Integrating Knowledge Graphs with Retrieval-Augmented Generation to Automate IoT Device Security Compliance

Mohammad M Islam
Dept. of Information Systems
University of Maryland Baltimore
County, Baltimore, MD, USA
im4@umbc.edu

Lavanya Elluri
Computer Information Systems
Texas A&M University-Central Texas,
Killeen, TX, USA
elluri@tamuct.edu

Karuna Pande Joshi
Dept. of Information Systems
University of Maryland Baltimore
County, Baltimore, MD, USA
kjoshi1@umbc.edu

Abstract - As IoT device adoption grows, ensuring cybersecurity compliance with IoT standards, like National Institute of Standards and Technology Interagency (NISTIR) 8259A, has become increasingly complex. These standards are typically presented in lengthy, text-based formats that are difficult to process and query automatically. We built a knowledge graph to address this challenge to represent the key concepts, relationships, and references within NISTIR 8259A. We further integrate this knowledge graph with Retrieval-Augmented Generation (RAG) techniques that can be used by large language models (LLMs) to enhance the accuracy and contextual relevance of information retrieval. Additionally, we evaluate the performance of RAG using both graph-based queries and vector database embeddings. Our framework, implemented in Neo4j, was tested using multiple LLMs, including LLAMA2, Mistral-7B, and GPT-4. Our findings show that combining knowledge graphs with RAG significantly improves query precision and contextual relevance compared to unstructured vector-based retrieval methods. While traditional rule-based compliance tools were not evaluated in this study, our results demonstrate the advantages of structured, graph-driven querying for security standards like NISTIR 8259A.

Keywords - Knowledge Graph, Retrieval-Augmented Generation (RAG), IoT Security, Cybersecurity Compliance, Neo4j, Large Language Models (LLMs), LangChain, Vector Database

I. Introduction

The growing adoption of Internet of Things (IoT) devices has brought an urgent need for automated and accurate cybersecurity compliance management. Frameworks like NISTIR 8259A provide baseline security capabilities and guidance for manufacturers to secure their devices. However, these documents are dense, complex, and not machine-processable, making manual compliance time-consuming and error-prone. Existing research has highlighted similar challenges in domains such as medical IoT security [1] and cloud compliance automation [2].

Knowledge graphs offer a structured way to represent relationships and entities embedded within these standards. When combined with Retrieval-Augmented Generation (RAG), they enable more contextual, accurate, and dynamic querying. Recent surveys and studies [3], [5], [8] demonstrate the increasing role of RAG systems and (LLMs) in transforming how complex knowledge is queried and retrieved.

In this paper, we present a comprehensive knowledge graph for NISTIR 8259A, designed to capture cybersecurity capabilities, rationales, and associated references. We integrated this graph with RAG techniques and evaluated its performance using Neo4j [9] as the graph database, coupled with various LLMs such as LLAMA2, Mistral-7B, and GPT-4. This builds upon prior work in ontology-based compliance systems [2], trustworthy RAG systems [3], and semantic knowledge graph querying [10]. Our comparative analysis demonstrates how this approach improves query precision, contextual relevance, and retrieval efficiency, offering a practical path toward automating IoT cybersecurity compliance.

II. Related Work

A. Cybersecurity Compliance Automation

Cybersecurity compliance is often complicated by the volume and complexity of regulatory standards. While frameworks like NISTIR 8228 and 8259 offer valuable guidance for IoT security, their text-heavy nature limits automation [1]. Manual compliance checks are still standard, which can be inefficient and error-prone [2], [6].

To address this, researchers have proposed more automated approaches. Dover et al. highlight how IoT introduces risks not fully addressed by traditional frameworks and proposes a quantitative method for assessing compliance based on NISTIR 8228 [1]. Others have explored translating regulatory texts into structured, machine-readable formats using semantic technologies, enabling automated querying and validation [4]. Although promising, integrating these solutions into fast-evolving areas like IoT remains challenging. Recent work points to AI-driven approaches such as knowledge graphs and RAG models as potential enablers of scalable, automated compliance.

Knowledge graphs (KGs) offer a structured way to represent cybersecurity rules and relationships, improving both retrieval and automation [5]. Frameworks like the Unified Cybersecurity Ontology (UCO) connect standards and threat data to support compliance analysis [5]. Joshi et al. further showed how KGs can encode security controls—like HIPAA—for automated validation in cloud systems [6]. Building on this, IoT-Reg integrates regulations such as NISTIR 8228, GDPR, and HIPAA, linking them to device capabilities to support real-time assessments [6]. However, KG development remains labor-intensive. Emerging work on dynamic, self-updating graphs aims to reduce manual upkeep as policies change [13].

RAG enhances large language models by retrieving external information during inference, improving accuracy and relevance [5]. Initially introduced by Lewis et al., RAG has been applied in domains like compliance and security policy enforcement [5], [8]. DepsRAG, developed by Alhanahnah et al., applies this concept to software security by retrieving compliance policies based on a graph of software dependencies, reducing manual effort [4]. Still, dynamic retrieval raises concerns about trust and consistency [12]. Hybrid approaches combining RAG with structured KGs are gaining attention for flexibility and reliability [12].

Unlike prior work that focuses solely on ontology representation or basic retrieval, our approach combines structured semantic reasoning (via Cypher queries on knowledge graphs) with generative LLMs like GPT-4, LLAMA-2, and Mistral-7B output, offering a practical, automated compliance pipeline. Our comparative analysis of different LLMs and hybrid retrieval systems against the NISTIR 8259A framework is a novel contribution not found in earlier literature.

III. Methodology

We have developed a novel approach that integrates Knowledge Graphs KGs and RAG, which can be used by LLMs to extract compliance policies from cybersecurity standards. Our methodology consists of the following key components:

A. NISTIR 8259 Cybersecurity Knowledge Graph

A knowledge graph is the foundation for structuring and representing cybersecurity compliance data. We have created a comprehensive KG based on NISTIR 8259A, a widely recognized framework for IoT device security. The KG captures the relationship between security capabilities, compliance rules, device attributes, and threat mitigation strategies.

1. Data Collection and Preprocessing

The first step involved extracting key security concepts, requirements, and recommendations from NISTIR 8259A and related IoT security compliance documents. Extracted information was structured into entities (e.g., device configuration, data protection, access control) and relationships (e.g., compliance dependencies, risk mitigation links, security best practices).

We derived these capabilities through a manual content analysis of the NISTIR 8259A document, identifying recurring themes and security objectives. Each section of the standard was reviewed by two independent researchers to extract core security functions. These were mapped to nodes and relationships in the KG schema using Neo4j modeling conventions.

2. Knowledge Graph Modeling

The knowledge graph was constructed using Neo4j, a graph database optimized for structured knowledge representation. (Fig. 1) illustrates this KG. The schema was designed based on the six core cybersecurity capabilities outlined in NISTIR 8259A. To ensure the accuracy and reliability of the knowledge graph, we used independent expert review to validate both its structure and content. Furthermore, we recognize the dynamic nature of cybersecurity knowledge and will continue to verify and refine the knowledge graph through systematic evaluation and iterative updates. These core capabilities are:

Device Configuration ensures that an IoT device's settings can be modified securely while restricting unauthorized changes. A knowledge graph links this capability to related controls, policies, and compliance requirements, enabling automated checks on configuration security.

Data Protection involves securing stored and transmitted data using cryptographic techniques. The knowledge graph establishes relationships between data protection mechanisms, cryptographic standards, and threat vectors to support real-time risk assessments.

Device Identification assigns unique logical and physical identifiers to IoT devices, which is crucial for authentication and tracking.

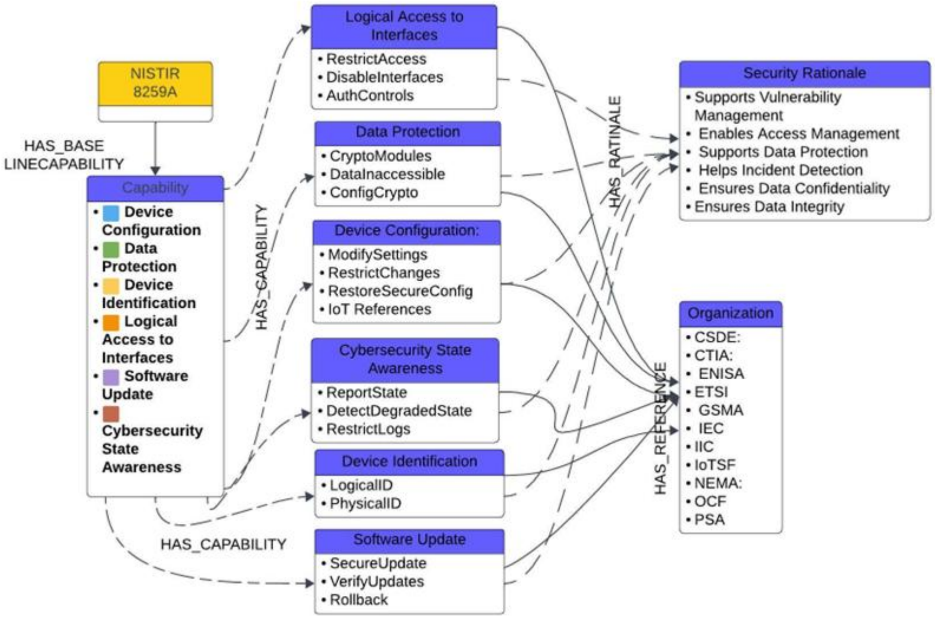

Fig. 1: Block Diagram for Knowledge Graph NIST 8259

The knowledge graph connects these identifiers to authentication protocols, access control mechanisms, and device management systems, ensuring robust identity verification.

Logical Access to Interfaces governs the restriction of access to the network and local interfaces, reducing the attack surface. The knowledge graphs access rules, authorization mechanisms, and potential attack pathways, allowing organizations to enforce least privilege principles.

Software Update capability ensures that IoT devices can be securely updated while preventing unauthorized tampering. A knowledge graph helps automate vulnerability management, patch deployment tracking, and rollback procedures by mapping update mechanisms to security best practices.

Cybersecurity State Awareness enables continuous monitoring and reporting of a device's security status. A knowledge graph integrates security logs, anomaly detection systems, and compliance frameworks, enhancing situational awareness and incident response.

By interlinking these capabilities with cybersecurity frameworks, regulatory standards, and industry best practices, a knowledge graph facilitates contextualized security intelligence, supports automated policy enforcement, and strengthens cybersecurity resilience within IoT ecosystems. Its inherent ability to perform semantic reasoning, threat detection, and compliance verification positions it as a critical tool for security analysts, regulatory bodies, and IoT manufacturers alike. As IoT networks continue to increase in scale and complexity, knowledge graphs offer a scalable, adaptive, and intelligent solution for managing cybersecurity, ensuring that devices remain secure, compliant, and resilient in the face of evolving threats. To validate and continually improve the integrity and relevance of our knowledge graph, we actively engage independent researchers who provide objective assessments and expert feedback throughout its development lifecycle.

B. Integrating Knowledge Graph Querying into a RAG-Based Compliance System

We incorporated RAG techniques into the compliance automation system to improve retrieval accuracy and contextual relevance. RAG enhances LLMs by dynamically retrieving relevant knowledge from structured sources (e.g., knowledge graphs and vector databases) before generating a response.

A key innovation of our methodology is using queries to interact with the knowledge graph (Fig. 2). Instead of static keyword-based retrieval, cypher-based queries allow for context-aware searches that leverage graph relationships to extract the most relevant compliance rules. Cypher query examples in Neo4j:

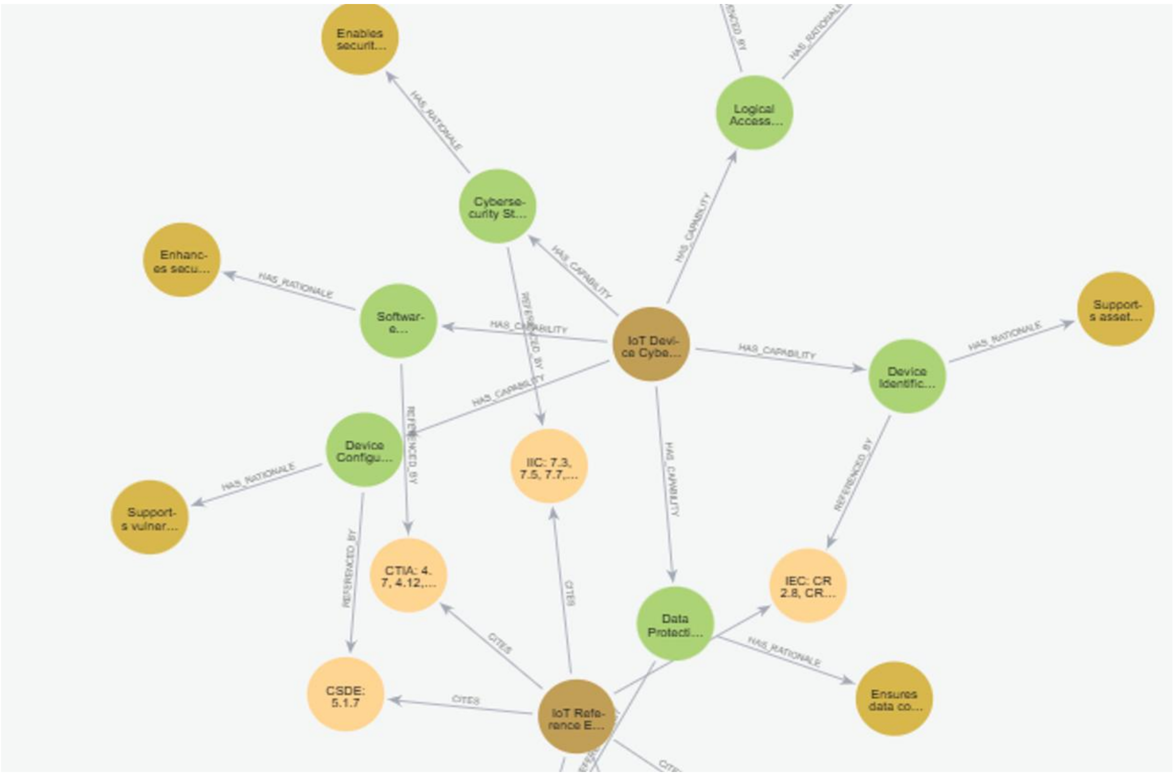

Fig. 2: Visualization of Knowledge Graph structure showing hierarchical relationships among NISTIR 8259A capabilities and linked compliance rules modeled in Neo4j.

Get list *all main capabilities*

$$\boldsymbol{MATCH}\ (\boldsymbol{c}: \boldsymbol{Capability})$$
$$\boldsymbol{RETURN\ c.name\ AS\ Capability}$$

Get all features under a specific capability (e.g., "Data Protection")

$$\boldsymbol{MATCH}\ (: \boldsymbol{Capability}\ \{\boldsymbol{name}: "\boldsymbol{Data\ Protection}"\}) - [: \boldsymbol{HAS\_FEATURE}] -> (\boldsymbol{f}: \boldsymbol{Feature})$$
$$\boldsymbol{RETURN\ f.name\ AS\ Feature}$$

These structured queries were then translated into natural language so that users could interact with the system more intuitively. In parallel, we also use vector similarity search to retrieve semantically relevant content from unstructured text, enhancing the model's context and improving the relevance of its responses.

C. Comparative Analysis of Querying Methods: Vector Database vs. Knowledge Graphs

To evaluate the effectiveness of our approach, we compared vector-based retrieval against knowledge graph-based retrieval using the same set of IoT security queries.

1. Vector Database-Based Retrieval

Vector-based retrieval RAF (Fig.3) was implemented using Facebook AI Similarity Search (FAISS), where compliance documents and security guidelines were embedded into high-dimensional vector representations. Queries were processed by finding the most semantically similar vector representations.

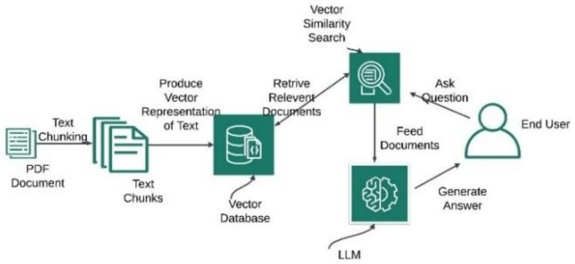

Fig. 3: RAG using a knowledge graph for LLM applications.

2. Knowledge Graph-Based Retrieval

KG-based retrieval was evaluated by executing cypher-based queries to extract compliance data. This approach preserved hierarchical relationships between security concepts. Knowledge graph coupled with RAG techniques (Fig 4) facilitates greater compliance retrieval accuracy and contextual relevance. With RAG, the framework can obtain relevant information dynamically from the knowledge graph.

RAG enhances the retrieval's precision while ensuring the nuances of the context between the IoT device security standards such as NISTIR 8259A [4].

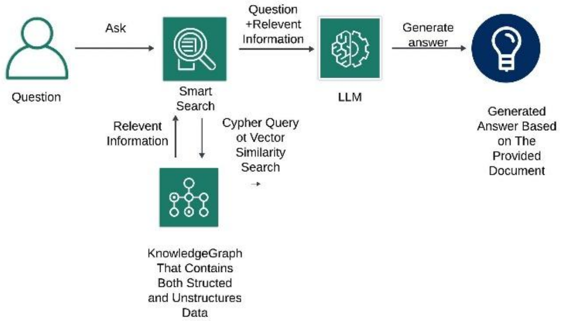
Question
Ask
Smart Search
Question +Relevent Information
LLM
Generate answer
Generated Answer Based on The Provided Document
Relevent Information
Cypher Query ot Vector Similarity Search
KnowledgeGraph That Contains Both Structed and Unstructures Data

Fig. 4: RAG using a knowledge graph for LLM applications.

By comparing query accuracy and response relevance, our results demonstrate that KG-based retrieval provides more context-aware compliance responses than vector retrieval, particularly for complex regulatory questions.

D. Evaluating Retrieval-Augmented Compliance Automation Using LLMs

We tested three different LLMs (GPT-4, LLAMA-2, and Mistral-7B) to analyze their effectiveness in retrieving and contextualizing security compliance information.

1. Experimental Setup

This study focuses on the "Data Protection" capability, beginning with directly querying LLMs to assess baseline performance. Next, we embed the capability and its query for vector-based retrieval of relevant reference documents. We also created a knowledge graph, which we also query through LLMs. To evaluate effectiveness, we tested LLAMA2, Mistral-7B, and GPT-4 across three experimental setups, each progressively adding retrieval augmentation to analyze the impact on the models' ability to extract and apply compliance information.

Phase-1: Baseline Performance Without Retrieval-Augmented Generation (No RAG).

The first phase establishes a baseline performance metric by evaluating how well the LLMs respond to compliance-related queries without retrieving external data.

Example Query: "In NIST.IR 8259A, what are the references for the Data Protection capability?"

Phase-2: Text-Based Retrieval-Augmented Generation
In the second phase, we incorporate text-based retrieval by leveraging a vector database to supply relevant regulatory documents before the models generate responses.

Phase-3: Structured Retrieval with Knowledge Graphs
The third phase evaluates the impact of structured knowledge retrieval by integrating a cybersecurity KG into the RAG system.

2. Performance Analysis

The dataset presents the performance of three language models—GPT-4, LLAMA-2, and Mistral-7B—across three distinct configurations: Vector Database RAG, Direct Prompting with LLMs, and Knowledge Graph RAG. Here's a more concise, scientific explanation of their performances:

Table 1: Retrieval Accuracy and Contextual Relevance of LLMs Using Prompt Direct LLMs Results

Model	**Retrieval Accuracy (%)**	**Contextual Relevance (%)**
GPT-4	2%	1%
LLAMA-2	0%	0%
Mistral-7B	0%	0%

Table 2: Retrieval Accuracy and Contextual Relevance of LLMs using Vector Database RAG Results

Model	**Retrieval Accuracy (%)**	**Contextual Relevance (%)**
GPT-4	83%	78%
LLAMA-2	60%	60%
Mistral-7B	55%	50%

Accuracy was measured by comparing retrieved answers against a set of ground truth statements derived from the NISTIR 8259A document.

Table 3: Retrieval Accuracy and Contextual Relevance of LLMs using Knowledge Graph RAG Results

Model	Retrieval Accuracy (%)	Contextual Relevance (%)
GPT-4	97%	95%
LLAMA-2	85%	80%
Mistral-7B	70%	70%

Contextual relevance was scored by two domain experts using a 5-point Likert scale, then converted to percentage agreement. Each model answered 20 compliance questions under three configurations.

The results highlight clear performance differences across methods. Direct prompting with LLMs showed minimal success, with GPT-4 achieving just 2% accuracy and the others scoring zero. However, adding vector-based retrieval significantly improved performance—The best results came from the Knowledge Graph RAG setup, where GPT-4 achieved 97% accuracy and 95% contextual relevance.

IV. Conclusion and Future Work

This study demonstrates the effectiveness of combining knowledge graphs with Retrieval-Augmented Generation (RAG) to enhance retrieval accuracy and contextual relevance for IoT security compliance. By structuring the NISTIR 8259A standard into a knowledge graph, we enable more targeted and meaningful query responses, particularly when integrated with large language models like LLAMA2, Mistral-7B, and GPT-4. While our evaluation focused on retrieval performance, the results suggest this approach can support future efforts toward compliance automation.

Future work will focus on dynamically updating knowledge graphs with evolving security policies, developing hybrid retrieval models that balance speed and precision, and fine-tuning LLMs on domain-specific compliance data to improve accuracy and adaptability. Together, these enhancements move us closer to scalable, AI-driven compliance solutions for the rapidly evolving IoT landscape.

Acknowledgements

This work was partially funded by the National Science Foundation (NSF) award 2310844, IUCRC Phase II UMBC: Center for Accelerated Real-Time Analytics (CARTA) and by NSF award 2348147. We express our gratitude to colleagues whose insights and expertise significantly contributed to the research.

References

[1] Dover, T. P. (2021). Evaluating medical IoT (MIoT) device security using NISTIR-8228 expectations. arXiv preprint arXiv:2104.03283.

[2] Dixit, S., Joshi, K. P., Choi, S. G., & Elluri, L. (2022, May). Semantically rich access control in cloud ehr systems based on ma-abe. In 2022 IEEE 8th intl conference on big data security on cloud (BigDataSecurity), IEEE intl conference on high performance and smart computing, (HPSC) and IEEE intl conference on intelligent data and security (IDS) (pp. 1-10). IEEE.

[3] Y. Zhou, et al., "Trustworthiness in retrieval-augmented generation systems: A survey," arXiv preprint arXiv:2409.10102, 2024.

[4] M. Alhanahnah, Y. Boshmaf, and B. Baudry, "DepsRAG: Towards Managing Software Dependencies Using Large Language Models," arXiv preprint arXiv:2405.20455, 2024.

[5] E. P. Lewis, P. Perez, A. Piktus, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems (NeurIPS), 2020.

[6] K. U. Echenim and K. P. Joshi, "IoT-Reg: A Comprehensive Knowledge Graph for Real-Time IoT Data Privacy Compliance," Proceedings of IEEE Big Data Conference, 2023.

[7] R. Jia, B. Zhang, S. J. Méndez, et al., "Leveraging Large Language Models for Semantic Query Processing in a Scholarly Knowledge Graph," arXiv preprint arXiv:2405.15374, 2024.

[8] W. Fan, Y. Ding, L. Ning, et al., "A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models," ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2024.

[9] "Neo4j: The Leading Graph Database for Connected Data," Neo4j Documentation. Available at: https://neo4j.com

[10] P. G. Omran, R. Jia, and J. Zhang, "Semantic Knowledge Graph Querying for AI," arXiv preprint arXiv:2405.15399, 2024.

[11] M. A. Nasir, M. Shafiq, A. Anwar, and F. R. Dogar, "Exploring Graph-Based Security Compliance Models for Cloud Environments," IEEE Transactions on Cloud Computing, vol. 12, no. 3, pp. 1453-1465, 2023.

[12] A. Singh, L. K. Joshi, and R. Gupta, "Ensuring Trustworthy AI-Based Compliance Verification: A Hybrid Approach," Journal of AI and Cybersecurity, vol. 8, no. 2, pp. 101-115, 2024.

[13] R. Agarwal, T. Elsaleh, and E. Tragos, "Gdpr-inspired iot ontology enabling semantic interoperability, federation of deployments and privacy preserving applications," in 2022 IEEE 8th World Forum on Internet of Things (WF-IoT). IEEE, 2022, pp. 1–8

[14] Oranekwu I, Elluri L, Batra G. Automated Knowledge Framework for IoT Cybersecurity Compliance. IEEE International Conference on Big Data (BigData) 2024.