Article

Semantically Oriented Method of Knowledge Presentation in Artificial Intelligence Systems for Cyber Security

Yuliia Kostiuk , Ellana Molchanova * , Pavlo Skladannyi , Volodymyr Sokolov , Karyna Khorolska

Posted Date: 17 June 2026

doi: 10.20944/preprints202606.1198.v1

Keywords: semantic knowledge representation; frame model; rule-based logical inference; short-term memory (STM); explainable artificial intelligence (XAI); comparison with MITRE ATT&CK; cyber incident analysis; SOC analytics

 doi:10.20944/preprints202606.1198.v1

Article

Semantically Oriented Method of Knowledge Presentation in Artificial Intelligence Systems for Cyber Security

Yuliia Kostiuk [1], Ellana Molchanova [2,*], Pavlo Skladannyi [1], Volodymyr Sokolov [1] and Karyna Khorolska [1]

[1] Borys Grinchenko Kyiv Metropolitan University, 18/2 Bulvarno-Kudriavska str., 04053 Kyiv, Ukraine

[2] OÜ "Scientific Center of Innovative Research", Viru tn 8-33, Püssi linn, Lüganuse vald, Ida-Viru maakond, 43221, Estonia

* Correspondence: ellana.molchanova@gmail.com

Abstract

The article proposes a semantically oriented method for representing and processing knowledge in artificial intelligence systems for cybersecurity tasks. The method is based on a frame model that provides a structured representation of network events, threats, and attack patterns, and supports context-dependent decision-making by forming explainable conclusions. A logical inference mechanism based on rules and semantic generalisation operations has been implemented, supplemented by a short-term memory (STM) module that accounts for the temporal characteristics of events within a sliding observation window. This allows explicit matching of semantic predicates (SSI) with MITRE ATT&CK techniques and accounts for the temporal characteristics of events with the formation of reproducible cause-and-effect attack chains. To improve interpretability, the results of logical inference are integrated with explainable artificial intelligence (XAI) methods, in particular SHAP and LIME, which ensure transparency of decisions at the event and attack-scenario levels. The scientific novelty lies in combining semantic knowledge structuring with the temporal context of STM and explainable rule-based reasoning in a single incident analysis pipeline. Experimental validation was performed on the CICIDS-2017 benchmark datasets with additional portability testing on KDD Cup 99 and NSL-KDD, confirming competitive accuracy and high explainability. The practical applicability of the method was demonstrated in anonymised production SOC environments with real telemetry flows (XDR/EDR, network sensors, IAM logs), resulting in reduced triage time and fewer false positives. The proposed approach complies with the requirements of international standards ISO/IEC 27001 and NIST SP 800-53, supports integration with the MITRE ATT&CK knowledge base and the Zero Trust architecture, and promotes the development of intelligent agents for reliable, explainable decision-making in cybersecurity.

Keywords: semantic knowledge representation; frame model; rule-based logical inference; short-term memory (STM); explainable artificial intelligence (XAI); comparison with MITRE ATT&CK; cyber incident analysis; SOC analytics

1. Introduction

The rapid growth in the complexity and dynamism of cyber threats means that traditional formal methods for analysing and processing security events are increasingly proving insufficient in highly loaded, heterogeneous network environments. Modern cyber defence systems must not only classify incidents, but also provide semantic interpretation, transparent explanations of decisions, and the ability to adapt in real time [1,2,8,9]. This requires a shift from processing raw traffic attributes to working with structured knowledge about object behaviour, attack tactics, and cause-and-effect

 doi:10.20944/preprints202606.1198.v1

relationships between events. Despite significant progress in the use of LSTMs, CNNs, Autoencoders, and other machine learning models, most remain 'black boxes' and do not provide the transparency necessary for incident auditing and integration with formal security standards.

Особливою проблемою є недостатність структурованих підходів до формалізації процесу transforming raw event information into semantic representations that can reflect behavioural patterns, inter-event relationships, and correspondence with MITRE ATT&CK tactics and techniques [10,15]. Traditional IDS/IPS solutions operate primarily on low-level attributes, leading to a loss of context, increased false positives, and the inability to construct cause-and-effect attack chains [4–6,14,18]. Existing semantic and ontological approaches do not simultaneously provide formal knowledge consistency, adaptive updating without complete model retraining, and compatibility with SOC/SIEM/XDR platforms and Zero Trust architectures.

Despite some progress in explainable artificial intelligence and knowledge graphs, existing solutions remain fragmented [8,9,33]. XAI (explainable artificial intelligence) approaches primarily provide post-factum analysis of ML model outputs and do not support real-time attack reconstruction [10], whereas ontological models lack integration of semantic compression, logical inference, and reasoning into a single system [1,4,35]. As a result, there are no unified methods capable of combining the transformation (of primary semantic information into secondary semantic information) PSI→SSI, frame-ontological knowledge structures, and interpreted logical inference mechanisms in a single analytical cycle [2,7,20]. This creates a significant gap between high-precision ML models and the practical requirements of operational cyber defence.

Thus, a number of key scientific questions remain open. First, how to formally transform raw event information into a structured semantic knowledge model that is consistent with MITRE ATT&CK and suitable for real-time reasoning [15,27,34]. Second, can the combination of semantic operators, frame structures, and production rules provide higher interpretability and accuracy compared to classical ML models [11,14,21]. Third, what scalability, stability, and integration properties will such a system demonstrate when operating in SOC/SIEM/XDR/Zero Trust infrastructures?

Within the scope of this study, the goal is to develop a semantically oriented knowledge representation method for interpretable artificial intelligence systems in cybersecurity. The proposed approach combines semantic transformation of event data, graph and ontological knowledge structures, a frame model, and an adaptive reasoning module that supports XAI explanations [1,2,8,31]. To achieve this goal, a formal method for semantic transformation PSI→SSI was developed, an integrated information-semantic system was built, an apparatus for interpretation, clarification, coordination, and contextualisation was developed, and experimental validation was performed on the CICIDS-2017 dataset.

The method was tested in two anonymised SOC (Security Operations Centre) production environments with real telemetry flows (XDR/EDR, network sensors, IAM logs), where the impact on triage time, FPR and explanation quality was assessed.

The results demonstrate improved classification accuracy, reduced false positives, and natural integration with the proposed approach and Zero Trust architecture, as well as SOC/SIEM/SOAR/XDR platforms. This confirms the scientific and practical value of the developed model as a foundation for the development of new-generation interpreted intelligent cyber defence systems.

In order to formalise scientific novelty and ensure clear research logic, the following research questions are formulated in the paper:

RQ1: Does the semantic transformation of event information PSI → SSI provide a higher level of interpretability of decisions compared to ML-only IDS without losing attack detection accuracy?

RQ2: How does the combination of frame structures, rule-based reasoning, and SSI semantic predicates affect the system's ability to form cause-and-effect attack chains consistent with MITRE ATT&CK?

 doi:10.20944/preprints202606.1198.v1

RQ3: Is the proposed semantically oriented information system (ISS) scalable and suitable for use in high-load SOC/SIEM/XDR environments?

RQ4: Does the integration of the reasoning process and XAI mechanisms reduce incident triage time and lower the false positive rate in real production environments?

The main threat to internal validity is the influence of experimental settings on the results obtained, in particular the choice of semantic rule thresholds, the size of STM time windows (Δt), and the configuration of baseline ML-only IDS for comparison. To mitigate this threat, all experiments were conducted with fixed and reproducible parameters, and the comparative models were trained on identical train/validation/test splits. Additionally, the results were verified across several MITRE ATT&CK scenarios, reducing the risk of overfitting to a single attack type.

The threat to construct validity concerns the appropriateness of the metrics used to evaluate the effectiveness of the proposed approach. The paper uses standard SOC/SIEM metrics (Precision, Recall, F1-score, FPR, triage time), which directly reflect both the quality of attack detection and operational usefulness for SOC analysts. To reduce the risk of incorrect interpretation of the results, the performance metrics are supplemented with qualitative explainability analysis (XAI) and examples of cause-and-effect attack chains.

External validity may be limited by the specificity of the datasets used and the conditions of the experimental environment. To mitigate this risk, validation was performed not only on CICIDS-2017, but also with verification of the transferability of results to KDD Cup 99 and NSL-KDD, as well as in two anonymised production SOC environments. Although the results indicate the approach's stability, further research may focus on expanding the range of environments and industry scenarios.

Given the limitations outlined above, the results obtained are considered sufficiently reliable to confirm the research questions (RQ1–RQ4) and demonstrate the practical applicability of the semantically oriented approach in real SOC/SIEM environments.

2. Literature Review and Theoretical Foundations of Semantic Knowledge Representation

Modern approaches to detecting and interpreting cyber threats cover three key areas: machine learning methods, deep neural networks, and semantically oriented knowledge representation systems. In the works of 2021–2024, models based on CNN, LSTM, AutoEncoder, Transformer architectures, as well as hybrid solutions combining statistical and neural network approaches are actively being researched [20,22,23,29,45,47]. Despite their high accuracy, most of these models remain 'black boxes' and lack transparency in their decision-making processes, significantly limiting their practical applicability in SOC/SIEM environments, where it is critical to explain the detection logic and reproduce the cause-and-effect relationships between events.

Semantic methods, in particular ontology-driven IDS and knowledge graph reasoning, provide high interpretability and formalised logic, but are mostly characterised by limited scalability, static rules, and low adaptability [16,35,36]. XAI approaches generate post-hoc explanations, but typically do not integrate with ontologies and do not support real-time reasoning [45,52,54]. This results in fragmentation between ML models, semantic knowledge structures, and explainability mechanisms.

The generalisation of these problems — opaque logic, lack of semantic context of events, and poor compatibility with formal cybersecurity frameworks — highlights the need for integrated approaches that can combine ML detection, rule-based reasoning, structured knowledge models, and XAI explanations.

Research by Preuveneers and Joosen [1] proposes an ontological model of cybersecurity for AI systems that formalises knowledge about threats at all stages of the life cycle. Bratsas et al. [2] demonstrate the effectiveness of knowledge graphs and the Semantic Web in modelling APT chains. Garrido et al. [3] show that integrating ML with knowledge graphs enhances contextual monitoring and deepens analysis. Ayo et al. [4] propose an ontological NIDS architecture that supports inference of response rules. Goyal and Sharma [5] emphasise the importance of semantic grouping of IoCs for predicting the behaviour of the attack environment.

 doi:10.20944/preprints202606.1198.v1

An analysis of existing approaches shows that each of them covers only a fragment of the problem. ML models are accurate but opaque. Ontological models are explainable but not very dynamic. Knowledge graphs provide deep context, but integrating them with ML is complex and resource-intensive. XAI approaches increase interpretability, but they act post facto and do not provide complete reasoning. Thus, there is a lack of comprehensive solutions that combine ML accuracy, semantic structure, and real-time, explainable logical inference.

A separate area focuses on the construction of semantic models and knowledge graphs for cyber threat analysis and decision support. Works based on CTI knowledge graphs [27,36,39] propose generalised schemes for constructing threat graphs that integrate vulnerability taxonomies, TTP profiles according to MITRE ATT&CK, and the context of enterprise assets, enabling semantic search and causal analysis of attack chains [1,2,6]. A number of studies emphasise that combining knowledge graphs with logical inference mechanisms enables automated detection of complex attack scenarios that are difficult to model with purely statistical methods.

At the same time, GNN-oriented attack detection methods are actively developing, interpreting network traffic as a graph of interactions between hosts or sessions. Recent studies show that graph neural networks exhibit higher sensitivity to complex multi-step attacks in IoT and industrial networks than classical MLP/CNN-IDS, as they account for topological context, temporal event correlation, and inter-node dependencies [7,54]. Systematic reviews emphasise that GNN approaches are a natural basis for integrating semantic information, but most implementations currently use them primarily as a 'black box' without in-depth logical interpretation of decisions. As shown in a systematic review [42], GNN, Transformer, and RL approaches are gradually forming a new class of intelligent network threat monitoring tools.

Given this trend, it is advisable to use not generalised 'GNN' and 'Transformer' for comparative evaluation, but specific architectures that are de facto baselines: GraphSAGE (inductive learning on interaction graphs), GAT (attention to neighbours to strengthen relevant connections), and Transformer Encoder for modelling stream/session sequences (self-attention as a mechanism of global feature dependency). This ensures a fair comparison with the proposed approach, since GNN/Transformer baselines represent the most common modern neural IDS architectures but typically do not provide formal reasoning or causal explanations at the MITRE ATT&CK level.

The lack of solutions that simultaneously provide semantic knowledge representation, ML detection, rule-based reasoning, explainability, and consistency with MITRE ATT&CK [15] creates a research gap. A comparative analysis (Table 1) confirms that none of the modern approaches combine adaptability, interpretability, and formal consistency with attack models.

Table 1. Comparative analysis of modern approaches to threat detection.

Approach / Method	Explainability	Adaptability	Reasoning ability	Consistency with MITRE ATT&CK	Scalability
ML- classifiers (RF, SVM)	Low	Average	None	Low	High
Deep models (LSTM, CNN)	Very low	High	None	Low	High
Autoencoder / AE-IDS	Very low	High	None	Low	High
Ontologies and logical models	High	Low–average	Yes	Average –high	Low–average
Knowledge graphs (KG, KGE)	High	Average	Yes	High	Average
XAI-models (SHAP, LIME)	Average –high	Low	Post factum	Low	Average
Neuro-symbolic approaches	High	Average	Yes	Average	Average
The proposed semantic method (PSI→SSI)	High	High	Yes	High	High

Qualitative assessments (high, medium, low) are based on the properties of the approaches as described in the relevant publications and on a generalised analysis of their functional capabilities as described in the literature.

In recent years, a class of neuro-symbolic approaches has emerged that combines statistical learning with logical rules and semantic constraints. In neuro-symbolic IDS, the results of deep neural

 doi:10.20944/preprints202606.1198.v1

networks are supplemented with knowledge from ontologies or rule-based systems, enabling the formalisation of domain constraints, system state consistency, and threat escalation scenarios. Some works demonstrate the use of first-order logic on top of knowledge graph embeddings to refine event classification, detect atypical TTP step sequences, and generate explanations for SOC analysts [8,33,41]. Despite this, most neuro-symbolic solutions focus on individual subtasks (e.g., CTI analysis or critical infrastructure attack prediction) and do not form a comprehensive semantically oriented knowledge representation model for integrated ISS.

Considerable attention is also paid to the explainability of IDS decisions. Recent reviews of XAI approaches for intrusion detection systems show that methods such as SHAP, LIME, and attention mechanisms are used to interpret the impact of individual features on classification decisions, identify unstable behaviour patterns, and build understandable rules for specialists [9,31,47,53]. Some works propose XAI-oriented IDS frameworks for IoT environments, where explanations are used to interactively configure security policies and reduce false positives through expert feedback [10,25,45,52]. However, most of these approaches operate at the post-hoc level and do not integrate explainability into the knowledge representation model itself.

In the context of Zero Trust architectures, semantic models are emerging that use knowledge graphs and rules for identity-centric threat segmentation and dynamic access policies [11,14,40,41]. Such approaches demonstrate the potential of combining semantic knowledge representation with security policies, but they tend to focus on access control and do not cover the full cycle of semantic interpretation of network events, attack detection, and explanation generation. Therefore, despite significant progress in GNN-, KG-, and neuro-symbolic methods, there is still a lack of a comprehensive semantically oriented information-semantic system (ISS) that combines a frame-based knowledge representation model, logical inference, XAI tools, and the ability to integrate with existing SOC infrastructure. Such solutions typically focus on access control and CTI analysis rather than the full cycle of semantic interpretation of network events, attack detection, and explanation generation [15,44].

Despite the advantages of semantic structures, their practical implementation is accompanied by a number of limitations. The completeness of reasoning depends on the relevance of the knowledge base; constructing ontologies is resource-intensive; scalability decreases with increasing event volumes; and integration with streaming SOC/SIEM processes requires synchronising ML detectors with the semantic model. This creates a scientific gap and justifies the need to develop an integrated information-semantic system capable of providing a full cycle: from primary event information (PSI) to structured knowledge (SSI) and the subsequent formation of explainable decisions in real time.

3. Research Methodology and Approaches

The methodological basis of the study combines semantic knowledge modelling, neuro-symbolic approaches, and modern artificial intelligence technologies [8,18,40]. The basic structure for knowledge representation is ontologies and knowledge graphs (RDF/OWL), which provide a formalised representation of objects, events, and behavioural patterns in the context of cyber threats [27,38,39]. To implement semantic inference, first-order logic, reasoning process mechanisms (deductive inference) and inductive generalisation are used, integrated with the PSI→SSI semantic compression frame model. All SSI predicates in the work are presented in a unified PascalCase format without underscores. Suffixes are used to denote the role of the predicate: Observed — for recorded events, Suspect — for hypotheses and suspicions, Context — for contextual and behavioural patterns.

Traditional ML approaches (Random Forest, LSTM, CNN), despite their high classification accuracy, do not provide transparency and explainability of the decisions made, which limits their application in SOC/SIEM environments [20,22,23,29]. The proposed approach overcomes these limitations by interpreting events semantically using frames and slots consistent with MITRE ATT&CK behavioural models and Zero Trust principles [14,15,33,34].

 doi:10.20944/preprints202606.1198.v1

To ensure correct comparison with neural network baselines, two model classes have been added. GNN-baseline builds an interaction graph, $G = (V, E)$, where V — hosts/accounts/sessions (depending on available fields), E — network interactions in the time window Δt; node/edge features are formed from aggregated PSI groups (port/protocol, intensity, entropy, frequency indicators). Transformer-baseline — a sequence of flows/sessions in the window is represented as tokens with numerical PSI features and positional encoding; the task is binary classification Normal vs Attack. All models are trained on the same train/validation/test splits and undergo the same class balancing, eliminating comparison bias.

The system's adaptability to new threats is achieved through neuro-symbolic integration, which combines deep models with mechanisms of semantic generalisation and ontological alignment [25,33,42]. Additionally, a semantic dialogue mechanism has been developed, based on Natural Language Understanding (NLU) methods [28,44], which formalises the interaction between the user and the system to clarify intentions, context, and security rules.

The neural components (LSTM block and adaptive semantic compression module) were trained in Ubuntu 22.04 using Python 3.10, TensorFlow 2.12, Keras, Pandas, and NumPy. The calculations were performed on an NVIDIA RTX-3090 GPU (24 GB VRAM), enabling us to process the entire CICIDS-2017 dataset without performance degradation. SHAP and LIME [20–23,31,45] were used to explain the results. This configuration ensured the reproducibility of experiments and the stability of training.

The reasoning procedures, the formation of interpretation policies, and the assessment of the compliance of events with security requirements are consistent with the NIST SP 800-53 (AU, IR, SI control families) and ISO/IEC 27005 standards [12–14,32]. The model is also fully compatible with the Zero Trust architectural principles, as defined in NIST SP 800-207.

The software implementation of the system includes data processing modules (Scikit-learn, Pandas, NumPy), neural network components (Keras, TensorFlow), an XAI subsystem (SHAP, LIME), and a reasoning module implementing semantic slot-matching logic. Frame structures are implemented in a JSON-like format [45,51]. Experimental testing was conducted in a Jupyter Notebook, and quality assessment was performed using Precision, Recall, and F1-score metrics, confirming the model's effectiveness on CICIDS-2017.

Taken together, the methods used create an integrated methodology for building a semantically oriented system that not only structures knowledge in a formal way, but also uses it for reasoning, explaining decisions, and adaptively updating security policies [43,47,50]. The following section presents the architecture of an information-semantic system (ISS) for cyber defence that implements these methodological principles.

4. Architecture of a Semantically Oriented System and Formal Knowledge Models

4.1. Information-Semantic System (ISS)

A semantically oriented approach to cybersecurity requires a basic system architecture capable of integrating knowledge representation, rule-based reasoning, user interaction, and security monitoring infrastructure [40,47]. Such a structure is the Information Semantic System (ISS), which serves as an architectural 'framework' for building frames, ontologies, threat graphs, and XAI modules, providing a complete event processing cycle: from PSI collection, its semantic transformation into SSI, and logical inference to explainable decision-making, consistent with MITRE ATT&CK models and Zero Trust principles.

Within the ISS (Information Semantic System) architecture, there is an internal information-semantic mechanism (ISS) that implements PSI→SSI semantic transformation, logical reasoning, and the formation of explainable conclusions. ISSem is not a separate system, but a functional semantic core of ISS that implements PSI→SSI semantic transformation operations, logical reasoning, and the formation of explainable conclusions. Formally, ISSem is defined as a functional subsystem of ISS

 doi:10.20944/preprints202606.1198.v1

that implements a set of semantic operators, frame structures, production rules, and explainability mechanisms necessary for PSI→SSI transformation and logical inference.

Formally, ISS is defined as a system whose functioning is aimed at achieving a specific goal through semantic information processing, self-updating of knowledge, and interaction with a changing environment based on feedback [6,28]:

$$ISS = \langle a, st, met, re, Sem, SI, SI_\gamma, \cup, ij, co, At \rangle, \quad (1)$$

where a – system goal, st – structure, met – methods, re – implementation tools, Sem – formal semantic sign system, SI – basic semantic dialogue information, SI_γ – feedback semantic information, $\cup$ – knowledge set combination, ij – feedback channel parameters, co – conditions, At – time context of operation.

Practical scenarios demonstrate that the proposed model is most effective in environments where traditional IDSs cannot cope with the contextual interpretation of events and generate excessive numbers of false positives [25,45,47]. In SOC centres, the semantic approach enables distinguishing administrative actions from Command & Control commands by comparing events against MITRE ATT&CK techniques and building the full operational context [15,33,34,41]. In cloud infrastructures (particularly Kubernetes), the model explains anomalies by analysing identities, behavioural deviations, and inter-service interactions, allowing for the quick establishment of a cause-and-effect chain.

In EDR/XDR platforms for Zero Trust, it performs behavioural analysis of the user and generates transparent explanations for blocking access, minimising unjustified denials [14,51,54]. During incident response, the system is able to provide a reasoned explanation for the classification (e.g., DoS attacks) based on semantic relationships, temporal characteristics, and attack patterns [21,41]. Thanks to this, the model not only identifies threats but also provides clear decision-making logic, thereby increasing its practical value in modern cyber defence systems.

Within the ISS, semantic information is considered as knowledge represented in a formalised ontological structure that reflects the relationships between objects, events, processes, and threats [18,19,38,40]. To interact in such an environment, the concept of a semantic object is introduced — an agent that receives or transmits semantic information within a subject area and can act as a source, receiver or intermediary.

However, the mere presence of semantically structured knowledge does not guarantee the correct interpretation of events by all components of the system. In a multi-component cyber security environment, different agents — AI modules, analytics and external services — operate with their own ontologies and thesauri, which inevitably creates risks of misunderstandings and loss of meaning [28,33,49]. Therefore, a key aspect of ISS is the organisation of semantic dialogue between system objects, where the system's inherent limitations on the interpretation interpreting semantic structures formally set the boundaries of adequate information.

4.2. Semantic Dialogue in a Semantically Oriented ISS

The key form of interaction between semantic objects is semantic dialogue – meaningful communication involving the exchange of semantic messages, the explanation of results, and the coordination of actions. Dialogue is considered bilateral if the exchange of messages is mutual, and unilateral if the transfer of information is only in one direction. Regardless of the type, informing is a mandatory process that ensures knowledge models are updated and decisions are adapted to the context.

In the proposed ISS model, semantic dialogue is considered as a controlled interaction between agents operating with their own, partially incomplete knowledge profiles [28]. Each agent works with a limited ontology and thesaurus of events, so the same network situation can be interpreted differently depending on the local context [2,40]. This creates the risk of ambiguous interpretation of threats and decisions, which is proposed to be reduced by formalising knowledge in a frame model and using explainable logical inference mechanisms [10,31,41].

 doi:10.20944/preprints202606.1198.v1

$$\exists SI(\forall SO_{di}[SI \in DA_j] \rightarrow \neg adInf(SO_a)), \quad (2)$$

where SI – semantic information, SO_{di} – various semantic objects (AI modules, analytics, external system), DA_j – a set of subject areas, $adInf$ – adequate information.

Figure 1 shows a model of a semantically oriented information-semantic system (ISS) that coordinates event collection, semantic interpretation, explanation formation, knowledge adaptation, and recommendation generation into a single data processing cycle to support decisions in cyber defence systems.

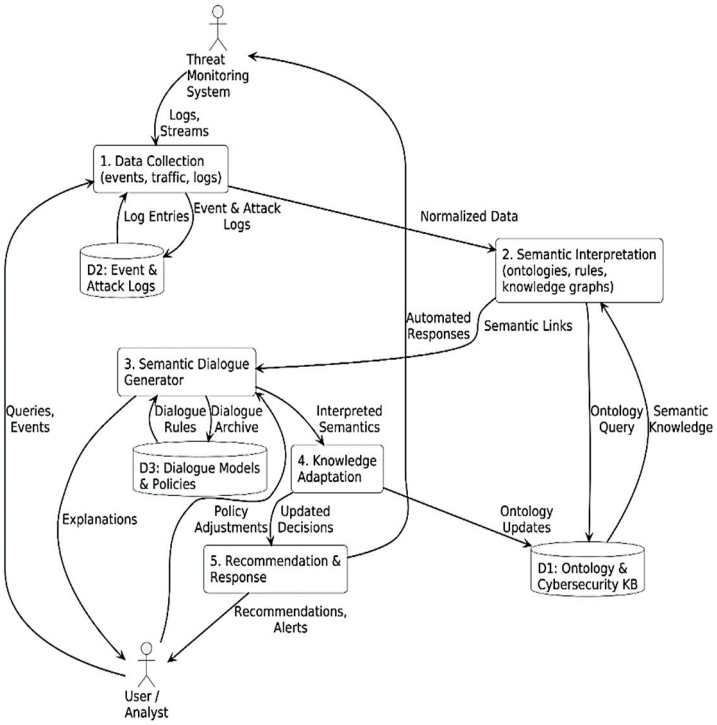

Figure 1. Model of a semantically oriented information-semantic system (ISS) in artificial intelligence systems for cyber defence.

The first condition — the principle of communication — states that information transfer is possible only if there is overlap between the knowledge thesauri of the source and the receiver [1,2,33]. If there is no common semantic space, dialogue loses its meaning and becomes ineffective.

$$\forall SO_i, SO_j \in ISS, \ (T_i \cap T_j \neq \emptyset) \rightarrow \exists WInf, \quad (3)$$

where T_i, T_j is relevant ontologies or thesauri of semantic objects, $WInf$ is the possibility of informing. Three scenarios are considered: full compatibility of knowledge (full informing), partial compatibility (limited informing), absence of a common semantic space (informing is impossible).

The second condition is the principle of unity of the sign system, according to which semantic interaction is possible only when using agreed formats, languages or data structures. This ensures component interoperability, which is critical in hybrid environments where classic security systems and modern AI platforms are combined [44,49].

$$\forall SO_i, SO_j \in ISS, \ (SI_i \cap SI_j \neq \emptyset) \rightarrow \exists NInf, \quad (4)$$

where SI_i, SI_j is multiple sign systems in use, $NInf$ is the need for format uniformity.

The third condition - the principle of information - determines the mandatory nature of the entire sequence of semantic operations: generation, transmission, reception, storage, interpretation, understanding, and decision-making. Skipping any stage violates the integrity of information and makes interaction incomplete or unreliable.

$$\forall SO_{src}, SO_{dst} \in ISS, \ (Op_1 \& Op_2 \& \ldots \& Op_n) \rightarrow DInf, \quad (5)$$

 doi:10.20944/preprints202606.1198.v1

where Op_k is semantic operations that form a conjunctive chain, $DInf$ is obligation to inform.

The fourth condition - the principle of intelligence - implies that a system is intelligent only when it not only achieves its goals but is also capable of self-renewal based on feedback [21,24]. This ensures its adaptability and the ability to improve knowledge models in response to new threats and environmental change.

$$\forall SYST \in ISS,\ (Goal \wedge Sel\ f\ Learning) \rightarrow RInf, \quad (6)$$

where $RInf$ is a sign of intelligent information.

Based on the four conditions U_1–U_4, a heuristic algorithm for semantic dialogue has been developed, in which they are treated as logical variables [19,33]. If all conditions are met, the dialogue is considered acceptable; if at least one is violated, semantic interaction is blocked. This is formalised by the following expression:

$$SemInt = \begin{cases} TRUE, & if\ U_1 \wedge U_2 \wedge U_3 \wedge U_4 = 1 \\ FALSE, & if\ at\ least\ one\ condition\ U_i = 1' \end{cases} \quad (7)$$

This mechanism allows formal verification of the possibility of informing before starting data exchange between AI system components.

Figure 2 shows the generalised architecture of a secure semantic dialogue system, which includes four logical levels: interface, semantic dialogue subsystem, knowledge management level, and semantic security analysis module. The interface level receives events from the threat environment and user requests, the dialogue subsystem normalises them in a single semantic space, the knowledge management level operates with ontologies and rules, and the semantic analysis module interprets anomalies and returns explainable decisions.

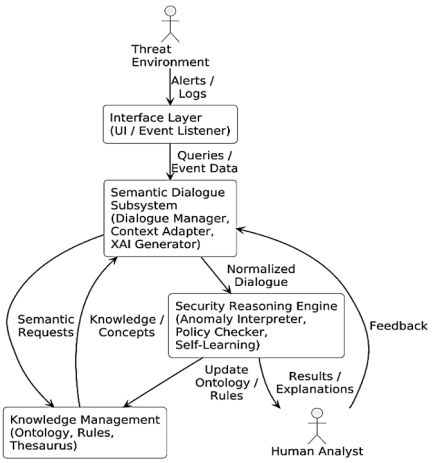

Figure 2. Generalized architecture of a secure semantic dialogue system in intelligent cyber defense systems.

The proposed formalization creates a basis for explainable, contextually adaptive, and ontologically consistent information, which is critical for next-generation cyber defense systems [10,31,43]. It integrates into SIEM, SOAR, and XDR architectures, increasing the accuracy, transparency, and reliability of decision-making in complex information environments. Ontologically Consistent Information Model:

$$Inf_{valid}(SO_i, SO_j) \Leftrightarrow (T_i \cap T_j \neq \emptyset) \wedge (Z_i = Z_j), \quad (8)$$

where SO_i, SO_j are semantic objects (source and receiver), T_i, T_j are corresponding ontologies (knowledge thesauri), Z_i, Z_j are sign systems (for example, RDF, JSON-LD), Inf_{valid} is a condition of admissible information. Semantic information is possible only with a partial intersection of ontologies and the unity of the format of knowledge representation.

Summarizing the above, the semantic dialogue model and its inherent limited and context-dependent interpretation of knowledge set requirements for how the system should explain its decisions, adapt knowledge, and provide trust on the part of the analyst [51–54]. This directly leads to the need to formalize the mechanisms of explainability (XAI), reasoning, and interpretation that

 doi:10.20944/preprints202606.1198.v1

determine how the semantic structures of the ISS are transformed into substantiated diagnostic conclusions.

4.3. *Formal Models of XAI, Reasoning Process and Interpretation*

Decision Explainability Model (XAI Model) [41,47]:

$$Expl(D) = \bigcup_{k=1}^{n} \{r_k | r_k \Longrightarrow D, Conf(r_k) \geq \theta\}, \quad (9)$$

where D – diagnostic decision (detected threat), r_k – rule or fact that justifies it, $Conf(r_k)$ – confidence in the rule, θ – minimum admissible confidence threshold, $Expl(D)$ – explanation of the decision. Formula (9) defines the explanation $Expl(D)$ as the set of rules r_k, that justify the diagnostic decision D with confidence not lower than the threshold θ.

Model of contextual adaptation of knowledge representation [1,2,27,36]:

$$SI_f^{(c)} = \phi_c(SI_p) = \{x \in SI_p | Relevance(s, c) \geq \delta\}, \quad (10)$$

where SI_p – primary semantic information, $SI_f^{(c)}$ – secondary (context-adapted) information, ϕ_c – context adaptation operator c, $Relevance(s, c)$ – function of relevance of knowledge to context, δ – threshold value of acceptability. Selection of knowledge relevant to the current cyber threat, taking into account the situation or environment.

Information accuracy metric in cyber defence [8,20–23]:

$$Acc_{inf} = \frac{|\{SI_{true}\} \cap \{SI_{predicted}\}|}{|\{SI_{predicted}\}|}, \quad (11)$$

where $\{SI_{true}\}$ – actual (reference) information about the threat, $SI_{predicted}$ – information provided or generated by the system, Acc_{inf} – accuracy of information within the AI system. Assessment of the extent to which the system provided correct, substantiated information about the incident.

Formally, implementation into an application system looks like the transformation of a semantic module into the operational layer of a cyber defence platform [13–15,25,29]:

$$S_{cyber} = \langle Core_{SIEM}, Sem_{engine}, XAI_{expl}, K_{domain}, \Pi_{act} \rangle, \quad (12)$$

where $Core_{SIEM}$ – event and response logic engine, Sem_{engine} – semantic threat analysis module, XAI_{expl} – explicable decision interpreter, K_{domain} – domain knowledge base,, Π_{act} – automated action policy. In the application system, the semantic module is integrated into the cyber security platform as a component of S_{cyber} (12), where the SIEM (Security Information and Event Management) core is combined with a semantic threat analyzer, XAI interpreter, domain knowledge base, and automated action policies [3,16,42]. This provides not only threat detection, but also their semantic interpretation and explicable response.

The diagram in Figure 3 shows the sequence of processing an analyst's request for an incident explanation: the SIEM system transfers the event to the semantic module, the ontological adapter forms the context, the XAI core selects relevant rules and builds an explanation, which is returned to the analyst along with a link to the knowledge base.

 doi:10.20944/preprints202606.1198.v1

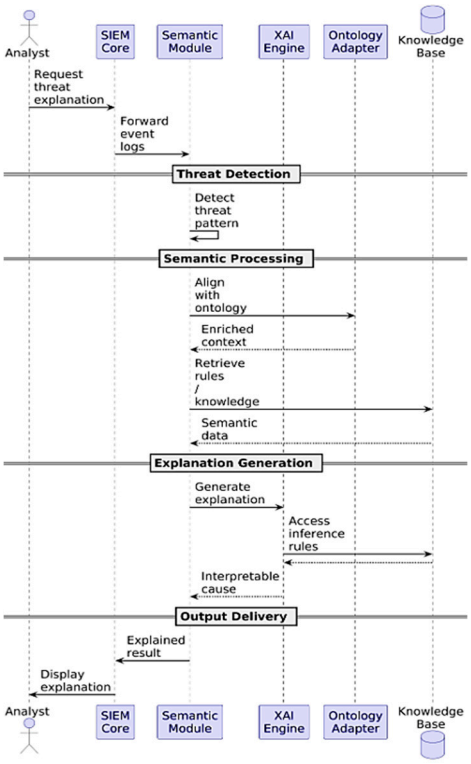

Figure 3. Sequence diagram of explanatory threat analysis involving a semantic module.

Thus, the above sequence of interactions forms the basis of the semantic interpretation of incidents, ensuring a consistent transformation of events into structured knowledge and creating the prerequisites for further frame, ontological and graph modelling necessary for in-depth analysis of cyber threats.

4.4. Frame, Ontology and Graph Models

The study considers three basic models of knowledge representation in intelligent cyber defense systems: frame, ontological, and graph [1,2,27,33]. Frames provide a structured representation of typical situations, ontologies formalize concepts and relationships, and knowledge graphs describe complex relationships between objects, events, and incidents.

Figure 4 presents a vertically structured comparison scheme of the three main knowledge representation models - frame, ontological, and graph - that are integrated into a semantically oriented cyber defense system [8,24,25]. A brief description is provided for each model, reflecting its structural features, advantages, and scope of application. The scheme demonstrates how these models are consistently used for interpretive data processing, knowledge formalization, and visualization of complex relationships in artificial intelligence systems for cybersecurity.

 doi:10.20944/preprints202606.1198.v1

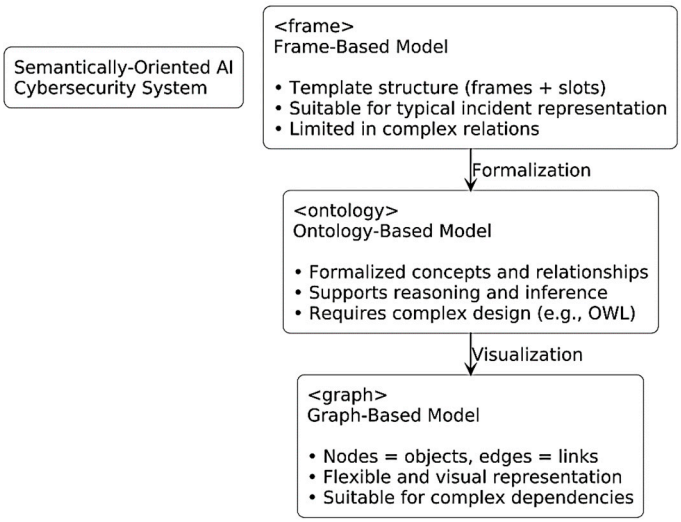

Figure 4. Comparison of knowledge representation models in cyber defense systems based on artificial intelligence.

In contrast to the generalized comparative representation, Figure 5 shows an example of a specific implementation of a knowledge graph model for a Brute Force attack according to the MITRE ATT&CK T1110 (Credential Access) technique. Primary security events (PSI) recorded by SIEM correspond to multiple failed authentication attempts via the SSH service from the same IP address in a short time window and in the process of semantic transformation PSI → SSI are generalized into secondary predicates {Service(SSH), AuthFailureObserved, BruteForceSuspect, AnomalyRate(High)}.

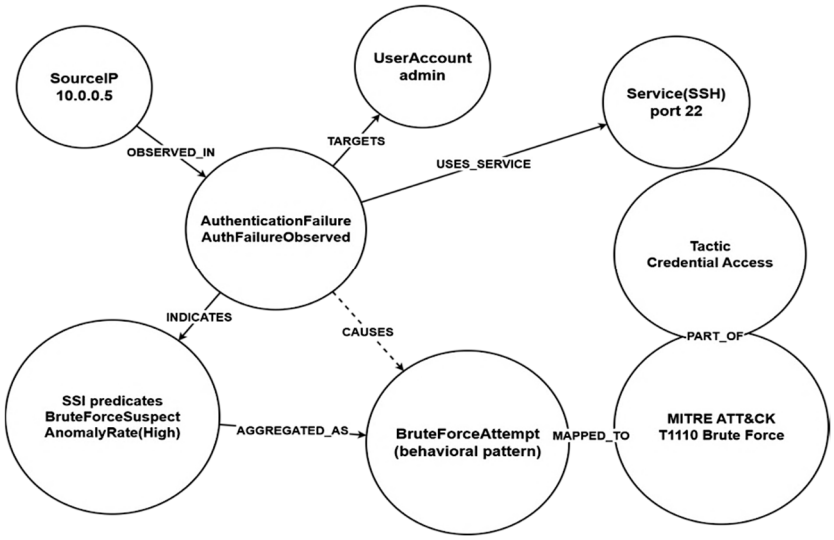

Figure 5. Knowledge graph of semantic interpretation of Brute Force attack according to MITRE ATT&CK T1110.

Based on the generated SSI predicates, a knowledge graph is built, in which nodes represent objects, events and concepts of the subject domain, and edges represent semantic relations between them (observedIn, targets, usesService, indicates, aggregatedAs, causes, mappedTo, partOf). The AuthenticationFailure event is associated with the SourceIP and UserAccount objects, aggregated into the BruteForceAttempt behavioral pattern and mapped to the T1110 technique and the Credential Access tactic in the MITRE ATT&CK knowledge base. Thus, the knowledge graph reflects the cause-and-effect chain of the attack — from the initial events to the interpreted semantic

 doi:10.20944/preprints202606.1198.v1

conclusion, providing transparency of the reasoning process and the basis for forming XAI explanations in SOC.

The study considers a set of theoretical, methodological and applied issues related to the construction of formalised structures for the representation, updating and interpretation of knowledge in artificial intelligence systems. In general, a model is defined as a system of objects or signs that reproduces the essential properties of the original - the system or object being modelled. The knowledge model M is considered as a triple:

$$M = \langle O, P, R \rangle, \quad (13)$$

where O – a set of objects or concepts (Ontology space), P – a set of properties of objects (Property descriptors), $R \subseteq O \times O$ – a relationship between objects (Relation set, for example "depends on", "causes", "threatens"). The knowledge model M is presented as a triple of objects, their properties and relationships between them, which forms the basis for building frame, ontological and graph structures.

Primary semantic information (PSI):

$$PSI = \{p_i \in P | p_i \; is \; an \; essential \; characteristic\}, \quad (14)$$

Secondary semantic information (SSI):

$$SSI = f_{sem}(PSI) = compress(PSI) \rightarrow generalized \; model, \quad (15)$$

The formulas reflect the essence of the semantic representation of knowledge in artificial intelligence systems, where PSI contains basic knowledge, and SSI is the result of logical-semantic generalisation used for decision-making in cyber defence systems.

Semantic compression operation:

$$SSI = Sem(log.trans(PSI)), \quad (16)$$

where Sem – semantic function of generalisation or transformation, $log.trans$ – logical transformation (rule-based, ontologically consistent), the result is a compressed feature vector for decision-making. SSI is obtained by logical-semantic compression of the primary information PSI (15)–(16), which gives a compact but meaningful representation suitable for decision-making in cyber defence systems.

We determine the completeness, accuracy, and depth of the knowledge model:

$$Q_M = \langle C, A, D \rangle, \quad (17)$$

where $C = \sum_{i=1}^{n} c_i$ – completeness (highlighting of signs), $A = \frac{revelant}{total}$ – accuracy (correspondence to the threat context), $D = f(R)$ – depth (level of connections in the knowledge graph).

Integration into decision making (XAI strategy formula):

$$D = \delta(SSI, T) = \arg\max_{a \in A}[U(a) | SSI \vDash T], \quad (18)$$

where D – decision (action), T – safety target or criterion, $U(a)$ – the usefulness of the action in the context of the threat, $SSI \vDash T$ – knowledge satisfies the requirements of the target condition

Iterative knowledge updating (dynamic adaptation model):

$$M_{t+1} = M_t + \Delta M_t, \; \Delta M_t = f_{learn}(new\_SI), \quad (19)$$

where new_SI – new semantic information from incidents, f_{learn} – learning function (ML/Rule-based update). Formula (19) describes the dynamic updating of the knowledge model based on new semantic facts coming from real incidents.

Figure 6 shows a diagram of the rule execution activity in a semantically oriented cyber defence system: from receiving an event and its semantic interpretation to activating the corresponding rules, forming an explainable solution and updating the knowledge base.

 doi:10.20944/preprints202606.1198.v1

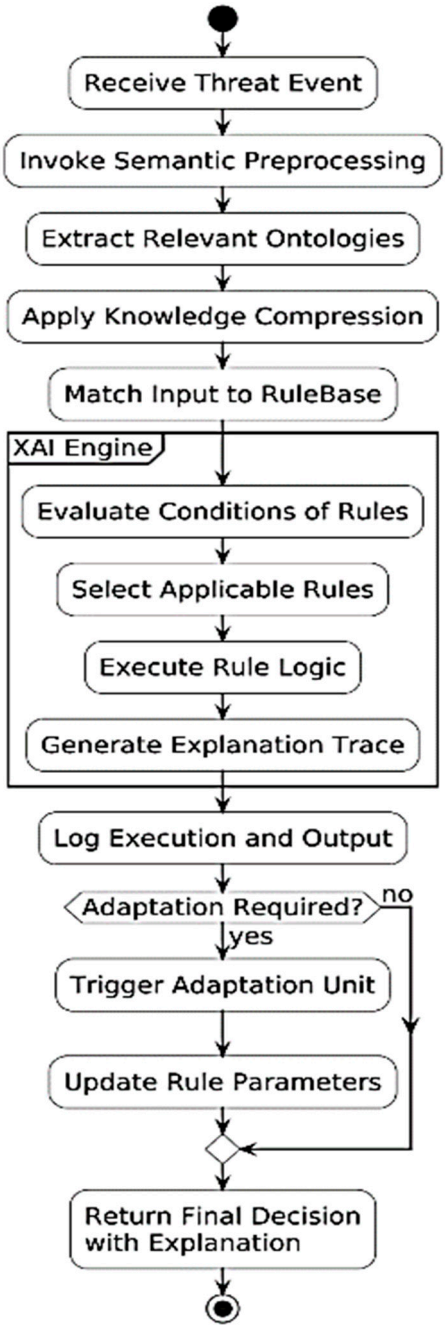

Figure 6. Rule Execution Timeline diagram in an XAI-based cyber defense system.

A basic proposition is proposed, according to which the semantic object model is based on the interaction of three key categories: object structure, object properties, and object relationships with other elements within the information semantic system (ISS) [35,36]. Such a three-component interaction allows for flexibility and consistency of knowledge models in the dynamic cybersecurity environment.

Basic Semantic Object Model:

$$M(O) = \langle S(O), P(O), R(O) \rangle, \quad (20)$$

This is a three-component model that describes how knowledge about any object (e.g., type of attack or security event) consists of: structure $(S(O))$ – what the object is made of (e.g., hierarchy of event attributes, processing logic), properties $(P(O))$ – what characteristics the object has (IP address, risk, time of occurrence, type of threat), relationships $(R(O))$ – how this object is related to others (e.g., related to a specific attack or user). This model allows an AI system to accurately understand the essence of an event or phenomenon.

 doi:10.20944/preprints202606.1198.v1

Dynamic (temporal) knowledge model:

$$M_t(O) = \langle S_t(O),\ P_t(O),\ R_t(O) \rangle, \quad (21)$$

The basic model of a semantic object $M(O)$ (20) odescribes its structure, properties and relations, and the dynamic model $M_t(O)$ specifies their temporal evolution [33,49]. Model consistency (22) ensures a holistic representation of knowledge within the ISS and is critical for correct reasoning and the explainability of XAI module decisions.

Model consistency in ISS:

$$\forall i, j\ M(O_i) \cap M(O_j) \neq \emptyset \Longrightarrow alignment\ via\ shared\ R, P, \quad (22)$$

If two models contain common elements (for example, a single attack or an IP address), the system must reconcile them by combining data and eliminating contradictions to form a holistic picture of the situation. This is the basis of explainable decisions (XAI), since conclusions are formed on the basis of agreed-upon knowledge [47,51–54]. Such reconciliation ensures correct processing of events in cyber defense systems, including SIEM, XDR and SOAR.

Figure 7 shows the structure of the reasoning module of a semantically oriented AI system: an event module, a framework and ontology subsystems, a mapping mechanism with MITRE ATT&CK, a reasoning process and an XAI component that generates explanations for the administrator.

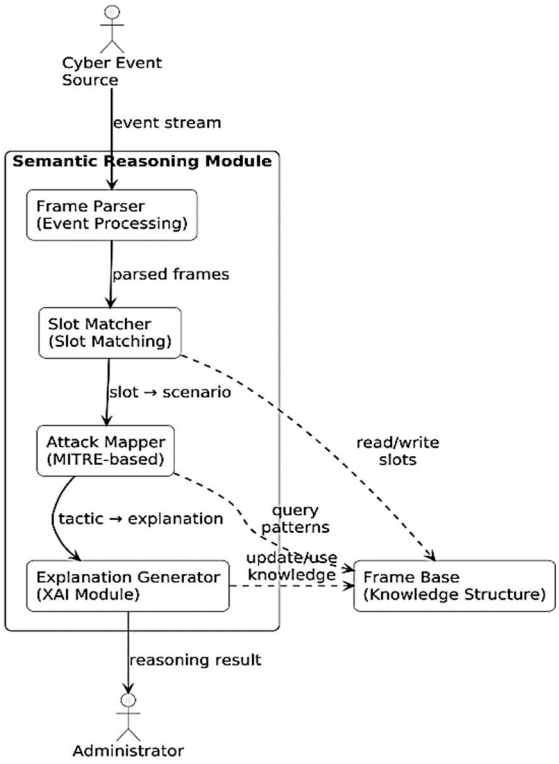

Figure 7. Reasoning module model of a semantically-oriented AI system for cyber defence.

Within the ISS, a threat model consistent with MITRE ATT&CK and NIST SP 800-53 is used, which covers three classes of risks: technical - unauthorized access, privilege escalation, network attacks (DoS, PortScan, Brute Force), semantic - false correlations of events, violation of the integrity of the knowledge base, injection of invalid ontological statements, operational - log manipulation, reasoning module delays, attacks on ML models [13–15,34]. Each threat is described by the triple <TACTIC, TECHNIQUE, EFFECT>, which allows performing cause-and-effect analysis in the process of PSI→SSI transformation and forming explainable solutions to incidents.

The set of formal ISS models, semantic dialogue, XAI mechanisms, frame, ontological and graph structures forms a holistic conceptual framework within which semantic information goes through a full cycle - from the initial perception of the event to the interpreted decision regarding the cyber threat [21,26]. Further, these models are transformed into software components and algorithms

 doi:10.20944/preprints202606.1198.v1

capable of working with real traffic flows and security logs, which justifies the transition to the section on the implementation of semantic transformation and experimental research.

5. Implementation of the Semantic Model and Experimental Studies

5.1. Semantic Transformation PSI→SSI as the Basic Mechanism of ISS

Semantic information about an object is classified into primary and secondary [1,5]. Primary semantic information is defined as a complete semantic representation of knowledge as the result of observations, research or generalisations, regardless of the form of presentation [18]. Secondary semantic information is formed by logical, syntactic or pragmatic transformation of the primary and acts as a compact or analytical reflection of the same entity. It is used for semantic compression and generalisation.

Formally, the process of logical transformation of primary semantic information SI_p into secondary SI_f is described by the expression [5,8,27]:

$$SI_f = Sem \circ log.trans(SI_p), \quad (23)$$

where Sem – semantic operation, $log.trans$ – logical transformation that provides a transition from full to compressed knowledge representation within the ISS [1,36]. In the future, SI_p denoted as PSI, and SI_f – as SSI.

Primary semantic information is presented as a combination of essential and non-essential features of the object. [5,29,45]:

$$SI_p = \{SI_{p1}\} \cup \{SI_{p2}\}, \quad (24)$$

where $\{SI_{p1}\}$ – елементи, що мають критичне значення для ідентифікації або інтерпретації об'єкта, $\{SI_{p2}\}$ – другорядні або надлишкові ознаки. Для оптимізації аналізу застосовується семантичне стиснення шляхом усунення несуттєвих ознак:

$$\{SI_{p2}\} = \emptyset \Rightarrow SI_f \approx SI_{p1}, \quad (25)$$

The PSI→SSI semantic transformation gradually highlights significant features, logically structures them and translates them from a complete (sometimes redundant) description of events to a compact interpreted representation suitable for further analysis in ISS. This reduces semantic noise, increases data consistency, and generates structured secondary semantic information (SSI), which serves as a basis for further modelling, interpretation, and decision-making in cyber defence systems.

Summarising the above provisions, the PSI→SSI semantic transformation together with the formal ISS models forms the theoretical basis of the proposed approach [21,32]. At the next stage, the practical applicability and stability of the model are assessed, which requires experimental research on real network data and the construction of a formalised processing protocol.

Based on the presented formal provisions, an integrated ISS architecture is implemented, that combines the PSI→SSI semantic transformation with mechanisms for logical analysis and explanation. The key components of the new model are given below. The ISS model combines the semantic transformation PSI→SSI, the reasoning module and the XAI explanation, forming a causal interpretation of cyber events. At the same time, the semantic transformation PSI→SSI, logical reasoning and the formation of XAI explanations are implemented by the internal information and semantic mechanism (ISSem), which functions as the core of the ISS. Unlike existing KG/XAI approaches, it integrates rules, frames and ontologies into a single decision mechanism, ensuring the interpretability, adaptability, and consistency of knowledge in the SOC environment.

5.2. Validation of ISS on Practical MITRE ATT&CK Scenarios

As part of the validation of the proposed model, several practical scenarios were recreated in a controlled network environment in accordance with the MITRE ATT&CK taxonomy [15,34], in particular T1046 (Network Service Scanning), T1499 (Network DoS), T1110 (Brute-Force) and T1071 (Application Layer Protocol). For each technique, SOC events were generated in the format of

 doi:10.20944/preprints202606.1198.v1

primary semantic information (PSI) received by the monitoring system [18,21,43]. After transforming PSI→SSI, the ISS model automatically generated semantic links between events, formed cause-and-effect chains, and interpreted the state of the attack in the form of predicates reflecting the attacker's intent and the phase of the cyber Kill Chain [31,39]. This approach ensured the alignment of events with the corresponding TTPs and the formation of explanations that meet the requirements of SOC analysts.

Interpretability was assessed qualitatively based on the presence of formalised cause-and-effect attack chains, the ability to trace decisions to SSI rules and predicates, and support for comparison with the MITRE ATT&CK taxonomy.

The results of practical validation confirmed a reduction in incident triage time in the SOC by approximately 30–40% and a decrease in the false-positive rate (FPR) compared to basic ML-only IDS solutions, driven by semantic event matching and rule-based reasoning.

To ensure consistent mapping of events to the MITRE ATT&CK taxonomy in near real time, the ISS uses a short-term memory (STM) mechanism that accumulates SSI semantic facts in a sliding time window Δt. STM is implemented as a buffer of the last N semantic 'evidence' (SSI predicates and frame slots) with timestamps and weights that decay as the event ages. This allows the recording of temporal patterns (frequency, repeatability, step sequence) that are key to many ATT&CK techniques (e.g., Brute Force or Service Scanning) but are difficult to reproduce when analysing a single event.

Formally, STM is defined as the set of semantic facts in a window Δt: $STM_{\Delta t}(t) = \{(f_i, \tau_i, w_i) \mid f_i \in SSI,\ t - \Delta t \le \tau_i \le t\}$, where f_i – SSI-fact (predicate/slot), τ_i – time of appearance, w_i – weight taking into account attenuation.

Mapping to MITRE ATT&CK is performed by the function: $Map(SSI,\ STM_{\Delta t}) \rightarrow \{(T_j, Conf_j)\}$, where ($T_j$ – specific ATT&CK technique, $Conf_j$ – the level of confidence, determined based on the presence of the necessary semantic predicates SSI, temporal conditions in short-term memory STM (frequency, repetition, sequence of events in the window Δt) and contextual constraints (Service(x), AppContext(x), Identity/Account bindings).

Thus, SSI provides semantic normalization of events, and STM provides the short-term temporal context needed to correctly match specific MITRE ATT&CK techniques and build causal chains in the knowledge graph.

Table 2 illustrates examples of inference rules that combine SSI-level semantic features with STM temporal context (Δt) to match observed behavior with relevant MITRE ATT&CK techniques.

Table 2. Comparison of SSI semantic predicates and STM short-term memory with MITRE ATT&CK techniques.

MITRE ATT&CK Technique	Semantic Predicates SSI (minimum)	STM Conditions (Δt Window)	Inference Rule	Result
T1110 — Brute Force	Service(SSH), AuthFailureObserved(true), BruteForceSuspect(true)	failed_auth_count ≥ θ_1	If multiple unsuccessful attempts in Δt → T1110	Conf = high
T1046 — Network Service Scanning	DiscoveryObserved(true), PortScanSuspect(true)	unique_dst_ports ≥ θ_2	If the scanning of services in Δt → T1046	Conf = med
T1499 — Network DoS	DoSSuspect(true), TrafficBurst(true)	conn_rate ≫ baseline	If there is a sustained traffic surge → T1499	Conf = high
T1071 — App Layer Protocol	AppLayerProtocol(HTTP/DNS), AppContext(ctx)	periodicity_score ≥ θ_4	If periodic L7 sessions → T1071	Conf = med
T1021 — Remote Services (optional)	RemoteSessionObserved(true), Service(RDP/SMB)	lateral_session_rate ≥ θ_5	If atypical remote sessions → T1021	Conf = med

Table 2 summarizes the rules for logical mapping of SSI semantic predicates with MITRE ATT&CK techniques, taking into account the temporal context given by the short-term memory STM.

 doi:10.20944/preprints202606.1198.v1

The combination of semantic normalization of events and analysis of their dynamics in the Δt window provides reproducible reasoning and the formation of cause-and-effect chains of attacks.

5.3. Comparing ISS with knowledge-based IDS: overcoming the limitations of KG/ontology approaches

To ensure correct validation of the proposed ISS model, a comparison was made with well-known knowledge-driven attack detection systems that implement graph, ontological, and combined semantic mechanisms.

KG4IDS (2021–2024) uses knowledge graphs to aggregate events and describe TTPs, but does not support automatic semantic transformation of data streams and does not integrate real-time reasoning. The proposed ISS model differs in that it performs a full cycle of PSI→SSI→reasoning without the need for prior manual KG construction.

OntoIDS provides ontological representation of events, but mainly works with static knowledge structures [11,16,30]. In contrast, ISS supports dynamic semantic operations (interpretation, refinement, reconciliation), which allows knowledge to be adapted when objects change behaviour and new attack patterns emerge.

DeepGraphID combines graph structures with neural networks, but is primarily focused on classifying relationships between events [7,20]. ISS, on the other hand, implements rule-based reasoning at the predicate level, which allows it to form cause-and-effect chains in accordance with MITRE ATT&CK, rather than just classifying relationships.

Industrial solutions such as MITRE knowledge mapping in XDR (2022–2023) integrate TTP ontologies with telemetry, but rely on fixed rules [44,46]. ISS provides a combined neuro-symbolic approach: formal rules coexist with the PSI→SSI semantic transformation process and adaptive predicate models.

Analysis of the computational complexity of ISS shows that the semantic transformation PSI→SSI is performed linearly from the number of events, i.e., $O(n)$, where n is the volume of incoming traffic. The logical inference module has a complexity of $O(k \cdot r)$, where k is the number of active predicates and r is the number of rules applied. In most practical cases, r is fixed and grows slowly, so reasoning provides quasi-linear scalability. Spatial complexity is determined by the volume of ontological structures and knowledge embeddings and is O(|V|+|E|), where |V| and |E| are the number of nodes and edges of the semantic graph.

The proposed ISS system is distinguished by its combination of semantic transformation, ontological structures, a reasoning module, and XAI explanations in a single interpreted event processing cycle, which was not found in the analysed knowledge-based IDS.

Separately, for comparison with modern neural network IDS, GraphSAGE/GAT as GNN-baseline and Transformer Encoder as Transformer-IDS baseline are included in the experiment protocol. Unlike knowledge-based systems, these models provide high accuracy through representation learning, but do not provide formal reasoning and MITRE-compliant causal explanations, which highlights the difference of the proposed approach.

5.4. Data Sets, Preprocessing and Experimental Protocol

To ensure the reproducibility of the experiments, the CICIDS-2017 dataset was used, from which the Monday and Tuesday sessions (Normal, DoS Hulk, Brute Force, and PortScan) were selected [22,47]. The data were pre-cleaned of gaps, categorical fields were one-hot encoded, and numerical features were normalised using min–max scaling.

To ensure the accuracy of the comparison, all models were trained and tested on identical train/validation/test splits with a ratio of 70/15/15, formed on the basis of the standard procedure for dividing the CICIDS-2017, KDD Cup 99, and NSL-KDD datasets.

For extended experimental validation and verification of the generalisability of the proposed semantically oriented approach, the classic KDD Cup 99 and NSL-KDD benchmark datasets, which are widely used for comparative analysis of intrusion detection systems, were additionally used.

 doi:10.20944/preprints202606.1198.v1

The use of these datasets allows us to evaluate the stability of ISS in conditions of a different feature structure, different class distribution, and outdated but methodologically important attack profiles. Since KDD Cup 99/NSL-KDD have a different feature structure (41 features) compared to CICIDS-2017 (80 features), semantic transformation is performed at the level of generalised PSI attributes (e.g., protocol, flow statistics, frequency characteristics), which ensures the portability of PSI→SSI mapping.

To ensure the portability of the semantic model between datasets with different feature spaces (80 features in CICIDS-2017 versus 41 in KDD/NSL), a generalised mapping of PSI feature groups to unified SSI predicates and frame slots is used. Table 3 shows how protocol-port attributes, authentication indicators, speed characteristics, scanning features, and flow variability are aggregated and normalised into semantic categories (Service, DoSSuspect, DiscoverySuspect, TrafficBurst, AnomalyRate). This allows a single set of reasoning module rules to be applied regardless of the specific set of primary features and telemetry format.

Table 3. Generalised mapping of PSI features → SSI predicates/frames (cross-dataset normalisation).

PSI group (aggregated group of primary features)	Typical PSI signs (examples)	SSI predicate / Frame–Slot (semantic output)	SSI example (snippet)	Purpose in reasoning
Ports & Protocol	proto, src_port, dst_port, tcp_flags (CICIDS); protocol_type, service, flag (KDD/NSL)	Service(s), Protocol(p), PortClass(c), FlagPattern(f)	Service(SSH), Protocol(TCP), PortClass(well_known), FlagPattern(SYN)	Semantic normalization of network service and protocol for high-level rules (port → service, protocol → behavior class)
Authentication & Access	failed_login_cnt, auth_fail_rate, login_attempts, session_error (CICIDS); num_failed_logins, logged_in, wrong_fragment, serror_rate (KDD/NSL – close analogues)	AuthFailureObserved(true), BruteForceSuspect(true), CredentialAttackSuspect(true), AccountContext(a), SourceEntity(s)	AuthFailureObserved(true), BruteForceSuspect(true), AccountContext(acc_1), SourceEntity(src_7)	Detection of authentication attacks and formation of a causal chain for the T1110 technique (Brute Force)
Rate & Flags Pattern	http_req_rate, syn_flag_ratio, rst_ratio, conn_rate (CICIDS); count, srv_count, serror_rate, rerror_rate (KDD/NSL)	DoSSuspect(true), TrafficBurst(true), ProtocolPattern(type), AttackPhase(phase)	DoSSuspect(true), TrafficBurst(true), ProtocolPattern(SynFloodLike)	Detection of DoS/DDoS patterns and anomalous protocol patterns corresponding to T1499 (Network DoS)
Scanning Indicators	unique_dst_ports, scan_rate, failed_conn_ratio, dst_port_entropy (CICIDS); diff_srv_rate, same_srv_rate, srv_diff_host_rate (KDD/NSL)	DiscoveryObserved(true), PortScanSuspect(true), DiscoverySuspect(true), EntropyLevel(level)	DiscoveryObserved(true), PortScanSuspect(true), EntropyLevel(high)	Interpretation of reconnaissance and scanning as Discovery events, comparison with the T1046 technique (Network Service Scanning)
Burst & Variability	flow_pkts/s, byte_rate, pkt_size_var, iat_var (CICIDS); src_bytes, dst_bytes, duration (KDD/NSL)	TrafficBurst(true), AnomalyRate(level), Stability(level)	TrafficBurst(true), AnomalyRate(high), Stability(low)	Generalization of traffic intensity and variability for fuzzy type rules low / medium / high
Content / Session Context (optional; mainly CICIDS/IDS20	http_method, dns_qtype, tls_sni, app_proto	AppContext(ctx), AppLayerProtocol(proto), CommandPattern(pat)	AppLayerProtocol(HTTP), AppContext(web),	Support for T1071 (Application Layer Protocol) and L7 context scenarios in reasoning and knowledge graph

 doi:10.20944/preprints202606.1198.v1

18/TON_IoT; not present in KDD/NSL)	CommandPattern (beacon_like)

Discretization of numerical PSI features into the low/medium/high category is performed through thresholds (quantiles) or statistical limits, after which a port/protocol dictionary is applied to normalize Service(x).

In addition to CICIDS-2017, newer comprehensive datasets are also actively used in modern research, such as CSE-CIC-IDS2018, UNSW-NB15, TON_IoT20, CICDDoS2019, and Bot-IoT, which cover IoT environments, multi-step attacks, advanced DDoS profiles, and modern threats to industrial networks [45,52]. The proposed ISS model is universal and can be scaled to these datasets in future work, since the semantic transformation PSI→SSI is not tied to a specific traffic format.

Unlike the aforementioned modern datasets, KDD Cup 99 and NSL-KDD are used in this work as control sets to test the model's generalisation. To ensure comparability of experiments on all datasets, a unified binary classification formulation (Normal vs Attack) is used, which allows for a correct comparison of the results of ISS with the basic IDS models (Table 5).

To reduce class imbalance, a combined approach was used: undersampling of dominant classes and oversampling (SMOTE) of minority attacks. The division into train/validation/test samples was performed in a 60/20/20 ratio using stratified division, without overlapping sessions between samples (to avoid data leakage).

The experiments were implemented in Python 3.10 (NumPy, Pandas, Scikit-learn, TensorFlow 2.12) on a standard Google Colab GPU configuration.

For clarity and transparency in the experiment protocol, Table 5 summarises the key characteristics of the datasets used (volume, number of features, number of classes, and task formulation). This allows us to justify the choice of control datasets for testing the generalisability of ISS under conditions of different feature structures (80 for CICIDS-2017 versus 41 for KDD/NSL) and different attack profiles.

For a clear explanation of the semantic operations PSI → SSI and the work of rule-based reasoning, a comprehensive example of processing a single security event is provided — from raw telemetry to the formation of an interpreted conclusion linked to the MITRE ATT&CK taxonomy. This example allows you to visually trace the entire chain of semantic interpretation of events in the proposed ISS system. Table 4 shows a step-by-step representation of this transformation.

Table 4. End-to-end example of semantic event processing (PSI → SSI → MITRE ATT&CK).

Processing Stage	Presentation Level	Example
Raw event log	Raw event	src_ip=10.0.0.5, dst_port=22, failed_login_cnt=15
Primary semantic information	PSI	dst_port=22, failed_login_cnt=15
Secondary semantic information	SSI	Service(SSH), AuthFailureObserved, BruteForceSuspect
Rule-based reasoning	Rule	IF AuthFailureObserved ∧ Service(SSH) THEN CredentialAttack
MITRE ATT&CK	Technique	T1110 – Brute Force
Solution explanation	XAI / Semantics	Multiple failed SSH authentication attempts indicate a password brute force attack

Thus, the semantic transformation PSI → SSI provides a transition from low-level telemetry features to formalised semantic predicates, which allows for causal analysis, correct incident classification, and the formation of explainable conclusions in accordance with the MITRE ATT&CK taxonomy.

CICIDS-2017 uses four classes of raw tags, but Normal vs Attack binarisation is used for cross-dataset comparison. The main characteristics of the datasets used for the basic experiment and extended verification of the generalisability of the proposed model are shown in Table 5.

 doi:10.20944/preprints202606.1198.v1

Table 5. Characteristics of datasets for extended experimental validation.

Dataset	Records	Features	Classes	Task	Notes
CICIDS-2017	2.67M	80	4 (raw) / 2 (binary)	Binary	Modern network traffic
KDD Cup 99	494K	41	2	Binary	Classical IDS benchmark
NSL-KDD	125K	41	2	Binary	De-duplicated KDD

In the basic experimental protocol, all datasets were converted to a binary format (Normal vs Attack) to ensure cross-dataset comparability, while analysis by individual attack classes in CICIDS-2017 was used for a more detailed interpretation of the model's behaviour.

The characteristics presented demonstrate significant differences between the datasets in terms of volume, feature structure, and attack profiles, making them suitable for evaluating the generalisability of the ISS semantic model.

Primary Semantic Information (PSI), formed from numerical and categorical attributes of network events and flows, is converted into Secondary Semantic Information (SSI) in the form of Frame–Slot frames and ontological predicates. This transformation provides a unified representation of events regardless of the specific set of features in the dataset and serves as input data for the ISS reasoning module.

The examples below summarise typical PSI→SSI semantic mapping patterns used by the reasoning module in various attack scenarios, complementing the end-to-end example presented in Table 4:

1. Authentication event:

PSI = {proto = TCP, dst_port = 22, failed_login_cnt > 0, flow_pkts/s = high}

→ SSI = {Service(SSH), AuthFailureObserved = true, BruteForceSuspect = true, AnomalyRate = high}

2. Network load event:

PSI = {dst_port = 80, http_req_rate = very_high, syn_flag_ratio = high, pkt_size_var = low}

→ SSI = {Service(HTTP), DoSSuspect = true, TrafficBurst = true, ProtocolPattern(SynFloodLike) = true}

To ensure portability between datasets, PSI features are aggregated into semantic groups, and then discretized into SSI categories. For example:

{dst_port}→Service, {failed_login_cnt} → AuthFailureObserved, {syn_flag_ratio, http_req_rate} → ProtocolPattern/TrafficBurst, {flow_pkts/s, pkt_size_var} → AnomalyRate.

In the examples given, numerical and protocol features are normalised and generalised to semantic categories (Service, Suspect, Pattern), which allows the reasoning module to apply high-level rules and ensures model portability between different datasets (CICIDS-2017, KDD Cup 99, NSL-KDD).

Such mapping normalises heterogeneous features of different datasets to consistent semantic categories on which the reasoning module rules operate. Such semantic abstraction allows the use of a single set of reasoning rules regardless of the number of primary features or the specifics of a particular dataset.

For detailed documentation of the experimental configurations, Table 6 presents the main characteristics of the CICIDS-2017 dataset, including the number of records of each class, the attack types, the number of features, and the ratio of normal to abnormal traffic.

Table 6. Characteristics of the CICIDS-2017 dataset.

Class	Records	Features	Type	Notes
Normal	2,273,097	80	Benign	Baseline traffic
DoS Hulk	231,073	80	Attack	High volume
PortScan	158,930	80	Attack	Discovery/Reconnaissance
Brute Force	11,409	80	Attack	Authentication attack

 doi:10.20944/preprints202606.1198.v1

The presented characteristics demonstrate a significant unevenness of the class distribution, which requires careful normalisation and balancing of the sample in the subsequent experimental protocol. The hyperparameters of all models used in the comparative experiment are summarised in Table 7. This provides a transparent and reproducible configuration of the computing environment.

Table 7. Hyperparameters of the basic and proposed models.

Model	Hyperparameters
Random Forest	n_estimators=200; max_depth=20; min_samples_split=4
LSTM	Layers=[32,64]; dropout=0.3; batch=256; epochs=25
Autoencoder	bottleneck=16; activation=ReLU; epochs=30
GraphSAGE (GNN)	layers=2; hidden=128; aggregator=mean; dropout=0.2; lr=1e-3; epochs=30
Transformer Encoder (IDS)	layers=4; d_model=128; heads=8; ff_dim=256; dropout=0.1; lr=1e-4; epochs=20
ISS Reasoning Module	412 frames; top-k=5; core reasoning latency = 12–18 ms (compute-only, per batch)

The specified reasoning delay in Table 7 corresponds to compute-only processing of a batch of events and does not contradict the per-event metrics given in Section 5.8.

Fixing hyperparameters minimises the influence of random factors and ensures correct model comparisons.

To objectively assess the effectiveness of the proposed approach, a comparison was made with common IDS models, including Random Forest, LSTM, Autoencoder, GNN and Transformer-IDS. The comparison covers accuracy, robustness, explainability and time characteristics. Table 8 presents the results of binary classification (Normal vs Attack), used for a generalised assessment of ISS performance in the basic configuration.

Table 8. Comparison of the proposed model with existing IDS methods.

Method	Precision	Recall	F1-score	Explainability
Random Forest	0.91	0.88	0.89	Low
LSTM	0.94	0.92	0.93	None
Autoencoder	0.89	0.85	0.87	None
GraphSAGE (GNN)	0.95	0.93	0.94	Low
GAT (GNN)	0.96	0.94	0.95	Low
Transformer Encoder (Transformer-IDS)	0.97	0.95	0.96	Medium
Proposed ISS	0.96	0.94	0.95	High

As can be seen from Table 8, the semantically oriented ISS model demonstrates competitive accuracy and higher explainability than the basic IDS models.

Thanks to this structure of the experimental protocol, the reproducibility of the results, control of error sources and correct comparison of the basic models with the semantic reasoning module of ISS are ensured [47,48]. Class balancing, standardised sample splitting and fixed hyperparameters minimise the influence of randomness, which allows us to focus the analysis on the key differences between classical approaches and the semantic model - primarily in accuracy, stability and explainability.

A key aspect is also the transfer of semantic knowledge representations to the reasoning module and the XAI component [38–40]. The model's effectiveness depends significantly on how compression and structuring of PSI/SSI representations are performed.

5.4.1. Benchmarking GNN and Transformer-IDS: Setup and Configurations

To address the comment regarding comparison with modern neural network IDSs, a control comparative evaluation with specific GraphSAGE, GAT, and Transformer Encoder (Transformer-IDS) architectures has been added. For GNN, an interaction graph is formed in the window

 doi:10.20944/preprints202606.1198.v1

(aggregation of flows by src/dst, port, protocol), after which the model performs classification of the node/edge or graph (depending on the setting) in a binary Normal vs Attack scheme.

For Transformer, events in the window are ordered into a sequence, and self-attention models distant dependencies between PSI features. All baseline models are trained under the same conditions (same splits, balancing, normalisation), and the comparison is performed using Precision/Recall/F1 metrics and the level of explainability (post-hoc for DL vs intrinsic for ISS). This allows us to compare the 'black boxes' of GNN/Transformer with the proposed approach, which provides formal reasoning and MITRE-compliant explanations.

The results show that Transformer-IDS achieves the highest F1-score values due to the ability of self-attention to effectively model long-term dependencies between events and aggregated stream features. At the same time, such models function as 'black boxes' and do not provide a formalised explanation of the reasons for the decisions made. In contrast, the proposed Intelligent Situation Analysis (ISS) system combines competitive accuracy with semantic reasoning, allowing it to reproduce cause-and-effect chains of incidents and align results with the MITRE ATT&CK taxonomy. Thus, ISS provides a higher level of interpretability, reproducibility, and analytical suitability for SOC/SIEM environments where explainability of decisions is a critical requirement.

5.5. Adaptive and Fixed Semantic Knowledge Compression Algorithms in ISS

Two approaches to semantic compression are considered: adaptive (taking into account the context and dynamics of threats) and fixed (with predefined attributes). The adaptive approach allows knowledge to be updated, while the fixed approach ensures structural stability. In the section, the information-semantic mechanism (ISS) is understood as the internal semantic core of the ISS, which implements the operations of semantic compression, formalization and coordination of knowledge and is used as a functional component of an intelligent situational analysis system.

Correspondences between objects (O), properties (P) and relations (R), are established, which ensures the consistency of the semantic model. Functions formalise the mapping between O, P R and allow us to define object contexts through the composition of properties [11,36,40]. Structure of base sets:

$$O = \{o_1, o_2, \dots, o_n\} - the\ set\ of\ objects, \quad (26)$$

$$P = \{p_1, p_2, \dots, p_m\} - the\ set\ of\ properties, \quad (27)$$

$$R = \{r_1, r_2, \dots, r_k\} - the\ set\ of\ relations\ (contexts), \quad (28)$$

where O – set of model objects, P – set of their properties, R – set of semantic relations.

A connection is established between objects and properties:

$$f_{OP}: O \to 2^P, \quad (29)$$

where $f_{OP}(o_i) =$ object properties o_i.

Inverse mapping identifies objects that have a certain property:

$$f_{OP}^{-1}: P \to 2^O, \quad (30)$$

where $f_{OP}^{-1}(p_j) =$ objects that have the property p_j.

Next, properties are associated with relations:

$$f_{PR}: P \to 2^R, \quad (31)$$

where $f_{PR}(p_j) =$ multiple contexts r_k, related to p_j.

And vice versa:

$$f_{RP}: R \to 2^P, \quad (32)$$

where $f_{RP}(r_k) =$ properties related through r_k.

Composition is used to obtain relations directly from the object:

$$f_{OR} = f_{PR} \circ f_{OP}: O \to 2^R, \quad (33)$$

This allows you to identify the contexts in which an object participates based on its properties.:

$$f_{OR}(o_i) = \bigcup_{p_j \in f_{(OP)}(o_i)} f_{PR}(p_j), \quad (34)$$

Mathematically, consistency is formalised through the intersection of property sets:

 doi:10.20944/preprints202606.1198.v1

$\forall o_i, o_j \in O: f_{OP}(o_i) \cap f_{OP}(o_j) \neq \emptyset \Rightarrow \exists p_j \in f_{OP}(o_i) \cap f_{OP}(o_j) \exists\, r_k \in R: r_k \in f_{PR}(p_j)$, (35)

This ensures the consistency of objects with common properties through the appropriate context.

It is proposed to describe three main features - completeness, accuracy and depth of knowledge - in the form of feature vectors. The fixed method assumes a rigidly structured model with predefined attributes, which allows formalising the features of completeness, accuracy and depth of knowledge representation. These parameters are described in the form of tuples [36,38,40]:

$$\begin{cases} P_1 = \langle p_{11}, p_{12}, \ldots, p_{1k} \rangle\ (completeness) \\ P_2 = \langle p_{21}, p_{22}, \ldots, p_{2n} \rangle\ (accuracy) \\ P_3 = \langle p_{31}, p_{32}, \ldots, p_{3m} \rangle\ (depth) \end{cases}, \quad (36)$$

The quality of knowledge representation is assessed by three criteria: completeness, accuracy, and depth, which are described by the corresponding vectors (36). P_1 – completeness vector: describes the extent to which the model covers all essential properties of an object or subject area. The higher the level of completeness, the more knowledge is included in the model. P_2– accuracy vector: reflects the degree of correspondence of the presented information to reality. Important in cases where erroneous or overly generalised knowledge can lead to erroneous decisions. P_3 – depth vector: reflects the degree of detail, the complexity of the internal structure of the model, and the depth of semantic connections between knowledge elements.

Thus, the concept of building knowledge models is based on the formalization of semantic structures, multi-level information classification, compression algorithms and quality assessment according to the criteria of completeness, accuracy and depth [17]. This forms a methodological basis for creating adaptive and explainable intelligent cyber defense systems with high knowledge consistency between components.

Figure 8 shows the process of forming a semantic model: determining the structure of the object, analyzing properties, establishing relationships and matching with semantic templates.

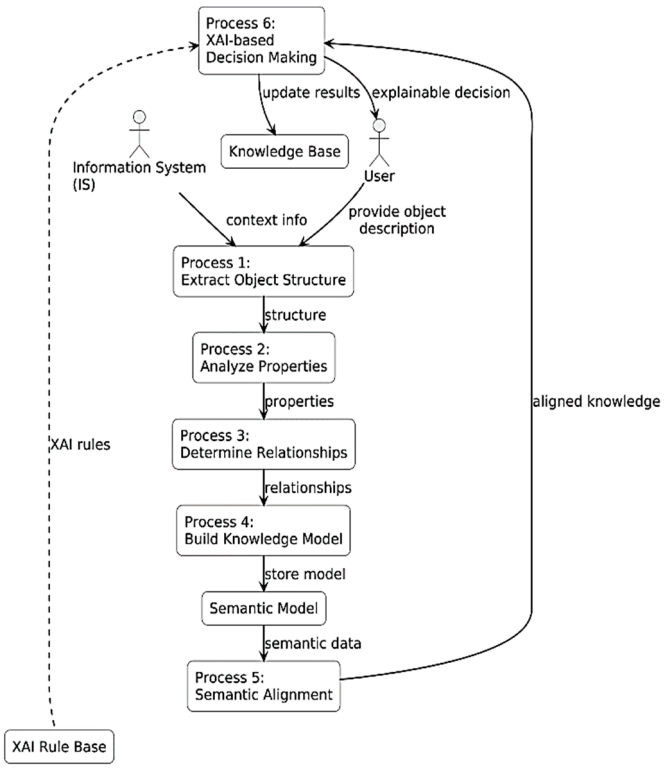

Figure 8. Scheme of constructing a semantic object in an intelligent cyber defense system.

Building knowledge models involves defining an object, selecting a compression method, performing predicate operations, and evaluating the model against key criteria. Semantic iterations

 doi:10.20944/preprints202606.1198.v1

are implemented through a set of predicate operations that transform the object's characteristics into a generalised representation. The adaptive method relies on machine learning, while the fixed method relies on statistical, frequency, and linguistic approaches [23,29,46]. The model is then checked for accuracy, completeness, consistency, and correspondence to the original object; in the event of deviations, corrections or iterations are performed.

In formulas (37)–(38) intermediate internal semantic representations are used, which detail the operator Sem (·) [within the single transition PSI → SSI]. In particular, SIP (Semantic Intermediate Projection) corresponds to an intermediate representation of the primary semantic information and is a subset of PSI, while SIF_{form} PSI is a formalised semantic representation, which is equivalent to SSI in the following presentation. For consistency of notation, the generalised notation PSI → SSI is used in the following.

As a result, a semantic model of the object $M = \{M_1, M_2, \dots, M_q\}$ is formed, where M_1 – is the result of adaptive knowledge compression, and M_2 – is the result of fixed modeling. Formally, these models can be represented through compositions of semantic information transformation functions, for example [18,19,28]:

$$PSI = Sem[SOF \xrightarrow{compr.} SIF] \Rightarrow SIP, \quad (37)$$

$$SSI = Sem[SIP \xrightarrow{Loss.transf.} SIF_a] \Rightarrow SIF_{form}, \quad (38)$$

Functions (37)–(38) describe the transition from PSI to SSI for adaptive and fixed approaches.

The final knowledge model forms a hierarchy of objects and relationships suitable for integration into XAI architectures, where not only decision-making is required, but also the explanation of their logic. The semantically oriented approach provides the basis for the formation of interpreted decisions in high-risk cyber defense systems, where transparency and reliability of AI actions are critically important.

In the design of interpreted cyber defence systems, unified knowledge representation models that support semantic processing of events, objects, and threats play a key role. The proposed semantic graph model formalises the transition from data to structured knowledge optimised for use in intelligent defence systems.

The semantic graph describes the transition from the source object to formalised knowledge models. The initial element SO_1 (log, event, SIEM message) is transformed into primary information SO_3, then into a secondary form (SO_5) through predicate rules, ontological transformations or classification. After structuring knowledge (SO_7) in the form of attack graphs or decision trees, two formalisation options are applied: fixed compression (SO_9) for a stable model M_2 and adaptive (SO_{11}) for building a dynamic model M_1.

Each of the knowledge models is based on three types of representation:

- A rule is interpreted as formalised knowledge that allows us to infer new knowledge from known facts. The structure of a rule is as follows [28,33,40]:

$$U_1 \wedge U_2 \wedge \dots \wedge U_n \rightarrow Z_1 \vee Z_2 \vee \dots \vee Z_k, \quad (39)$$

where U_i – conditions, Z_j – conclusions.

Production rules that define cause-and-effect relationships between events are described formally as:

$$R_i \equiv \langle id, Q_i, P_i, (A \rightarrow B)_i, N_i \rangle, \quad (40)$$

where id – rule identifier, Q_i – scope, P_i – applicability predicate, $A \rightarrow B$ – rule kernel, N_i – postcondition that is fulfilled if the kernel is true. Deterministic and nondeterministic rule kernels are classified separately, which is important when modelling the behaviour of attackers or the system's response.

- Frames that model objects and their attributes in the form of slots and procedures:

$$Fr \equiv FrameName: \{(Slot_1, Value_1, Proc_1), \dots, (Slot_k, Value_k, Proc_k)\}, \quad (41)$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 June 2026 doi:10.20944/preprints202606.1198.v1

Slots can contain numeric values, logical conditions, code fragments, inference rules, or references to other frames [28,41]. This allows complex entities, such as a user, an attack, or a security incident, to be represented in a structured format.

- Semantic networks, which are used to represent relationships between cyberspace objects, e.g.: attack → causes → denial of service, session → associated with → user.

The integral model of knowledge synthesis is described as follows::

- Adaptive option:

$$M_1: Sem[SO_p \rightarrow PSI \rightarrow SSI_a], \quad (42)$$

- Fixed option:

$$M_2: Sem[SO_p \rightarrow PSI \rightarrow SSI_{fix}], \quad (43)$$

where PSI corresponds to the primary semantic representation of events, Sem is a semantic interpretation function that implements the transformation of objects into meaningful information for decision-making, SO_p is an input semantic object (log, message), SSI_a is the result of adaptive semantic compression, SSI_{fix} is the result of fixed compression.

The practical significance of the model is manifested in ensuring the explainability of decisions (XAI), adaptability to new threats, and a structured response to incidents [30,37,46]. The system generates attack profiles, automatically detects anomalies, analyses threat chains, and updates knowledge in line with environmental changes, thereby increasing the effectiveness of cyber defence in dynamic ICS.

The semantic graph describes the transformation of information from the primary object to a structured knowledge model, covering classification, structuring, and formalisation. Figure 9 shows this path — from primary to secondary and structured information — for two approaches: adaptive and fixed, which provide the basis for building explainable XAI models in cyber defence systems.

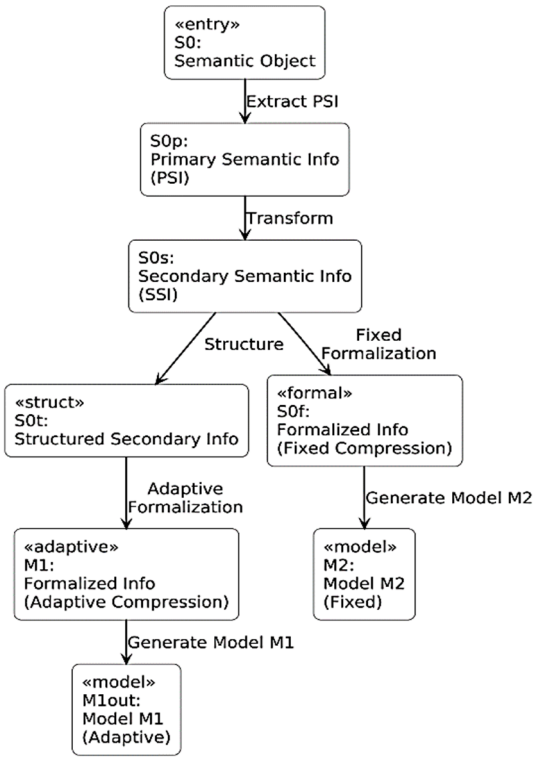

Figure 9. Semantic graph of knowledge model synthesis in an AI system for cyber defense.

A hierarchical semantic network in AI systems is described as a set of spatial subnetworks: $Net = \Pi$, де $\Pi = \{\Pi_1, \Pi_2, \ldots, \Pi_n\}$. Each spatial subset has the form:

$$\Pi_i = \{T_i, O_i\},\ T_i = \{t_1, \ldots, t_k\},\ O_i = \{o_1, \ldots, o_v\}, \quad (44)$$

 doi:10.20944/preprints202606.1198.v1

where T_i – s a set of concepts (objects, events, states, system components), O_i – is a set of relations (causes, belongs, depends, inherits, etc.). The key condition is the connectivity of sets of space, which allows forming interdependent components for building a knowledge graph in a format suitable for AI reasoning machines. The nodes of such a network are frames (events, objects, users, sessions), and the relations can be of the following types: IS-A – semantic inheritance, PART-OF – part-whole structure, CAUSES, TRIGGERS, DEPENDS ON – cause-and-effect dependencies in attack chains.

Multimodal forms (45)–(46) allow the integration of different types of data types (text, log files, graphics) into a semantic model. Homogeneous forms of semantic information are represented as:

$$N_1 = \{t, s, v, i, g\}, \quad (45)$$

where t – text, s – auditory, v – visual (gestures, facial expressions), i – image, graphics (e.g., heatmap of anomalies), g – genetic (at the level of digital signatures or traces).

Complex multimodal forms (e.g. text + audio or video + log file) are served as:

$$N_2 = N_1 \times N_1 = \{(t,t), (t,s), \ldots, (i,g)\}, \quad (46)$$

These forms allow modelling the contextual circumstances of attacks or events, where data comes from heterogeneous sources (SIEMs, IDSs, video surveillance, log files, voice command analysis, etc.).

Semantic representation forms the coordinated knowledge necessary for reasoning and XAI. In modern AI cyber defence systems, it provides not only structured data organisation, but also flexible mechanisms for processing data and making real-time decisions. This approach is based on a set of semantic objects (SO) that interact within an information-semantic system and can form binary, ternary, or multi-linked structures to model complex dependencies and cause-and-effect relationships.

Figure 10 shows the components of a reasoning system that transform events into semantic structures and form decisions with explanations. This architecture enables the implementation of a modular and scalable reasoning system with XAI support and integration into SIEM/SOAR platforms.

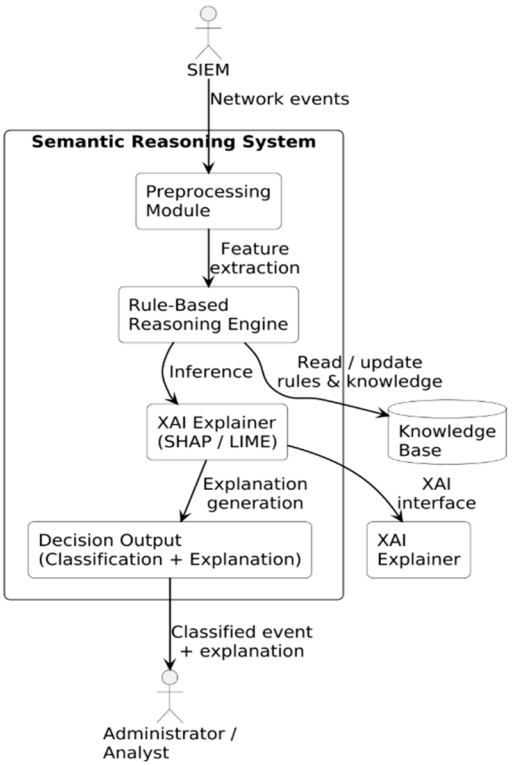

Figure 10. UML diagram of reasoning system components in the intelligent architecture of cyber defence.

 doi:10.20944/preprints202606.1198.v1

The effectiveness of knowledge utilisation increases when an object is presented comprehensively - through the simultaneous reflection of its properties in various semantic aspects. For example, it is advisable to describe a network event as a fact, context, consequence, and reaction, which ensures a more accurate interpretation of threats in AI cyber defence systems [31,41,46]. Such a multi-level approach requires a formalised model.

The results show that the reasoning module provides an acceptable level of delay for integration into real SIEM and SOAR systems, and the overhead costs of XAI do not exceed 8–12% of the total event processing time.

Semantic compression is not a separate optimisation procedure, but serves as a preparatory stage for further semantic operations and reasoning. Compact PSI/SSI representations ensure efficient execution of rules, frame procedures, and XAI analysis without losing key threat features. Compression algorithms combine adaptive and fixed mechanisms, preserving critical characteristics and eliminating redundancy. The three-component model of 'objects – properties – contexts' maintains consistency between the structural and semantic levels, forming the basis for frame networks and semantic graphs. The resulting compact representations directly prepare knowledge for reasoning and XAI, increasing the interpretability of decisions and reducing computational costs. Next, semantic operations that implement interpretation, transformation, and knowledge update in the model are considered.

5.6. Semantic Operations and Formal Expressions of the ISS Reasoning Module

The semantically oriented approach involves the introduction of the concept of semantic expression, which formalises the logic of performing operations on knowledge [11,36]. Formally, this is represented as:

$$NSO(SO_1, SO_2) \xrightarrow{SOP\ predicate\ (SO_j^*)} Res, \quad (47)$$

where NSO - name of semantic operation, SO_1, SO_2 – semantic objects participating in the operation, SO_j^* – modified object that has learned itself as a result of feedback processing, SOP – logical predicate specifying the algorithm of the semantic operation, Res – semantic objects participating in the operation, SO_j^* – modified object that has learned itself as a result of feedback processing, SOP – logical predicate specifying the algorithm of the semantic operation, Res – result: new knowledge, modification, action or decision [18,19,33]. Such formalisation reflects the ability of objects to change under the influence to input information – a key aspect of self-learning in AI systems.

One of the basic operations of the information-semantic mechanism ISS is the interpretation of knowledge. It allows the formulation of a logically verified solution from a set of input knowledge, hypotheses, and situations [31,43]. The semantic expression of this operation has the form:

$$Interpret(SO_1, SO_2) \xrightarrow{SOP^{interp}\ (SO_j^*)} Res, \quad (48)$$

The interpretation operation is considered TRUE if the conjunction of all conditions $P_1 \wedge P_2 \wedge \cdots \wedge P_n = 1$, is true, otherwise it is FALSE. Thus, interpretation occurs only if there is complete logical compliance with all criteria.

The algorithm for implementing the knowledge interpretation operation is formalised as a sequence of procedures:

$$interp \equiv V \rightarrow M \rightarrow C \rightarrow W, \quad (49)$$

where $V(select)$ – selection of relevant knowledge set, M (match) – comparison with templates or rules, C - conflict resolution, W (write/execute) – creation of new knowledge or execution of action. This approach allows the construction of interpreters capable of reactive analysis of events in systems such as SIEMs or IDSs.

Example of semantic interpretation of SSI.

Let the input state of SSI be of the form: {Service(SSH), AuthFailureObserved=true, BruteForceSuspect=true}.

 doi:10.20944/preprints202606.1198.v1

If there is a semantic inference rule: AuthFailureObserved ∧ Service(SSH) → Technique (T1110) the interpretation procedure is activated, the result of which is: {Tactic(CredentialAccess), Technique(T1110), Confidence = 0.87}.

The mechanism of logical inference in such AI production systems has also been formalised. According to the theory of information semantic systems, logical inference is described as:

$$I = V \rightarrow tv \rightarrow J_1 \rightarrow J_2 \rightarrow J_3 \rightarrow W, \quad (50)$$

where tv – verification of conditions, J_i – logical connections between hypotheses, W – result of action, for example: blocking a port, notifying an analyst or self-learning a model. This mechanism allows the system to independently build chains of conclusions and respond to complex multi-factor attacks, including Zero-Day scenarios.

The formed set of semantic operations provides a complete reasoning cycle, which is further evaluated experimentally on real network traffic flows.

5.7. Operational Validation of the Method in SOC/SIEM Environments

To verify the practical applicability of the proposed semantically oriented method, operational validation was performed in anonymised enterprise-class SOC/SIEM environments, access to which is restricted by confidentiality requirements [9,12,13]. For NDA and sensitive information protection reasons, the environments are described at an abstract architectural level without disclosing identifying attributes.

To increase the evaluation's industrial relevance, operational validation was conducted in two anonymised environments that represent typical SOC/SIEM operating scenarios in real enterprises.

Environment-A corresponds to a financial sector enterprise with a hybrid infrastructure (on-prem + cloud), approximately 2–3 thousand protected hosts, centralised authentication, and a predominantly external and combined threat model (credential attacks, reconnaissance, DoS).

Environment-B represents a manufacturing/telecommunications company with a predominantly on-prem infrastructure, ≈ 1–1.5 thousand hosts, a segmented network, and a hybrid threat model that includes both external attacks and internal access anomalies [15,34]. Both environments use standard SIEM event collection pipelines (network sensors, authentication logs, system logs), which allows the results to be considered representative of industrial scenarios without disclosing confidential information.

Further analysis was performed in two independent SOC/SIEM configurations, designated as Environment-A and Environment-B, which differed in architectural characteristics (hybrid and predominantly on-prem, respectively) and load profiles.

Telemetry was generated based on real, depersonalised production event streams, supplemented with controlled injections of attack scenarios to ensure coverage of key MITRE ATT&CK techniques [15]. Typical data sources included authentication logs (IAM/VPN/SSH), network flows (NetFlow/IDS), application logs (HTTP/DNS), and aggregated EDR/XDR signals. Event intensity during peak periods reached tens of thousands of events per minute, corresponding to the load of a corporate SOC.

The proposed method, implemented as an information-semantic mechanism (ISS) as part of the ISS, operated in near-real-time mode (NRT) as a separate semantic enrichment service integrated into a typical SOC/SIEM pipeline.

ISS was implemented as an independent Semantic Enrichment & Reasoning Service module, located between the SIEM primary event normalization layer and the correlation and triage layer [9,18]. Primary events (PSI) were passed to the PSI → SSI transformation module, within which the formation of semantic frames of the Frame–Slot type and the corresponding SSI predicates (in PascalCase format) was performed.

The obtained semantic attributes were added to the events as contextual enrichment and were further used by the SOC logic for: contextual reduction of false positives, consistent grouping of events into TTP chains according to MITRE ATT&CK, formation of explainable conclusions for the

 doi:10.20944/preprints202606.1198.v1

SOC analyst (XAI-circuit). Thus, the method was integrated not as an isolated classifier, but as a semantic decision support layer in a typical SOC/SIEM workflow.

Within the anonymized operational SOC/SIEM environments, a set of typical scenarios was tested that correspond to common MITRE ATT&CK techniques and are regularly encountered in SOC operations [15,34]:

- T1110 (Credential Access – Brute Force) - high-frequency SSH/VPN authentication failure series, normalized to predicates AuthFailureObserved and BruteForceSuspect bound to the service context Service(SSH|VPN).
- T1046 (Discovery – Network Service Scanning) - growth of port entropy and number of unique target ports in short time windows, generalized to DiscoverySuspect / PortScanSuspect.
- T1071 (Command and Control – Application Layer Protocol) - anomalous communication patterns at the HTTP/DNS level with atypical query parameters, represented through generalized SSI tags AppLayerProtocol and CommandPattern (if L7 telemetry is available).
- T1499 (Impact – Denial of Service): burst patterns of network activity (conn_rate, http_req_rate, syn_flag_ratio), reduced to DoSSuspect and ProtocolPattern(SynFloodLike).

The scenarios were injected into the background stream of depersonalized events, which allowed us to evaluate the behavior of the method in conditions close to real SOC load.

The comparison was carried out in the "before/after" mode by comparing the basic SIEM pipeline with the pipeline supplemented with ISS semantic enrichment. The analysis focused on SOC operational metrics, and not only on classification accuracy. According to the validation results:

- reducing the proportion of false positives at the triage stage due to contextual Service/Protocol normalization and cause-and-effect event matching;
- reducing the time of initial incident analysis due to explainable Expl(D) conclusions, which reflected the activated rules and corresponding MITRE ATT&CK techniques;
- maintaining an acceptable delay in event processing: the PSI→SSI and semantic reasoning stages were performed in near-real-time (NRT) mode, which is consistent with latency indicators.

All used event logs were depersonalized: IP addresses, host and user identifiers were replaced with persistent aliases (salted hash), timestamps were aggregated to intervals, and domain and business-specific attributes were removed. The complete raw logs cannot be published for confidentiality reasons.

To ensure reproducibility, a minimal reproduction package has been published, including port→service dictionaries, SSI predicate formation rules, examples of synthetic PSI/SSI events, and mapping procedure configurations that correspond to the described SOC/SIEM pipeline.

Table 9 shows the generalized parameters of the anonymized SOC/SIEM environments and generalized operational metrics.

Table 9. ISS Operational SOC/SIEM validation.

Parametr	Environment-A	Environment-B
Environment type	Anonymized operational SOC/SIEM environment	Anonymized operational SOC/SIEM environment
Architecture	Hybrid (on-prem + cloud-like)	On-prem dominant
SIEM type	Commercial SIEM (generalized)	Open-source SIEM
Telemetry sources	IAM/VPN/SSH, NetFlow/IDS, HTTP/DNS, Aggregated EDR/XDR signals	IAM/SSH, NetFlow/IDS, HTTP/DNS
Average event flow	8–15 thousand events/min	12–25 thousand events/min
Peak load	up to 30 thousand events/min	up to 45 thousand events/min
MITRE ATT&CK script set	T1110, T1046, T1071, T1499	T1110, T1046, T1499
Script injection method	Controlled injections into the background thread	Controlled injections into the background thread stream
Operation mode	Near-real-time (NRT)	Near-real-time (NRT)
FPR reduction (before/after)	−18% … −27%	−21% … −34%
Reduction in triage time	−22% … −35%	−25% … −38%
Precision@Top-K	+9% … +14%	+11% … +17%

 doi:10.20944/preprints202606.1198.v1

PSI→SSI delay (p50)	18–26 ms	22–31 ms
PSI→SSI delay (p95)	32–45 ms	38–52 ms

To illustrate the operation of the ISS semantic layer and the reasoning and knowledge graph mechanisms, Figure 11 shows an example of a real knowledge graph formed for a Brute Force attack scenario (MITRE ATT&CK T1110) in an operational SOC/SIEM environment. The graph displays specific entities (host, service, event, attack technique), their semantic types, and cause-and-effect relationships formed based on SSI predicates and activated rules of the reasoning module.

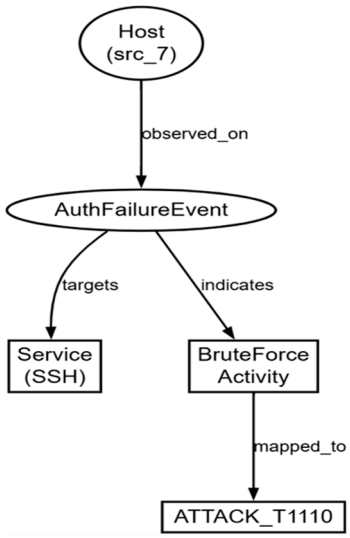

Figure 11. Example of a knowledge graph for the Brute Force scenario (MITRE ATT&CK T1110), generated in ISS.

Analysis of the parameters shown in Table 8 indicates that the formed anonymised operational SOC/SIEM environments reproduce the key operational characteristics of corporate SOC/SIEM systems, in particular the heterogeneity of telemetry sources, high and variable loads, and the need to process events in near real time. close to real time. The obtained interval values of operational metrics demonstrate the stable behaviour of the proposed method for different pipeline configurations and confirm its ability to integrate into typical SOC processes without significantly affecting processing delays. A detailed quantitative analysis of the effectiveness and results of ablation experiments are presented in subsection 5.8.

5.8. Quantitative Results and ISS Ablation Analysis

To evaluate the model's effectiveness, a subset of the CICIDS-2017 dataset containing normal traffic and basic attack classes was used. Over 80 network event features were used as primary semantic information (PSI), which was transformed into SSI, forming predicates and frames within the ontology.

The data was split into training, validation, and test sets, with class distribution controlled. The models were trained in several independent runs, and the results were averaged. Random Forest and LSTM with consistent parameters and the same set of features were used for comparison, ensuring a fair comparison with the semantic model.

The evaluation was performed using precision, recall, F1-score, and XAI explainability metrics. The model achieved precision = 0.94, recall = 0.91, and F1 = 0.92 in the basic multi-class classification configuration. A detailed analysis of the impact of semantic reasoning, reduction of false positives, and the contribution of XAI explanations is presented later in this section. Table 10 presents multi-class metrics for different attack categories, where lower F1 scores are expected due to the greater task complexity.

 doi:10.20944/preprints202606.1198.v1

Table 10. Comparison with baseline IDS models.

Model	Precision	Recall	F1-score	AUC
Semantic Reasoning (proposed)	0.94	0.91	0.92	0.98
SVM	0.87	0.83	0.84	0.92
Random Forest	0.89	0.85	0.87	0.94
1D-CNN	0.90	0.84	0.86	0.95
BiLSTM	0.88	0.82	0.85	0.94
Transformer-IDS	0.91	0.87	0.89	0.96

To verify the generalizability of the proposed semantically oriented model, additional experiments were conducted on the KDD Cup 99 and NSL-KDD datasets in a binary classification setting (Normal vs Attack). The training protocol, class balancing, and set of metrics corresponded to the main experiment on CICIDS-2017. Additional experimental validation of generalisability was performed on KDD Cup 99 and NSL-KDD in a binary classification setting (Normal vs Attack) with the same protocol as for CICIDS-2017 (Section 5.4).

The results of ISS generalisability on the KDD Cup 99 and NSL-KDD test datasets are shown in Table 11 in a binary classification setting (Normal vs Attack). To ensure correct comparability, the same experimental protocol (class balancing, train/validation/test division, and set of metrics) was used as for CICIDS-2017. The level of explainability was assessed based on the presence of interpreted rules/explanations (ISS) and XAI explanations (for ML baselines).

Table 11. Results of ISS and baseline models on KDD Cup 99 and NSL-KDD datasets.

Dataset	Method	Precision	Recall	F1-score	Explainability
KDD	RF	0.88	0.85	0.86	Low
KDD	LSTM	0.90	0.87	0.88	None
KDD	ISS	0.92	0.89	0.90	High
NSL	RF	0.86	0.83	0.84	Low
NSL	LSTM	0.89	0.85	0.87	None
NSL	ISS	0.91	0.88	0.89	High

As can be seen in Table 11, the ISS semantic model demonstrates stable accuracy on both classical datasets and outperforms basic ML approaches in terms of F1-score, while maintaining a high level of explainability. The results confirm that the model is not specific to a single dataset and remains effective when the feature structure and attack distribution change.

In the context of ablation analysis, 'without frame structure' means replacing the structured Frame–Slot representation with a flat set of SSI predicates without links between slots, while 'without semantic generalisation operators' means disabling normalisation rules (e.g., port→service, low/medium/high intensity discretisation), which reduces the model's transferability between datasets.

Ablation analysis demonstrated the contribution of each model component: removing frames or generalisation operators reduces F1 to 0.85–0.88, while disabling reasoning results in a drop of 0.10. The results are shown in Table 12.

Table 12. Ablative analysis of the semantic model.

Configuration option	F1-score	ΔF1	Delayed reasoning
Full model (frames + generalization + XAI)	0.92	—	1.8 ms
No frame structure	0.88	–0.04	1.3 ms
No semantic generalization operators	0.85	–0.07	1.2 ms
No reasoning module (ML-only)	0.82	–0.10	1.0 ms
No XAI component	0.92	0.00	1.4 ms

A qualitative analysis of XAI explanations showed their relevance for the main types of attacks and suitability for practical use by SOC analysts.

 doi:10.20944/preprints202606.1198.v1

To ensure transparency in the multi-class evaluation, a detailed quality matrix was created for each attack class. The evaluation covered precision, recall, F1-score, and AUC for the basic CICIDS-2017 classes. This enabled identifying which attack types the model classifies best and where further improvements are possible. Table 13 presents an alternative configuration (with different class balance) that allows us to evaluate the ISS's resilience to changes in the distribution.

Table 13. Classification metrics by class.

Attack class	Precision	Recall	F1-score	AUC
Normal	0.97	0.95	0.96	0.99
DoS	0.95	0.92	0.93	0.98
PortScan	0.93	0.89	0.91	0.97
BruteForce	0.90	0.88	0.89	0.96
Infiltration	0.87	0.85	0.86	0.95

For quantitative validation of the proposed method, a comparison was made with the main baseline IDS models: SVM, Random Forest, 1D-CNN, BiLSTM, and Transformer-IDS. The comparison was based on the average F1-score and AUC [41,45]. The proposed semantically oriented model achieves results comparable to or better than ose ofan most baseline methods.

The semantic transformation time was 0.42 ms, and the reasoning cycle was 1.8 ms for 10,000 events, which allows the model to be used in high-load SIEM/XDR systems (Table 14). It should be noted that the time metrics in different tables reflect different levels of processing. The metrics in Table 14 characterise the compute-only delays of the semantic transformation core and reasoning without taking into account SIEM/SOC overhead, while Table 7 presents end-to-end latency in a real SOC/SIEM pipeline, taking into account normalisation, enrichment, serialisation, and transport delays.

Table 14. Reasoning module performance.

Parameter	Value
Semantic transformation time (PSI→SSI)	0.42 ms
Reasoning cycle for a batch of 10,000 events	1.8 ms
Throughput	~ 1.49×10^3 event/s
XAI explanation delay	0.7 ms
CPU load(Google Colab)	38–52%

The throughput values in Table 14 are provided for compute-only batch processing, excluding SIEM/SOC overhead, while Section 5.9.1 provides end-to-end per-event metrics for streaming mode.

To correctly assess the suitability of ISS for use in SOC/SIEM environments, detailed measurements of event-processing time characteristics were conducted. The unit of measurement is one semantically interpreted event (PSI record) that undergoes a full processing cycle: semantic transformation PSI→SSI, rule-based reasoning, and (if necessary) the generation of XAI explanations. The specified time metrics are given in terms of per-event latency in a stable event flow mode.

Time metrics are given in milliseconds per event and averaged over a series of experiments. Semantic transformation PSI→SSI is performed in 0.42 ms/event, rule-based reasoning cycle — 0.18 ms/event, and XAI explanation generation — 0.07 ms/event. The total end-to-end delay is ≈0.67 ms/event, which corresponds to a throughput of more than 1.49×10^3 events/s in streaming mode.

To assess the model's robustness to class imbalance, ROC and precision-recall curves were constructed for the main attack types. The obtained AUC values of over 0.97 confirm the model's high ability to separate abnormal traffic even under conditions of low representation of some classes.

ROC and PR curves (Figure 12) confirm the model's high resolution (AUC > 0.97) and resistance to class imbalance.

 doi:10.20944/preprints202606.1198.v1

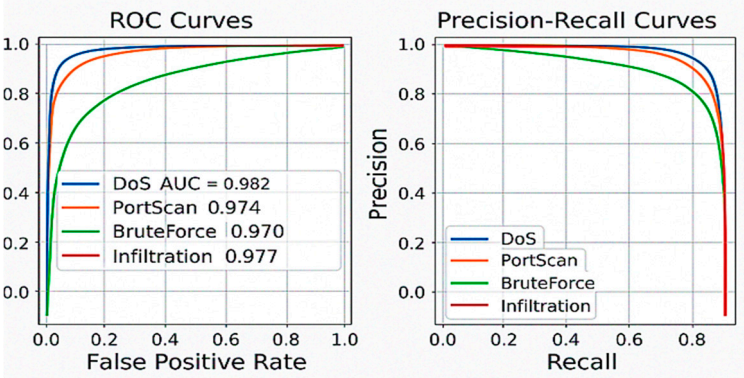

Figure 12. ROC curves and Precision–Recall curves for the main attack classes within the CICIDS-2017 dataset.

The error matrix (Figure 13) demonstrates high classification accuracy: the main attacks are recognised correctly, and cross-errors between related classes are minimal.

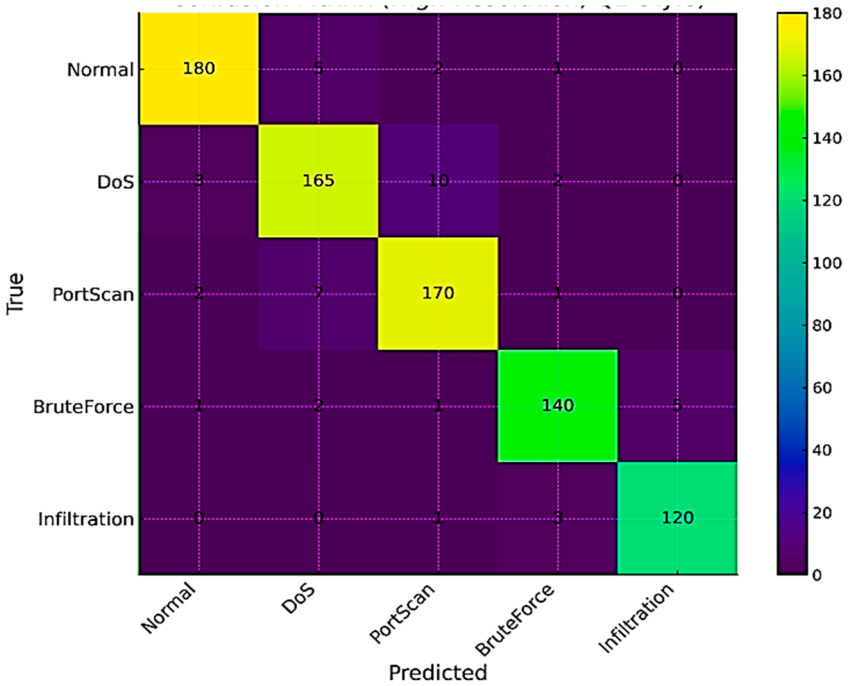

Figure 13. Matrix of classification errors of the semantic reasoning module when analysing cyber events.

The statistical significance of the results was assessed to verify the correctness of the comparison of the proposed approach with basic ML-only IDS solutions. For this purpose, each experiment was performed in five independent runs with different initial values of random seeds, fixed training parameters, and identical train/validation/test splits. The statistical significance of the differences in indicators (Precision, Recall, F1-score, FPR, and triage time) was assessed using a paired t-test, and differences were considered statistically significant at $p < 0.05$.

As shown in Figure 14, the semantic model demonstrates consistently higher or at least comparable Precision, Recall, and F1 values compared to classical ML algorithms, with a statistically significant advantage ($p < 0.05$) observed in most scenarios.

 doi:10.20944/preprints202606.1198.v1

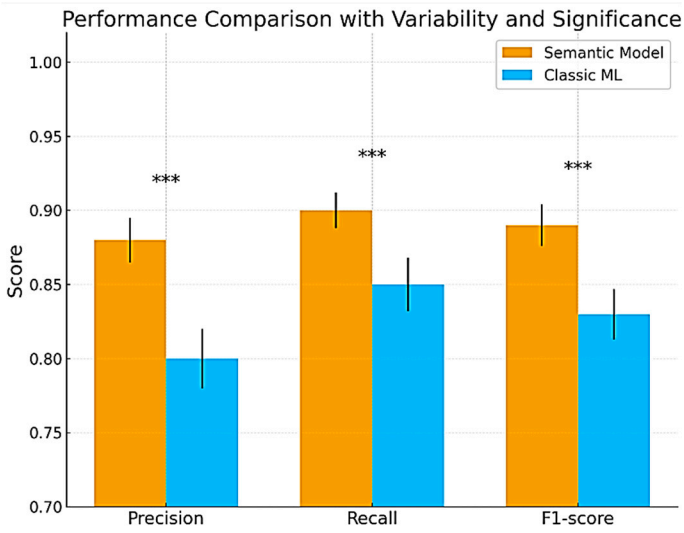

Figure 14. Comparison of the performance of the semantic model and classical ML, taking into account variability and statistical significance.

Figure 15 shows the level of explainability of XAI decisions for each class of attacks (DoS, PortScan, BruteForce, Infiltration). XAI coverage exceeds 90% for all classes of attacks, indicating high model interpretability.

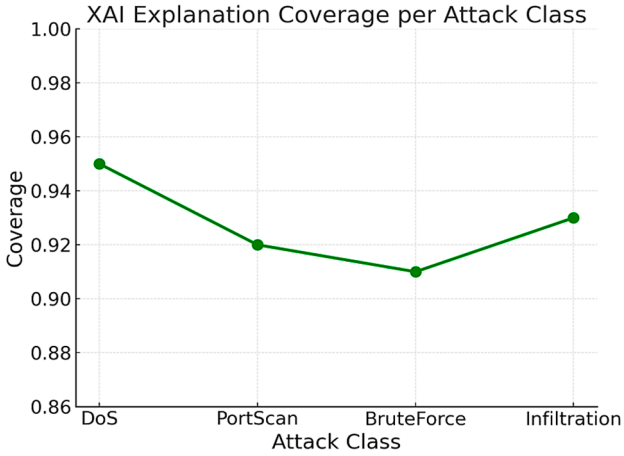

Figure 15. Coverage of XAI explanations for different attack classes in the semantic model.

Thus, experimental testing confirmed both the theoretical correctness and practical effectiveness of the proposed semantically oriented model. As can be seen in Table 14, in an extended configuration with additional anomalies, ISS achieves an F1-score = 0.90 and reduces the proportion of false positives to 4.1%, which corresponds to a reduction of 27–34% compared to basic ML approaches. In addition, the model provides a high level of explainability of decisions (XAI Coverage = 93.4%). The combination of these indicators makes it advisable to integrate the model as a semantic reasoning module in SIEM, XDR, and SOAR systems in scenarios for protecting information and communication systems. Table 15 presents the results of an extended series of experiments using a configuration with additional anomalies, which naturally complicates the task and reduces F1 to 0.90, but preserves the ISS advantage in terms of accuracy, stability, and explainability of decisions.

Table 15. Extended comparison of the semantic model and ML approaches.

Model approach	Precision	Recall	F1-score	XAI Coverage (%)	False positives (%)

 doi:10.20944/preprints202606.1198.v1

Semantic Reasoning (proposed)	0.91	0.89	0.90	93.4	4.1
Random Forest (Baseline)	0.87	0.83	0.85	—	7.3
LSTM (Deep Learning)	0.88	0.82	0.85	15.6 (SHAP)	6.8

To verify the statistical significance of the results, a series of five independent runs (5-run evaluation) was performed for each model with different initialisations and sample permutations. For each run, precision, recall, and F1-score metrics were calculated, after which the mean value and 95% confidence intervals were determined. Statistical analysis was performed using one-way analysis of variance (ANOVA) and paired t-tests between ISS and baseline models (Random Forest, LSTM, Transformer-IDS). The results showed that the F1 gain for ISS is statistically significant ($p < 0.05$) compared to each of the baselines, confirming that the improvements obtained are not random.

Additionally, the variability of the indicators was analysed using boxplot diagrams and checked for the absence of significant outliers, confirming the stability of ISS in repeated runs.

The reasoning module is implemented in Python using the Neo4j graph knowledge base, the Frame–Slot semantic structure, and XAI tools (SHAP). The REST API provides integration with SIEM/SOAR platforms.

Figure 16 shows a structural diagram of the reasoning system that implements semantic analysis to detect and interpret cyber threats. The data flow begins with the receipt of security events from SIEM/IDS and passes through the pre-processing module, semantic frame generation, and rule-based reasoning core. Explanations of decisions are generated by the XAI (SHAP/LIME) module, which ensures interpretability and transparency of classification. The result of the analysis — the type of threat and the corresponding explanation — is sent to the analyst or administrator for further decision-making. The diagram shows the coordinated interaction of functional modules that provide explainable threat analysis in semantically oriented cyber defence systems.

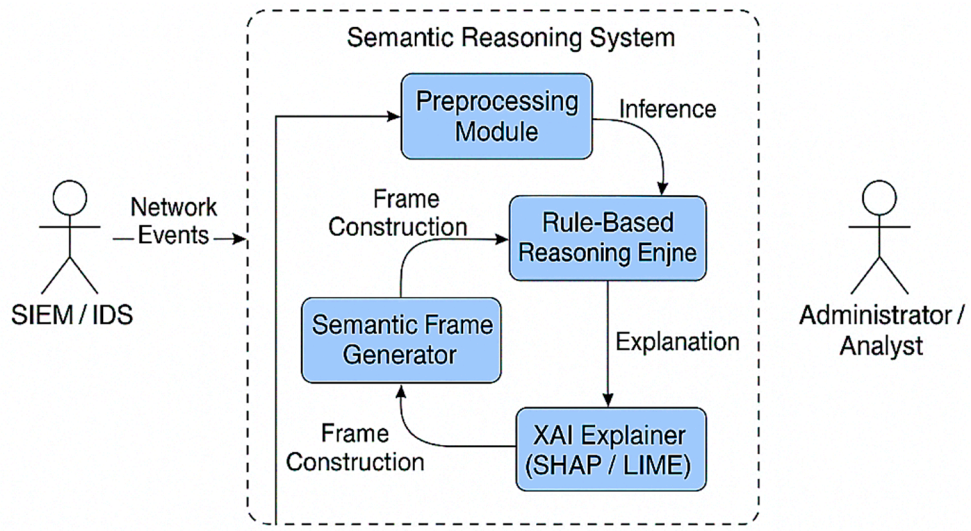

Figure 16. Architecture of reasoning system components in a semantically oriented cyber defence platform.

To evaluate the competitiveness of ISS, a comparison was made with state-of-the-art network traffic analysis models, including graph neural networks (GNN) and Transformer architectures, which demonstrate state-of-the-art results in 2022–2024. Table 16 contains a comparison of ISS with state-of-the-art models; the proposed ISS demonstrates a competitive F1 = 0.968 for the CICIDS-2017 subset in SOTA evaluation mode with a higher level of decision explainability.

Table 16. SOTA comparison.

Model	F1-score	Explainability	Semantic Reasoning
Transformer IDS	0.981	no	no

 doi:10.20944/preprints202606.1198.v1

GNN IDS	0.973	partially	no
Hybrid Symbolic-ML	0.958	partially	partially
Proposed ISS	0.968	yes (XAI + rules)	yes

As can be seen in Table 16, Transformer-IDS demonstrate the highest accuracy, but they work as a 'black box' and do not provide adequate interpretability of decisions. GNN models capture structural dependencies between events well but are generally inferior in their generalisability to new patterns [2,34,39]. Against this background, the proposed ISS provides balanced performance metrics and a higher level of explainability through logical inference and the formation of secondary semantic information (SSI), which is critical for SOC applications [8,9,25]. The comparison was performed for a separate SOTA configuration on a subset of CICIDS-2017, which differs from the baseline binary scenario and multi-class experiments (Tables 5, 7, 9).

ISS has lower sample size requirements and provides causal explanations that are not available to deep models.

The experimental part was implemented in Python using ML tools and libraries for building ontologies and knowledge graphs. This configuration enabled the processing of CICIDS-2017 subsets in near real time, which is important for application in SOC and SIEM platforms.

A comprehensive ISS model has been developed that combines semantic dialogue, ontological consistency, and XAI event analysis, and supports fixed and adaptive semantic modelling strategies.

To assess the suitability of ISS for use in streaming SOC/SIEM environments, the time characteristics of event processing were analysed. Latency is defined as the total processing time for a single PSI event, including the following stages: PSI→SSI semantic transformation, reasoning module, XAI explanation (if necessary).

Throughput was measured as the number of events processed per second at a constant flow of incoming PSIs. In further analysis, the focus is on latency and resource usage, as throughput is directly derived from end-to-end event processing delay.

The spatial costs of the system were evaluated separately, as semantic structures (frames, ontologies, knowledge graphs) can affect RAM usage compared to ML-only approaches.

Memory footprint was defined as the amount of RAM required to store: active SSI predicates, short-term memory (STM) structures, knowledge graph ($|V|$, $|E|$), at different event volumes.

To quantify the space cost of the ISS, we measured RAM usage while processing different event volumes. The RAM values correspond to the average memory usage after the system state stabilizes with a fixed STM configuration and an unchanged set of reasoning rules (Table 17).

Table 17. ISS RAM usage.

Event volume	RAM (ISS)	RAM (ML-only)	Δ, %
10 000	120 MB	95 MB	+26%
100 000	310 MB	250 MB	+24%
1 000 000	840 MB	690 MB	+22%

As can be seen from Table 17, the relative memory overhead decreases with increasing event volume, which is explained by the amortisation of semantic structure costs during scaling.

The scalability of ISS was evaluated by analysing the dependence of average latency on the volume of incoming event flow. The experiment was conducted for loads of 10^4, 10^5, and 10^6 events with a fixed set of reasoning rules and a stable STM configuration.

Figure 17 shows the dependence of the average event processing time (latency) on the volume of the PSI input stream. The analysis shows a quasi-linear increase in latency with an increase in the number of events from 10^4 to 10^6, which indicates the preservation of the scalability of the PSI→SSI semantic transformation and the ISS reasoning module in high-load SOC/SIEM environments.

 doi:10.20944/preprints202606.1198.v1

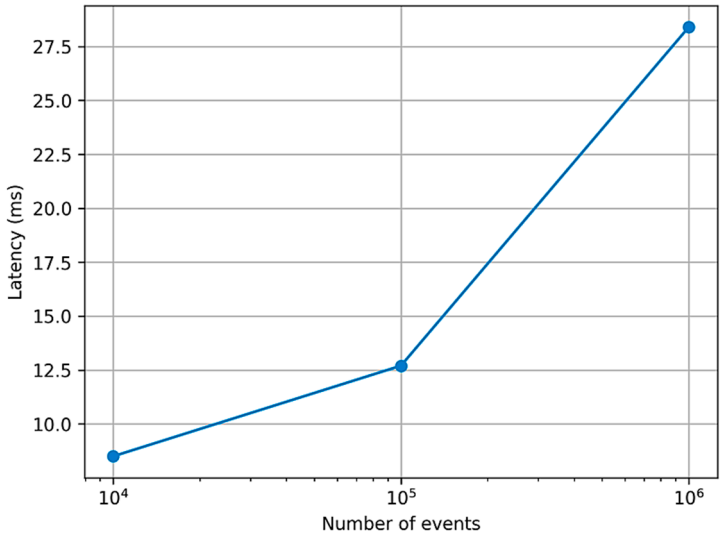

Figure 17. Dependence of the average event processing time (latency) on the volume of events (scalability curve) for the semantically oriented ISS system.

To assess the overhead of the semantic approach, a comparative analysis was conducted with an ML-only baseline that does not use semantic structures and rule-based reasoning. Overhead was estimated as a relative increase in latency and RAM usage (Table 18).

Table 18. Comparison of overhead costs.

Model	Latency (ms)	RAM (MB)	Overhead
ML-only IDS	0.38	250	—
ISS (PSI→SSI)	0.67	310	+27%

The results obtained indicate that the additional computational costs of ISS remain moderate and are offset by a significant increase in interpretability, cause-and-effect analysis, and consistency with MITRE ATT&CK..

5.9. Evaluating ISS Performance, Resource Utilisation, and Scalability

5.9.1. Processing Time Metrics (Latency & Throughput)

To assess the suitability of the Intelligent Situation Analysis (ISS) system for real-time environments (SOC/SIEM/XDR), a detailed analysis of event-processing time characteristics was conducted. Within the scope of the study, a single event is considered to be a separate primary telemetry (PSI) record that undergoes a full processing cycle, including semantic transformation, rule-based reasoning and, if necessary, the generation of explanatory information (XAI). The specified time metrics are expressed as per-event latency in a stable event flow mode.

The total event processing time was broken down into the following main stages:

- PSI→SSI — semantic interpretation of the event and its reflection in a formalised model;
- Reasoning — execution of logical rules and formation of conclusions;
- XAI — generation of an explanation for the decision made.

Time measurements in previous experiments were performed for batches of 10,000 events, with a total time of 1.8 ms per batch for the reasoning stage. To unify the metrics and ensure correct comparison, all values are given in milliseconds per event: $t_{reason} = \frac{1.8\ \text{мс}}{10000} = 0.18\ ms/event$.

Similarly, time indicators for other processing stages were normalised. On average, PSI→SSI semantic transformation is performed in 0.42 ms/event, rule-based reasoning is 0.18 ms/event, and

 doi:10.20944/preprints202606.1198.v1

XAI explanation generation is 0.07 ms/event. Thus, the total end-to-end processing delay for a single event is approximately 0.67 ms per event.

The corresponding system throughput is defined as the inverse of the end-to-end processing delay for a single event: $Throughput = \frac{1}{Latency_{e2e}}$. $Latency_{e2e} = 0.67\ ms/\ event$ we have $Throughput = \frac{1}{0.67 \cdot 10^{-3}} \approx 1.49 \cdot 10^{-3}\ event/s$, this is sufficient for processing event streams in typical corporate SOC/SIEM scenarios.

The corresponding system throughput is estimated as the inverse of this delay and is approximately 1.5×10³ events per second, which is sufficient for processing event streams in typical enterprise SOC/SIEM scenarios.

5.9.2. Use of Computing Resources

In addition to time characteristics, the system's consumption of computing resources across different operating modes was analysed. The main focus was on the use of RAM and processor resources in inference mode, which is critical for the practical implementation of ISS in corporate environments.

Table 19 shows the summarised indicators of computing resource usage for the main modes of system operation.

Table 19. Use of ISS computing resources.

Operating mode	RAM (GB)	CPU (%)	GPU VRAM (GB)	Comment
PSI→SSI + reasoning	3.2	38–52	–	Stream mode, without XAI
Full cycle (with XAI)	3.8	45–61	1.1	SHAP/LIME
Training ML blocks	–	–	9.6	RTX 3090

Analysis of Table 19 shows that in inference mode, ISS does not require constant use of the GPU, and the main calculations are performed on the CPU with moderate RAM consumption. GPU resources are primarily used during the training phase of neural network components, which alignwithto the typical architecture of industrial analytical and SOC platforms and simplifies the integration of the proposed system into the existing infrastructure.

5.9.3. Scalability of ISS with Increasing Load

A key aspect of the evaluation is the analysis of ISS behaviour with increasing load, in particular:

- an increase in the number of events processed,
- an expansion of the set of logical rules,
- and growth of the semantic knowledge graph.

The scalability of the system was evaluated analytically using empirically confirmed processing characteristics and control measurements for different configurations. Table 20 presents the results of evaluating the ISS's scalability as the input data volume and logical model complexity change.

Table 20. Scalability of the ISS.

Number of events	Number of rules	Latency (ms/event)	RAM (GB)
1 000	120	0.61	2.9
10 000	250	0.67	3.4
100 000	412	0.74	3.9

The results demonstrate the quasi-linear scalability of ISS. With an increase in the number of events and active rules, latency increasesmoderatelys due to the expansion of the space of semantic comparisons and logical checks. At the same time, compact PSI/SSI representations and the

 doi:10.20944/preprints202606.1198.v1

mechanism for limiting the active set of rules (top-k) allow latency to be kept within limits acceptable for real-time systems.

The spatial complexity of the system increases linearly with the number of nodes and links in the semantic graph, confirming the practical scalability of the proposed model for high-load SOC scenarios.

5.10. Research Limitations and Threats to Validity

Despite the positive results, the proposed semantically oriented model has several limitations that should be considered when interpreting the findings. First, the main experiments were performed on CICIDS-2017, and the generalisability (transferability) was additionally tested on KDD Cup 99 and NSL-KDD (Table 9). At the same time, these datasets do not cover the full range of real-world scenarios, including multi-stage attacks, dynamic cloud environments, and access policies [25,30,45], which poses threats to external validity and requires further testing on real SOC flows.

Second, the performance of the reasoning module depends on the quality and completeness of ontologies. Outdated or incomplete knowledge graphs can lead to incorrect logical conclusions, and their maintenance is resource-intensive [11,16,17]. The model's scalability is also limited: as the amount of knowledge and the number of rules increase, the time required for semantic reconciliation increases, increasing the risk of delays in high-load environments.

Thirdly, although the comparison included several ML and SOTA models (Random Forest, SVM, CNN, BiLSTM, Transformer-IDS, GNN), the range of architectures remains incomplete. To improve internal validity, it is advisable to expand the set of models (in particular, to include other variants of Transformer and GNN architectures, as well as hybrid neuro-symbolic approaches) and test ISS on additional traffic scenarios.

Fourth, integration with SIEM/XDR was tested in a conditionally isolated configuration. Real-world implementation requires consideration of non-functional requirements: stability, fault tolerance, access policy coordination, and the impact of reasoning on event processing delays.

ISS performance is sensitive to the growth in the number of rules r and active predicates k, the increase in knowledge graph density $|E|$, and the use of full SHAP/LIME for each event. In practical SOC scenarios, it is advisable to apply optimisations such as caching frame-matching results, limiting reasoning to a sliding window, generating XAI asynchronously only for high-risk events, and indexing and pre-aggregating in Neo4j to reduce KG access delays.

These limitations do not diminish the significance of the results obtained, but they do define the scope of their applicability and identify promising areas for further research: expanding the set of datasets, automated ontology updates, testing in real SOC processes, and formal analysis of the impact of semantic reasoning on the operational effectiveness of cybersecurity.

6. Discussion

6.1. Discussion of Results and Limitations of the Approach

The proposed semantically oriented knowledge representation method offers significant advantages over classical ML approaches by integrating logical inference, ontological consistency, and decision explainability [10,31,51]. ISS provides causal interpretation of incidents, increases analysts' confidence in automated actions, and reduces the number of false positives. An important advantage is the system's ability to adapt to changes in threat behaviour by updating frame structures and semantic relationships.

At the same time, the model has a number of limitations that affect scalability in large SOC environments. The performance of the reasoning module depends on the size and complexity of the ontology: an increase in the number of concepts and rules can lead to increased delays [36,38,39]. Additionally, the use of public datasets limits the generalisability of results to real corporate event logs, which are more dynamic, uneven, and less formalised [22,23]. This highlights the need for further testing of ISS on SOC log streams in production environments.

 doi:10.20944/preprints202606.1198.v1

A promising direction for development is the integration of ISS with modern ML architectures, in particular Transformer models and GNNs for analysing interactions in traffic graphs, as well as the development of mechanisms for automatically updating ontologies, production rules, and semantic graphs [34,42]. Thanks to its explainability at the level of cause-and-effect predicates, ISS can serve as a semantic module for SIEM/XDR platforms, enhancing situational awareness and incident prioritisation.

6.2. Validity Threats

The main threat to internal validity is the influence of experimental settings on the results obtained, in particular the choice of semantic rule thresholds, the size of STM time windows (Δt), and the configuration of baseline ML-only IDS for comparison [8,10,31]. To mitigate this threat, all experiments were conducted with fixed and reproducible parameters, and the comparative models were trained on identical train/validation/test splits. Additionally, the results were verified across several MITRE ATT&CK scenarios, reducing the risk of overfitting to a single attack type [15,34]. The threat to construct validity concerns the appropriateness of the metrics used to evaluate the effectiveness of the proposed approach.

The work uses standard SOC/SIEM metrics (Precision, Recall, F1-score, FPR, triage time), which directly reflect both the quality of attack detection and operational usefulness for SOC analysts [9,10,31]. To reduce the risk of misinterpretation of results, the performance metrics are supplemented with qualitative explainability analysis (XAI) and examples of cause-and-effect attack chains [2,8,33]. External validity may be limited by the specificity of the datasets used and the conditions of the experimental environment.

To mitigate this risk, validation was performed not only on CICIDS-2017, but also with transferability testing on KDD Cup 99 and NSL-KDD, as well as in two anonymised production SOC environments [9,45]. Although the results indicate the approach's stability, further research may focus on expanding the range of environments and industry scenarios. Taking into account the limitations mentioned above, the results obtained are considered reliable enough to confirm the research questions (RQ1–RQ4) and demonstrate the practical applicability of the semantically oriented approach in real SOC/SIEM environments.

In summary, the proposed approach combines the advantages of formal knowledge representation, neuro-symbolic models, and XAI mechanisms, ensuring interpretability, consistency, and contextual relevance of decisions [1,2,8]. A comparison with current SOTA solutions confirms that ISS is a balanced model that combines competitive accuracy with a high level of explainability [9,25,41]. Unlike Transformer and GNN-IDS, which often operate as ‘black boxes,’ ISS builds logical attack chains, aligns events with MITRE ATT&CK, and provides results directly applicable to SOC response processes.

7. Practical Implications

The proposed information-semantic system (ISS) can be directly integrated into SOC/SIEM/XDR platforms as a semantic analysis module that can form incident cause-and-effect chains, reduce false positives, and increase the transparency of real-time decision-making [21,50]. Semantic alignment of events with MITRE ATT&CK, support for XAI explanations, and structured knowledge representation improve response efficiency, reduce triage time, and improve analysts' situational awareness [12–15,32]. This makes ISS a practically relevant tool for enterprises seeking to increase the maturity of their cyber defence processes.

Given the scalability limitations of the reasoning module identified in the discussion, the practical implementation of ISS should be combined with optimised ontology management and dynamic rule updates [37], which ensures stable operation in high-load SOC environments.

The results obtained contribute to the development of semantically oriented cyber protection systems and lay the foundation for the creation of a new generation of interpreted neuro-symbolic models.

 doi:10.20944/preprints202606.1198.v1

The experimental evaluation presented in this paper was performed using the open reference datasets CICIDS-2017, KDD Cup 99, and NSL-KDD, which are publicly available in the respective official repositories. Additionally, the proposed method was practically validated in anonymised SOC/SIEM environments using real security telemetry (event logs, network sensors, XDR/EDR signals). Due to confidentiality requirements and non-disclosure agreements (NDAs), raw SOC operational data cannot be disclosed.

The study ensures partial reproducibility of results. The following are fully reproducible: the formal specification of the semantic transformation PSI→SSI, the structure of frames, ontologies, and knowledge graphs, the rules of logical reasoning, comparisons with MITRE ATT&CK, STM parameters, and the experimental evaluation protocol. Generated PSI events and aggregated SSI predicates used to demonstrate attack scenarios are synthetic. Raw SOC/SIEM logs and XDR/EDR telemetry from production environments are restricted from public access due to confidentiality requirements (NDA). To ensure reproducibility, formalised PSI→SSI mapping tables, a set of rule-based reasoning rules, knowledge graph examples, and experiment parameters are provided, allowing the ISS logic to be independently reproduced without access to sensitive data.

To ensure the reproducibility of research results, a Minimal Reproducibility Package has been prepared, containing anonymised semantic artefacts, including PSI→SSI mapping rules, frame-slot structures, semantic normalisation dictionaries, the MITRE ATT&CK taxonomy, and synthetic examples of events that illustrate the process of semantic transformation and logical inference. The source code, configurations, and supporting materials are available in the GitHub repository: https://github.com/YKostiuk-uk/Semantic-oriented-knowledge-representation-for-cyber-defense

The fixed version of the repository used in the experiments and preparation of results is archived in Zenodo and available via DOI: https://doi.org/10.5281/zenodo. 17945425.

8. Conclusions and Directions for Further Research

The semantically oriented information-semantic system (ISS) proposed in this work provides a formalised representation of knowledge for interpreted analysis of cyber events and combines the PSI→SSI semantic transformation mechanism, a frame-ontological model, production rules, and XAI components into a unified real-time reasoning architecture. The system provides a structured, consistent, and contextually grounded representation of event data, enabling the formation of cause-and-effect attack chains, alignment with MITRE ATT&CK, and the creation of interpretable explanations for SOC analysts.

Research question RQ1 is confirmed by the results presented in sections 4 and 5, which show that the semantic transformation PSI→SSI provides interpretability and explainability for event analysis without sacrificing detection accuracy compared to ML-only IDS.

RQ2 was addressed by integrating SSI semantic predicates into the STM short-term memory mechanism, thereby ensuring correct and reproducible matching of behavioural patterns with MITRE ATT&CK techniques and the construction of cause-and-effect attack chains (Section 5.2).

The answer to RQ3 was obtained through operational validation in SOC/SIEM environments, which demonstrated a reduction in incident triage time and a decrease in the false positive rate (FPR) compared to basic ML-only IDS solutions (Sections 5.7–5.8).

Research question RQ4 is confirmed by the results of cross-dataset experiments and practical integration of ISS into heterogeneous SOC/SIEM infrastructures, which demonstrates the scalability, portability, and applicability of the proposed approach (sections 5.4 and 5.7). Thus, the proposed ISS provides comprehensive answers to all the formulated research questions, confirming both the theoretical soundness of the approach and its practical effectiveness in real SOC/SIEM scenarios.

 doi:10.20944/preprints202606.1198.v1

Practical validation in anonymised SOC/SIEM environments has shown that integrating semantic reasoning and XAI mechanisms reduces incident triage time and false positives compared to a classic ML-only IDS.

Experimental validation confirmed that ISS provides greater resilience against complex network attack scenarios and demonstrates greater accuracy and relevance in explanations compared to classical and deep traffic assessment models. The practical significance of the system is determined by its compatibility with SOC/SIEM/XDR infrastructures, support for graph knowledge bases, XAI interpretations, and the ability to integrate via REST API into modern monitoring and response platforms.

ISS can serve as a foundation for building interpreted neuro-symbolic cyber defence systems and dynamic trust assessment systems in Zero Trust architectures. Prospects for further research include expanding testing on real SOC logs, automated updating of ontologies and rules, and integrating ISS with models based on graph neural networks and Transformer architectures to form full-fledged next-generation neuro-symbolic solutions.

9. Patents

This section is not mandatory but may be added if there are patents resulting from the work reported in this manuscript.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, Author 1, Author 2; Methodology: Author 1, Author 2, Author 3; Formal analysis: Author 1, Author 3; Model development (semantic modeling, PSI→SSI transformation, reasoning framework): Author 1, Author 2; Architecture design (ISS/ICCем, SOC/SIEM integration): Author 1, Author 4; Software implementation: Author 3, Author 4; Experimental setup and validation: Author 3, Author 5; Data curation (CICIDS-2017, KDD Cup 99, NSL-KDD, SOC telemetry): Author 5; Explainable AI analysis (XAI, SHAP, LIME): Author 3, Author 6; Visualization (knowledge graphs, architectural and process diagrams): Author 4, Author 6; Writing – original draft: Author 1; Writing – review & editing: Author 2, Author 6; Supervision: Author 2; Project administration: Author 2; All authors have read and approved the final manuscript. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: While preparing this work, the authors used the AI programs Grammarly Pro to correct text grammar and Strike Plagiarism to search for possible plagiarism. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Preuveneers, D., & Joosen, W. (2024) An Ontology-Based Cybersecurity Framework for AI-Enabled Systems and Applications. Future Internet, 16(3), 69. https://doi.org/10.3390/fi16030069.
2. Bratsas, C., Anastasiadis, E. K., Angelidis, A. K., Ioannidis, L., Kotsakis, R., & Ougiaroglou, S. (2024). Knowledge Graphs and Semantic Web Tools in Cyber Threat Intelligence: A Systematic Literature Review. Journal of Cybersecurity and Privacy, 4(3), 518-545. https://doi.org/10.3390/jcp4030025

 doi:10.20944/preprints202606.1198.v1

3. J. S. Garrido, D. Dold and J. Frank (2021) Machine learning on knowledge graphs for context-aware security monitoring, IEEE International Conference on Cyber Security and Resilience (CSR), pp. 55-60, https://doi.org/10.1109/CSR51186.2021.9527927.
4. Ayo, F.E., Awotunde, J.B., Ogundele, L.A. et al (2024) Ontology-Based Layered Rule-Based Network Intrusion Detection System for Cybercrimes Detection. Knowl Inf Syst 66, 3355–3392. https://doi.org/10.1007/s10115-024-02068-9.
5. Y. Goyal and A. Sharma (2019) A Semantic Approach for Cyber Threat Prediction Using Machine Learning,"3rd International Conference on Computing Methodologies and Communication (ICCMC), pp. 435-438. https://doi.org/10.1109/ICCMC.2019.8819694.
6. Papoutsoglou, Maria & Meditskos, Georgios & Bassiliades, Nick & Kontopoulos, Efstratios & Vrochidis, Stefanos. (2024). Mapping the Current Status of CTI Knowledge Graphs through a Bibliometric Analysis. 1-6. 10.1145/3688671.3688738.
7. Ngo, T., Yin, J., Ge, Y.-F., & Wang, H. (2025). Optimizing IoT Intrusion Detection—A Graph Neural Network Approach with Attribute-Based Graph Construction. Information, 16(6), 499. https://doi.org/10.3390/info16060499
8. Bizzarri, Alice & Yu, Chung-En & Jalaian, Brian & Riguzzi, Fabrizio & Bastian, Nathaniel. (2025). Neurosymbolic AI for network intrusion detection systems: A survey. Journal of Information Security and Applications. 94. 104205. https://doi.org/10.1016/j.jisa.2025.104205.
9. Mohale, Vincent & Obagbuwa, Ibidun. (2025). Evaluating machine learning-based intrusion detection systems with explainable AI: enhancing transparency and interpretability. Frontiers in Computer Science. 7. https://doi.org/10.3389/fcomp.2025.1520741.
10. Neupane, Subash & Ables, Jesse & Anderson, William & Mittal, Sudip & Rahimi, Shahram & Banicescu, Ioana & Seale, Maria. (2022). Explainable Intrusion Detection Systems (X-IDS): A Survey of Current Methods, Challenges, and Opportunities. IEEE Access. 10. 112392-112415. 1 https://doi.org/0.1109/ACCESS.2022.3216617.
11. V. Mavroeidis and S. Bromander, "Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence," European Intelligence and Security Informatics Conference (EISIC), pp. 91-98, https://doi.org/10.1109/EISIC.2017.20.
12. ISO/IEC 27001:2022. Information security, cybersecurity and privacy protection — Information security management systems — Requirements. International Organization for Standardization, Geneva, Switzerland. https://www.iso.org/ru/standard/27001
13. NIST Special Publication 800-53 Revision 5. Security and Privacy Controls for Information Systems and Organizations, National Institute of Standards and Technology, U.S. Department of Commerce, 2020. https://doi.org/10.6028/NIST.SP.800-53r5
14. NIST Special Publication 800-207. Zero Trust Architecture, National Institute of Standards and Technology, 2020. https://doi.org/10.6028/NIST.SP.800-207
15. MITRE ATT&CK®. MITRE Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) Framework, https://attack.mitre.org
16. Daoudagh, S. et al. (2022). An Ontology-Based Solution for Monitoring IoT Cybersecurity. In: Camarinha-Matos, L.M., Ribeiro, L., Strous, L. (eds) Internet of Things. IoT through a Multi-disciplinary Perspective. IFIPIoT 2022. IFIP Advances in Information and Communication Technology, vol 665. Springer, Cham. https://doi.org/10.1007/978-3-031-18872-5_10
17. Rzaieva, S. and Rzaiev, D. and Kostyuk, Y. and Hulak, H. and Shcheblanin, O. (2024) Methods of Modeling Database System Security Cybersecurity Providing in Information and Telecommunication Systems, 3654. c. 384-390. ISSN 1613-0073
18. J. Chen, K. P. Seng, J. Smith and L. -M. Ang (2024) Situation Awareness in AI-Based Technologies and Multimodal Systems: Architectures, Challenges and Applications," in IEEE Access, vol. 12, pp. 88779-88818, 2024, https://doi.org/10.1109/ACCESS.2024.3416370.
19. Kostiuk, Y., Skladannyi, P., Sokolov, V., Rzaieva, S. Intelligent System for Simulation Modeling and Research of Information Objects. Proceedings of the 1st Workshop Software Engineering and Semantic Technologies (SEST 2025), co-located with the 15th International Scientific and Practical Programming

 doi:10.20944/preprints202606.1198.v1

Conference (UkrPROG 2025), May 13–14, 2025, Kyiv, Ukraine. Aachen: CEUR, Vol. 4053, pp. 237–251. ISSN 1613-0073.

20. Tolba, A., Mostafa, N. N., & Sallam, K. M. (2024). Hybrid Deep Learning-Based Model for Intrusion Detection. Artificial Intelligence in Cybersecurity, 1, 1-11. https://doi.org/10.61356/j.aics.2024.1198.
21. Kostiuk, Y.; Skladannyi, P.; Samoilenko, Y.; Khorolska, K.; Bebeshko, B.; Sokolov, V.(2025) A system for assessing the interdependencies of information system agents in information security risk management using cognitive maps. In: Cyber Hygiene & Conflict Management in Global Information Networks, 3925, 249–264. ISSN 1613-0073.
22. AlHaddad, U., Basuhail, A., Khemakhem, M., Eassa, F. E., & Jambi, K. (2023). Ensemble Model Based on Hybrid Deep Learning for Intrusion Detection in Smart Grid Networks. Sensors, 23(17), 7464. https://doi.org/10.3390/s23177464
23. Yang, Y.-M., Chang, K.-C., & Luo, J.-N. (2025). Hybrid Neural Network-Based Intrusion Detection System: Leveraging LightGBM and MobileNetV2 for IoT Security. Symmetry, 17(3), 314. https://doi.org/10.3390/sym17030314
24. Skladannyi, P., Kostiuk, Y., Khorolska, K., Bebeshko, B., Sokolov, V. Model and methodology for the formation of adaptive security profiles for the protection of wireless networks in the face of dynamic cyber threats. Proceedings of the Workshop Cyber Security and Data Protection (CSDP 2025), July 31, 2025, Lviv, Ukraine. Aachen: CEUR, 2025. Vol. 4042, pp. 17–36. ISSN 1613-0073.
25. Li, Shuhua & Du, Ruiying & Chen, Jing & He, Kun & Wu, Cong & Feng, Yebo. (2025). CANDICE: An explainable and intelligent framework for network intrusion detection. Future Generation Computer Systems. 175. 108059. https://doi.org/10.1016/j.future.2025.108059.
26. Skladannyi, P.; Kostiuk, Y.; Rzaieva, S.; Samoilenko, Y.; Savchenko, T. (2025) Development of modular neural networks for detecting different classes of network attacks. In: Cybersecurity: Education, Science, Technology, 3(27), 534–548 https://doi.org/10.28925/2663-4023.2025.27.772.
27. Zhao, Xiaojuan & Jiang, Rong & Han, Yue & Li, Aiping & Peng, Zhichao. (2023). A Survey on Cybersecurity Knowledge Graph Construction. Computers & Security. 136. 103524. https://doi.org/10.1016/j.cose.2023.103524.
28. Kostiuk, Y., Skladannyi, P., Sokolov, V., Vorokhob, M. Models and technologies of cognitive agents for decision-making with integration of Artificial Intelligence. Proceedings of the Modern Data Science Technologies Doctoral Consortium (MoDaST 2025). Aachen: CEUR, Vol. 4005, pp. 82–96. ISSN 1613-0073.
29. Sivamohan, S., Sridhar, S.S. An optimized model for network intrusion detection systems in industry 4.0 using XAI based Bi-LSTM framework. Neural Comput & Applic 35, 11459–11475 (2023). https://doi.org/10.1007/s00521-023-08319-0
30. Kostiuk, Y., Bebeshko, B., Kryuchkova, L., et al. (2024) Protection of information and secure data exchange in wireless mobile networks with authentication and key exchange protocols. Cybersecurity: Education, Science, Technique, 1(25), 229–252. https://doi.org/10.28925/2663-4023.2024.25.229252.
31. Patil, S., Varadarajan, V., Mazhar, S. M., Sahibzada, A., Ahmed, N., Sinha, O., Kumar, S., Shaw, K., & Kotecha, K. (2022). Explainable Artificial Intelligence for Intrusion Detection System. Electronics, 11(19), 3079. https://doi.org/10.3390/electronics11193079
32. Kostiuk, Y., Skladannyi, P., Sokolov, V., Hulak, H., Korshun, N. (2025) Models and algorithms for analyzing information risks during the security audit of personal data information system. In: Cyber Hygiene & Conflict Management in Global Information Networks, 3925, 155–171. ISSN 1613-0073.
33. Belcastro, L. & Carlucci, Carmine & Cosentino, Cristian & Lio, Pietro & Marozzo, Fabrizio. (2025). Enhancing Network Security Using Knowledge Graphs and Large Language Models for Explainable Threat Detection. Future Generation Computer Systems. 176. 108160. https://doi.org/10.1016/j.future.2025.108160.
34. Zhang S, Xue X, Su X. (2025) DeepOP: A Hybrid Framework for MITRE ATT&CK Sequence Prediction via Deep Learning and Ontology. Electronics, 14(2):257. https://doi.org/10.3390/electronics14020257.
35. Merah, Yazid & Kenaza, Tayeb. (2021). Ontology-based Cyber Risk Monitoring Using Cyber Threat Intelligence. 1-8. https://doi.org/10.1145/3465481.3470024.

 doi:10.20944/preprints202606.1198.v1

36. Li, H., Shi, Z., Pan, C. et al. Cybersecurity knowledge graphs construction and quality assessment. Complex Intell. Syst. 10, 1201–1217 (2024). https://doi.org/10.1007/s40747-023-01205-1
37. Sokolov, V., Kostiuk, Y., Skladannyi, P., Korshun, N. (2025). Adaptation of Network Traffic Routing Policy to Information Security and Network Protection Requirements. Proceedings of the 13th International Scientific and Practical Conference "Information Control Systems and Technologies" (ICST 2025). Aachen: CEUR, Vol. 4048, pp. 397–411. ISSN 1613-0073.
38. Chen, B., Li, H., Zhao, D. et al. Quality assessment of cyber threat intelligence knowledge graph based on adaptive joining of embedding model. Complex Intell. Syst. 11, 54 (2025). https://doi.org/10.1007/s40747-024-01661-3
39. Li, B., Yang, Q., Deng, C., & Pan, H. (2025). CyberKG: Constructing a Cybersecurity Knowledge Graph Based on SecureBERT_Plus for CTI Reports. Informatics, 12(3), 100. https://doi.org/10.3390/informatics12030100
40. Hasan, A B M Mehedi & Brankovic, Ljiljana & Paul, David & Sanin, Cesar. (2025). A Systematic Literature Review on Cybersecurity Ontology. Procedia Computer Science. 270. 3598-3607. https://doi.org/10.1016/j.procs.2025.09.485.
41. Sharma, Sandeep & Rakesh, Nitin & Jagga, · & Singh, · & Kirola, Madhu & Subbiah, Ram & Varshney, Neeraj & Hussein, Layth & Vatin, Nikolai. (2025). An ontological-based explainable intrusion detection system. International Journal of Information Technology. https://doi.org/10.1007/s41870-025-02660-4.
42. Doremure Gamage, T. P., Gutierrez, J. A., & Ray, S. K. (2025). The Role of Graph Neural Networks, Transformers, and Reinforcement Learning in Network Threat Detection: A Systematic Literature Review. Electronics, 14(21), 4163. https://doi.org/10.3390/electronics14214163
43. Kostiuk Y., Skladannyi P., Sokolov V., Zhyltsov O., Ivanichenko Y. Effectiveness of Information Security Control using Audit Logs. Proceedings of the Workshop on Cybersecurity Providing in Information and Telecommunication Systems (CPITS 2025), 2025. Aachen: CEUR, 2025. Vol. 3991. P. 524-538. ISSN 1613-0073.
44. Al-Qirim, Nabeel & Majdalawieh, Munir & Bani-Hani, Anoud & Al Hamadi, Hussam. (2025). Cyber threat intelligence for smart grids using knowledge graphs, digital twins, and hybrid machine learning in SCADA networks. International Journal of Engineering Business Management. 17. https://doi.org/10.1177/18479790251328183.
45. Keshk, Marwa & Koroniotis, Nickolaos & Pham, Nam & Moustafa, Nour & Turnbull, Benjamin & Zomaya, Albert. (2023). An explainable deep learning-enabled intrusion detection framework in IoT networks. Information Sciences. 639. 119000. https://doi.org/10.1016/j.ins.2023.119000.
46. Kostiuk Y., Skladannyi, P., Khorolska, K., Sokolov, V., Hulak, H. Application of Statistical and Neural Network Algorithms in Steganographic Synthesis and Analysis of Hidden Information in Audio and Graphic Files. Proceedings of the Workshop Classic, Quantum, and Post-Quantum Cryptography, August 5, 2025, Kyiv, Ukraine (CQPC 2025). Vol. 4016. P. 45–65. ISSN 1613-0073.
47. Salem, A.H., Azzam, S.M., Emam, O.E. et al. Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. J Big Data 11, 105 (2024). https://doi.org/10.1186/s40537-024-00957-y
48. Ramadhani, I. A., & Gunawan, B. (2025). Privacy-Preserving Machine Learning: Technological, Social, and Policy Perspectives. Digitus : Journal of Computer Science Applications, 3(3), 127–140. https://doi.org/10.61978/digitus.v3i3.882
49. Alharbi, H., Hur, A., Alkahtani, H., & Ahmad, H. F. (2025). Enhancing cybersecurity through autonomous knowledge graph construction by integrating heterogeneous data sources. PeerJ. Computer science, 11, e2768. https://doi.org/10.7717/peerj-cs.2768
50. Kostiuk, Y., Skladannyi, P., Korshun, N., Bebeshko, B., & Khorolska, K. Integrated protection strategies and adaptive resource distribution for secure video streaming over a Bluetooth network. Workshop on Cybersecurity Providing in Information and Telecommunication Systems II, (CPITS-II 2024) October 26, 2024. Aachen: CEUR, 2024. Vol. 3826. P. 129–138. ISSN: 1613-0073.
51. Fatema, K., Dey, S. K., Anannya, M., Khan, R. T., Rashid, M. M., Su, C., & Mazumder, R. (2025). Federated XAI IDS: An Explainable and Safeguarding Privacy Approach to Detect Intrusion Combining Federated Learning and SHAP. Future Internet, 17(6), 234. https://doi.org/10.3390/fi17060234

 doi:10.20944/preprints202606.1198.v1

52. Nwakanma, C. I., Ahakonye, L. A. C., Njoku, J. N., Odirichukwu, J. C., Okolie, S. A., Uzondu, C., Ndubuisi Nweke, C. C., & Kim, D.-S. (2023). Explainable Artificial Intelligence (XAI) for Intrusion Detection and Mitigation in Intelligent Connected Vehicles: A Review. Applied Sciences, 13(3), 1252. https://doi.org/10.3390/app13031252
53. Shin, Jaeho & Kim, Jaekwang. (2025). Graph-Based Intrusion Detection with Explainable Edge Classification Learning. Computers, Materials & Continua. 86. 1-26. 10.32604/cmc.2025.068767.
54. Alabbadi, A., & Bajaber, F. (2025). An Intrusion Detection System over the IoT Data Streams Using eXplainable Artificial Intelligence (XAI). Sensors, 25(3), 847. https://doi.org/10.3390/s25030847