Research Papers on 'Cybersecurity'

Google Scholar

Semantic Scholar

IEEE Xplore

PubMed

Paper 1:

A Framework for Evaluating Emerging Cyberattack Capabilities of AI

Date: 2025-03-31

Time: 10:35:02

Authors:

Mikel Rodriguez, Raluca Ada Popa, Four Flynn, Lihao Liang, Allan Dafoe, Anna Wang

Summary:

- As frontier AI models become more capable, evaluating their potential to enable cyberattacks is crucial for ensuring the safe development of Artificial General Intelligence (AGI). Current cyber evaluation efforts are often ad-hoc, lacking systematic analysis of attack phases and guidance on targeted defenses. This work introduces a novel evaluation framework that addresses these limitations by: (1) examining the end-to-end attack chain, (2) identifying gaps in AI threat evaluation, and (3) helping defenders prioritize targeted mitigations and conduct AI-enabled adversary emulation for red teaming. Our approach adapts existing cyberattack chain frameworks for AI systems. We analyzed over 12,000 real-world instances of AI use in cyberattacks catalogued by Google's Threat Intelligence Group. Based on this analysis, we curated seven representative cyberattack chain archetypes and conducted a bottleneck analysis to pinpoint potential AI-driven cost disruptions. Our benchmark comprises 50 new challenges spanning various cyberattack phases. Using this benchmark, we devised targeted cybersecurity model evaluations, report on AI's potential to amplify offensive capabilities across specific attack phases, and offer recommendations for prioritizing defenses. We believe this represents the most comprehensive AI cyber risk evaluation framework published to date.

Click here for more

Paper 2:

What Makes an Evaluation Useful? Common Pitfalls and Best Practices

Date: 2025-03-30

Time: 12:51:47

Authors:

Gil Gekker, Meirav Segal, Dan Lahav, Omer Nevo

Summary:

- Following the rapid increase in Artificial Intelligence (AI) capabilities in recent years, the AI community has voiced concerns regarding possible safety risks. To support decision-making on the safe use and development of AI systems, there is a growing need for high-quality evaluations of dangerous model capabilities. While several attempts to provide such evaluations have been made, a clear definition of what constitutes a "good evaluation" has yet to be agreed upon. In this practitioners' perspective paper, we present a set of best practices for safety evaluations, drawing on prior work in model evaluation and illustrated through cybersecurity examples. We first discuss the steps of the initial thought process, which connects threat modeling to evaluation design. Then, we provide the characteristics and parameters that make an evaluation useful. Finally, we address additional considerations as we move from building specific evaluations to building a full and comprehensive evaluation suite.

Click here for more

Paper 3:

From Content Creation to Citation Inflation: A GenAI Case Study

Date: 2025-03-30

Time: 12:17:26

Authors:

Haitham S. Al-Sinani, Chris J. Mitchell

Summary:

- This paper investigates the presence and impact of questionable, AI-generated academic papers on widely used preprint repositories, with a focus on their role in citation manipulation. Motivated by suspicious patterns observed in publications related to our ongoing research on GenAI-enhanced cybersecurity, we identify clusters of questionable papers and profiles. These papers frequently exhibit minimal technical content, repetitive structure, unverifiable authorship, and mutually reinforcing citation patterns among a recurring set of authors. To assess the feasibility and implications of such practices, we conduct a controlled experiment: generating a fake paper using GenAI, embedding citations to suspected questionable publications, and uploading it to one such repository (ResearchGate). Our findings demonstrate that such papers can bypass platform checks, remain publicly accessible, and contribute to inflating citation metrics like the H-index and i10-index. We present a detailed analysis of the mechanisms involved, highlight systemic weaknesses in content moderation, and offer recommendations for improving platform accountability and preserving academic integrity in the age of GenAI.

Click here for more

Paper 4:

Teams of LLM Agents can Exploit Zero-Day Vulnerabilities

Date: 2025-03-30

Time: 00:26:48

Authors:

Yuxuan Zhu, Antony Kellermann, Akul Gupta, Philip Li, Richard Fang,
Rohan Bindu, Daniel Kang

Summary:

- LLM agents have become increasingly sophisticated, especially in the realm of cybersecurity. Researchers have shown that LLM agents can exploit real-world vulnerabilities when given a description of the vulnerability and toy capture-the-flag problems. However, these agents still perform poorly on real-world vulnerabilities that are unknown to the agent ahead of time (zero-day vulnerabilities). In this work, we show that teams of LLM agents can exploit real-world, zero-day vulnerabilities. Prior agents struggle with exploring many different vulnerabilities and long-range planning when used alone. To resolve this, we introduce HPTSA, a system of agents with a planning agent that can launch subagents. The planning agent explores the system and determines which subagents to call, resolving long-term planning issues when trying different vulnerabilities. We construct a benchmark of 14 real-world vulnerabilities and show that our team of agents improve over prior agent frameworks by up to 4.3X.

Click here for more

Paper 5:

Large Language Models are Unreliable for Cyber Threat

Intelligence

Date: 2025-03-29

Time: 18:09:36

Authors:

Emanuele Mezzi, Fabio Massacci, Katja Tuma

Summary:

- Several recent works have argued that Large Language Models (LLMs) can be used to tame the data deluge in the cybersecurity field, by improving the automation of Cyber Threat Intelligence (CTI) tasks. This work presents an evaluation methodology that other than allowing to test LLMs on CTI tasks when using zero-shot learning, few-shot learning and fine-tuning, also allows to quantify their consistency and their confidence level. We run experiments with three state-of-the-art LLMs and a dataset of 350 threat intelligence reports and present new evidence of potential security risks in relying on LLMs for CTI. We show how LLMs cannot guarantee sufficient performance on real-size reports while also being inconsistent and overconfident. Few-shot learning and fine-tuning only partially improve the results, thus posing doubts about the possibility of using LLMs for CTI scenarios, where labelled datasets are lacking and where confidence is a fundamental factor.

Click here for more

Paper 6:

Training Large Language Models for Advanced Typosquatting Detection

Date: 2025-03-28

Time: 13:16:27

Authors:

Jackson Welch

Summary:

- Typosquatting is a long-standing cyber threat that exploits human error in typing URLs to deceive users, distribute malware, and conduct phishing attacks. With the proliferation of domain names and new Top-Level Domains (TLDs), typosquatting techniques have grown more sophisticated, posing significant risks to individuals, businesses, and national cybersecurity infrastructure. Traditional detection methods primarily focus on well-known impersonation patterns, leaving gaps in identifying more complex attacks. This study introduces a novel approach leveraging large language models (LLMs) to enhance typosquatting detection. By training an LLM on character-level transformations and pattern-based heuristics rather than domain-specific data, a more adaptable and resilient detection mechanism develops. Experimental results indicate that the Phi-4 14B model outperformed other tested models when properly fine tuned achieving a 98% accuracy rate with only a few thousand training samples. This research highlights the potential of LLMs in cybersecurity applications, specifically in mitigating domain-based deception tactics, and provides insights into optimizing machine learning strategies for threat detection.

Click here for more

Paper 7:

Federated Intrusion Detection System Based on Unsupervised Machine Learning

Date: 2025-03-28

Time: 01:01:58

Authors:

Maxime Gourceyraud, Rim Ben Salem, Christopher Neal, Frédéric Cuppens,
Nora Boulahia Cuppens

Summary:

- Recent Intrusion Detection System (IDS) research has increasingly moved towards the adoption of machine learning methods. However, most of these systems rely on supervised learning approaches, necessitating a fully labeled training set. In the realm of network intrusion detection, the requirement for extensive labeling can become impractically burdensome. Moreover, while IDS training could benefit from inter-company knowledge sharing, the sensitive nature of cybersecurity data often precludes such cooperation. To address these challenges, we propose an IDS architecture that utilizes unsupervised learning to reduce the need for labeling. We further facilitate collaborative learning through the implementation of a federated learning framework. To enhance privacy beyond what current federated clustering models offer, we introduce an innovative federated K-means++ initialization technique. Our findings indicate that transitioning from a centralized to a federated setup does not significantly diminish performance.

Click here for more

Paper 8:

Debate-Driven Multi-Agent LLMs for Phishing Email Detection

Date: 2025-03-27

Time: 23:18:14

Authors:

Ngoc Tuong Vy Nguyen, Felix D Childress, Yunting Yin

Summary:

- Phishing attacks remain a critical cybersecurity threat. Attackers constantly refine their methods, making phishing emails harder to detect. Traditional detection methods, including rule-based systems and supervised machine learning models, either rely on predefined patterns like blacklists, which can be bypassed with slight modifications, or require large datasets for training and still can generate false positives and false negatives. In this work, we propose a multi-agent large language model (LLM) prompting technique that simulates debates among agents to detect whether the content presented on an email is phishing. Our approach uses two LLM agents to present arguments for or against the classification task, with a judge agent adjudicating the final verdict based on the quality of reasoning provided. This debate mechanism enables the models to critically analyze contextual cue and deceptive patterns in text, which leads to improved classification accuracy. The proposed framework is evaluated on multiple phishing email datasets and demonstrate that mixed-agent configurations consistently outperform homogeneous configurations. Results also show that the debate structure itself is sufficient to yield accurate decisions without extra prompting strategies.

Click here for more

Paper 9:

An Industry Interview Study of Software Signing for Supply Chain Security

Date: 2025-03-27

Time: 21:00:56

Authors:

Kelechi G. Kalu, Tanya Singla, Chinenye Okafor, Santiago Torres-Arias, James C. Davis

Summary:

- Many software products are composed of components integrated from other teams or external parties. Each additional link in a software product's supply chain increases the risk of the injection of malicious behavior. To improve supply chain provenance, many cybersecurity frameworks, standards, and regulations recommend the use of software signing. However, recent surveys and measurement studies have found that the adoption rate and quality of software signatures are low. We lack in-depth industry perspectives on the challenges and practices of software signing.   To understand software signing in practice, we interviewed 18 experienced security practitioners across 13 organizations. We study the challenges that affect the effective implementation of software signing in practice. We also provide possible impacts of experienced software supply chain failures, security standards, and regulations on software signing adoption. To summarize our findings: (1) We present a refined model of the software supply chain factory model highlighting practitioner's signing practices; (2) We highlight the different challenges-technical, organizational, and human-that hamper software signing implementation; (3) We report that experts disagree on the importance of signing; and (4) We describe how internal and external events affect the adoption of software signing. Our work describes the considerations for adopting software signing as one aspect of the broader goal of improved software supply chain security.

Click here for more

Paper 10:

Redefining Network Topology in Complex Systems: Merging Centrality   Metrics, Spectral Theory, and Diffusion Dynamics

Date: 2025-03-27

Time: 17:21:46

Authors:

Arsh Jha

Summary:

- This paper introduces a novel framework that combines traditional centrality measures with eigenvalue spectra and diffusion processes for a more comprehensive analysis of complex networks. While centrality measures such as degree, closeness, and betweenness have been commonly used to assess nodal importance, they provide limited insight into dynamic network behaviors. By incorporating eigenvalue analysis, which evaluates network robustness and connectivity through spectral properties, and diffusion processes that model information flow, this framework offers a deeper understanding of how networks function under dynamic conditions. Applied to synthetic networks, the

approach identifies key nodes not only by centrality but also by their role in diffusion dynamics and vulnerability points, offering a multi-dimensional view that traditional methods alone cannot. This integrated analysis enables a more precise identification of critical nodes and potential weaknesses, with implications for improving network resilience in fields ranging from epidemiology to cybersecurity. Keywords: Centrality measures, eigenvalue spectra, diffusion processes, network analysis, network robustness, information flow, synthetic networks.

Click here for more

Paper 11:

Intelligent IoT Attack Detection Design via ODLLM with Feature Ranking-based Knowledge Base

Date: 2025-03-27

Time: 16:41:57

Authors:

Satvik Verma, Qun Wang, E. Wes Bethel

Summary:

- The widespread adoption of Internet of Things (IoT) devices has introduced significant cybersecurity challenges, particularly with the increasing frequency and sophistication of Distributed Denial of Service (DDoS) attacks. Traditional machine learning (ML) techniques often fall short in detecting such attacks due to the complexity of blended and evolving patterns. To address this, we propose a novel framework leveraging On-Device Large Language Models (ODLLMs) augmented with fine-tuning and knowledge base (KB) integration for intelligent IoT network attack detection. By implementing feature ranking techniques and constructing both long and short KBs tailored to model capacities, the proposed framework ensures efficient and accurate detection of DDoS attacks while overcoming computational and privacy limitations. Simulation results demonstrate that the optimized framework achieves superior accuracy across diverse attack types, especially when using compact models in edge computing environments. This work provides a scalable and secure solution for real-time IoT security, advancing the applicability of edge intelligence in cybersecurity.

Click here for more

Paper 12:

Graph Anomaly Detection in Time Series: A Survey

Date: 2025-03-27

Time: 15:47:29

Authors:

Thi Kieu Khanh Ho, Ali Karami, Narges Armanfard

Summary:

- With the recent advances in technology, a wide range of systems continue to collect a large amount of data over time and thus generate time series. Time-Series Anomaly Detection (TSAD) is an important task in various time-series applications such as e-commerce, cybersecurity, vehicle maintenance, and healthcare monitoring. However, this task is very challenging as it requires considering both the intra-variable dependency (relationships within a variable over time) and the inter-variable dependency (relationships between multiple variables) existing in time-series data. Recent graph-based approaches have made impressive progress in tackling the challenges of this field. In this survey, we conduct a comprehensive and up-to-date review of TSAD using graphs, referred to as G-TSAD. First, we explore the significant potential of graph representation for time-series data and and its contributions to facilitating anomaly detection. Then, we review state-of-the-art graph anomaly detection techniques, mostly leveraging deep learning architectures, in the context of time series. For each method, we discuss its strengths, limitations, and the specific applications where it excels. Finally, we address both the technical and application challenges currently facing the field, and suggest potential future directions for advancing research and improving practical outcomes.

Click here for more

Paper 13:

Prompt, Divide, and Conquer: Bypassing Large Language Model Safety   Filters via Segmented and Distributed Prompt Processing

Date: 2025-03-27

Time: 15:19:55

Authors:

Johan Wahréus, Ahmed Hussain, Panos Papadimitratos

Summary:

- Large Language Models (LLMs) have transformed task automation and content generation across various domains while incorporating safety

filters to prevent misuse. We introduce a novel jailbreaking framework that employs distributed prompt processing combined with iterative refinements to bypass these safety measures, particularly in generating malicious code. Our architecture consists of four key modules: prompt segmentation, parallel processing, response aggregation, and LLM-based jury evaluation. Tested on 500 malicious prompts across 10 cybersecurity categories, the framework achieves a 73.2% Success Rate (SR) in generating malicious code. Notably, our comparative analysis reveals that traditional single-LLM judge evaluation overestimates SRs (93.8%) compared to our LLM jury system (73.2%), with manual verification confirming that single-judge assessments often accept incomplete implementations. Moreover, we demonstrate that our distributed architecture improves SRs by 12% over the non-distributed approach in an ablation study, highlighting both the effectiveness of distributed prompt processing and the importance of robust evaluation methodologies in assessing jailbreak attempts.

Click here for more

Paper 14:

Advancing CAN Network Security through RBM-Based Synthetic Attack Data Generation for Intrusion Detection Systems

Date: 2025-03-27

Time: 13:33:55

Authors:

Huacheng Li, Jingyong Su, Kai Wang

Summary:

- The rapid development of network technologies and industrial intelligence has augmented the connectivity and intelligence within the automotive industry. Notably, in the Internet of Vehicles (IoV), the Controller Area Network (CAN), which is crucial for the communication of electronic control units but lacks inbuilt security measures, has become extremely vulnerable to severe cybersecurity threats. Meanwhile, the efficacy of Intrusion Detection Systems (IDS) is hampered by the scarcity of sufficient attack data for robust model training. To overcome this limitation, we introduce a novel methodology leveraging the Restricted Boltzmann Machine (RBM) to generate synthetic CAN attack data, thereby producing training datasets with a more balanced sample distribution. Specifically, we design a CAN Data Processing Module for transforming raw CAN data into an RBM-trainable format, and a Negative Sample Generation Module to generate data reflecting the distribution of CAN data frames denoting network intrusions. Experimental results show the generated data significantly improves IDS performance, with CANet accuracy rising from 0.6477 to 0.9725 and EfficientNet from 0.1067 to 0.1555. Code is available at https://github.com/wangkai-tech23/CANDataSynthetic.

Click here for more

Paper 15:

Reasoning Under Threat: Symbolic and Neural Techniques for Cybersecurity Verification

Date: 2025-03-27

Time: 11:41:53

Authors:

Sarah Veronica

Summary:

- Cybersecurity demands rigorous and scalable techniques to ensure system correctness, robustness, and resilience against evolving threats. Automated reasoning, encompassing formal logic, theorem proving, model checking, and symbolic analysis, provides a foundational framework for verifying security properties across diverse domains such as access control, protocol design, vulnerability detection, and adversarial modeling. This survey presents a comprehensive overview of the role of automated reasoning in cybersecurity, analyzing how logical systems, including temporal, deontic, and epistemic logics are employed to formalize and verify security guarantees. We examine SOTA tools and frameworks, explore integrations with AI for neural-symbolic reasoning, and highlight critical research gaps, particularly in scalability, compositionality, and multi-layered security modeling. The paper concludes with a set of well-grounded future research directions, aiming to foster the development of secure systems through formal, automated, and explainable reasoning techniques.

Click here for more

Paper 16:

Predictable Interval MDPs through Entropy Regularization

Date: 2025-03-27

Time: 10:26:03

Authors:

Menno van Zutphen, Giannis Delimpaltadakis, Maurice Heemels, Duarte

Antunes

Summary:

- Regularization of control policies using entropy can be instrumental in adjusting predictability of real-world systems. Applications benefiting from such approaches range from, e.g., cybersecurity, which aims at maximal unpredictability, to human-robot interaction, where predictable behavior is highly desirable. In this paper, we consider entropy regularization for interval Markov decision processes (IMDPs). IMDPs are uncertain MDPs, where transition probabilities are only known to belong to intervals. Lately, IMDPs have gained significant popularity in the context of abstracting stochastic systems for control design. In this work, we address robust minimization of the linear combination of entropy and a standard cumulative cost in IMDPs, thereby establishing a trade-off between optimality and predictability. We show that optimal deterministic policies exist, and devise a value-iteration algorithm to compute them. The algorithm solves a number of convex programs at each step. Finally, through an illustrative example we show the benefits of penalizing entropy in IMDPs.

Click here for more

Paper 17:

Locally Optimal Solutions for Integer Programming Games

Date: 2025-03-26

Time: 18:45:12

Authors:

Pravesh Koirala, Mel Krusniak, Forrest Laine

Summary:

- Integer programming games (IPGs) are n-person games with integer strategy spaces. These games are used to model non-cooperative combinatorial decision-making and are used in domains such as cybersecurity and transportation. The prevalent solution concept for IPGs, Nash equilibrium, is difficult to compute and even showing whether such an equilibrium exists is known to be Sp2-complete. In this work, we introduce a class of relaxed solution concepts for IPGs called locally optimal integer solutions (LOIS) that are simpler to obtain than pure Nash equilibria. We demonstrate that LOIS are not only faster and more readily scalable in large-scale games but also support desirable features such as equilibrium enumeration and selection. We also show that these solutions can model a broader class

of problems including Stackelberg, Stackelberg-Nash, and generalized IPGs. Finally, we provide initial comparative results in a cybersecurity game called the Critical Node game, showing the performance gains of LOIS in comparison to the existing Nash equilibrium solution concept.

Click here for more

Paper 18:

A Computational Model for Ransomware Detection Using Cross-Domain Entropy Signatures

Date: 2025-03-26

Time: 15:56:28

Authors:

Michael Mannon, Evan Statham, Quentin Featherstone, Sebastian Arkwright, Clive Fenwick, Gareth Willoughby

Summary:

- Detecting encryption-driven cyber threats remains a large challenge due to the evolving techniques employed to evade traditional detection mechanisms. An entropy-based computational framework was introduced to analyze multi-domain system variations, enabling the identification of malicious encryption behaviors through entropy deviations. By integrating entropy patterns across file operations, memory allocations, and network transmissions, a detection methodology was developed to differentiate between benign and ransomware-induced entropy shifts. A mathematical model was formulated to quantify entropy dynamics, incorporating time-dependent variations and weighted domain contributions to enhance anomaly detection. Experimental evaluations demonstrated that the proposed approach achieved high accuracy across diverse ransomware families while maintaining low false positive rates. Computational efficiency analysis indicated minimal processing overhead, suggesting feasibility for real-time implementation in security-sensitive environments. The study highlighted entropy fluctuations as a useful indicator for identifying malicious encryption processes, reinforcing entropy-driven methodologies as a viable component of cybersecurity strategies.

Click here for more

Paper 19:

Decentralized Entropy-Driven Ransomware Detection Using Autonomous Neural Graph Embeddings

Date: 2025-03-26

Time: 15:54:51

Authors:

Ekaterina Starchenko, Hugo Bellinghamshire, David Pickering, Tristan Weatherspoon, Nathaniel Berkhamstead, Elizabeth Green, Magnus Rothschild

Summary:

- The increasing sophistication of cyber threats has necessitated the development of advanced detection mechanisms capable of identifying and mitigating ransomware attacks with high precision and efficiency. A novel framework, termed Decentralized Entropy-Driven Detection (DED), is introduced, leveraging autonomous neural graph embeddings and entropy-based anomaly scoring to address the limitations of traditional methods. The framework operates on a distributed network of nodes, eliminating single points of failure and enhancing resilience against targeted attacks. Experimental results demonstrate its ability to achieve detection accuracy exceeding 95\%, with false positive rates maintained below 2\% across diverse ransomware variants. The integration of graph-based modeling and machine learning techniques enables the framework to capture complex system interactions, facilitating the identification of subtle anomalies indicative of ransomware activity. Comparative analysis against existing methods highlights its superior performance in terms of detection rates and computational efficiency. Case studies further validate its effectiveness in real-world scenarios, showcasing its ability to detect and mitigate ransomware attacks within minutes of their initiation. The proposed framework represents a significant step forward in cybersecurity, offering a scalable and adaptive solution to the growing challenge of ransomware detection.

Click here for more

Paper 20:

Decentralized Entropy-Based Ransomware Detection Using Autonomous Feature Resonance

Date: 2025-03-26

Time: 15:48:36

Authors:

Barnaby Quince, Levi Gareth, Sophie Larkspur, Thaddeus Wobblethorn, Thomas Quibble

Summary:

- The increasing sophistication of cyber threats has necessitated the development of advanced detection mechanisms capable of identifying malicious activities with high precision and efficiency. A novel

approach, termed Autonomous Feature Resonance, is introduced to address the limitations of traditional ransomware detection methods through the analysis of entropy-based feature interactions within system processes. The proposed method achieves an overall detection accuracy of 97.3\%, with false positive and false negative rates of 1.8\% and 2.1\%, respectively, outperforming existing techniques such as signature-based detection and behavioral analysis. Its decentralized architecture enables local processing of data, reducing latency and improving scalability, while a self-learning mechanism ensures continuous adaptation to emerging threats. Experimental results demonstrate consistent performance across diverse ransomware families, including LockBit 3.0, BlackCat, and Royal, with low detection latency and efficient resource utilization. The method's reliance on entropy as a distinguishing feature provides robustness against obfuscation techniques, making it suitable for real-time deployment in high-throughput environments. These findings highlight the potential of entropy-based approaches to enhance cybersecurity frameworks, offering a scalable and adaptive solution for modern ransomware detection challenges.

Click here for more

Paper 21:

Power Networks SCADA Communication Cybersecurity, A Qiskit Implementation

Date: 2025-03-26

Time: 09:40:31

Authors:

Hillol Biswas

Summary:

- The cyber-physical system of electricity power networks utilizes supervisory control and data acquisition systems (SCADA), which are inherently vulnerable to cyber threats if usually connected with the internet technology (IT). Power system operations are conducted through communication systems that are mapped to standards, protocols, ports, and addresses. Real-time situational awareness is a standard term with implications and applications in both power systems and cybersecurity. In the plausible quantum world (Q-world), conventional approaches will likely face new challenges. The unique art of transmitting a quantum state from one place, Alice, to another, Bob, is known as quantum communication. Quantum communication for SCADA

communication in a plausible quantum era thus obviously entails wired communication through optical fiber networks complying with the typical cybersecurity criteria of confidentiality, integrity, and availability for classical internet technology unless a quantum internet (qinternet) transpires practically. When combined with the reverse order of AIC for operational technology, the cybersecurity criteria for power networks' critical infrastructure drill down to more specific sub-areas. Unlike other communication modes, such as information technology (IT) in broadband internet connections, SCADA for power networks, one of the critical infrastructures, is intricately intertwined with operations technology (OT), which significantly increases complexity. Though it is desirable to have a barrier called a demilitarized zone (DMZ), some overlap is inevitable. This paper highlights the opportunities and challenges in securing SCADA communication in the plausible quantum computing and communication regime, along with a corresponding integrated Qiskit implementation for possible future framework development.

Click here for more

Paper 22:

Advancing Vulnerability Classification with BERT: A Multi-Objective Learning Model

Date: 2025-03-26

Time: 06:04:45

Authors:

Himanshu Tiwari

Summary:

- The rapid increase in cybersecurity vulnerabilities necessitates automated tools for analyzing and classifying vulnerability reports. This paper presents a novel Vulnerability Report Classifier that leverages the BERT (Bidirectional Encoder Representations from Transformers) model to perform multi-label classification of Common Vulnerabilities and Exposures (CVE) reports from the National Vulnerability Database (NVD). The classifier predicts both the severity (Low, Medium, High, Critical) and vulnerability types (e.g., Buffer Overflow, XSS) from textual descriptions. We introduce a custom training pipeline using a combined loss function-Cross-Entropy for severity and Binary Cross-Entropy with Logits for types-integrated into a Hugging Face Trainer subclass. Experiments on recent NVD data demonstrate promising results, with decreasing evaluation loss across epochs. The system is deployed via a REST API and a Streamlit UI, enabling real-time vulnerability analysis. This work contributes a scalable, open-source solution for cybersecurity practitioners to automate vulnerability triage.

Click here for more

Paper 23:

Knowledge Transfer from LLMs to Provenance Analysis: A Semantic-Augmented Method for APT Detection

Date: 2025-03-25

Time: 20:11:36

Authors:

Fei Zuo, Junghwan Rhee, Yung Ryn Choe

Summary:

- Advanced Persistent Threats (APTs) have caused significant losses across a wide range of sectors, including the theft of sensitive data and harm to system integrity. As attack techniques grow increasingly sophisticated and stealthy, the arms race between cyber defenders and attackers continues to intensify. The revolutionary impact of Large Language Models (LLMs) has opened up numerous opportunities in various fields, including cybersecurity. An intriguing question arises: can the extensive knowledge embedded in LLMs be harnessed for provenance analysis and play a positive role in identifying previously unknown malicious events? To seek a deeper understanding of this issue, we propose a new strategy for taking advantage of LLMs in provenance-based threat detection. In our design, the state-of-the-art LLM offers additional details in provenance data interpretation, leveraging their knowledge of system calls, software identity, and high-level understanding of application execution context. The advanced contextualized embedding capability is further utilized to capture the rich semantics of event descriptions. We comprehensively examine the quality of the resulting embeddings, and it turns out that they offer promising avenues. Subsequently, machine learning models built upon these embeddings demonstrated outstanding performance on real-world data. In our evaluation, supervised threat detection achieves a precision of 99.0%, and semi-supervised anomaly detection attains a precision of 96.9%.

Click here for more

Paper 24:

Substation Bill of Materials: A Novel Approach to Managing Supply Chain Cyber-risks on IEC 61850 Digital Substations

Date: 2025-03-25

Time: 13:28:36

Authors:

Xabier Yurrebaso, Fernando Ibañez, Ángel Longueira-Romero

Summary:

- Smart grids have undergone a profound digitization process, integrating new data-driven control and supervision techniques, resulting in modern digital substations (DS). Attackers are more focused on attacking the supply chain of the DS, as they a comprise a multivendor environment. In this research work, we present the Substation Bill of Materials (Subs-BOM) schema, based on the CycloneDX specification, that is capable of modeling all the IEDs in a DS and their relationships from a cybersecurity perspective. The proposed Subs-BOM allows one to make informed decisions about cyber risks related to the supply chain, and enables managing multiple DS at the same time. This provides energy utilities with an accurate and complete inventory of the devices, the firmware they are running, and the services that are deployed into the DS. The Subs-BOM is generated using the Substation Configuration Description (SCD) file specified in the IEC 61850 standard as its main source of information. We validated the Subs-BOM schema against the Dependency-Track software by OWASP. This validation proved that the schema is correctly recognized by CycloneDX-compatible tools. Moreover, the Dependency-Track software could track existing vulnerabilities in the IEDs represented by the Subs-BOM.

Click here for more

Paper 25:

Hierarchical Polysemantic Feature Embedding for Autonomous Ransomware Detection

Date: 2025-03-25

Time: 13:17:09

Authors:

Sergei Nikitka, Sebastian Harringford, Charlotte Montgomery, Algernon Braithwaite, Matthew Kowalski

Summary:

- The evolution of ransomware requires the development of more sophisticated detection methodologies capable of identifying malicious

behaviors beyond traditional signature-based and heuristic techniques. The proposed Hierarchical Polysemantic Feature Embedding framework introduces a structured approach to ransomware detection through hyperbolic feature representations that capture hierarchical dependencies within executable behaviors. By embedding ransomware-relevant features into a non-Euclidean space, the framework maintains a well-defined decision boundary, ensuring improved generalization across previously unseen ransomware variants. Experimental evaluations demonstrated that the framework consistently outperformed conventional machine learning-based models, achieving higher detection accuracy while maintaining low false positive rates. The structured clustering mechanism employed within the hyperbolic space enabled robust classification even in the presence of obfuscation techniques, delayed execution strategies, and polymorphic transformations. Comparative analysis highlighted the limitations of existing detection frameworks, particularly in their inability to dynamically adapt to evolving ransomware tactics. Computational efficiency assessments indicated that the proposed method maintained a balance between detection performance and processing overhead, making it a viable candidate for real-world cybersecurity applications. The ability to detect emerging ransomware families without requiring extensive retraining demonstrated the adaptability of hierarchical embeddings in security analytics.

Click here for more

Paper 26:

Hierarchical Entropic Diffusion for Ransomware Detection: A Probabilistic Approach to Behavioral Anomaly Isolation

Date: 2025-03-25

Time: 13:14:37

Authors:

Vasili Iskorohodov, Maximilian Ravensdale, Matthias von Holstein, Hugo Petrovic, Adrian Yardley

Summary:

- The increasing complexity of cryptographic extortion techniques has necessitated the development of adaptive detection frameworks capable of identifying adversarial encryption behaviors without reliance on predefined signatures. Hierarchical Entropic Diffusion (HED) introduces a structured entropy-based anomaly classification mechanism that systematically tracks fluctuations in entropy evolution to differentiate between benign cryptographic processes and unauthorized encryption attempts. The integration of hierarchical clustering, entropy profiling, and probabilistic diffusion modeling refines detection granularity, ensuring that encryption anomalies are identified despite obfuscation strategies or incremental execution methodologies. Experimental evaluations demonstrated that HED maintained high classification accuracy across diverse ransomware families, outperforming traditional heuristic-based and signature-driven approaches while reducing false positive occurrences. Comparative analysis highlighted that entropy-driven anomaly segmentation improved detection efficiency under variable system workload conditions, ensuring real-time classification feasibility. The computational overhead associated with entropy anomaly detection remained within operational constraints, reinforcing the suitability of entropy-driven classification for large-scale deployment. The ability to identify adversarial entropy manipulations before encryption completion contributes to broader cybersecurity defenses, offering a structured methodology for isolating unauthorized cryptographic activities within heterogeneous computing environments. The results further emphasized that entropy evolution modeling facilitates predictive anomaly detection, enhancing resilience against encryption evasion techniques designed to circumvent traditional detection mechanisms.

Click here for more

Paper 27:

Red Teaming with Artificial Intelligence-Driven Cyberattacks: A Scoping Review

Date: 2025-03-25

Time: 13:14:19

Authors:

Mays Al-Azzawi, Dung Doan, Tuomo Sipola, Jari Hautamäki, Tero Kokkonen

Summary:

- The progress of artificial intelligence (AI) has made sophisticated methods available for cyberattacks and red team activities. These AI attacks can automate the process of penetrating a target or collecting sensitive data. The new methods can also accelerate the execution of the attacks. This review article examines the use of AI technologies in cybersecurity attacks. It also tries to describe typical targets for such attacks. We employed a scoping review methodology to analyze

articles and identify AI methods, targets, and models that red teams can utilize to simulate cybercrime. From the 470 records screened, 11 were included in the review. Various cyberattack methods were identified, targeting sensitive data, systems, social media profiles, passwords, and URLs. The application of AI in cybercrime to develop versatile attack models presents an increasing threat. Furthermore, AI-based techniques in red team use can provide new ways to address these issues.

Click here for more

Paper 28:

Leveraging VAE-Derived Latent Spaces for Enhanced Malware Detection with Machine Learning Classifiers

Date: 2025-03-24

Time: 14:44:55

Authors:

Bamidele Ajayi, Basel Barakat, Ken McGarry

Summary:

- This paper assesses the performance of five machine learning classifiers: Decision Tree, Naive Bayes, LightGBM, Logistic Regression, and Random Forest using latent representations learned by a Variational Autoencoder from malware datasets. Results from the experiments conducted on different training-test splits with different random seeds reveal that all the models perform well in detecting malware with ensemble methods (LightGBM and Random Forest) performing slightly better than the rest. In addition, the use of latent features reduces the computational cost of the model and the need for extensive hyperparameter tuning for improved efficiency of the model for deployment. Statistical tests show that these improvements are significant, and thus, the practical relevance of integrating latent space representation with traditional classifiers for effective malware detection in cybersecurity is established.

Click here for more

Paper 29:

ThreatCrawl: A BERT-based Focused Crawler for the Cybersecurity Domain

Date: 2025-03-24

Time: 09:14:21

Authors:

Philipp Kuehn, Mike Schmidt, Markus Bayer, Christian Reuter

Summary:

- Publicly available information contains valuable information for Cyber Threat Intelligence (CTI). This can be used to prevent attacks that have already taken place on other systems. Ideally, only the initial attack succeeds and all subsequent ones are detected and stopped. But while there are different standards to exchange this information, a lot of it is shared in articles or blog posts in non-standardized ways. Manually scanning through multiple online portals and news pages to discover new threats and extracting them is a time-consuming task. To automize parts of this scanning process, multiple papers propose extractors that use Natural Language Processing (NLP) to extract Indicators of Compromise (IOCs) from documents. However, while this already solves the problem of extracting the information out of documents, the search for these documents is rarely considered. In this paper, a new focused crawler is proposed called ThreatCrawl, which uses Bidirectional Encoder Representations from Transformers (BERT)-based models to classify documents and adapt its crawling path dynamically. While ThreatCrawl has difficulties to classify the specific type of Open Source Intelligence (OSINT) named in texts, e.g., IOC content, it can successfully find relevant documents and modify its path accord ingly. It yields harvest rates of up to 52%, which are, to the best of our knowledge, better than the current state of the art. The results and source code will be made publicly available upon acceptance.

Click here for more

Paper 30:

The Human-Machine Identity Blur: A Unified Framework for Cybersecurity Risk Management in 2025

Date: 2025-03-24

Time: 00:37:14

Authors:

Kush Janani

Summary:

- The modern enterprise is facing an unprecedented surge in digital identities, with machine identities now significantly outnumbering human identities. This paper examines the cybersecurity risks emerging from what we define as the "human-machine identity blur" - the point at which human and machine identities intersect, delegate authority,

and create new attack surfaces. Drawing from industry data, expert insights, and real-world incident analysis, we identify key governance gaps in current identity management models that treat human and machine entities as separate domains. To address these challenges, we propose a Unified Identity Governance Framework based on four core principles: treating identity as a continuum rather than a binary distinction, applying consistent risk evaluation across all identity types, implementing continuous verification guided by zero trust principles, and maintaining governance throughout the entire identity lifecycle. Our research shows that organizations adopting this unified approach experience a 47 percent reduction in identity-related security incidents and a 62 percent improvement in incident response time. We conclude by offering a practical implementation roadmap and outlining future research directions as AI-driven systems become increasingly autonomous.

Click here for more

Paper 31:

Exploring Energy Landscapes for Minimal Counterfactual Explanations: Applications in Cybersecurity and Beyond

Date: 2025-03-23

Time: 19:48:37

Authors:

Spyridon Evangelatos, Eleni Veroni, Vasilis Efthymiou, Christos Nikolopoulos, Georgios Th. Papadopoulos, Panagiotis Sarigiannidis

Summary:

- Counterfactual explanations have emerged as a prominent method in Explainable Artificial Intelligence (XAI), providing intuitive and actionable insights into Machine Learning model decisions. In contrast to other traditional feature attribution methods that assess the importance of input variables, counterfactual explanations focus on identifying the minimal changes required to alter a model's prediction, offering a ``what-if'' analysis that is close to human reasoning. In the context of XAI, counterfactuals enhance transparency, trustworthiness and fairness, offering explanations that are not just interpretable but directly applicable in the decision-making processes.   In this paper, we present a novel framework that integrates perturbation theory and statistical mechanics to generate minimal counterfactual explanations in explainable AI. We employ a local Taylor expansion of a Machine Learning model's predictive

function and reformulate the counterfactual search as an energy minimization problem over a complex landscape. In sequence, we model the probability of candidate perturbations leveraging the Boltzmann distribution and use simulated annealing for iterative refinement. Our approach systematically identifies the smallest modifications required to change a model's prediction while maintaining plausibility. Experimental results on benchmark datasets for cybersecurity in Internet of Things environments, demonstrate that our method provides actionable, interpretable counterfactuals and offers deeper insights into model sensitivity and decision boundaries in high-dimensional spaces.

Click here for more

Paper 32:

Confronting Catastrophic Risk: The International Obligation to Regulate Artificial Intelligence

Date: 2025-03-23

Time: 06:24:45

Authors:

Bryan Druzin, Anatole Boute, Michael Ramsden

Summary:

- While artificial intelligence (AI) holds enormous promise, many experts in the field are warning that there is a non-trivial chance that the development of AI poses an existential threat to humanity. Existing regulatory initiative do not address this threat but merely instead focus on discrete AI-related risks such as consumer safety, cybersecurity, data protection, and privacy. In the absence of regulatory action to address the possible risk of human extinction by AI, the question arises: What legal obligations, if any, does public international law impose on states to regulate its development. Grounded in the precautionary principle, we argue that there exists an international obligation to mitigate the threat of human extinction by AI. Often invoked in relation to environmental regulation and the regulation of potentially harmful technologies, the principle holds that in situations where there is the potential for significant harm, even in the absence of full scientific certainty, preventive measures should not be postponed if delayed action may result in irreversible consequences. We argue that the precautionary principle is a general principle of international law and, therefore, that there is a positive obligation on states under the right to life within international human rights law to proactively take regulatory action to mitigate the potential existential risk of AI. This is significant because, if an international obligation to regulate the development of AI can be established under international law, then the basic legal framework would be in place to address this evolving threat.

Click here for more

Paper 33:

Assessing the influence of cybersecurity threats and risks on the adoption and growth of digital banking: a systematic literature review

Date: 2025-03-23

Time: 03:14:45

Authors:

Md. Waliullah, Md Zahin Hossain George, Md Tarek Hasan, Md Khorshed Alam, Mosa Sumaiya Khatun Munira, Noor Alam Siddiqui

Summary:

- The rapid digitalization of banking services has significantly transformed financial transactions, offering enhanced convenience and efficiency for consumers. However, the increasing reliance on digital banking has also exposed financial institutions and users to a wide range of cybersecurity threats, including phishing, malware, ransomware, data breaches, and unauthorized access. This study systematically examines the influence of cybersecurity threats on digital banking security, adoption, and regulatory compliance by conducting a comprehensive review of 78 peer-reviewed articles published between 2015 and 2024. Using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology, this research critically evaluates the most prevalent cyber threats targeting digital banking platforms, the effectiveness of modern security measures, and the role of regulatory frameworks in mitigating financial cybersecurity risks. The findings reveal that phishing and malware attacks remain the most commonly exploited cyber threats, leading to significant financial losses and consumer distrust. Multi-factor authentication (MFA) and biometric security have been widely adopted to combat unauthorized access, while AI-driven fraud detection and blockchain technology offer promising solutions for securing financial transactions. However, the integration of third-party FinTech solutions introduces additional security risks, necessitating stringent regulatory oversight and cybersecurity protocols. The study also highlights that compliance with global cybersecurity regulations, such as GDPR, PSD2, and GLBA, enhances digital banking security by enforcing strict authentication measures, encryption protocols, and real-time fraud monitoring.

Click here for more

Paper 34:

EXPLICATE: Enhancing Phishing Detection through Explainable AI and LLM-Powered Interpretability

Date: 2025-03-22

Time: 23:37:35

Authors:

Bryan Lim, Roman Huerta, Alejandro Sotelo, Anthonie Quintela, Priyanka Kumar

Summary:

- Sophisticated phishing attacks have emerged as a major cybersecurity threat, becoming more common and difficult to prevent. Though machine learning techniques have shown promise in detecting phishing attacks, they function mainly as "black boxes" without revealing their decision-making rationale. This lack of transparency erodes the trust of users and diminishes their effective threat response. We present EXPLICATE: a framework that enhances phishing detection through a three-component architecture: an ML-based classifier using domain-specific features, a dual-explanation layer combining LIME and SHAP for complementary feature-level insights, and an LLM enhancement using DeepSeek v3 to translate technical explanations into accessible natural language. Our experiments show that EXPLICATE attains 98.4 % accuracy on all metrics, which is on par with existing deep learning techniques but has better explainability. High-quality explanations are generated by the framework with an accuracy of 94.2 % as well as a consistency of 96.8\% between the LLM output and model prediction. We create EXPLICATE as a fully usable GUI application and a light Chrome extension, showing its applicability in many deployment situations. The research shows that high detection performance can go hand-in-hand with meaningful explainability in security applications. Most important, it addresses the critical divide between automated AI and user trust in phishing detection systems.

Click here for more

Paper 35:

Detecting and Mitigating DDoS Attacks with AI: A Survey

Date: 2025-03-22

Time: 21:54:23

Authors:

Alexandru Apostu, Silviu Gheorghe, Andrei Hîji, Nicolae Cleju, Andrei P■tra■cu, Cristian Rusu, Radu Ionescu, Paul Irofti

Summary:

- Distributed Denial of Service attacks represent an active cybersecurity research problem. Recent research shifted from static rule-based defenses towards AI-based detection and mitigation. This comprehensive survey covers several key topics. Preeminently, state-of-the-art AI detection methods are discussed. An in-depth taxonomy based on manual expert hierarchies and an AI-generated dendrogram are provided, thus settling DDoS categorization ambiguities. An important discussion on available datasets follows, covering data format options and their role in training AI detection methods together with adversarial training and examples augmentation. Beyond detection, AI based mitigation techniques are surveyed as well. Finally, multiple open research directions are proposed.

Click here for more

Paper 36:

Aportes para el cumplimiento del Reglamento (UE) 2024/1689 en robótica y sistemas autónomos

Date: 2025-03-22

Time: 11:04:42

Authors:

Francisco J. Rodríguez Lera, Yoana Pita Lorenzo, David Sobrín Hidalgo, Laura Fernández Becerra, Irene González Fernández, Jose Miguel Guerrero Hernández

Summary:

- Cybersecurity in robotics stands out as a key aspect within Regulation (EU) 2024/1689, also known as the Artificial Intelligence Act, which establishes specific guidelines for intelligent and automated systems. A fundamental distinction in this regulatory framework is the difference between robots with Artificial Intelligence (AI) and those that operate through automation systems without AI, since the former are subject to stricter security requirements due to their learning and autonomy capabilities. This work analyzes cybersecurity tools applicable to advanced robotic systems, with special emphasis on the protection of knowledge bases in cognitive architectures. Furthermore, a list of basic tools is proposed to guarantee the security, integrity, and resilience of these systems, and a practical case is presented, focused on the analysis of robot knowledge management, where ten evaluation criteria are defined to ensure compliance with the regulation and reduce risks in human-robot interaction (HRI) environments.

Click here for more

Paper 37:

Erasing Conceptual Knowledge from Language Models

Date: 2025-03-22

Time: 04:42:36

Authors:

Rohit Gandikota, Sheridan Feucht, Samuel Marks, David Bau

Summary:

- In this work, we propose Erasure of Language Memory (ELM), an approach for concept-level unlearning built on the principle of matching the distribution defined by an introspective classifier. Our key insight is that effective unlearning should leverage the model's ability to evaluate its own knowledge, using the model itself as a classifier to identify and reduce the likelihood of generating content related to undesired concepts. ELM applies this framework to create targeted low-rank updates that reduce generation probabilities for concept-specific content while preserving the model's broader capabilities. We demonstrate ELM's efficacy on biosecurity, cybersecurity, and literary domain erasure tasks. Comparative analysis shows that ELM achieves superior performance across key metrics, including near-random scores on erased topic assessments, maintained coherence in text generation, preserved accuracy on unrelated benchmarks, and robustness under adversarial attacks. Our code, data, and trained models are available at https://elm.baulab.info

Click here for more

Paper 38:

HCAST: Human-Calibrated Autonomy Software Tasks

Date: 2025-03-21

Time: 17:54:01

Authors:

David Rein, Joel Becker, Amy Deng, Seraphina Nix, Chris Canal, Daniel

O'Connel, Pip Arnott, Ryan Bloom, Thomas Broadley, Katharyn Garcia

Summary:

- To understand and predict the societal impacts of highly autonomous AI systems, we need benchmarks with grounding, i.e., metrics that directly connect AI performance to real-world effects we care about. We present HCAST (Human-Calibrated Autonomy Software Tasks), a benchmark of 189 machine learning engineering, cybersecurity, software engineering, and general reasoning tasks. We collect 563 human baselines (totaling over 1500 hours) from people skilled in these domains, working under identical conditions as AI agents, which lets us estimate that HCAST tasks take humans between one minute and 8+ hours. Measuring the time tasks take for humans provides an intuitive metric for evaluating AI capabilities, helping answer the question "can an agent be trusted to complete a task that would take a human X hours?" We evaluate the success rates of AI agents built on frontier foundation models, and we find that current agents succeed 70-80% of the time on tasks that take humans less than one hour, and less than 20% of the time on tasks that take humans more than 4 hours.

Click here for more

Paper 39:

CVE-Bench: A Benchmark for AI Agents' Ability to Exploit Real-World Web Application Vulnerabilities

Date: 2025-03-21

Time: 17:32:32

Authors:

Yuxuan Zhu, Antony Kellermann, Dylan Bowman, Philip Li, Akul Gupta, Adarsh Danda, Richard Fang, Conner Jensen, Eric Ihli, Jason Benn

Summary:

- Large language model (LLM) agents are increasingly capable of autonomously conducting cyberattacks, posing significant threats to existing applications. This growing risk highlights the urgent need for a real-world benchmark to evaluate the ability of LLM agents to exploit web application vulnerabilities. However, existing benchmarks fall short as they are limited to abstracted Capture the Flag competitions or lack comprehensive coverage. Building a benchmark for real-world vulnerabilities involves both specialized expertise to reproduce exploits and a systematic approach to evaluating unpredictable threats. To address this challenge, we introduce CVE-Bench, a real-world cybersecurity benchmark based on critical-severity

Common Vulnerabilities and Exposures. In CVE-Bench, we design a sandbox framework that enables LLM agents to exploit vulnerable web applications in scenarios that mimic real-world conditions, while also providing effective evaluation of their exploits. Our evaluation shows that the state-of-the-art agent framework can resolve up to 13% of vulnerabilities.

Click here for more

Paper 40:

Zero Trust Architecture: A Systematic Literature Review

Date: 2025-03-21

Time: 10:52:22

Authors:

Muhammad Liman Gambo, Ahmad Almulhem

Summary:

- The increasing complexity of digital ecosystems and evolving cybersecurity threats have highlighted the limitations of traditional perimeter-based security models, leading to the growing adoption of Zero Trust Architecture (ZTA). ZTA operates on the principle of "never trust, always verify", enforcing continuous authentication, conditional access, dynamic trust evaluation, and the principle of least privilege to enhance security across diverse domains. This study applies the PRISMA framework to analyze 10 years of research (2016-2025) on ZTA, presenting a systematic literature review (SLR) that synthesizes its applications, enabling technologies, and associated challenges. It provides a detailed taxonomy that organizes ZTA's application domains, together with the emerging technologies that facilitate its implementation, and critically examines the barriers to ZTA adoption. Additionally, the study traces the historical evolution of ZTA alongside notable events and publications trends while highlighting some potential factors for the surge over the past few years. This comprehensive analysis serves as a practical guide for researchers and practitioners seeking to leverage ZTA for stronger, more adaptive security frameworks in a rapidly shifting threat landscape.

Click here for more

Paper 41:

EVSOAR: Security Orchestration, Automation and Response via EV Charging Stations

Date: 2025-03-21

Time: 09:48:29

Authors:

Tadeu Freitas, Erick Silva, Rehana Yasmin, Ali Shoker, Manuel E. Correia, Rolando Martins, Paulo Esteves-Verissimo

Summary:

- Vehicle cybersecurity has emerged as a critical concern, driven by the innovation in the automotive industry, e.g., automomous, electric, or connnected vehicles. Current efforts to address these challenges are constrained by the limited computational resources of vehicles and the reliance on connected infrastructures. This motivated the foundation of Vehicle Security Operations Centers (VSOCs) that extend IT-based Security Operations Centers (SOCs) to cover the entire automotive ecosystem, both the in-vehicle and off-vehicle scopes. Security Orchestration, Automation, and Response (SOAR) tools are considered key for impelementing an effective cybersecurity solution. However, existing state-of-the-art solutions depend on infrastructure networks such as 4G, 5G, and WiFi, which often face scalability and congestion issues. To address these limitations, we propose a novel SOAR architecture EVSOAR that leverages the EV charging stations for connectivity and computing to enhance vehicle cybersecurity. Our EV-specific SOAR architecture enables real-time analysis and automated responses to cybersecurity threats closer to the EV, reducing the cellular latency, bandwidth, and interference limitations. Our experimental results demonstrate a significant improvement in latency, stability, and scalability through the infrastructure and the capacity to deploy computationally intensive applications, that are otherwise infeasible within the resource constraints of individual vehicles.

Click here for more

Paper 42:

The Future of IPTV: Security, AI Integration, 5G, and Next-Gen Streaming

Date: 2025-03-20

Time: 20:27:19

Authors:

Georgios Giannakopoulos, Peter Adegbenro, Maria Antonnette Perez

Summary:

- The evolution of Internet Protocol Television (IPTV) has transformed the landscape of digital broadcasting by leveraging high-speed internet connectivity to deliver high-quality multimedia content. IPTV provides a dynamic and interactive television experience through managed networks, ensuring superior Quality of Service (QoS) compared to open-network Internet TV. This study explores the technical infrastructure of IPTV, including its network architecture, data compression techniques, and the role of protocols such as IGMP and RTSP. It also examines security challenges, including encryption, digital rights management (DRM), and authentication mechanisms that safeguard IPTV services from unauthorized access and piracy. Moreover, the paper analyzes the distinctions between IPTV and open-network Internet TV, highlighting their respective advantages and limitations in terms of service control, bandwidth optimization, and content security. The integration of artificial intelligence (AI) and machine learning (ML) in IPTV enhances personalized content recommendations and predictive analytics, leading to improved user engagement and efficient network management. Additionally, emerging technologies such as 5G and cloud-based IPTV services are explored for their potential to further revolutionize the industry. While IPTV presents a robust alternative to traditional broadcasting, challenges such as bandwidth constraints, cybersecurity threats, and regulatory compliance remain significant. The study concludes that IPTV's future success will depend on advancements in network infrastructure, AI-driven optimizations, and strategic regulatory adaptations. As IPTV continues to evolve, hybrid models integrating IPTV and open-network streaming services are expected to enhance content accessibility, security, and overall user experience.

Click here for more

Paper 43:

Cultivating Cybersecurity: Designing a Cybersecurity Curriculum for the Food and Agriculture Sector

Date: 2025-03-20

Time: 16:17:11

Authors:

George Grispos, Logan Mears, Larry Loucks, William Mahoney

Summary:

- As technology increasingly integrates into farm settings, the food and agriculture sector has become vulnerable to cyberattacks. However, previous research has indicated that many farmers and food producers lack the cybersecurity education they require to identify and mitigate the growing number of threats and risks impacting the industry. This paper presents an ongoing research effort describing a cybersecurity initiative to educate various populations in the farming and

agriculture community. The initiative proposes the development and delivery of a ten-module cybersecurity course, to create a more secure workforce, focusing on individuals who, in the past, have received minimal exposure to cybersecurity education initiatives.

Click here for more

Paper 44:

Investigating The Implications of Cyberattacks Against Precision Agricultural Equipment

Date: 2025-03-20

Time: 16:10:35

Authors:

Mark Freyhof, George Grispos, Santosh K. Pitla, William Mahoney

Summary:

- As various technologies are integrated and implemented into the food and agricultural industry, it is increasingly important for stakeholders throughout the sector to identify and reduce cybersecurity vulnerabilities and risks associated with these technologies. However, numerous industry and government reports suggest that many farmers and agricultural equipment manufacturers do not fully understand the cyber threats posed by modern agricultural technologies, including CAN bus-driven farming equipment. This paper addresses this knowledge gap by attempting to quantify the cybersecurity risks associated with cyberattacks on farming equipment that utilize CAN bus technology. The contribution of this paper is twofold. First, it presents a hypothetical case study, using real-world data, to illustrate the specific and wider impacts of a cyberattack on a CAN bus-driven fertilizer applicator employed in row-crop farming. Second, it establishes a foundation for future research on quantifying cybersecurity risks related to agricultural machinery.

Click here for more

Paper 45:

DroidTTP: Mapping Android Applications with TTP for Cyber Threat Intelligence

Date: 2025-03-20

Time: 05:38:24

Authors:

Dincy R Arikkat, Vinod P., Rafidha Rehiman K. A., Serena Nicolazzo, Marco Arazzi, Antonino Nocera, Mauro Conti

Summary:

- The widespread adoption of Android devices for sensitive operations

like banking and communication has made them prime targets for cyber threats, particularly Advanced Persistent Threats (APT) and sophisticated malware attacks. Traditional malware detection methods rely on binary classification, failing to provide insights into adversarial Tactics, Techniques, and Procedures (TTPs). Understanding malware behavior is crucial for enhancing cybersecurity defenses. To address this gap, we introduce DroidTTP, a framework mapping Android malware behaviors to TTPs based on the MITRE ATT&CK framework. Our curated dataset explicitly links MITRE TTPs to Android applications. We developed an automated solution leveraging the Problem Transformation Approach (PTA) and Large Language Models (LLMs) to map applications to both Tactics and Techniques. Additionally, we employed Retrieval-Augmented Generation (RAG) with prompt engineering and LLM fine-tuning for TTP predictions. Our structured pipeline includes dataset creation, hyperparameter tuning, data augmentation, feature selection, model development, and SHAP-based model interpretability. Among LLMs, Llama achieved the highest performance in Tactic classification with a Jaccard Similarity of 0.9583 and Hamming Loss of 0.0182, and in Technique classification with a Jaccard Similarity of 0.9348 and Hamming Loss of 0.0127. However, the Label Powerset XGBoost model outperformed LLMs, achieving a Jaccard Similarity of 0.9893 for Tactic classification and 0.9753 for Technique classification, with a Hamming Loss of 0.0054 and 0.0050, respectively. While XGBoost showed superior performance, the narrow margin highlights the potential of LLM-based approaches in TTP classification.

Click here for more

Paper 46:

Cybersecurity in Vehicle-to-Grid (V2G) Systems: A Systematic Review

Date: 2025-03-19

Time: 22:55:40

Authors:

Mohammad A Razzaque, Shafiuzzaman K Khadem, Sandipan Patra, Glory Okwata, Md. Noor-A-Rahim

Summary:

- This paper presents a systematic review of recent advancements in V2G cybersecurity, employing the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework for detailed searches across three journal databases and included only peer-reviewed studies published between 2020 and 2024 (June). We identified and reviewed 133 V2G cybersecurity studies and found five important insights on existing V2G cybersecurity research. First, most studies (103 of 133) focused on protecting V2G systems against cyber threats, while only seven studies addressed the recovery aspect of the CRML (Cybersecurity Risk Management Lifecycle) function. Second, existing studies have adequately addressed the security of EVs and EVCS (EV charging stations) in V2G systems (112 and 81 of 133 studies, respectively). However, none have focused on the linkage between the behaviour of EV users and the cybersecurity of V2G systems. Third, physical access, control-related vulnerabilities, and user behaviour-related attacks in V2G systems are not addressed significantly. Furthermore, existing studies overlook vulnerabilities and attacks specific to AI and blockchain technologies. Fourth, blockchain, artificial intelligence (AI), encryption, control theory, and optimisation are the main technologies used, and finally, the inclusion of quantum safety within encryption and AI models and AI assurance (AIA) is in a very early stage; only two and one of 133 studies explicitly addressed quantum safety and AIA through explainability. By providing a holistic perspective, this study identifies critical research gaps and outlines future directions for developing robust end-to-end cybersecurity solutions to safeguard V2G systems and support global sustainability goals.

Click here for more

Paper 47:

Cyber Threats in Financial Transactions -- Addressing the Dual Challenge of AI and Quantum Computing

Date: 2025-03-19

Time: 20:16:27

Authors:

Ahmed M. Elmisery, Mirela Sertovic, Andrew Zayin, Paul Watson

Summary:

- The financial sector faces escalating cyber threats amplified by artificial intelligence (AI) and the advent of quantum computing. AI is being weaponized for sophisticated attacks like deepfakes and AI-driven malware, while quantum computing threatens to render current encryption methods obsolete. This report analyzes these threats, relevant frameworks, and possible countermeasures like quantum cryptography. AI enhances social engineering and phishing attacks via personalized content, lowers entry barriers for cybercriminals, and introduces risks like data poisoning and adversarial AI. Quantum computing, particularly Shor's algorithm, poses a fundamental threat to current encryption standards (RSA and ECC), with estimates suggesting cryptographically relevant quantum computers could emerge within the next 5-30 years. The "harvest now, decrypt later" scenario highlights the urgency of transitioning to quantum-resistant cryptography. This is key. Existing legal frameworks are evolving to address AI in cybercrime, but quantum threats require new initiatives. International cooperation and harmonized regulations are crucial. Quantum Key Distribution (QKD) offers theoretical security but faces practical limitations. Post-quantum cryptography (PQC) is a promising alternative, with ongoing standardization efforts. Recommendations for international regulators include fostering collaboration and information sharing, establishing global standards, supporting research and development in quantum security, harmonizing legal frameworks, promoting cryptographic agility, and raising awareness and education. The financial industry must adopt a proactive and adaptive approach to cybersecurity, investing in research, developing migration plans for quantum-resistant cryptography, and embracing a multi-faceted, collaborative strategy to build a resilient, quantum-safe, and AI-resilient financial ecosystem

Click here for more

Paper 48:

A Peek Behind the Curtain: Using Step-Around Prompt Engineering to Identify Bias and Misinformation in GenAI Models

Date: 2025-03-19

Time: 13:47:28

Authors:

Don Hickerson, Mike Perkins

Summary:

- This research examines the emerging technique of step-around prompt engineering in GenAI research, a method that deliberately bypasses AI safety measures to expose underlying biases and vulnerabilities in GenAI models. We discuss how Internet-sourced training data introduces unintended biases and misinformation into AI systems, which can be revealed through the careful application of step-around techniques. Drawing parallels with red teaming in cybersecurity, we argue that step-around prompting serves a vital role in identifying and addressing potential vulnerabilities while acknowledging its dual nature as both a research tool and a potential security threat. Our findings highlight three key implications: (1) the persistence of Internet-derived biases in AI training data despite content filtering, (2) the effectiveness of step-around techniques in exposing these biases when used responsibly, and (3) the need for robust safeguards against malicious applications of these methods. We conclude by proposing an ethical framework for using step-around prompting in AI research and development, emphasizing the importance of balancing system improvements with security considerations.

Click here for more

Paper 49:

ELTEX: A Framework for Domain-Driven Synthetic Data Generation

Date: 2025-03-19

Time: 09:46:54

Authors:

Arina Razmyslovich, Kseniia Murasheva, Sofia Sedlova, Julien Capitaine, Eugene Dmitriev

Summary:

- We present ELTEX (Efficient LLM Token Extraction), a domain-driven framework for generating high-quality synthetic training data in specialized domains. While Large Language Models (LLMs) have shown impressive general capabilities, their performance in specialized domains like cybersecurity remains limited by the scarcity of domain-specific training data. ELTEX addresses this challenge by systematically integrating explicit domain indicator extraction with dynamic prompting to preserve critical domain knowledge throughout the generation process. We demonstrate ELTEX's effectiveness in the context of blockchain-related cyberattack detection, where we fine-tune Gemma-2B using various combinations of real and ELTEX-generated data. Our results show that the ELTEX-enhanced model achieves performance competitive with GPT-4 across both standard classification metrics and uncertainty calibration, while requiring significantly fewer computational resources. We release a curated synthetic dataset of social media texts for cyberattack detection in blockchain. Our work demonstrates that domain-driven synthetic data generation can effectively bridge the performance gap between resource-efficient

models and larger architectures in specialized domains.

Click here for more

Paper 50:

Assessment of Cyberattack Detection-Isolation Algorithm for CAV Platoons Using SUMO

Date: 2025-03-18

Time: 18:29:58

Authors:

Sanchita Ghosh, Tanushree Roy

Summary:

- A Connected Autonomous Vehicle (CAV) platoon in an evolving real-world driving environment relies strongly on accurate vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication for its safe and efficient operation. However, a cyberattack on this communication network can corrupt the appropriate control actions, tamper with system measurement, and drive the platoon to unsafe or undesired conditions. As a first step toward practicable resilience against such V2V-V2I attacks, in this paper, we implemented a unified V2V-V2I cyberattack detection scheme and a V2I isolation scheme for a CAV platoon under changing driving conditions in Simulation of Urban MObility (SUMO). The implemented algorithm utilizes vehicle-specific residual generators that are designed based on analytical disturbance-to-state stability, robustness, and sensitivity performance constraints. Our case studies include two driving scenarios where highway driving is simulated using the Next-Generation Simulation (NGSIM) data and urban driving follows the benchmark EPA Urban Dynamometer Driving Schedule (UDDS). The results validate the applicability of the algorithm to ensure CAV cybersecurity and demonstrate the promising potential for practical test-bed implementation in the future.

Click here for more

Paper 51:

A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU

Date: 2025-03-17

Time: 10:18:52

Authors:

Farhad Mortezapour Shiri, Thinagaran Perumal, Norwati Mustapha,

Raihani Mohamed

Summary:

- Deep learning (DL) has emerged as a powerful subset of machine learning (ML) and artificial intelligence (AI), outperforming traditional ML methods, especially in handling unstructured and large datasets. Its impact spans across various domains, including speech recognition, healthcare, autonomous vehicles, cybersecurity, predictive analytics, and more. However, the complexity and dynamic nature of real-world problems present challenges in designing effective deep learning models. Consequently, several deep learning models have been developed to address different problems and applications. In this article, we conduct a comprehensive survey of various deep learning models, including Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Temporal Convolutional Networks (TCN), Transformer, Kolmogorov-Arnold networks (KAN), Generative Models, Deep Reinforcement Learning (DRL), and Deep Transfer Learning. We examine the structure, applications, benefits, and limitations of each model. Furthermore, we perform an analysis using three publicly available datasets: IMDB, ARAS, and Fruit-360. We compared the performance of six renowned deep learning models: CNN, RNN, Long Short-Term Memory (LSTM), Bidirectional LSTM, Gated Recurrent Unit (GRU), and Bidirectional GRU alongside two newer models, TCN and Transformer, using the IMDB and ARAS datasets. Additionally, we evaluated the performance of eight CNN-based models, including VGG (Visual Geometry Group), Inception, ResNet (Residual Network), InceptionResNet, Xception (Extreme Inception), MobileNet, DenseNet (Dense Convolutional Network), and NASNet (Neural Architecture Search Network), for image classification tasks using the Fruit-360 dataset.

Click here for more

Paper 52:

Cancermorphic Computing Toward Multilevel Machine Intelligence

Date: 2025-03-17

Time: 02:23:29

Authors:

Rosalia Moreddu, Michael Levin

Summary:

- Despite their potential to address crucial bottlenecks in computing architectures and contribute to the pool of biological inspiration for

engineering, pathological biological mechanisms remain absent from computational theory. We hereby introduce the concept of cancer-inspired computing as a paradigm drawing from the adaptive, resilient, and evolutionary strategies of cancer, for designing computational systems capable of thriving in dynamic, adversarial or resource-constrained environments. Unlike known bioinspired approaches (e.g., evolutionary and neuromorphic architectures), cancer-inspired computing looks at emulating the uniqueness of cancer cells survival tactics, such as somatic mutation, metastasis, angiogenesis and immune evasion, as parallels to desirable features in computing architectures, for example decentralized propagation and resource optimization, to impact areas like fault tolerance and cybersecurity. While the chaotic growth of cancer is currently viewed as uncontrollable in biology, randomness-based algorithms are already being successfully demonstrated in enhancing the capabilities of other computing architectures, for example chaos computing integration. This vision focuses on the concepts of multilevel intelligence and context-driven mutation, and their potential to simultaneously overcome plasticity-limited neuromorphic approaches and the randomness of chaotic approaches. The introduction of this concept aims to generate interdisciplinary discussion to explore the potential of cancer-inspired mechanisms toward powerful and resilient artificial systems.

Click here for more

Paper 53:

Enforcing Cybersecurity Constraints for LLM-driven Robot Agents for Online Transactions

Date: 2025-03-17

Time: 01:01:10

Authors:

Shraddha Pradipbhai Shah, Aditya Vilas Deshpande

Summary:

- The integration of Large Language Models (LLMs) into autonomous robotic agents for conducting online transactions poses significant cybersecurity challenges. This study aims to enforce robust cybersecurity constraints to mitigate the risks associated with data breaches, transaction fraud, and system manipulation. The background focuses on the rise of LLM-driven robotic systems in e-commerce, finance, and service industries, alongside the vulnerabilities they introduce. A novel security architecture combining blockchain

technology with multi-factor authentication (MFA) and real-time anomaly detection was implemented to safeguard transactions. Key performance metrics such as transaction integrity, response time, and breach detection accuracy were evaluated, showing improved security and system performance. The results highlight that the proposed architecture reduced fraudulent transactions by 90%, improved breach detection accuracy to 98%, and ensured secure transaction validation within a latency of 0.05 seconds. These findings emphasize the importance of cybersecurity in the deployment of LLM-driven robotic systems and suggest a framework adaptable to various online platforms.

Click here for more

Paper 54:

From ML to LLM: Evaluating the Robustness of Phishing Webpage Detection Models against Adversarial Attacks

Date: 2025-03-15

Time: 11:39:42

Authors:

Aditya Kulkarni, Vivek Balachandran, Dinil Mon Divakaran, Tamal Das

Summary:

- Phishing attacks attempt to deceive users into stealing sensitive information, posing a significant cybersecurity threat. Advances in machine learning (ML) and deep learning (DL) have led to the development of numerous phishing webpage detection solutions, but these models remain vulnerable to adversarial attacks. Evaluating their robustness against adversarial phishing webpages is essential. Existing tools contain datasets of pre-designed phishing webpages for a limited number of brands, and lack diversity in phishing features. To address these challenges, we develop PhishOracle, a tool that generates adversarial phishing webpages by embedding diverse phishing features into legitimate webpages. We evaluate the robustness of three existing task-specific models -- Stack model, VisualPhishNet, and Phishpedia -- against PhishOracle-generated adversarial phishing webpages and observe a significant drop in their detection rates. In contrast, a multimodal large language model (MLLM)-based phishing detector demonstrates stronger robustness against these adversarial attacks but still is prone to evasion. Our findings highlight the vulnerability of phishing detection models to adversarial attacks, emphasizing the need for more robust detection approaches. Furthermore, we conduct a user study to evaluate whether PhishOracle-

generated adversarial phishing webpages can deceive users. The results show that many of these phishing webpages evade not only existing detection models but also users. We also develop the PhishOracle web app, allowing users to input a legitimate URL, select relevant phishing features and generate a corresponding phishing webpage. All resources will be made publicly available on GitHub.

Click here for more

Paper 55:

Accelerating Sparse Tensor Decomposition Using Adaptive Linearized Representation

Date: 2025-03-15

Time: 00:49:55

Authors:

Jan Laukemann, Ahmed E. Helal, S. Isaac Geronimo Anderson, Fabio Checconi, Yongseok Soh, Jesmin Jahan Tithi, Teresa Ranadive, Brian J Gravelle, Fabrizio Petrini, Jee Choi

Summary:

- High-dimensional sparse data emerge in many critical application domains such as healthcare and cybersecurity. To extract meaningful insights from massive volumes of these multi-dimensional data, scientists employ unsupervised analysis tools based on tensor decomposition (TD) methods. However, real-world sparse tensors exhibit highly irregular shapes and data distributions, which pose significant challenges for making efficient use of modern parallel processors. This study breaks the prevailing assumption that compressing sparse tensors into coarse-grained structures or along a particular dimension/mode is more efficient than keeping them in a fine-grained, mode-agnostic form. Our novel sparse tensor representation, Adaptive Linearized Tensor Order (ALTO), encodes tensors in a compact format that can be easily streamed from memory and is amenable to both caching and parallel execution. In contrast to existing compressed tensor formats, ALTO constructs one tensor copy that is agnostic to both the mode orientation and the irregular distribution of nonzero elements. To demonstrate the efficacy of ALTO, we propose a set of parallel TD algorithms that exploit the inherent data reuse of tensor computations to substantially reduce synchronization overhead, decrease memory footprint, and improve parallel performance. Additionally, we characterize the major execution bottlenecks of TD methods on the latest Intel Xeon Scalable processors and introduce

dynamic adaptation heuristics to automatically select the best algorithm based on the sparse tensor characteristics. Across a diverse set of real-world data sets, ALTO outperforms the state-of-the-art approaches, achieving more than an order-of-magnitude speedup over the best mode-agnostic formats. Compared to the best mode-specific formats, ALTO achieves 5.1X geometric mean speedup at a fraction (25%) of their storage costs.

Click here for more

Paper 56:

Hacking Cryptographic Protocols with Advanced Variational Quantum Attacks

Date: 2025-03-14

Time: 15:36:05

Authors:

Borja Aizpurua, Pablo Bermejo, Josu Etxezarreta Martinez, Roman Orus

Summary:

- Here we introduce an improved approach to Variational Quantum Attack Algorithms (VQAA) on crytographic protocols. Our methods provide robust quantum attacks to well-known cryptographic algorithms, more efficiently and with remarkably fewer qubits than previous approaches. We implement simulations of our attacks for symmetric-key protocols such as S-DES, S-AES and Blowfish. For instance, we show how our attack allows a classical simulation of a small 8-qubit quantum computer to find the secret key of one 32-bit Blowfish instance with 24 times fewer number of iterations than a brute-force attack. Our work also shows improvements in attack success rates for lightweight ciphers such as S-DES and S-AES. Further applications beyond symmetric-key cryptography are also discussed, including asymmetric-key protocols and hash functions. In addition, we also comment on potential future improvements of our methods. Our results bring one step closer assessing the vulnerability of large-size classical cryptographic protocols with Noisy Intermediate-Scale Quantum (NISQ) devices, and set the stage for future research in quantum cybersecurity.

Click here for more

Paper 57:

Cost-effective Deep Learning Infrastructure with NVIDIA GPU

Date: 2025-03-14

Time: 09:54:36

Authors:

Aatiz Ghimire, Shahnawaz Alam, Siman Giri, Madhav Prasad Ghimire

Summary:

- The growing demand for computational power is driven by advancements in deep learning, the increasing need for big data processing, and the requirements of scientific simulations for academic and research purposes. Developing countries like Nepal often struggle with the resources needed to invest in new and better hardware for these purposes. However, optimizing and building on existing technology can still meet these computing demands effectively. To address these needs, we built a cluster using four NVIDIA GeForce GTX 1650 GPUs. The cluster consists of four nodes: one master node that controls and manages the entire cluster, and three compute nodes dedicated to processing tasks. The master node is equipped with all necessary software for package management, resource scheduling, and deployment, such as Anaconda and Slurm. In addition, a Network File Storage (NFS) system was integrated to provide the additional storage required by the cluster. Given that the cluster is accessible via ssh by a public domain address, which poses significant cybersecurity risks, we implemented fail2ban to mitigate brute force attacks and enhance security. Despite the continuous challenges encountered during the design and implementation process, this project demonstrates how powerful computational clusters can be built to handle resource-intensive tasks in various demanding fields.

Click here for more

Paper 58:

Phishsense-1B: A Technical Perspective on an AI-Powered Phishing Detection Model

Date: 2025-03-13

Time: 23:03:09

Authors:

SE Blake

Summary:

- Phishing is a persistent cybersecurity threat in today's digital landscape. This paper introduces Phishsense-1B, a refined version of the Llama-Guard-3-1B model, specifically tailored for phishing detection and reasoning. This adaptation utilizes Low-Rank Adaptation (LoRA) and the GuardReasoner finetuning methodology. We outline our LoRA-based fine-tuning process, describe the balanced dataset comprising phishing and benign emails, and highlight significant

performance improvements over the original model. Our findings indicate that Phishsense-1B achieves an impressive 97.5% accuracy on a custom dataset and maintains strong performance with 70% accuracy on a challenging real-world dataset. This performance notably surpasses both unadapted models and BERT-based detectors. Additionally, we examine current state-of-the-art detection methods, compare prompt-engineering with fine-tuning strategies, and explore potential deployment scenarios.

Click here for more

Paper 59:

Change-Point Detection in Dynamic Networks with Missing Links

Date: 2025-03-13

Time: 12:29:04

Authors:

Farida Enikeeva, Olga Klopp

Summary:

- Structural changes occur in dynamic networks quite frequently and its detection is an important question in many situations such as fraud detection or cybersecurity. Real-life networks are often incompletely observed due to individual non-response or network size. In the present paper we consider the problem of change-point detection at a temporal sequence of partially observed networks. The goal is to test whether there is a change in the network parameters. Our approach is based on the Matrix CUSUM test statistic and allows growing size of networks. We show that the proposed test is minimax optimal and robust to missing links. We also demonstrate the good behavior of our approach in practice through simulation study and a real-data application.

Click here for more

Paper 60:

Optimal Security Response to Network Intrusions in IT Systems

Date: 2025-03-13

Time: 08:27:31

Authors:

Kim Hammar

Summary:

- Cybersecurity is one of the most pressing technological challenges of

our time and requires measures from all sectors of society. A key measure is automated security response, which enables automated mitigation and recovery from cyber attacks. Significant strides toward such automation have been made due to the development of rule-based response systems. However, these systems have a critical drawback: they depend on domain experts to configure the rules, a process that is both error-prone and inefficient. Framing security response as an optimal control problem shows promise in addressing this limitation but introduces new challenges. Chief among them is bridging the gap between theoretical optimality and operational performance. Current response systems with theoretical optimality guarantees have only been validated analytically or in simulation, leaving their practical utility unproven.   This thesis tackles the aforementioned challenges by developing a practical methodology for optimal security response in IT infrastructures. It encompasses two systems. First, it includes an emulation system that replicates key components of the target infrastructure. We use this system to gather measurements and logs, based on which we identify a game-theoretic model. Second, it includes a simulation system where game-theoretic response strategies are optimized through stochastic approximation to meet a given objective, such as mitigating potential attacks while maintaining operational services. These strategies are then evaluated and refined in the emulation system to close the gap between theoretical and operational performance. We prove structural properties of optimal response strategies and derive efficient algorithms for computing them. This enables us to solve a previously unsolved problem: demonstrating optimal security response against network intrusions on an IT infrastructure.

Click here for more

Paper 61:

Older adults' safety and security online: A post-pandemic exploration of attitudes and behaviors

Date: 2025-03-12

Time: 20:21:49

Authors:

Edgar Pacheco

Summary:

- Older adults' growing use of the internet and related technologies, further accelerated by the COVID-19 pandemic, has prompted not only a critical examination of their behaviors and attitudes about online threats but also a greater understanding of the roles of specific characteristics within this population group. Based on survey data and using descriptive and inferential statistics, this empirical study delves into this matter. The behaviors and attitudes of a group of older adults aged 60 years and older (n=275) regarding different dimensions of online safety and cybersecurity are investigated. The results show that older adults report a discernible degree of concern about the security of their personal information. Despite the varied precautions taken, most of them do not know where to report online threats. What is more, regarding key demographics, the study found some significant differences in terms of gender and age group, but not disability status. This implies that older adults do not seem to constitute a homogeneous group when it comes to attitudes and behaviors regarding safety and security online. The study concludes that support systems should include older adults in the development of protective measures and acknowledge their diversity. The implications of the results are discussed and some directions for future research are proposed.

Click here for more

Paper 62:

Auspex: Building Threat Modeling Tradecraft into an Artificial Intelligence-based Copilot

Date: 2025-03-12

Time: 17:54:18

Authors:

Andrew Crossman, Andrew R. Plummer, Chandra Sekharudu, Deepak Warrier, Mohammad Yekrangian

Summary:

- We present Auspex - a threat modeling system built using a specialized collection of generative artificial intelligence-based methods that capture threat modeling tradecraft. This new approach, called tradecraft prompting, centers on encoding the on-the-ground knowledge of threat modelers within the prompts that drive a generative AI-based threat modeling system. Auspex employs tradecraft prompts in two processing stages. The first stage centers on ingesting and processing system architecture information using prompts that encode threat modeling tradecraft knowledge pertaining to system decomposition and description. The second stage centers on chaining the resulting system analysis through a collection of prompts that encode tradecraft knowledge on threat identification, classification, and mitigation. The two-stage process yields a threat matrix for a system that specifies threat scenarios, threat types, information security categorizations and potential mitigations. Auspex produces formalized threat model output in minutes, relative to the weeks or months a manual process takes. More broadly, the focus on bespoke tradecraft prompting, as opposed to fine-tuning or agent-based add-ons, makes Auspex a lightweight, flexible, modular, and extensible foundational system capable of addressing the complexity, resource, and standardization limitations of both existing manual and automated threat modeling processes. In this connection, we establish the baseline value of Auspex to threat modelers through an evaluation procedure based on feedback collected from cybersecurity subject matter experts measuring the quality and utility of threat models generated by Auspex on real banking systems. We conclude with a discussion of system performance and plans for enhancements to Auspex.

Click here for more

Paper 63:

Automatic Association of Quality Requirements and Quantifiable Metrics   for Cloud Security Certification

Date: 2025-03-12

Time: 15:06:45

Authors:

John Bianchi, Shuya Dong, Luca Petrillo, Marinella Petrocchi

Summary:

- The European Cybersecurity Certification Scheme for Cloud Services (EUCS) is one of the first cybersecurity schemes in Europe, defined by the European Union Agency for Cybersecurity (ENISA). It aims to encourage cloud providers to strengthen their cybersecurity policies in order to receive an official seal of approval from European authorities. EUCS defines a set of security requirements that the cloud provider must meet, in whole or in part, in order to achieve the security certification. The requirements are written in natural language and cover every aspect of security in the cloud environment, from logging access to protecting the system with anti-malware tools to training staff. Operationally, each requirement is associated with one or more evaluable metrics. For example, a requirement to monitor access attempts to a service will have associated metrics that take into account the number of accesses, the number of access attempts, who is accessing, and what resources are being used. Partners in the European project Medina, which ended in October 2023, defined 163 metrics and manually mapped them to 70 EUCS requirements. Manual mapping is intuitively a long and costly process in terms of human resources. This paper proposes an approach based on Sentence Transformers to automatically associate requirements and metrics. In terms of correctness of associations, the proposed method achieves a Normalized Discounted Cumulative Gain of 0.640, improving a previous experiment by 0.146 points.

Click here for more

Paper 64:

Quantum Computing and Cybersecurity Education: A Novel Curriculum for Enhancing Graduate STEM Learning

Date: 2025-03-12

Time: 13:26:54

Authors:

Suryansh Upadhyay, Koustubh Phalak, Jungeun Lee, Kathleen Mitchell Hill, Swaroop Ghosh

Summary:

- Quantum computing is an emerging paradigm with the potential to transform numerous application areas by addressing problems considered intractable in the classical domain. However, its integration into cyberspace introduces significant security and privacy challenges. The exponential rise in cyber attacks, further complicated by quantum capabilities, poses serious risks to financial systems and national security. The scope of quantum threats extends beyond traditional software, operating system, and network vulnerabilities, necessitating a shift in cybersecurity education. Traditional cybersecurity education, often reliant on didactic methods, lacks hands on, student centered learning experiences necessary to prepare students for these evolving challenges. There is an urgent need for curricula that address both classical and quantum security threats through experiential learning. In this work, we present the design and evaluation of EE 597: Introduction to Hardware Security, a graduate level course integrating hands-on quantum security learning with classical security concepts through simulations and cloud-based quantum hardware. Unlike conventional courses focused on quantum threats to cryptographic systems, EE 597 explores security challenges specific to quantum computing itself. We employ a mixed-methods evaluation using pre and post surveys to assess student learning outcomes and engagement. Results indicate significant improvements in students' understanding of quantum and hardware security, with strong positive feedback on course structure and remote instruction (mean scores: 3.33 to 3.83 on a 4 point scale).

Click here for more

Paper 65:

Detecting and Preventing Data Poisoning Attacks on AI Models

Date: 2025-03-12

Time: 11:55:01

Authors:

Halima I. Kure, Pradipta Sarkar, Ahmed B. Ndanusa, Augustine O. Nwajana

Summary:

- This paper investigates the critical issue of data poisoning attacks on AI models, a growing concern in the ever-evolving landscape of artificial intelligence and cybersecurity. As advanced technology systems become increasingly prevalent across various sectors, the need for robust defence mechanisms against adversarial attacks becomes paramount. The study aims to develop and evaluate novel techniques for detecting and preventing data poisoning attacks, focusing on both theoretical frameworks and practical applications. Through a comprehensive literature review, experimental validation using the CIFAR-10 and Insurance Claims datasets, and the development of innovative algorithms, this paper seeks to enhance the resilience of AI models against malicious data manipulation. The study explores various methods, including anomaly detection, robust optimization strategies, and ensemble learning, to identify and mitigate the effects of poisoned data during model training. Experimental results indicate that data poisoning significantly degrades model performance, reducing classification accuracy by up to 27% in image recognition tasks (CIFAR-10) and 22% in fraud detection models (Insurance Claims dataset). The proposed defence mechanisms, including statistical anomaly detection and adversarial training, successfully mitigated poisoning effects, improving model robustness and restoring accuracy levels by an average of 15-20%. The findings further demonstrate that ensemble learning techniques provide an additional layer of resilience, reducing false positives and false negatives caused by

adversarial data injections.

Click here for more

Paper 66:

SSD: A State-based Stealthy Backdoor Attack For Navigation System in UAV   Route Planning

Date: 2025-03-12

Time: 10:19:58

Authors:

Zhaoxuan Wang, Yang Li, Jie Zhang, Xingshuo Han, Kangbo Liu, Lyu Yang, yuan Zhou, Tianwei Zhang, Quan Pan

Summary:

- Unmanned aerial vehicles (UAVs) are increasingly employed to perform high-risk tasks that require minimal human intervention. However, UAVs face escalating cybersecurity threats, particularly from GNSS spoofing attacks. While previous studies have extensively investigated the impacts of GNSS spoofing on UAVs, few have focused on its effects on specific tasks. Moreover, the influence of UAV motion states on the assessment of network security risks is often overlooked. To address these gaps, we first provide a detailed evaluation of how motion states affect the effectiveness of network attacks. We demonstrate that nonlinear motion states not only enhance the effectiveness of position spoofing in GNSS spoofing attacks but also reduce the probability of speed-related attack detection. Building upon this, we propose a state-triggered backdoor attack method (SSD) to deceive GNSS systems and assess its risk to trajectory planning tasks. Extensive validation of SSD's effectiveness and stealthiness is conducted. Experimental results show that, with appropriately tuned hyperparameters, SSD significantly increases positioning errors and the risk of task failure, while maintaining 100% stealth across three state-of-the-art detectors.

Click here for more

Paper 67:

Automated Consistency Analysis of LLMs

Date: 2025-03-10

Time: 18:14:34

Authors:

Aditya Patwardhan, Vivek Vaidya, Ashish Kundu

Summary:

- Generative AI (Gen AI) with large language models (LLMs) are being

widely adopted across the industry, academia and government. Cybersecurity is one of the key sectors where LLMs can be and/or are already being used. There are a number of problems that inhibit the adoption of trustworthy Gen AI and LLMs in cybersecurity and such other critical areas. One of the key challenge to the trustworthiness and reliability of LLMs is: how consistent an LLM is in its responses? In this paper, we have analyzed and developed a formal definition of consistency of responses of LLMs. We have formally defined what is consistency of responses and then develop a framework for consistency evaluation. The paper proposes two approaches to validate consistency: self-validation, and validation across multiple LLMs. We have carried out extensive experiments for several LLMs such as GPT4oMini, GPT3.5, Gemini, Cohere, and Llama3, on a security benchmark consisting of several cybersecurity questions: informational and situational. Our experiments corroborate the fact that even though these LLMs are being considered and/or already being used for several cybersecurity tasks today, they are often inconsistent in their responses, and thus are untrustworthy and unreliable for cybersecurity.

Click here for more

Paper 68:

Operationalizing Cybersecurity Knowledge: Design, Implementation & Evaluation of a Knowledge Management System for CACAO Playbooks

Date: 2025-03-10

Time: 15:24:39

Authors:

Orestis Tsirakis, Konstantinos Fysarakis, Vasileios Mavroeidis, Ioannis Papaefstathiou

Summary:

- Modern cybersecurity threats are growing in complexity, targeting increasingly intricate & interconnected systems. To effectively defend against these evolving threats, security teams utilize automation & orchestration to enhance response efficiency and consistency. In that sense, cybersecurity playbooks are key enablers, providing a structured, reusable, and continuously improving approach to incident response, enabling organizations to codify requirements, domain expertise, and best practices and automate decision-making processes to the extent possible. The emerging Collaborative Automated Course of Action Operations (CACAO) standard defines a common machineprocessable schema for cybersecurity playbooks, facilitating interoperability for their exchange and ensuring the ability to orchestrate and automate cybersecurity operations. However, despite its potential and the fact that it is a relatively new standardization work, there is a lack of tools to support its adoption and, in particular, the management & lifecycle development of CACAO playbooks, limiting their practical deployment. Motivated by the above, this work presents the design, development, and evaluation of a Knowledge Management System (KMS) for managing CACAO cybersecurity playbooks throughout their lifecycle, providing essential tools to streamline playbook management. Using open technologies & standards, the proposed approach fosters standards-based interoperability & enhances the usability of state-of-the-art cybersecurity orchestration & automation primitives. To encourage adoption, the resulting implementation is released as open-source, which, to the extent of our knowledge, comprises the first publicly available & documented work in this domain, supporting the broader uptake of CACAO playbooks & promoting the widespread use of interoperable automation and orchestration mechanisms in cybersecurity operations.

Click here for more

Paper 69:

Creating Cybersecurity Regulatory Mechanisms, as Seen Through EU and US Law

Date: 2025-03-10

Time: 12:33:33

Authors:

Kaspar Rosager Ludvigsen

Summary:

- Because digital devices and systems are widely used in all aspects of society, the risk of adversaries creating cyberattacks on a similar level remains high. As such, regulation of these aspects must follow, which is the domain of cybersecurity. Because this topic is worldwide, different jurisdictions should take inspiration from successful techniques elsewhere, with the European Union and the US being the most experienced and long-standing. What can be derived from their approaches separately to be used in other democratic jurisdictions, and what happens when we compare them with this pragmatic approach in mind? Cybersecurity is oddly enough quite well understood in most jurisdictions worldwide. However, concept comprehension cannot enforce

or create compliance, hence the need for good regulatory approaches. The comparative legal analysis of the EU and the US show that there are large differences in definitions and enforcement, but some concepts are repeated in both jurisdictions. These can be further refined to become derivable principles, which can be used to inspire legislation in any democratic jurisdiction. They are: Voluntary Cooperation, Adaptable Definitions, Strong-arm Authorities, Mandated Computer Emergency Response Teams, and Effective Sanctions. These 5 principles are not exhaustive but combine classic regulatory and practical lessons from these two jurisdictions.

Click here for more

Paper 70:

Resource Constraint Estimation of In-Quantum Implementated Mini-AES

Date: 2025-03-09

Time: 11:44:11

Authors:

Syed Shahmir, Ghulam Murtaza, Ala-Al-Fuqaha, Saif Al-Kuwari, Tasawar Abbas

Summary:

- The advancement in quantum technology has brought the implementation of Grover's Search attack on cybersecurity algorithms much closer to reality. For such tasks, the Hilbert space has to be designed to make the relation between the input and output of these cryptographic algorithms in quantum circuits. Also, these algorithms are tested on current quantum simulators such as Qiskit, where memory constraints and limited processing power are hurdles. Here, we present an easy-to-implement method using a conventional arithmetic number field approach which can be applied to quantum circuits by CNOT and Toffoli gates, while focusing on resource constraints. Also, we give a Python code that can generate a Qiskit code for the quantum implementation of similar cryptographic S-boxes using the CNOT and 3-Tofflie gates by using simple logic presented in this paper.

Click here for more

Paper 71:

Deep Learning-Driven Malware Classification with API Call Sequence Analysis and Concept Drift Handling

Date: 2025-03-08

Time: 15:10:45

Authors:

Bishwajit Prasad Gond, Durga Prasad Mohapatra

Summary:

- Malware classification in dynamic environments presents a significant challenge due to concept drift, where the statistical properties of malware data evolve over time, complicating detection efforts. To address this issue, we propose a deep learning framework enhanced with a genetic algorithm to improve malware classification accuracy and adaptability. Our approach incorporates mutation operations and fitness score evaluations within genetic algorithms to continuously refine the deep learning model, ensuring robustness against evolving malware threats. Experimental results demonstrate that this hybrid method significantly enhances classification performance and adaptability, outperforming traditional static models. Our proposed approach offers a promising solution for real-time malware classification in ever-changing cybersecurity landscapes.

Click here for more

Paper 72:

Double Backdoored: Converting Code Large Language Model Backdoors to Traditional Malware via Adversarial Instruction Tuning Attacks

Date: 2025-03-07

Time: 00:46:35

Authors:

Md Imran Hossen, Sai Venkatesh Chilukoti, Liqun Shan, Sheng Chen, Yinzhi Cao, Xiali Hei

Summary:

- Instruction-tuned Large Language Models designed for coding tasks are increasingly employed as AI coding assistants. However, the cybersecurity vulnerabilities and implications arising from the widespread integration of these models are not yet fully understood due to limited research in this domain. This work investigates novel techniques for transitioning backdoors from the AI/ML domain to traditional computer malware, shedding light on the critical intersection of AI and cyber/software security. To explore this intersection, we present MalInstructCoder, a framework designed to comprehensively assess the cybersecurity vulnerabilities of instruction-tuned Code LLMs. MalInstructCoder introduces an automated data poisoning pipeline to inject malicious code snippets into benign code, poisoning instruction fine-tuning data while maintaining

functional validity. It presents two practical adversarial instruction tuning attacks with real-world security implications: the clean prompt poisoning attack and the backdoor attack. These attacks aim to manipulate Code LLMs to generate code incorporating malicious or harmful functionality under specific attack scenarios while preserving intended functionality. We conduct a comprehensive investigation into the exploitability of the code-specific instruction tuning process involving three state-of-the-art Code LLMs: CodeLlama, DeepSeek-Coder, and StarCoder2. Our findings reveal that these models are highly vulnerable to our attacks. Specifically, the clean prompt poisoning attack achieves the ASR@1 ranging from over 75% to 86% by poisoning only 1% (162 samples) of the instruction fine-tuning dataset. Similarly, the backdoor attack achieves the ASR@1 ranging from 76% to 86% with a 0.5% poisoning rate. Our study sheds light on the critical cybersecurity risks posed by instruction-tuned Code LLMs and highlights the urgent need for robust defense mechanisms.

Click here for more

Paper 73:

Matrix Factorization for Inferring Associations and Missing Links

Date: 2025-03-06

Time: 18:22:46

Authors:

Ryan Barron, Maksim E. Eren, Duc P. Truong, Cynthia Matuszek, James Wendelberger, Mary F. Dorn, Boian Alexandrov

Summary:

- Missing link prediction is a method for network analysis, with applications in recommender systems, biology, social sciences, cybersecurity, information retrieval, and Artificial Intelligence (AI) reasoning in Knowledge Graphs. Missing link prediction identifies unseen but potentially existing connections in a network by analyzing the observed patterns and relationships. In proliferation detection, this supports efforts to identify and characterize attempts by state and non-state actors to acquire nuclear weapons or associated technology - a notoriously challenging but vital mission for global security. Dimensionality reduction techniques like Non-Negative Matrix Factorization (NMF) and Logistic Matrix Factorization (LMF) are effective but require selection of the matrix rank parameter, that is, of the number of hidden features, k, to avoid over/under-fitting. We

introduce novel Weighted (WNMFk), Boolean (BNMFk), and Recommender (RNMFk) matrix factorization methods, along with ensemble variants incorporating logistic factorization, for link prediction. Our methods integrate automatic model determination for rank estimation by evaluating stability and accuracy using a modified bootstrap methodology and uncertainty quantification (UQ), assessing prediction reliability under random perturbations. We incorporate Otsu threshold selection and k-means clustering for Boolean matrix factorization, comparing them to coordinate descent-based Boolean thresholding. Our experiments highlight the impact of rank k selection, evaluate model performance under varying test-set sizes, and demonstrate the benefits of UQ for reliable predictions using abstention. We validate our methods on three synthetic datasets (Boolean and uniformly distributed) and benchmark them against LMF and symmetric LMF (symLMF) on five real-world protein-protein interaction networks, showcasing an improved prediction performance.

Click here for more

Paper 74:

Guidelines for Applying RL and MARL in Cybersecurity Applications

Date: 2025-03-06

Time: 09:46:16

Authors:

Vasilios Mavroudis, Gregory Palmer, Sara Farmer, Kez Smithson Whitehead, David Foster, Adam Price, Ian Miles, Alberto Caron, Stephen Pasteris

Summary:

- Reinforcement Learning (RL) and Multi-Agent Reinforcement Learning (MARL) have emerged as promising methodologies for addressing challenges in automated cyber defence (ACD). These techniques offer adaptive decision-making capabilities in high-dimensional, adversarial environments. This report provides a structured set of guidelines for cybersecurity professionals and researchers to assess the suitability of RL and MARL for specific use cases, considering factors such as explainability, exploration needs, and the complexity of multi-agent coordination. It also discusses key algorithmic approaches, implementation challenges, and real-world constraints, such as data scarcity and adversarial interference. The report further outlines open research questions, including policy optimality, agent

cooperation levels, and the integration of MARL systems into operational cybersecurity frameworks. By bridging theoretical advancements and practical deployment, these guidelines aim to enhance the effectiveness of AI-driven cyber defence strategies.

Click here for more

Paper 75:

Detecting and Deterring Manipulation in a Cognitive Hierarchy

Date: 2025-03-06

Time: 09:39:06

Authors:

Nitay Alon, Joseph M. Barnby, Stefan Sarkadi, Lion Schulz, Jeffrey S. Rosenschein, Peter Dayan

Summary:

- Social agents with finitely nested opponent models are vulnerable to manipulation by agents with deeper reasoning and more sophisticated opponent modelling. This imbalance, rooted in logic and the theory of recursive modelling frameworks, cannot be solved directly. We propose a computational framework, $\aleph$-IPOMDP, augmenting model-based RL agents' Bayesian inference with an anomaly detection algorithm and an out-of-belief policy. Our mechanism allows agents to realize they are being deceived, even if they cannot understand how, and to deter opponents via a credible threat. We test this framework in both a mixed-motive and zero-sum game. Our results show the $\aleph$ mechanism's effectiveness, leading to more equitable outcomes and less exploitation by more sophisticated agents. We discuss implications for AI safety, cybersecurity, cognitive science, and psychiatry.

Click here for more

Paper 76:

Unsupervised anomaly detection on cybersecurity data streams: a case   with BETH dataset

Date: 2025-03-06

Time: 07:45:48

Authors:

Evgeniy Eremin

Summary:

- In modern world the importance of cybersecurity of various systems is increasing from year to year. The number of information security events generated by information security tools grows up with the development of the IT infrastructure. At the same time, the cyber threat landscape does not remain constant, and monitoring should take into account both already known attack indicators and those for which there are no signature rules in information security products of various classes yet. Detecting anomalies in large cybersecurity data streams is a complex task that, if properly addressed, can allow for timely response to atypical and previously unknown cyber threats. The possibilities of using of offline algorithms may be limited for a number of reasons related to the time of training and the frequency of retraining. Using stream learning algorithms for solving this task is capable of providing near-real-time data processing. This article examines the results of ten algorithms from three Python stream machine-learning libraries on BETH dataset with cybersecurity events, which contains information about the creation, cloning, and destruction of operating system processes collected using extended eBPF. ROC-AUC metric and total processing time of processing with these algorithms are presented. Several combinations of features and the order of events are considered. In conclusion, some mentions are given about the most promising algorithms and possible directions for further research are outlined.

Click here for more

Paper 77:

PacketCLIP: Multi-Modal Embedding of Network Traffic and Language for Cybersecurity Reasoning

Date: 2025-03-05

Time: 18:58:58

Authors:

Ryozo Masukawa, Sanggeon Yun, Sungheon Jeong, Wenjun Huang, Yang Ni, Ian Bryant, Nathaniel D. Bastian, Mohsen Imani

Summary:

- Traffic classification is vital for cybersecurity, yet encrypted traffic poses significant challenges. We present PacketCLIP, a multi-modal framework combining packet data with natural language semantics through contrastive pretraining and hierarchical Graph Neural Network (GNN) reasoning. PacketCLIP integrates semantic reasoning with efficient classification, enabling robust detection of anomalies in encrypted network flows. By aligning textual descriptions with packet behaviors, it offers enhanced interpretability, scalability, and practical applicability across diverse security scenarios. PacketCLIP achieves a 95% mean AUC, outperforms baselines by 11.6%, and reduces model size by 92%, making it ideal for real-time anomaly detection. By bridging advanced machine learning techniques and practical cybersecurity needs, PacketCLIP provides a foundation for scalable, efficient, and interpretable solutions to tackle encrypted traffic classification and network intrusion detection challenges in resource-constrained environments.

Click here for more

Paper 78:

Enhancing Cybersecurity in Critical Infrastructure with LLM-Assisted Explainable IoT Systems

Date: 2025-03-05

Time: 04:53:07

Authors:

Ashutosh Ghimire, Ghazal Ghajari, Karma Gurung, Love K. Sah, Fathi Amsaad

Summary:

- Ensuring the security of critical infrastructure has become increasingly vital with the proliferation of Internet of Things (IoT) systems. However, the heterogeneous nature of IoT data and the lack of human-comprehensible insights from anomaly detection models remain significant challenges. This paper presents a hybrid framework that combines numerical anomaly detection using Autoencoders with Large Language Models (LLMs) for enhanced preprocessing and interpretability. Two preprocessing approaches are implemented: a traditional method utilizing Principal Component Analysis (PCA) to reduce dimensionality and an LLM-assisted method where GPT-4 dynamically recommends feature selection, transformation, and encoding strategies. Experimental results on the KDDCup99 10% corrected dataset demonstrate that the LLM-assisted preprocessing pipeline significantly improves anomaly detection performance. The macro-average F1 score increased from 0.49 in the traditional PCA-based approach to 0.98 with LLM-driven insights. Additionally, the LLM generates natural language explanations for detected anomalies, providing contextual insights into their causes and implications. This framework highlights the synergy between numerical AI models and LLMs, delivering an accurate, interpretable, and efficient solution for IoT cybersecurity in critical infrastructure.

Click here for more

Paper 79:

AttackSeqBench: Benchmarking Large Language Models' Understanding of Sequential Patterns in Cyber Attacks

Date: 2025-03-05

Time: 04:25:21

Authors:

Javier Yong, Haokai Ma, Yunshan Ma, Anis Yusof, Zhenkai Liang, Ee-Chien Chang

Summary:

- The observations documented in Cyber Threat Intelligence (CTI) reports play a critical role in describing adversarial behaviors, providing valuable insights for security practitioners to respond to evolving threats. Recent advancements of Large Language Models (LLMs) have demonstrated significant potential in various cybersecurity applications, including CTI report understanding and attack knowledge graph construction. While previous works have proposed benchmarks that focus on the CTI extraction ability of LLMs, the sequential characteristic of adversarial behaviors within CTI reports remains largely unexplored, which holds considerable significance in developing a comprehensive understanding of how adversaries operate. To address this gap, we introduce AttackSeqBench, a benchmark tailored to systematically evaluate LLMs' capability to understand and reason attack sequences in CTI reports. Our benchmark encompasses three distinct Question Answering (QA) tasks, each task focuses on the varying granularity in adversarial behavior. To alleviate the laborious effort of QA construction, we carefully design an automated dataset construction pipeline to create scalable and well-formulated QA datasets based on real-world CTI reports. To ensure the quality of our dataset, we adopt a hybrid approach of combining human evaluation and systematic evaluation metrics. We conduct extensive experiments and analysis with both fast-thinking and slow-thinking LLMs, while highlighting their strengths and limitations in analyzing the sequential patterns in cyber attacks. The overarching goal of this work is to provide a benchmark that advances LLM-driven CTI report understanding and fosters its application in real-world cybersecurity operations. Our dataset and code are available at https://github.com/Javiery3889/AttackSeqBench .

Click here for more

Paper 80:

Network Anomaly Detection for IoT Using Hyperdimensional Computing on NSL-KDD

Date: 2025-03-04

Time: 22:19:26

Authors:

Ghazal Ghajari, Ashutosh Ghimire, Elaheh Ghajari, Fathi Amsaad

Summary:

- With the rapid growth of IoT devices, ensuring robust network security has become a critical challenge. Traditional intrusion detection systems (IDSs) often face limitations in detecting sophisticated attacks within high-dimensional and complex data environments. This paper presents a novel approach to network anomaly detection using hyperdimensional computing (HDC) techniques, specifically applied to the NSL-KDD dataset. The proposed method leverages the efficiency of HDC in processing large-scale data to identify both known and unknown attack patterns. The model achieved an accuracy of 91.55% on the KDDTrain+ subset, outperforming traditional approaches. These comparative evaluations underscore the model's superior performance, highlighting its potential in advancing anomaly detection for IoT networks and contributing to more secure and intelligent cybersecurity solutions.

Click here for more

Paper 81:

A Kolmogorov-Arnold Network for Explainable Detection of Cyberattacks on EV Chargers

Date: 2025-03-04

Time: 05:06:39

Authors:

Ahmad Mohammad Saber, Max Mauro Dias Santos, Mohammad Al Janaideh, Amr Youssef, Deepa Kundur

Summary:

- The increasing adoption of Electric Vehicles (EVs) and the expansion of charging infrastructure and their reliance on communication expose Electric Vehicle Supply Equipment (EVSE) to cyberattacks. This paper presents a novel Kolmogorov-Arnold Network (KAN)-based framework for detecting cyberattacks on EV chargers using only power consumption measurements. Leveraging the KAN's capability to model nonlinear, high-dimensional functions and its inherently interpretable architecture, the framework effectively differentiates between normal

and malicious charging scenarios. The model is trained offline on a comprehensive dataset containing over 100,000 cyberattack cases generated through an experimental setup. Once trained, the KAN model can be deployed within individual chargers for real-time detection of abnormal charging behaviors indicative of cyberattacks. Our results demonstrate that the proposed KAN-based approach can accurately detect cyberattacks on EV chargers with Precision and F1-score of 99% and 92%, respectively, outperforming existing detection methods. Additionally, the proposed KANs's enable the extraction of mathematical formulas representing KAN's detection decisions, addressing interpretability, a key challenge in deep learning-based cybersecurity frameworks. This work marks a significant step toward building secure and explainable EV charging infrastructure.

Click here for more

Paper 82:

Survey Perspective: The Role of Explainable AI in Threat Intelligence

Date: 2025-03-03

Time: 21:39:15

Authors:

Nidhi Rastogi, Devang Dhanuka, Amulya Saxena, Pranjal Mairal, Le Nguyen

Summary:

- The increasing reliance on AI-based security tools in Security Operations Centers (SOCs) has transformed threat detection and response, yet analysts frequently struggle with alert overload, false positives, and lack of contextual relevance. The inability to effectively analyze AI-generated security alerts lead to inefficiencies in incident response and reduces trust in automated decision-making. In this paper, we show results and analysis of our investigation of how SOC analysts navigate AI-based alerts, their challenges with current security tools, and how explainability (XAI) integrated into their security workflows has the potential to become an effective decision support. In this vein, we conducted an industry survey. Using the survey responses, we analyze how security analysts' process, retrieve, and prioritize alerts. Our findings indicate that most analysts have not yet adopted XAI-integrated tools, but they express high interest in attack attribution, confidence scores, and feature contribution explanations to improve interpretability, and

triage efficiency. Based on our findings, we also propose practical design recommendations for XAI-enhanced security alert systems, enabling AI-based cybersecurity solutions to be more transparent, interpretable, and actionable.

Click here for more

Paper 83:

PhishVQC: Optimizing Phishing URL Detection with Correlation Based Feature Selection and Variational Quantum Classifier

Date: 2025-03-03

Time: 18:28:01

Authors:

Md. Farhan Shahriyar, Gazi Tanbhir, Abdullah Md Raihan Chy, Mohammed Abdul Al Arafat Tanzin, Md. Jisan Mashrafi

Summary:

- Phishing URL detection is crucial in cybersecurity as malicious websites disguise themselves to steal sensitive infor mation. Traditional machine learning techniques struggle to per form well in complex real-world scenarios due to large datasets and intricate patterns. Motivated by quantum computing, this paper proposes using Variational Quantum Classifiers (VQC) to enhance phishing URL detection. We present PhishVQC, a quantum model that combines quantum feature maps and vari ational ansatzes such as RealAmplitude and EfficientSU2. The model is evaluated across two experimental setups with varying dataset sizes and feature map repetitions. PhishVQC achieves a maximum macro average F1-score of 0.89, showing a 22% improvement over prior studies. This highlights the potential of quantum machine learning to improve phishing detection accuracy. The study also notes computational challenges, with execution wall times increasing as dataset size grows.

Click here for more

Paper 84:

Forecasting Frontier Language Model Agent Capabilities

Date: 2025-03-03

Time: 17:11:16

Authors:

Govind Pimpale, Axel Højmark, Jérémy Scheurer, Marius Hobbhahn

Summary:

- As Language Models (LMs) increasingly operate as autonomous agents,

accurately forecasting their capabilities becomes crucial for societal preparedness. We evaluate six forecasting methods that predict downstream capabilities of LM agents. We use "one-step" approaches that predict benchmark scores from input metrics like compute or model release date directly or "two-step" approaches that first predict an intermediate metric like the principal component of cross-benchmark performance (PC-1) and human-evaluated competitive Elo ratings. We evaluate our forecasting methods by backtesting them on a dataset of 38 LMs from the OpenLLM 2 leaderboard. We then use the validated two-step approach (Release Date$\to$Elo$\to$Benchmark) to predict LM agent performance for frontier models on three benchmarks: SWE-Bench Verified (software development), Cybench (cybersecurity assessment), and RE-Bench (ML research engineering). Our forecast predicts that by the beginning of 2026, non-specialized LM agents with low capability elicitation will reach a success rate of 54% on SWE-Bench Verified, while state-of-the-art LM agents will reach an 87% success rate. Our approach does not account for recent advances in inference-compute scaling and might thus be too conservative.

Click here for more

Paper 85:

Quantum machine learning algorithms for anomaly detection: A review

Date: 2025-03-03

Time: 16:01:45

Authors:

Sebastiano Corli, Lorenzo Moro, Daniele Dragoni, Massimiliano Dispenza, Enrico Prati

Summary:

- The advent of quantum computers has justified the development of quantum machine learning algorithms , based on the adaptation of the principles of machine learning to the formalism of qubits. Among such quantum algorithms, anomaly detection represents an important problem crossing several disciplines from cybersecurity, to fraud detection to particle physics. We summarize the key concepts involved in quantum computing, introducing the formal concept of quantum speed up. The review provides a structured map of anomaly detection based on quantum machine learning. We have grouped existing algorithms according to the different learning methods, namely quantum supervised, quantum unsupervised and quantum reinforcement learning, respectively. We

provide an estimate of the hardware resources to provide sufficient computational power in the future. The review provides a systematic and compact understanding of the techniques belonging to each category. We eventually provide a discussion on the computational complexity of the learning methods in real application domains.

Click here for more

Paper 86:

Jailbreaking Generative AI: Empowering Novices to Conduct Phishing Attacks

Date: 2025-03-03

Time: 10:51:10

Authors:

Rina Mishra, Gaurav Varshney, Shreya Singh

Summary:

- The rapid advancements in generative AI models, such as ChatGPT, have introduced both significant benefits and new risks within the cybersecurity landscape. This paper investigates the potential misuse of the latest AI model, ChatGPT-4o Mini, in facilitating social engineering attacks, with a particular focus on phishing, one of the most pressing cybersecurity threats today. While existing literature primarily addresses the technical aspects, such as jailbreaking techniques, none have fully explored the free and straightforward execution of a comprehensive phishing campaign by novice users using ChatGPT-4o Mini. In this study, we examine the vulnerabilities of AI-driven chatbot services in 2025, specifically how methods like jailbreaking and reverse psychology can bypass ethical safeguards, allowing ChatGPT to generate phishing content, suggest hacking tools, and assist in carrying out phishing attacks. Our findings underscore the alarming ease with which even inexperienced users can execute sophisticated phishing campaigns, emphasizing the urgent need for stronger cybersecurity measures and heightened user awareness in the age of AI.

Click here for more

Paper 87:

The Road Less Traveled: Investigating Robustness and Explainability in CNN Malware Detection

Date: 2025-03-03

Time: 10:42:00

Authors:

Matteo Brosolo, Vinod Puthuvath, Mauro Conti

Summary:

- Machine learning has become a key tool in cybersecurity, improving both attack strategies and defense mechanisms. Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated high accuracy in detecting malware images generated from binary data. However, the decision-making process of these black-box models remains difficult to interpret. This study addresses this challenge by integrating quantitative analysis with explainability tools such as Occlusion Maps, HiResCAM, and SHAP to better understand CNN behavior in malware classification. We further demonstrate that obfuscation techniques can reduce model accuracy by up to 50%, and propose a mitigation strategy to enhance robustness. Additionally, we analyze heatmaps from multiple tests and outline a methodology for identification of artifacts, aiding researchers in conducting detailed manual investigations. This work contributes to improving the interpretability and resilience of deep learning-based intrusion detection systems

Click here for more

Paper 88:

SynGhost: Invisible and Universal Task-agnostic Backdoor Attack via Syntactic Transfer

Date: 2025-03-03

Time: 06:34:48

Authors:

Pengzhou Cheng, Wei Du, Zongru Wu, Fengwei Zhang, Libo Chen, Zhuosheng Zhang, Gongshen Liu

Summary:

- Although pre-training achieves remarkable performance, it suffers from task-agnostic backdoor attacks due to vulnerabilities in data and training mechanisms. These attacks can transfer backdoors to various downstream tasks. In this paper, we introduce $\mathtt{maxEntropy}$, an entropy-based poisoning filter that mitigates such risks. To overcome the limitations of manual target setting and explicit triggers, we propose $\mathtt{SynGhost}$, an invisible and universal task-agnostic backdoor attack via syntactic transfer, further exposing vulnerabilities in pre-trained language models (PLMs). Specifically, $\mathtt{SynGhost}$ injects multiple syntactic backdoors into the pre-training space through corpus poisoning, while preserving the PLM's

pre-training capabilities. Second, $\mathtt{SynGhost}$ adaptively selects optimal targets based on contrastive learning, creating a uniform distribution in the pre-training space. To identify syntactic differences, we also introduce an awareness module to minimize interference between backdoors. Experiments show that $\mathtt{SynGhost}$ poses significant threats and can transfer to various downstream tasks. Furthermore, $\mathtt{SynGhost}$ resists defenses based on perplexity, fine-pruning, and $\mathtt{maxEntropy}$. The code is available at https://github.com/Zhou-CyberSecurity-AI/SynGhost.

Click here for more

Paper 89:

Detecting Unsuccessful Students in Cybersecurity Exercises in Two Different Learning Environments

Date: 2025-03-02

Time: 18:15:48

Authors:

Valdemar Švábenský, Kristián Tká■ik, Aubrey Birdwell, Richard Weiss, Ryan S. Baker, Pavel ■eleda, Jan Vykopal, Jens Mache, Ankur Chattopadhyay

Summary:

- This full paper in the research track evaluates the usage of data logged from cybersecurity exercises in order to predict students who are potentially at risk of performing poorly. Hands-on exercises are essential for learning since they enable students to practice their skills. In cybersecurity, hands-on exercises are often complex and require knowledge of many topics. Therefore, students may miss solutions due to gaps in their knowledge and become frustrated, which impedes their learning. Targeted aid by the instructor helps, but since the instructor's time is limited, efficient ways to detect struggling students are needed. This paper develops automated tools to predict when a student is having difficulty. We formed a dataset with the actions of 313 students from two countries and two learning environments: KYPO CRP and EDURange. These data are used in machine learning algorithms to predict the success of students in exercises deployed in these environments. After extracting features from the data, we trained and cross-validated eight classifiers for predicting the exercise outcome and evaluated their predictive power. The contribution of this paper is comparing two approaches to feature

engineering, modeling, and classification performance on data from two learning environments. Using the features from either learning environment, we were able to detect and distinguish between successful and struggling students. A decision tree classifier achieved the highest balanced accuracy and sensitivity with data from both learning environments. The results show that activity data from cybersecurity exercises are suitable for predicting student success. In a potential application, such models can aid instructors in detecting struggling students and providing targeted help. We publish data and code for building these models so that others can adopt or adapt them.

Click here for more

Paper 90:

The Good, the Bad, and the (Un)Usable: A Rapid Literature Review on Privacy as Code

Date: 2025-03-02

Time: 17:05:13

Authors:

Nicolás E. Díaz Ferreyra, Sirine Khelifi, Nalin Arachchilage, Riccardo Scandariato

Summary:

- Privacy and security are central to the design of information systems endowed with sound data protection and cyber resilience capabilities. Still, developers often struggle to incorporate these properties into software projects as they either lack proper cybersecurity training or do not consider them a priority. Prior work has tried to support privacy and security engineering activities through threat modeling methods for scrutinizing flaws in system architectures. Moreover, several techniques for the automatic identification of vulnerabilities and the generation of secure code implementations have also been proposed in the current literature. Conversely, such as-code approaches seem under-investigated in the privacy domain, with little work elaborating on (i) the automatic detection of privacy properties in source code or (ii) the generation of privacy-friendly code. In this work, we seek to characterize the current research landscape of Privacy as Code (PaC) methods and tools by conducting a rapid literature review. Our results suggest that PaC research is in its infancy, especially regarding the performance evaluation and usability assessment of the existing approaches. Based on these findings, we outline and discuss prospective research directions concerning empirical studies with software practitioners, the curation of benchmark datasets, and the role of generative AI technologies.

Click here for more

Paper 91:

CAGN-GAT Fusion: A Hybrid Contrastive Attentive Graph Neural Network for Network Intrusion Detection

Date: 2025-03-02

Time: 17:01:00

Authors:

Md Abrar Jahin, Shahriar Soudeep, M. F. Mridha, Raihan Kabir, Md Rashedul Islam, Yutaka Watanobe

Summary:

- Cybersecurity threats are growing, making network intrusion detection essential. Traditional machine learning models remain effective in resource-limited environments due to their efficiency, requiring fewer parameters and less computational time. However, handling short and highly imbalanced datasets remains challenging. In this study, we propose the fusion of a Contrastive Attentive Graph Network and Graph Attention Network (CAGN-GAT Fusion) and benchmark it against 15 other models, including both Graph Neural Networks (GNNs) and traditional ML models. Our evaluation is conducted on four benchmark datasets (KDD-CUP-1999, NSL-KDD, UNSW-NB15, and CICIDS2017) using a short and proportionally imbalanced dataset with a constant size of 5000 samples to ensure fairness in comparison. Results show that CAGN-GAT Fusion demonstrates stable and competitive accuracy, recall, and F1-score, even though it does not achieve the highest performance in every dataset. Our analysis also highlights the impact of adaptive graph construction techniques, including small changes in connections (edge perturbation) and selective hiding of features (feature masking), improving detection performance. The findings confirm that GNNs, particularly CAGN-GAT Fusion, are robust and computationally efficient, making them well-suited for resource-constrained environments. Future work will explore GraphSAGE layers and multiview graph construction techniques to further enhance adaptability and detection accuracy.

Click here for more

Paper 92:

CyberCScope: Mining Skewed Tensor Streams and Online Anomaly

Detection in Cybersecurity Systems

Date: 2025-03-02

Time: 12:17:24

Authors:

Kota Nakamura, Koki Kawabata, Shungo Tanaka, Yasuko Matsubara, Yasushi Sakurai

Summary:

- Cybersecurity systems are continuously producing a huge number of time-stamped events in the form of high-order tensors, such as {count; time, port, flow duration, packet size, . . . }, and so how can we detect anomalies/intrusions in real time? How can we identify multiple types of intrusions and capture their characteristic behaviors? The tensor data consists of categorical and continuous attributes and the data distributions of continuous attributes typically exhibit skew. These data properties require handling skewed infinite and finite dimensional spaces simultaneously. In this paper, we propose a novel streaming method, namely CyberCScope. The method effectively decomposes incoming tensors into major trends while explicitly distinguishing between categorical and skewed continuous attributes. To our knowledge, it is the first to compute hybrid skewed infinite and finite dimensional decomposition. Based on this decomposition, it streamingly finds distinct time-evolving patterns, enabling the detection of multiple types of anomalies. Extensive experiments on large-scale real datasets demonstrate that CyberCScope detects various intrusions with higher accuracy than state-of-the-art baselines while providing meaningful summaries for the intrusions that occur in practice.

Click here for more

Paper 93:

Transforming Cyber Defense: Harnessing Agentic and Frontier AI for Proactive, Ethical Threat Intelligence

Date: 2025-02-28

Time: 20:23:35

Authors:

Krti Tallam

Summary:

- In an era marked by unprecedented digital complexity, the cybersecurity landscape is evolving at a breakneck pace, challenging traditional defense paradigms. Advanced Persistent Threats (APTs) reveal inherent vulnerabilities in conventional security measures and underscore the urgent need for continuous, adaptive, and proactive strategies that seamlessly integrate human insight with cutting edge AI technologies. This manuscript explores how the convergence of agentic AI and Frontier AI is transforming cybersecurity by reimagining frameworks such as the cyber kill chain, enhancing threat intelligence processes, and embedding robust ethical governance within automated response systems. Drawing on real-world data and forward looking perspectives, we examine the roles of real time monitoring, automated incident response, and perpetual learning in forging a resilient, dynamic defense ecosystem. Our vision is to harmonize technological innovation with unwavering ethical oversight, ensuring that future AI driven security solutions uphold core human values of fairness, transparency, and accountability while effectively countering emerging cyber threats.

Click here for more

Paper 94:

Unmasking Stealthy Attacks on Nonlinear DAE Models of Power Grids

Date: 2025-02-28

Time: 15:27:58

Authors:

Abdallah Alalem Albustami, Ahmad F. Taha, Elias Bou-Harb

Summary:

- Smart grids are inherently susceptible to various types of malicious cyberattacks that have all been documented in the recent literature. Traditional cybersecurity research on power systems often utilizes simplified models that fail to capture the interactions between dynamic and steady-state behaviors, potentially underestimating the impact of cyber threats. This paper presents the first attempt to design and assess stealthy false data injection attacks (FDIAs) against nonlinear differential algebraic equation (NDAE) models of power networks. NDAE models, favored in industry for their ability to accurately capture both dynamic and steady-state behaviors, provide a more accurate representation of power system behavior by coupling dynamic and algebraic states. We propose novel FDIA strategies that simultaneously evade both dynamic and static intrusion detection systems while respecting the algebraic power flow and operational constraints inherent in NDAE models. We demonstrate how the coupling

between dynamic and algebraic states in NDAE models significantly restricts the attacker's ability to manipulate state estimates while maintaining stealthiness. This highlights the importance of using more comprehensive power system models in cybersecurity analysis and reveals potential vulnerabilities that may be overlooked in simplified representations. The proposed attack strategies are validated through simulations on the IEEE 39-bus system.

Click here for more

Paper 95:

Toward interoperable representation and sharing of disinformation incidents in cyber threat intelligence

Date: 2025-02-28

Time: 12:37:32

Authors:

Felipe Sánchez González, Javier Pastor-Galindo, José A. Ruipérez-Valiente

Summary:

- A key countermeasure in cybersecurity has been the development of standardized computational protocols for modeling and sharing cyber threat intelligence (CTI) between organizations, enabling a shared understanding of threats and coordinated global responses. However, while the cybersecurity domain benefits from mature threat exchange frameworks, there has been little progress in the automatic and interoperable sharing of knowledge about disinformation campaigns. This paper proposes an open-source disinformation threat intelligence framework for sharing interoperable disinformation incidents. This approach relies on i) the modeling of disinformation incidents with the DISARM framework (MITRE ATT&CK-based TTP modeling of disinformation attacks), ii) a custom mapping to STIX2 standard representation (computational data format), and iii) an exchange architecture (called DISINFOX) capable of using the proposed mapping with a centralized platform to store and manage disinformation incidents and CTI clients which consume the gathered incidents. The microservice-based implementation validates the framework with more than 100 real-world disinformation incidents modeled, stored, shared, and consumed successfully. To the best of our knowledge, this work is the first academic and technical effort to integrate disinformation threats in the CTI ecosystem.

Click here for more

Paper 96:

Cyber Defense Reinvented: Large Language Models as Threat Intelligence Copilots

Date: 2025-02-28

Time: 07:16:09

Authors:

Xiaoqun Liu, Jiacheng Liang, Qiben Yan, Muchao Ye, Jinyuan Jia, Zhaohan Xi

Summary:

- The exponential growth of cyber threat knowledge, exemplified by the expansion of databases such as MITRE-CVE and NVD, poses significant challenges for cyber threat analysis. Security professionals are increasingly burdened by the sheer volume and complexity of information, creating an urgent need for effective tools to navigate, synthesize, and act on large-scale data to counter evolving threats proactively. However, conventional threat intelligence tools often fail to scale with the dynamic nature of this data and lack the adaptability to support diverse threat intelligence tasks. In this work, we introduce CYLENS, a cyber threat intelligence copilot powered by large language models (LLMs). CYLENS is designed to assist security professionals throughout the entire threat management lifecycle, supporting threat attribution, contextualization, detection, correlation, prioritization, and remediation. To ensure domain expertise, CYLENS integrates knowledge from 271,570 threat reports into its model parameters and incorporates six specialized NLP modules to enhance reasoning capabilities. Furthermore, CYLENS can be customized to meet the unique needs of different or ganizations, underscoring its adaptability. Through extensive evaluations, we demonstrate that CYLENS consistently outperforms industry-leading LLMs and state-of-the-art cybersecurity agents. By detailing its design, development, and evaluation, this work provides a blueprint for leveraging LLMs to address complex, data-intensive cybersecurity challenges.

Click here for more