasonic.com

***Abstract*—Security Operations Centers (SOCs) have increasingly adopted Large Language Models (LLMs) to support cyberattack analysis, yet existing LLM usage often lacks knowledge required for accurate protocol-level explanations. In this study, we propose PAIEL, an LLM-based framework that integrates semantic context of protocol-level knowledge and structured context as external knowledge to generate accurate and faithful explanations for each protocol from raw packets, thereby supporting SOC analyst operations. Through extensive experiments, we show that PAIEL outperforms common LLM baselines in terms of both human and automatic evaluations by considering protocol specifications. Our results also indicate that both structured context and semantic context are necessary to generate effective explanations. We also conduct an evaluation of PAIEL as a real-world application by providing it with SOC analysts, and then demonstrate that PAIEL is practical in the real world.**

I. Introduction

A. Background

Security Operations Centers (SOCs) face increasing analytical demands as cyberattacks grow in volume. SOCs must continuously monitor diverse data, including low-level telemetry such as packets, to detect cyberattacks. SOC operations involve multiple stages: alert triage, evidence collection, threat investigation, and incident response [1]. In each stage, analysts must interpret evidence from packets to understand attack behaviors and make operational decisions [2]–[4]. Modern SOCs employ both heuristic-based [5], [6] and machine learning (ML)-based [7]–[11] tools to handle massive alert volumes [12]–[14]. These approaches often provide limited explanations [15], forcing analysts to shoulder the cognitive cost of interpreting protocol-level evidence and deciding appropriate responses manually [16], [17].

However, understanding novel attacks and emerging threats unknown to SOC analysts from massive and complex telemetry is non-trivial and presents numerous challenges. For example, OT (Operational Technology) protocols used in industrial control systems (ICS) environments exhibit highly complex structures: they involve numerous service types, object models, and vendor-specific variations. Consequently, accurate interpretation of their packets requires substantial domain expertise. It is more serious in protocol-level analysis from a single packet because analysts must interpret protocol background from raw packet alone. Protocol-level understanding in such OT protocols imposes a high cognitive cost on SOC analysts, especially for novice analysts. A typical approach to analyzing novel attacks is to leverage machine learning models that learn from known attacks to infer unknown ones [8]–[10], [18]. Nevertheless, typical machine learning models have inherent limitations in providing decision-relevant explanations, and hence may be insufficient to support complex decisions in SOCs with organizational constraints [15].

As a means to address the above limitation, SOCs have recently introduced Large Language Models (LLMs) for the purpose of interpreting cyber attacks for analysts [19]. LLMs learn semantic structures from large amounts of text data and can potentially generate appropriate responses based on input [20]–[22]. Yet, simply applying LLMs is often insufficient for protocol-level analysis. This is because the protocol-level knowledge of packets is governed by protocol specifications. These specifications typically lie outside the LLM's training distribution, which makes such knowledge difficult to infer without explicit protocol context.

To the best of our knowledge, the existing LLM applications that provide accurate explanation generation for cyberattacks [23]–[25] require pre-summarized knowledge files (e.g., CVE files [23], [25] or incident tickets [24]). Unfortunately, when such pre-summarized knowledge is no longer provided with an LLM, it may be unable to provide accurate explanations. While there is an existing method that directly handles raw network traffic [26], it primarily focuses on traffic generation and classification rather than interpreting packets. In SOC workflows, however, analysts must understand the protocol-level knowledge of single packets for decision making, e.g., distinguishing between legitimate control commands and malicious payloads. Therefore, systems without protocol-level explanation may be unable to support SOC analysis [20].

Based on the above background, designing LLM applications that generate protocol-level explanations from raw

Workshop on Security Operation Center Operations and Construction (WOSOC) 2026
23 February 2026, San Diego, CA, USA
ISBN 978-1-970672-03-9
https://dx.doi.org/10.14722/wosoc.2026.23013
www.ndss-symposium.org

packets, i.e., without requiring analysts to provide protocol-level knowledge and fine-tune models for specific environments, remains an open challenge. To address this challenge, we propose an LLM-based explanation framework to support protocol-level analysis as a fundamental task of SOCs. Accordingly, we examine the following two research questions:

RQ1 Can faithful and accurate protocol-level explanations be generated from a single packet?

RQ2 Do such faithful and accurate protocol-level explanations support security analysts in actual operation?

B. Contributions

In this paper, we propose *PAIEL (Packet Analysis and Insight Extraction using LLM)*, an LLM-based framework to generate protocol-level explanations for SOC analysts. PAIEL directly takes a single packet as input and then generates explanations grounded in protocol specifications. We demonstrate that faithful and accurate protocol-level explanations can be generated from a single packet across various protocols that require different service knowledge, indicating that packet explanation is generalized beyond a single protocol environment. Our main technical contributions are as follows:

(i) To realize the first contribution, we design a new framework, PAIEL. It integrates two complementary sources of context, i.e., *structured context* and *semantic context*. Structured contexts anchor explanations in protocol-defined knowledge, while semantic context provides a broader operational background. Neither context alone is sufficient. (See Section IV.)

(ii) Extensive experiments show that different characteristics are obtained by the above contexts. Structured context improves metrics in human evaluation, whereas semantic context improves a metric in automatic evaluation. Both contexts are necessary, and therefore, PAIEL outperforms common LLM baselines by virtue of introducing them. (See Section V.)

(iii) We also conduct an evaluation of PAIEL on actual SOC operation as a real-world application. We provide PAIEL with several SOC analysts and then confirm if PAIEL is practical for SOC analysis in the real world. (See Section VI.)

II. Related Work

In this section, we review three strands of related work: applications of LLMs in SOC workflows, LLM-based interpretation and summarization of raw packets, and LLM-integrated explanations of mitigation and incident response. We then describe the position of PAIEL within this landscape.

A. LLMs in SOC workflows

LLMs have been increasingly adopted in SOCs to assist with incident management and threat analysis since these tasks often impose high stress on SOC analysts [27], [28]. LocalIntel [23] aggregates global and local incident knowledge of LLMs to generate attack reports, and Xpert [24] supports incident management through interaction between incident-ticket databases and LLMs. The framework by Albanse et al. [25] is a human-machine collaboration using LLMs for the design of SOC workflows. However, inputs of the above works are often curated incident information instead of raw packets. According to the recent survey [29], context-aware and human-centered assistance align with SOC workflows. PAIEL focuses on SOC workflows by analyzing packets to support both novice and veteran analysts.

B. LLM-based Packet Interpretation and Summarization

Recent studies have motivated direct applications of LLMs to network traffic, such as packet interpretation. NetGPT [26] is an early work, feeding raw network traffic to an LLM to generate natural-language descriptions. However, such approaches rely on the LLM's implicit knowledge of networking protocols without explicit grounding, and hence tend to generate superficial explanations due to insufficient protocol context [30]. More recently, TrafficLLM [31] maps traffic to a unified representation and instruction-tunes LLMs for detection. PAIEL is similar to NetGPT because of its analysis from a single packet, and is more advanced by integrating protocol specifications with Retrieval-Augmented Generation (RAG). We compare the performance with a NetGPT-like method as a common LLM baseline in our experiments.

C. LLM-integrated Explanation of Mitigation and Response

Explanation generation in security systems is a key focus and LLMs enable it. DoLLM [32] applies LLMs to flow-level feature analysis for attack detection, but operate on aggregated traffic statistics instead of packets. ContextBuddy [33] learns analysts' context-selection patterns via imitation learning to recommend relevant context. However, such an approach neither interprets protocol-level knowledge nor provides protocol-level explanations. In contrast, PAIEL generates protocol-level explanations for why a specific packet triggers an alert, based on protocol specifications and retrieved contextual knowledge.

D. Positioning of PAIEL

Table I summarizes the key differences between PAIEL and the above related work. The related work [26], [31] that directly interprets raw packets make limited use of explicit external knowledge: for instance, they often focus on incident summarization and knowledge retrieval instead of protocol-level explanations The design of PAIEL differs from the related work, because it combines structured context derived from protocol specifications with semantic context to generate protocol-level explanations directly from raw packets. To the best of our knowledge, few studies discuss such a design.

III. Problem Statement

In this section, we describe the problem setting of this paper in detail. In security operations centers (SOCs), security alerts are often linked to specific packets. Accordingly, our task is to explain why such packets cause a security alert, including articulating their protocol-level knowledge and security-relevant implications, through packets. We describe a task formalization and why it is challenging.

TABLE I

Comparison with related work on LLM-based cybersecurity applications. "Network" indicates whether the method can handle raw packets; "Knowledge" indicates whether it incorporates external databases or domain knowledge; "SOC Apps" indicates the type of application provided.

Method	Network	Knowledge	SOC Apps
LocalIntel [23]		✓	Incident response
Xpert [24]		✓	Incident response
Albanse et al. [25]		✓	Design of workflow
NetGPT [26]	✓		Traffic analysis
TrafficLLM [31]	✓		Traffic detection
DoLLM [32]			Flow-level detection
ContextBuddy [33]		✓	Alert investigation
PAIEL	✓	✓	Packet analysis

A. Task Formulation

Let $\mathcal{P}$ denote the space of packets and $\mathcal{E}$ the space of natural-language explanations. Each packet $p \in \mathcal{P}$ consists of protocol header fields, application-level payload data, and possibly lightweight metadata. The task is to learn a mapping $f : \mathcal{P} \to \mathcal{E}$ that generates an explanation, including the protocol-level knowledge and security-relevant implications, from packets, as well as providing clarifying context to support SOC analysts. We focus on protocol-level explanation from a single packet, focusing on analysis with sufficient protocol information and public specifications for the target protocols.

B. Why This Task Is Challenging and Our Approach

The use of a single packet often lacks protocol-level knowledge required for effective retrieval, making both direct LLM applications unreliable. It is also the reason why NetGPT [26] tends to generate superficial explanations.

We argue that faithful and accurate explanations from packets require contextual grounding beyond their packet fields to avoid superficial explanations and hallucination. Accordingly, our approach integrates explicit protocol knowledge with an LLM to generate explanations that emphasize security-relevant implications for analysts' decision making.

IV. Proposed Method: PAIEL

In this section, we propose PAIEL, a protocol-level explanation framework that addresses the lack of protocol-level knowledge by interpreting single packets. We first introduce its key idea and then outline the overall workflow. Subsequent subsections detail the context extraction components used in PAIEL as the main module, and then describe its extension and applications of PAIEL. This section is identical to the first technical contribution.

A. Key Idea

The key idea of PAIEL is to integrate two complementary sources of context, i.e., structured context derived from protocol specifications using a service database and semantic context retrieved from external knowledge using RAG. By unifying these contexts into a single prompt, PAIEL enables an LLM to generate faithful and accurate explanations for single packets. This unified approach addresses the challenge in protocol-level explanation from a single packet with sufficient knowledge, such as the intended behavior and protocol-defined constraints. To realize the above approach, PAIEL employs a context extraction module that combines a service database with RAG as retrieval, thereby overcoming the limited grounding and unreliable behavior observed in the retrieval [34]. Additionally, PAIEL incorporates a context compression module as an extension to mitigate noisy and over-retrieval results.

B. PAIEL Workflow

Figure 1(a) summarizes the PAIEL workflow, which proceeds in three steps: packet representation construction, context extraction, and explanation generation.

1) Packet Representation Construction: PAIEL first converts each raw packet into a packet representation that supports both context retrieval and explanation generation. During this step, PAIEL extracts standard header fields (e.g., addresses and ports) as well as protocol control fields, which identify the protocol-defined service or operation, for use in structured context extraction. PAIEL also extracts payload content and sanitizes it to remove noise and sensitive artifacts. The resulting text serves as a primary signal for semantic context extraction and, possibly, context compression. Device metadata about communicating endpoints is possibly included to further ground the explanation in the operational environment. Such metadata is optional and dataset-dependent.

2) Context Extraction: As shown in Figure 1(b), PAIEL combines structured context from the service database with semantic context retrieved from external knowledge using RAG in order to generate a prompt to an LLM. The semantic context is optionally refined by the context compression module described in Section IV-D. We detail the context extraction components in the next subsection as the main module.

3) Explanation Generation: The contexts extracted from the previous module are unified and incorporated into the final prompt. Using this prompt, the LLM generates a protocol-level explanation that describes the packet's knowledge and highlights security-relevant implications.

C. Details of Context Extraction

Context extraction explicitly grounds single packets in protocol specifications and complementary sources of context to construct a prompt. PAIEL derives structured context from a service database for protocols and augments it with semantic context retrieved from external knowledge using RAG, so that the resulting explanations remain accurate and faithful to protocol-defined behavior while incorporating security-relevant implications.

1) Structured Context Extraction Using Service Database: This section presents a method for extracting structured context from a single packet for protocol-level explanation. Many protocols, such as ICS protocols, expose explicit service and operation identifiers in their control fields. These identifiers specify the intended syntax of a message for each protocol. We leverage them to derive structured context for protocol-level

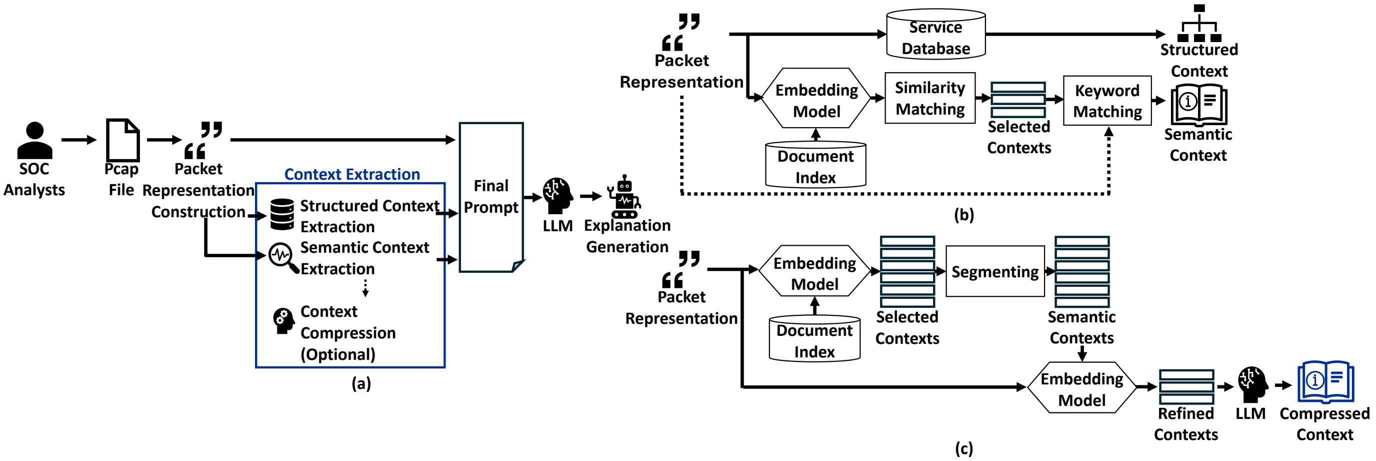

Fig. 1. Overview of the proposed PAIEL method. (a) Overall workflow of PAIEL. (b) Structured context extraction from the service database and semantic context extraction using RAG, which together form the context used in PAIEL. (c) Context compression pipeline, which replaces the semantic context extraction in (b) for the alternative PAIEL configuration.

explanations, allowing the LLM to interpret packets based on protocol-defined syntax more than superficial field values.

Specifically, for a given packet, service and operation identifiers are parsed from the protocol header and control fields. These identifiers are used to retrieve a corresponding syntactic description from a service database. This description constitutes the structured context for the protocol-level explanation. In doing so, the service database is constructed by exploiting protocol specifications, such as standardized service definitions and explicit identifier descriptions. This process does not require semantic interpretation because the structured context is derived from protocol fields and the protocol-defined syntax. The above approach makes the retrieval in RAG stable [34], and thus, can yield stable protocol-level explanations.

2) Semantic Context Extraction Using RAG: To complement the structured context, we incorporate semantic context that provides a broader operational and security-related implications. PAIEL retrieves the semantic context from external knowledge sources using a RAG approach.

To preserve relevant semantics during retrieval, PAIEL applies document-based chunking to external knowledge sources using RAG and builds a context database over these contexts. The chunked documents are indexed using dense vector representations to enable similarity-based retrieval. For each packet, PAIEL retrieves a small set of the most similar contexts from the document index as selected contexts. PAIEL further applies keyword matching between the packet representation and the selected contexts to emphasize lexically relevant content. Among the selected contexts, PAIEL determines the one with the highest keyword overlap with the packet representation and includes it in the LLM prompt as semantic context.

D. Extension: Context Compression

To address noise and over-retrieval results, PAIEL can introduce context compression as an optional step that refines the retrieved semantic context into a query-focused form. Figure 1(b) illustrates how structured and semantic contexts are constructed and combined within PAIEL. It mitigates noise and over-retrieval, which makes protocol-level knowledge obscure. PAIEL compresses retrieved context via a two-stage pipeline that narrows candidates and distills relevant content.

In the first stage, PAIEL forms a packet representation from protocol fields and, when available, incorporates the parsed service identifier. Using this query, PAIEL obtains a broad set of selected contexts. The selected contexts are segmented into semantic contexts, which are used to re-retrieve refined contexts and reduce over-retrieval results.

In the second stage, PAIEL distills a query-focused representation by extracting sparse keywords from packet fields and their values to guide the distillation step of the LLM. The LLM is instructed to generate output supported only by the refined contexts and to remove redundant information. The distilled contexts are then integrated into the final prompt as a compressed context, making protocol-level explanation more accurate and faithful while reducing noise in the explanations.

E. Applications

PAIEL is suitable for application-layer protocols with explicit service and operation definitions, and can extract service information without customization to each protocol. As our primary application, we focus on ICS/IoT protocols, such as BACnet [35], because they often contain diverse service information. We are supposed to apply PAIEL to the BACnet protocol [35], which is widely used in building automation systems. A BACnet packet consists of a Network Protocol Data Unit (NPDU) and an Application Protocol Data Unit (APDU), which together encode routing information and service-level information. PAIEL extracts service information from APDUs and retrieves relevant background information for explanations. The resulting context is provided to the LLM together with the packet payload in order to generate protocol-level explanations from a single packet.

As another application, we also describe an application of PAIEL to DNP3 [36] to demonstrate that the design generalizes beyond a single protocol. DNP3 packets encode application-layer function codes and object-based representations for control and monitoring operations. PAIEL applies the same extraction and retrieval pipeline to DNP3 traffic without introducing customization to each protocol.

V. Experiments

This section evaluates whether PAIEL can generate accurate and faithful protocol-level explanations from a single packet. Our extensive experiments are designed to answer RQ1 and RQ2. This section is also the second technical contribution.

A. Experimental Setting

We describe datasets, methods to be evaluated, prompts, and evaluation metrics in our experiments. Unless otherwise noted, BACnet protocol specifications are used as structured contextual knowledge throughout the experiments.

1) Datasets: We used datasets derived from two industrial control protocols, BACnet and DNP3, based on the applications in Section IV-E. We used both public and private datasets for the experiments. Due to ethical reasons, we conduct automatic evaluation on only the public dataset because it utilizes an external LLM service.

a) Public datasets: The public dataset consists of 50 BACnet pcap files extracted from the public wireshark data [37]. These files are publicly available and support repeatable evaluation under a shared experimental protocol. We also evaluate PAIEL on the DNP3 protocol [38], which is widely used in power and energy systems. For DNP3, we use a publicly available pcap file obtained from the ICS-Security-Tools repository [39]. Specifically, a database for the structured context of DNP3 is built by deterministically mapping protocol-defined Function Codes and Object Group/Variation identifiers to the corresponding specification descriptions in IEEE Std 1815-2012.

b) Private dataset: The private dataset consists of 19 BACnet pcap files containing security alerts collected from an operational building network. This reflects realistic alert scenarios in which a small number of packets trigger investigation. The captures are sanitized to replace identifying information with local placeholders. Packet timestamps range from February 2009 to March 2024.

2) Methods to Be Evaluated: We design baselines to isolate the contribution of each context source: no context, structured context only, semantic context only, and their integration. We then evaluate three baselines and two PAIEL variants. Table II compares the baseline configurations and the proposed PAIEL in terms of available contextual sources.

TABLE II
Comparison of baseline configurations and the proposed PAIEL in terms of available contextual sources.

Approach	Packet	Structured Ctx.	Semantic Ctx.
Vanilla LLM	✓	–	–
Database-only	✓	✓	–
RAG-only	✓	–	✓
PAIEL (Proposed)	✓	✓	✓

Vanilla LLM (no external context): In this baseline method, we directly prompt the LLM with packets without any external context, which is similar to NetGPT [26]. We implement it in full scratch instead of the code of NetGPT.

RAG-only baseline (with only semantic context): This baseline method uses standard retrieval-augmented generation without incorporating the service database. It follows the same document processing and retrieval configuration as PAIEL.

Database-only baseline (with only structured context): This baseline method uses only structured context derived from a service database. It isolates the effect of protocol-based structured context without semantic augmentation.

PAIEL-base (without context compression): This method integrates structured and semantic contexts as described in Section IV-C2, but does not apply the context compression.

PAIEL (with context compression): This method applies the context compression module to the PAIEL-base before prompting the LLM as described in Section IV-D.

Each method is implemented with the following settings. For all methods, we use Mixtral-8x7B-Instruct-v0.1, quantized to 4 bits, as the underlying LLM. Unless otherwise noted, the generation temperature is set to 0.0, with a repetition penalty of 1.1, and the maximum number of generated tokens to 2048. Each packet is converted into a normalized packet representation that includes standard header fields and protocol control fields required to identify the relevant service or operation. Device metadata is included when available.

Structured Context: We constructed a service database with the specification documents, in which each service and operation is associated with the specification descriptions. Each entry is then provided to the LLM as the structured context. Device metadata is included for the BACnet datasets.

Semantic Context: The specification documents are segmented into chunks and indexed for retrieval. We extract a set of keyword metadata from each chunk using an LLM and attach it as indexing information. Unless otherwise noted, all the methods using RAG utilize the same retrieval settings.

Context Compression: PAIEL applies a two-stage context compression process to fit within a fixed prompt. In the first stage, a broader set of candidate context segments is retrieved to prioritize recall (k=7). After integrating the segments with the structured contexts, the compressed context is re-segmented. The second stage refines the candidates to a smaller subset (k=4) that is most relevant to the query, and only the refined subset is included in the final prompt. For PAIEL (with context compression), these values are fixed regardless of protocol and packet. When the selected context spans multiple segments, we employ a tree-based aggregation module to consolidate relevant information under the input.

3) Prompts: We use three prompt types for explanation generation, compression, and faithfulness evaluation.

Explanation Prompts: The prompt template for explanation generation is as follows, where `\n\n` means the line break in command prompt:

> [INST]Given the packet contents, extract the part of the context *AS IS* which explains the terms or is related to the terms in packet.[/INST]\n Packet: {packet}\n Context: {doc}\n Instruction: Given the packet contents, only return extracted text *AS IS*\n Extracted text:\n\n

Compression Prompts: The prompt template for introducing the context compression module is as follows:

> Context information is below.\n ——————\n {retriever + service_db + document text}\n ——————\n Given the context information and no prior knowledge, return an answer only from context. Based on query remove any redundant information. If no relevant information is available return nothing.\n Query: Return the **part of the context** about the terms definitions individually or related to other terms below in understandable format:\n objectidentifier\n presentValue\n outOfService\n Answer:\n\n

Automatic Evaluation Prompts: The prompt template for the automatic evaluation is as follows:

> Query: \n [INST] You are a chatbot which explains BACnet packets.[/INST]\n Context: \n /** Answer **/\n Query : Explain the packet below using the context.\n Packet:\n {packet}

The query contains only the question text, while the packet content is included in the context with the retrieved documents.

4) Evaluation Metrics: We evaluate PAIEL using the following human evaluation and automatic evaluation. Each evaluation contains individual metrics.

a) Human Evaluation: We conduct a human evaluation by one of the authors, who is a network expert, to assess how the generated explanations are accurate from a human perspective. Human evaluation is reported using the following two metrics, where only the private dataset is evaluated:

Answer Relevancy (AR). This evaluates how explanations are relevant to security analysts. Explanations may sometimes include irrelevant information. A human expert assesses whether explanations are relevant for understanding packets. This metric follows control relevance [40] and coverage [41].

Control Accuracy (CA). This evaluates whether the statements in explanations are accurate for each packet and protocol. A human expert assesses whether explanations contain hallucinations and misleading claims on a 1–5 Likert scale.

b) Automatic Evaluation: We also conduct an automatic evaluation with an LLM to assess how the generated explanations are faithful from a systematic perspective. We use the following metric based on RAGAS [42], where only the public datasets are evaluated.

Faithfulness. This evaluates how the generated explanations are supported by context [42]. It is computed as $\frac{|V|}{|S|}$, where $|V|$ is the number of statements supported by the context and $|S|$ is the total number of statements in the explanation. We measure this metric using an LLM and normalize the score to the range $(0, 1)$, where a higher score is better. Faithfulness is computed using `GPT-4o` of Azure OpenAI[1] as the LLM, which tends to make hallucinations less and judgments reliable.

TABLE III

EVALUATION RESULTS ON EACH DATASET. BACNET IS A PRIVATE DATASET FOR HUMAN EVALUATION AND A PUBLIC DATASET FOR AUTOMATIC EVALUATION. ON THE OTHER HAND, DNP3 IS ONLY A PUBLIC DATASET FOR AUTOMATIC EVALUATION. WE DID NOT MEASURE FAITHFULNESS OF VANILLA LLM BECAUSE IT UNSTABLY COMPUTES THE SCORE DUE TO THE LACK OF RAG.

	BACnet			DNP3
Method	**AR**	**CA**	**Faithfulness**	**Faithfulness**
Vanilla LLM (no context)	3.18	4.86	–	–
RAG-only baseline	3.26	4.60	0.844	0.675
Database-only baseline	3.23	**4.89**	0.812	0.631
PAIEL-base	3.63	**4.89**	0.863	0.690
PAIEL	**3.82**	4.84	**0.876**	**0.695**

B. Results

We show the overall results with the metrics described above across all the methods. Table III shows the evaluation results for each protocol. These results demonstrate that PAIEL can generate accurate and faithful protocol-level explanations. Across AR and CA on human evaluation, both PAIEL-base and PAIEL improve the performance: for instance, PAIEL-base increases AR from 3.26 for RAG-only and 3.23 for Database-only to 3.63, and PAIEL further improves AR to 3.82. CA remains high when structured context is available: Database-only and PAIEL-base achieve the highest CA (4.89), while PAIEL slightly decreases to 4.84.

These results contain an ablation study for each module and hence reveal a trade-off between structured and semantic context. Database-only baseline achieves the highest score for CA but yields lower Faithfulness with 0.812 than RAG-only baseline with 0.844, which provides higher Faithfulness but reduces CA to 4.60. Integrating both contexts resolves this phenomenon: PAIEL-base matches the best CA with 4.89 while improving Faithfulness to 0.863 and AR to 3.63. Finally, the context compression module further improves Faithfulness with 0.876 and AR with 3.82 at a minor cost in CA. It suggests that the context compression makes protocol-level explanation more accurate and faithful.

Table III also shows that PAIEL achieves the highest Faithfulness for DNP3. Database-only baseline is a low score, and RAG-only baseline provides limited improvement. In contrast, PAIEL consistently yields the highest scores across both protocols. These results indicate that PAIEL, which integrates complementary sources of context, generalizes the performance across protocols with different domains.

C. Distributional Analysis of Faithfulness

To understand the impact of PAIEL on the generated explanation in more detail, we analyze the distribution of Faithfulness scores using boxplots. Figure 2a and Figure 2b compare RAG-only baseline and PAIEL on two protocols, i.e., BACnet and DNP3, under the metric described in Section V-A4b.

[1] https://azure.microsoft.com/en-us/products/ai-foundry/models/openai

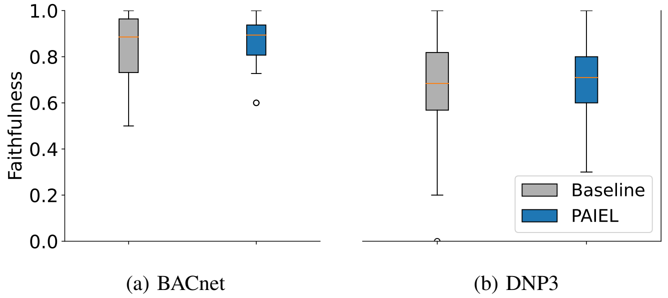

Fig. 2. Distribution of Faithfulness scores for the baseline and PAIEL. These results are obtained from the automatic evaluation.

According to the figures, the median Faithfulness remains largely comparable across both protocols. Meanwhile, PAIEL consistently raises the shorter tail of the distribution, which means improvements on the generated explanations compared with RAG-only baseline. This result is important for SOC analysts. Explanations with persistently low faithfulness require a high cognitive cost for SOC analysts, such as prompting repeated queries and extra verification about protocol specifications [2], [15]. Namely, an upward shift of the lower bound by PAIEL benefits human analysts.

D. Qualitative Analysis

We present a case illustrating the characteristic difference between Vanilla LLM and PAIEL-base as a qualitative analysis. Similar phenomena were observed across other packets. We quote excerpts verbatim from the generated explanations and omit irrelevant sentences for the sake of convenience.

1) Case Study 1. Specification grounding prevents field misinterpretation: Based on Table III, we analyze a representative packet where protocol grounding affects the interpretation of a payload field with the following packet:

Case 1 – Packet Excerpt (sanitized):
Service Choice: atomicWriteFile (7)
ObjectIdentifier: file, 1
record access
File Start Record: (Signed) 0
Record Count: (Unsigned) 1
Record Data: 12

Vanilla LLM tends to paraphrase field names but may inaccurately interpret fields regardless of the protocol specification. In this case, it interprets the literal value “12” as length (e.g., “12-bytes data”), which is plausible in contexts for a typical protocol but not justified by BACnet. In contrast, PAIEL-base leverages the structured context for `atomicWriteFile` and `record access`, allowing the model to interpret the packet as a file-write request and `Record Data` as the data content to be written (as value `12`). This illustrates how explicit protocol grounding reduces inaccurate protocol-level explanations. We show concrete output for each method below.

Vanilla LLM output (excerpted):
“The ADPU contains ’Record Data’ ... In this case, the data is 12 bytes long.”

PAIEL-base output (excerpted):
“Record Count: 1 ... Record Data: 12 ... the data content to be written.”

2) Case Study 2. Context compression improves coverage but may make explanations unessential: We next compare PAIEL-base and PAIEL to illustrate how the context compression affects the generated explanations, which helps interpret the slight differences in scores of the human evaluation in Table III. We focus on the following packet:

Case 2 – Packet Excerpt (sanitized):
Frame 1: Confirmed-REQ
Service Choice: readProperty (12)
ObjectIdentifier: load-control, 0
Property Identifier: expected-shed-level (214)
Frame 2: Complex-ACK
Service Choice: readProperty (12)
ObjectIdentifier: load-control, 0
Property Identifier: expected-shed-level (214)
shed level: (Unsigned) 0

PAIEL-base typically generates protocol-level explanations that focus on the request/response knowledge in the packet fields. PAIEL, by contrast, may incorporate a broader range of retrieved contents and distilled background. This can improve explanations. However, it will also make explanations redundant and unessential for interpreting behavior caused by the packet. We show concrete output for each method below.

PAIEL-base output (excerpted):
“Frame 1 requests the property value ... Frame 2 acknowledges and returns shed level = 0.”

PAIEL output (excerpted):
“Additional information from the standard: ... (background on load control / shed level)”

VI. Evaluation in Real World

In this section, we conduct an evaluation of PAIEL on SOC analysis in the real world in order to identify whether PAIEL is practical for SOC analysts in the real world. This section is identical to the third technical contribution.

A. Evaluation Design

We provide PAIEL with seven SOC analysts in the authors’ country, who are experts in ICS as the application described in Section IV-E, in order to use it in their SOC operation. The participants have diverse backgrounds, including novices and

TABLE IV

EVALUATION RESULTS ON SOC ANALYSIS IN THE REAL WORLD. EACH SCORE REPRESENTS THE AVERAGE ON A LIKERT-SCALE BY SEVEN SOC ANALYSTS, WHERE "5" IS "STRONGLY AGREE" AND "1" IS "STRONGLY DISAGREE".

Evaluation Category	**Evaluation Item**	**Avg. (Vanilla LLM)**	**Avg. (PAIEL)**
Overall Rating	Desire to Use Summarization	2.57	**4.29**
	Recommendation to Other Fields	3.43	**3.86**
	Support for Novices	2.86	**3.71**
	Support for Veterans	3.57	**4.29**
Quality of Information	Accuracy and Reliability of Output	3.50	**3.83**
	Readability and Comprehensibility of Output	**3.71**	3.29
Impact on Analysis Speed	Overall Analysis Time Reduction Effect	3.86	**4.43**
	Impact Understanding Speed	3.86	**4.57**
	Operational Understanding Speed	4.00	**4.43**
Impact on Analysis Quality	Smoothness of Report Creation	3.71	**4.43**
	Discrimination Between Normal/Abnormal Traffic	**3.43**	3.29

veterans, although we omit the details due to space limitations. In doing so, we conducted a preliminary survey to inform the purpose, procedure, and consent of this evaluation with open-ended questions. The task of the participants was to compare explanations generated by Vanilla LLM and PAIEL using their packet files, including security alerts, with access to specification documents and online resources. This evaluation was conducted using a 5-point Likert scale with accompanying free-text comments, following best practices in security research [43], although we omit the questionnaire details due to space limitations. Evaluation criteria include overall rating, which assesses the usefulness of the explanations and participants' willingness to use them in real-world tasks. Additional evaluation criteria assess the quality of information generated by LLMs, their impact on analysis speed, and their impact on the analysis quality of security alerts.

B. Results

The results are shown in Table IV. It indicates that PAIEL achieves higher scores than Vanilla LLM in most categories: remarkably, for all the evaluation items among the overall rating and the impact of the analysis speed, all the participants, including both novice and veteran analysts, rated significantly higher scores. Vanilla LLM achieves a higher score for the readability and comprehensibility of output because explanations with detailed content sometimes overwhelm several users. Although both methods receive moderate scores for the discrimination between normal/abnormal traffic, this indicates that protocol-level explanations alone still remain insufficient. We also received a positive comment for PAIEL as follows, although most comments are omitted due to space limitations: *"The summary information provided by PAIEL is effectively generated from detailed ICS protocol information and device data, aiding in understanding the relationship between equipment behavior and network events. As a result, it facilitates the interpretation of on-site conditions."* We received this comment from a novice analyst. While novice analysts are often challenged by the analysis of security alerts [44], PAIEL can facilitate the interpretation of packets according to the comment. Overall, this evaluation demonstrates that PAIEL is practical for SOC analysts in the real world.

VII. ETHICAL CONSIDERATIONS AND LIMITATIONS

For ethical considerations of the extensive experiments in Section V, we used publicly available datasets in the automatic evaluation because of the use of an external LLM service, and we collected neither personally identifiable information nor sensitive information. Before human evaluation, potentially identifying information contained in free-text responses, such as names and system identifiers, was anonymized and access to the data was restricted to the members involved in the human evaluation.

Likewise, for ethical considerations of the evaluation in the real world in Section VI, all the participants were informed in advance of the evaluation purpose and procedures, and informed consent was obtained prior to participation. All processing with their private packets was utilized in local models, and no private data was transmitted to external services.

There are three limitations of this paper. First, this study evaluated PAIEL on only two protocols, thereby limiting the generalizability to other protocols. Second, the service database must currently be manually constructed from protocol specifications, which may cover neither undocumented content nor incomplete specifications. Third, the current evaluation in the real world involved seven SOC analysts from a single country, and the results might be biased by the authors' culture. We are in the process of investigating PAIEL on diverse environments to overcome the above limitations.

VIII. CONCLUSION

This paper proposed PAIEL to address the challenge of generating faithful and accurate protocol-level explanations from a single packet. This task is important for effective SOC analysis but remains challenging due to the lack of protocol-level knowledge. PAIEL solves it by integrating structured context and semantic context. We demonstrated that PAIEL outperforms the common LLM baselines in terms of the answer relevancy and the control accuracy in human evaluation and the faithfulness in the automatic evaluation through extensive experiments. Importantly, these advantages indicate that both structured context and semantic contexts are necessary to generate accurate and faithful protocol-level explanations. We then conducted an evaluation of PAIEL by providing it with SOC analysts in the real world, and confirmed if PAIEL is practical in actual SOC operations. Future work includes further evaluation of a broader range of protocols, automating the construction of a service database, and real-world deployments into diverse SOCs.

References

[1] J. L. Tilbury and S. Flowerday, "Humans and automation: Augmenting security operation centers," *J. Cybersecur. Priv.*, vol. 4, no. 3, pp. 388–409, 2024. [Online]. Available: https://doi.org/10.3390/jcp4030020

[2] B. A. AlAhmadi, L. Axon, and I. Martinovic, "99% false positives: A qualitative study of SOC analysts' perspectives on security alarms," in *31st USENIX Security Symposium, USENIX Security 2022, Boston, MA, USA, August 10-12, 2022*, K. R. B. Butler and K. Thomas, Eds. USENIX Association, 2022, pp. 2783–2800. [Online]. Available: https://www.usenix.org/conference/usenixsecurity22/presentation/alahmadi

[3] L. Yang, Z. Chen, C. Wang, Z. Zhang, S. Booma, P. Cao, C. Adam, A. Withers, Z. Kalbarczyk, R. K. Iyer, and G. Wang, "True attacks, attack attempts, or benign triggers? an empirical measurement of network alerts in a security operations center," in *33rd USENIX Security Symposium, USENIX Security 2024, Philadelphia, PA, USA, August 14-16, 2024*, D. Balzarotti and W. Xu, Eds. USENIX Association, 2024. [Online]. Available: https://www.usenix.org/conference/usenixsecurity24/presentation/yang-limin

[4] S. Oesch, R. A. Bridges, J. M. Smith, J. M. Beaver, J. R. Goodall, K. M. T. Huffer, C. Miles, and D. Scofield, "An assessment of the usability of machine learning based tools for the security operations center," in *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), iThings/GreenCom/CPSCom/SmartData/Cybermatics 2020, Rhodes Island, Greece, November 2-6, 2020*. IEEE, 2020, pp. 634–641. [Online]. Available: https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics50389.2020.00111

[5] R. van der Kleij, J. M. Schraagen, B. Cadet, and H. Young, "Developing decision support for cybersecurity threat and incident managers," *Comput. Secur.*, vol. 113, p. 102535, 2022. [Online]. Available: https://doi.org/10.1016/j.cose.2021.102535

[6] A. Naseer, H. Naseer, A. Ahmad, S. B. Maynard, and A. M. Siddiqui, "Real-time analytics, incident response process agility and enterprise cybersecurity performance: A contingent resource-based analysis," *Int. J. Inf. Manag.*, vol. 59, p. 102334, 2021. [Online]. Available: https://doi.org/10.1016/j.ijinfomgt.2021.102334

[7] R. O. Andrade and S. G. Yoo, "Cognitive security: A comprehensive study of cognitive science in cybersecurity," *J. Inf. Secur. Appl.*, vol. 48, 2019. [Online]. Available: https://doi.org/10.1016/j.jisa.2019.06.008

[8] T. Ban, S. Ndichu, T. Takahashi, and D. Inoue, "Combat security alert fatigue with ai-assisted techniques," in *CSET '21: Cyber Security Experimentation and Test Workshop, Virtual, 9 August 2021*. ACM, 2021, pp. 9–16. [Online]. Available: https://doi.org/10.1145/3474718.3474723

[9] P. Baroni, F. Cerutti, D. Fogli, M. Giacomin, F. Gringoli, G. Guida, and P. Sullivan, "Self-aware effective identification and response to viral cyber threats," in *13th International Conference on Cyber Conflict, CyCon 2021, Tallinn, Estonia, May 25-28, 2021*, T. Jancárková, L. Lindström, G. Visky, and P. Zotz, Eds. IEEE, 2021, pp. 353–370. [Online]. Available: https://doi.org/10.23919/CyCon51939.2021.9468294

[10] G. González-Granadillo, S. González-Zarzosa, and R. Diaz, "Security information and event management (siem): Analysis, trends, and usage in critical infrastructures," *Sensors*, vol. 21, no. 14, pp. 1–28, 2021.

[11] M. Husák, L. Sadlek, S. Spacek, M. Lastovicka, M. Javorník, and J. Komárková, "CRUSOE: A toolset for cyber situational awareness and decision support in incident handling," *Comput. Secur.*, vol. 115, p. 102609, 2022. [Online]. Available: https://doi.org/10.1016/j.cose.2022.102609

[12] F. B. Kokulu, A. Soneji, T. Bao, Y. Shoshitaishvili, Z. Zhao, A. Doupé, and G. Ahn, "Matched and mismatched socs: A qualitative study on security operations center issues," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019, London, UK, November 11-15, 2019*, L. Cavallaro, J. Kinder, X. Wang, and J. Katz, Eds. ACM, 2019, pp. 1955–1970. [Online]. Available: https://doi.org/10.1145/3319535.3354239

[13] T. Yen, A. Oprea, K. Onarlioglu, T. Leetham, W. K. Robertson, A. Juels, and E. Kirda, "Beehive: large-scale log analysis for detecting suspicious activity in enterprise networks," in *Annual Computer Security Applications Conference, ACSAC '13, New Orleans, LA, USA, December 9-13, 2013*, C. N. P. Jr., Ed. ACM, 2013, pp. 199–208. [Online]. Available: https://doi.org/10.1145/2523649.2523670

[14] C. Zhong, J. Yen, P. Liu, and R. F. Erbacher, "Learning from experts' experience: Toward automated cyber security data triage," *IEEE Syst. J.*, vol. 13, no. 1, pp. 603–614, 2019. [Online]. Available: https://doi.org/10.1109/JSYST.2018.2828832

[15] A. I. Hauptman, B. G. Schelble, N. J. McNeese, and K. C. Madathil, "Adapt and overcome: Perceptions of adaptive autonomous agents for human-ai teaming," *Comput. Hum. Behav.*, vol. 138, p. 107451, 2023. [Online]. Available: https://doi.org/10.1016/j.chb.2022.107451

[16] W. U. Hassan, S. Guo, D. Li, Z. Chen, K. Jee, Z. Li, and A. Bates, "Nodoze: Combatting threat alert fatigue with automated provenance triage," in *26th Annual Network and Distributed System Security Symposium, NDSS 2019, San Diego, California, USA, February 24-27, 2019*. The Internet Society, 2019. [Online]. Available: https://www.ndss-symposium.org/ndss-paper/nodoze-combatting-threat-alert-fatigue-with-automated-provenance-triage/

[17] J. Happa, I. Agrafiotis, M. Helmhout, T. Bashford-Rogers, M. Goldsmith, and S. Creese, "Assessing a decision support tool for SOC analysts," *DTRAP*, vol. 2, no. 3, pp. 22:1–22:35, 2021. [Online]. Available: https://doi.org/10.1145/3430753

[18] M. A. Gurabi, M. U. Mansoor, R. Matzutt, A. Mandal, and S. Decker, "A conceptual framework to leverage heuristics for effective human-machine collaboration in incident handling," in *Innovative Security Solutions for Information Technology and Communications - 17th International Conference, SecITC 2024, Bucharest, Romania, November 21-22, 2024, Revised Selected Papers*, ser. Lecture Notes in Computer Science, L. Morogan, P. B. Roenne, and I. Bica, Eds., vol. 15595. Springer, 2024, pp. 18–35. [Online]. Available: https://doi.org/10.1007/978-3-031-87760-5_3

[19] F. Jalalvand, M. B. Chhetri, S. Nepal, and C. Paris, "Alert prioritisation in security operations centres: A systematic survey on criteria and methods," *ACM Comput. Surv.*, vol. 57, no. 2, pp. 42:1–42:36, 2025. [Online]. Available: https://doi.org/10.1145/3695462

[20] Y. Chen, M. Cui, D. Wang, Y. Cao, P. Yang, B. Jiang, Z. Lu, and B. Liu, "A survey of large language models for cyber threat detection," *Comput. Secur.*, vol. 145, p. 104016, 2024. [Online]. Available: https://doi.org/10.1016/j.cose.2024.104016

[21] I. H. Sarker, *Generative AI and Large Language Modeling in Cybersecurity*. Cham: Springer Nature Switzerland, 2024, pp. 79–99. [Online]. Available: https://doi.org/10.1007/978-3-031-54497-2_5

[22] H. Xu, S. Wang, N. Li, K. Wang, Y. Zhao, K. Chen, T. Yu, Y. Liu, and H. Wang, "Large language models for cyber security: A systematic literature review," *CoRR*, vol. abs/2405.04760, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2405.04760

[23] S. Mitra, S. Neupane, T. Chakraborty, S. Mittal, A. Piplai, M. Gaur, and S. Rahimi, "Localintel: Generating organizational threat intelligence from global and local cyber knowledge," in *Foundations and Practice of Security - 17th International Symposium, FPS 2024, Montréal, QC, Canada, December 9-11, 2024, Revised Selected Papers, Part II*, ser. Lecture Notes in Computer Science, K. Adi, S. Bourdeau, C. Durand, V. V. T. Tong, A. Dulipovici, Y. Kermarrec, and J. García-Alfaro, Eds., vol. 15533. Springer, 2024, pp. 63–78. [Online]. Available: https://doi.org/10.1007/978-3-031-87496-3_5

[24] Y. Jiang, C. Zhang, S. He, Z. Yang, M. Ma, S. Qin, Y. Kang, Y. Dang, S. Rajmohan, Q. Lin, and D. Zhang, "Xpert: Empowering incident management with query recommendations via large language models," in *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering, ICSE 2024, Lisbon, Portugal, April 14-20, 2024*. ACM, 2024, pp. 92:1–92:13. [Online]. Available: https://doi.org/10.1145/3597503.3639081

[25] M. Albanese, X. Ou, K. Lybarger, D. Lende, and D. B. Goldgof, "Towards ai-driven human-machine co-teaming for adaptive and agile

cyber security operation centers," *CoRR*, vol. abs/2505.06394, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2505.06394

[26] X. Meng, C. Lin, Y. Wang, and Y. Zhang, "Netgpt: Generative pretrained transformer for network traffic," *CoRR*, vol. abs/2304.09513, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2304.09513

[27] S. Nepal, J. Hernandez, R. Lewis, A. Chaudhry, B. Houck, E. Knudsen, R. Rojas, B. Tankus, H. Prafullchandra, and M. Czerwinski, "Burnout in cybersecurity incident responders: Exploring the factors that light the fire," *Proceedings of the ACM on Human-Computer Interaction*, vol. 8, no. CSCW1, pp. 1–35, 2024.

[28] K. Thimmaraju, S. I. Rispens, and G.-J. Ahn, "Human performance in security operations: a survey on burnout, well-being and flow state among practitioners," in *Proc. of WOSOC 2025*, 2025, pp. 1–10.

[29] R. Singh, S. Tariq, F. Jalalvand, M. B. Chhetri, S. Nepal, C. Paris, and M. Lochner, "Llms in the SOC: an empirical study of human-ai collaboration in security operations centres," *CoRR*, vol. abs/2508.18947, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2508.18947

[30] M. A. Ferrag, M. Ndhlovu, N. Tihanyi, L. C. Cordeiro, M. Debbah, T. Lestable, and N. S. Thandi, "Revolutionizing cyber threat detection with large language models: A privacy-preserving bert-based lightweight model for iot/iiot devices," *IEEE Access*, vol. 12, pp. 23 733–23 750, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3363469

[31] T. Cui, X. Lin, S. Li, M. Chen, Q. Yin, Q. Li, and K. Xu, "Trafficllm: Enhancing large language models for network traffic analysis with generic traffic representation," *CoRR*, vol. abs/2504.04222, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2504.04222

[32] Q. Li, Y. Zhang, Z. Jia, Y. Hu, L. Zhang, J. Zhang, Y. Xu, Y. Cui, Z. Guo, and X. Zhang, "Dollm: How large language models understanding network flow data to detect carpet bombing ddos," *CoRR*, vol. abs/2405.07638, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2405.07638

[33] R. Singh, M. B. Chhetri, S. Nepal, and C. Paris, "Contextbuddy: Ai-enhanced contextual insights for security alert investigation (applied to intrusion detection)," *CoRR*, vol. abs/2506.09365, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2506.09365

[34] S. Barnett, S. Kurniawan, S. Thudumu, Z. Brannelly, and M. Abdelrazek, "Seven failure points when engineering a retrieval augmented generation system," in *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI, CAIN 2024, Lisbon, Portugal, April 14-15, 2024*, J. Cleland-Huang, J. Bosch, H. Muccini, and G. A. Lewis, Eds. ACM, 2024, pp. 194–199. [Online]. Available: https://doi.org/10.1145/3644815.3644945

[35] S. T. Bushby, *BACnetTM: a standard communication infrastructure for intelligent buildings*. ASHRAE, 1997, vol. 6, no. 5, pp. 529–540. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0926580597000290

[36] *IEEE Standard for Electric Power Systems Communications—Distributed Network Protocol (DNP3)*, IEEE Std. 1815-2012, 2012.

[37] Steve Karg, "Wireshark Sample Captures," accessed: 2025-12-25, MIT License Copyright (c) 2023 kargs-net Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions: The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software. THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE. [Online]. Available: https://github.com/kargs-net/kargs-net.github.io

[38] IEEE, "IEEE Standard for Electric Power Systems Communications – Distributed Network Protocol (DNP3)," https://standards.ieee.org/ieee/1815/5414/, IEEE, 2012.

[39] ITI, "ICS-Security-Tools," accessed: 2025-12-25, Licensed Under CC-BY-4.0. [Online]. Available: https://github.com/ITI/ICS-Security-Tools/

[40] J. He, W. Kryscinski, B. McCann, N. Rajani, and C. Xiong, "Ctrlsum: Towards generic controllable text summarization," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, Y. Goldberg, Z. Kozareva, and Y. Zhang, Eds. Association for Computational Linguistics, 2022, pp. 5879–5915. [Online]. Available: https://doi.org/10.18653/v1/2022.emnlp-main.396

[41] S. Wang, E. Khramtsova, S. Zhuang, and G. Zuccon, "Feb4rag: Evaluating federated search in the context of retrieval augmented generation," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, G. H. Yang, H. Wang, S. Han, C. Hauff, G. Zuccon, and Y. Zhang, Eds. ACM, 2024, pp. 763–773. [Online]. Available: https://doi.org/10.1145/3626772.3657853

[42] S. ES, J. James, L. E. Anke, and S. Schockaert, "Ragas: Automated evaluation of retrieval augmented generation," in *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2024 - System Demonstrations, St. Julians, Malta, March 17-22, 2024*, N. Aletras and O. D. Clercq, Eds. Association for Computational Linguistics, 2024, pp. 150–158. [Online]. Available: https://aclanthology.org/2024.eacl-demo.16

[43] E. M. Redmiles, Y. Acar, S. Fahl, and M. L. Mazurek, "A summary of survey methodology best practices for security and privacy researchers," University of Maryland, Tech. Rep. CS-TR-5055, 2017. [Online]. Available: https://doi.org/10.13016/M22K2W

[44] F. Hahn, S. Cherry, K. Shashwat, L. Buldrini, D. Lende, and X. Ou, "Tools make me snore: A next-gen framework for training soc analysts non-perishable skills," in *Proc. of WOSOC 2025*, 2025, pp. 1–10.