Universal Library of Innovative Research and Studies Research Article

ISSN: 3065-0003

Open Access | PP: 01-06

DOI: https://doi.org/10.70315/uloap.ulirs.2023.001

Fundamental Principles for Building an Effective Detection Engineering System

Alina Gaifulina

Manager, Cyber Fusion Center and Incident Response at MorganFranklin Consulting, Prague, Czech Republic.

Abstract

The article presents a comprehensive analysis of the principles and architectures of detection engineering as a methodological foundation for building modern systems for identifying and preventing cyberattacks. The study employs an interdisciplinary approach that integrates machine learning, provenance data graph analysis, and the MITRE ATT&CK taxonomy. The analysis is based on recent international publications reflecting the shift from signature-based methods to context-oriented models capable of adaptive self-learning and feedback with monitoring centers. The key components of the detection engineering cycle are examined, including data standardization under the STIX 2.1 format, correlation of system calls with MITRE tactics, the use of directed graphs for behavioral modeling, and the implementation of adaptive thresholds in classification algorithms. Particular attention is given to identifying implementation barriers related to log heterogeneity, lack of realistic datasets, and the dependence of models on fixed metrics. The novelty of the study lies in formulating the principles of building a detection engineering pipeline that unites machine learning, behavioral analytics, and organizational mechanisms of SOC within a single adaptive security framework. The practical significance of the research consists in justifying approaches that reduce false positives, improve the interpretability of detections, and enhance system resilience to APT-class attacks. The article will be useful for researchers and practitioners in cybersecurity, developers of analytical platforms, and specialists involved in the design of monitoring and incident response centers.

Keywords: *Detection System, Data Analysis, Information Standardization, Behavior Modeling, Automation, Security, Monitoring.*

INTRODUCTION

Modern cybersecurity systems are undergoing profound changes driven by the increasing complexity of corporate infrastructures, the proliferation of cloud services, and the growing number of distributed components. Against this backdrop, traditional attack detection systems are gradually losing their effectiveness. They scale poorly, fail to provide a holistic context, and generate a high level of false positives [2]. The scalability of data, the heterogeneity of logs, and the rapid evolution of attacker tactics are leading to a decrease in the accuracy and speed of incident response. Under these conditions, it is necessary to transition from disparate monitoring tools to holistic detection engineering—a system based on standardized threat models, multi-layered data correlation, and transparent decision-making principles.

The relevance of this research is determined by the fact that existing intrusion detection systems are unable to effectively cope with the growing volumes of telemetry and the complexity of modern cyberattacks, including targeted campaigns and supply chain attacks. The problem is exacerbated by the lack of a unified ontological framework that integrates data from different levels, from operating system kernel events to MITRE ATT&CK tactics. As a result, analysts face information overload, and corporate SOCs face a loss of stability and explainability of detections [5]. Ensuring reliable, reproducible, and contextually meaningful detection is becoming a key factor in the effectiveness of the entire defense system.

According to international recommendations from NIST SP 800-94, the MITRE D3FEND initiative, and analytical forecasts from Gartner Cybersecurity Trends 2024, global cybersecurity practice is shifting from reactive monitoring to engineering-oriented detection methods that provide adaptive learning and contextual threat correlation.

The problem lies in the fact that most existing approaches still focus on technical aspects—thresholds, signatures, machine learning algorithms—while ignoring the methodology for building a unified detection engineering cycle: from data normalization to the visualization of cause-and-effect relationships. Principles are needed that can unite machine learning, MITRE taxonomies, and graph-based data provenance models into a coherent analytical loop.

The research hypothesis is that the integration of machine learning and data provenance analysis methods into a single analytical loop allows for an increase in the accuracy, reproducibility, and interpretability of threat detection processes compared to traditional signature-based and threshold-based approaches.

The purpose of this study is to analyze the fundamental principles for building an effective detection engineering system that combines machine learning models, the MITRE ATT&CK taxonomy, and intrusion detection system architectures based on data provenance analysis; to identify the key mechanisms that increase the accuracy, explainability, and resilience of detection; and to determine the development directions for detection engineering as a methodological basis for modern incident monitoring and response centers.

MATERIALS AND METHODS

The methodological foundation of this study is formed at the intersection of cyber threat analysis theory, detection engineering, and applied methods for building intelligent monitoring systems. The research is based on modern works that examine various approaches to integrating machine learning, MITRE ATT&CK taxonomies, and graph-based data provenance models into a single analytical loop.

The study by Disha R. A. [1] conducted a comparative analysis of machine learning models for attack detection systems, showing that the use of a Gini Impurity-based Weighted Random Forest (GIWRF) method provides the highest classification accuracy while reducing the number of features. The work by Georgiadou A. [2] proposed a cultural-organizational model for assessing MITRE ATT&CK risks, linking the maturity of a corporate security culture with the completeness of implemented protective measures, which expanded the methodology for evaluating the effectiveness of monitoring centers. The study by González-Granadillo G. [3] systematized the development trends of SIEM platforms in critical infrastructures, identifying key requirements for the standardization of data sources and the automation of event correlation.

The work by Khraisat A. [4] provided a review of methods and datasets for building attack detection systems, proposing a classification of IDS technologies based on architecture and the type of algorithm used. The study by Pahlevan M. [5] developed a model for the secure exchange of threat intelligence based on blockchain technology and the TAXII 2.1 standard, ensuring integrity and trust in the transmission of information between participants. The work by Rosso M. [6] presented the SAIBERSOC methodology, describing the operational structure of security monitoring centers and the role of detection engineering in shaping their technological maturity.

The study by Sacher-Boldewin D. [7] revealed the intelligent lifecycle of active cyber defense processes, including the phases of analysis, learning, and adaptation of detection models. The work by Son V. N. [8] demonstrated the application of the MITRE ATT&CK taxonomy for describing malware behavior, which allowed for the formalization of attack patterns at the system event level. The study by Xiong W. [9] developed a methodology for modeling cyber threats based on the MITRE ATT&CK Enterprise matrix, enabling the mapping of attack scenarios to corporate system vulnerabilities. A significant contribution to the development of the theoretical basis was made by the study of Zipperle M. [10], which systematized approaches to building detection systems based on data provenance analysis (PIDS), defining their architectural modules and scalability problems.

Thus, the methodological strategy of the research is based on the synthesis of principles from machine learning, MITRE ATT&CK taxonomies, data provenance analysis systems, and security monitoring center architectures. This approach has made it possible to identify key areas for improving detection engineering—data unification to ensure the compatibility of telemetry sources, contextual modeling of behavior to enhance the explainability of detections and restore cause-and-effect relationships, and the implementation of adaptive metrics that provide for the dynamic adjustment of thresholds and the resilience of analytical models to changes in the threat structure.

RESULTS

The analysis conducted showed that modern detection system architectures are evolving within two complementary approaches, based on machine learning models and MITRE ATT&CK taxonomies. The first group is aimed at improving classification accuracy and reducing the number of false positives through statistical analysis and feature selection, while the second is focused on enhancing explainability and contextual accuracy by mapping observed events to attacker tactics and techniques. The combination of these areas forms the basis for building integrated detection engineering systems that unite the data level and the threat semantics level.

The study by Disha R. A. [1] proposed the Gini Impurity-based Weighted Random Forest (GIWRF) method, which optimizes the performance of attack detection models. On the UNSW-NB15 and TON_IoT datasets, the authors showed that GIWRF allows for the reduction of the number of features to 20 and 10, respectively, without loss of accuracy. The best result was achieved using a Decision Tree classifier, which provided an F1 score of 0.98 and a reduction in the false-positive rate. This approach demonstrated the effectiveness of weighted ensembles with a limited number of relevant features and emphasized the importance of interpretability as a key factor in detection engineering.

The study by Son V. N. [8] proposed an architecture based on the integration of Sysmon system events with MITRE ATT&CK tactics, including T1055 (Process Injection) and T1547 (Boot or Logon Autostart Execution). To describe the relationships between events and techniques, Sigma rules were used, providing standardized data correlation at the level of security monitoring centers. This allowed for a shift from recording individual anomalies to a contextual analysis of attacker behavior and reduced the detection time for multi-stage attacks. Table 1 examines the difference between architectures based on machine learning methods and systems implementing the MITRE taxonomic approach.

Table 1. Comparison of detection models in ML- and MITRE-oriented systems (Compiled by the author based on sources: [1, 3, 8])

Parameter	GIWRF	Sysmon + MITRE
Data type	Feature vectors (UNSW-NB15, TON_IoT)	System event logs (ETW / Sysmon)
Method	Weighted Random Forest	Rule-based correlation with MITRE ATT&CK
Objective	Reduce false-positive rate and improve F1-score	Correlate process behavior with ATT&CK TTPs
Best result	F1 = 0.98 (Decision Tree)	Early detection of T1055 and T1547 techniques

An analysis of the data in Table 1 shows that models based on machine learning provide high classification accuracy when high-quality datasets are available but are limited in the explainability of their results. In contrast, MITRE-oriented solutions allow for the detailed analysis of attacker behavioral patterns, increasing the transparency of the analysis, but require manual rule updates and depend on the completeness of the knowledge base.

The combination of these approaches in a single detection engineering loop is seen as a promising direction, capable of combining the metric efficiency of ML models with the semantic accuracy of MITRE taxonomies. As noted in the studies by Georgiadou A. [2] and Zipperle M. [10], hybrid architectures allow for the formation of cause-and-effect relationships between events and adversary tactics, thereby increasing the adaptability and resilience of detection systems to new threat scenarios.

The development of Provenance-based Intrusion Detection Systems (PIDS) has become a logical stage in the evolution of detection engineering in response to the growing complexity of attacks and volumes of telemetry [6]. Unlike classic signature-based solutions, PIDS capture the cause-and-effect relationships between processes, files, and network objects, enabling the reconstruction of the event timeline and the analytical traceability of an attacker's actions.

The study by Xiong, W. [9] systematized scientific works and proposed a structural model of PIDS that includes four interconnected modules—data collection, graph summarization, intrusion detection, and benchmark datasets. This typology made it possible to identify the bottlenecks of each lifecycle stage and to reveal the systemic limitations hindering practical implementation. For instance, the volume of data collected at the operating system kernel level in modern experiments reaches several terabytes, which causes an overload of computing nodes and reduces the efficiency of real-time processing. Table 2 examines the distribution of existing solutions by architectural modules and the main scalability problems.

Table 2. Main PIDS approaches by architectural module (Compiled by the author based on sources: [4, 9, 10])

Module	Example systems	Key objective	Scalability issue
Data collection	CamFlow, Sysmon, Provmon	Collect provenance data from OS kernel	High system load (>40%)
Graph summarization	NodeMerge, LogApprox	Compress event graphs (DAG)	Semantic loss
Intrusion detection	DeepLog, ProvDetector	Detect APT anomalies	Threshold dependency
Benchmark datasets	DARPA TC E3, LANL	Replicate real-world attack traces	Insufficient documentation

As can be seen from the data in Table 2, the main load falls on the data collection module, which is implemented using low-level monitoring tools that provide maximum event completeness but at the same time create a load on system resources. Following this, the graph summarization module addresses the task of reducing the size of the directed acyclic graph, but methods like NodeMerge and LogApprox lead to a loss of semantics in the relationships between objects [7].

The intrusion detection module uses machine learning algorithms (e.g., DeepLog, ProvDetector) to identify anomalies in process behavior. However, such approaches remain dependent on rigidly defined thresholds, which limits their adaptation to new attack scenarios. Similar conclusions were reached in the study by Disha R. A. [1], which showed that even with high classification accuracy (F1 = 0.98), the effectiveness of the models depends on the correct selection of features and training conditions.

Furthermore, according to the observations of Georgiadou A. [2], the risk of errors increases in the absence of standardized mechanisms for data correlation, which makes the use of MITRE ATT&CK taxonomies as a normalization layer important. The study by González-Granadillo G. [3] emphasizes that such integration is possible only in the presence of unified event log formats and centralized SIEM platforms that ensure data compatibility between sources. The benchmark datasets module, as noted by Rosso M. [6], plays a key role in evaluating the correctness of detectors. However, even test environments such as DARPA TC E3 and LANL are characterized by limited documentation and insufficient realism, which is confirmed in the analytical review by Zipperle M. [10]. This hinders the replication of attacks and the verification of the results obtained.

Consequently, the evolution of PIDS architectures reveals a contradiction between the depth of analytics and the computational stability of the systems. On the one hand, the high detail of provenance data allows for the reconstruction of complex attack scenarios; on the other hand, it causes problems of scalability and loss of semantics during graph compression. The comparative-analytical meta-analysis

of publications has shown that combining the advantages of both approaches is possible by forming hybrid models that unite graph analysis with the principles of detection engineering. As shown in the works of Xiong W. [9] and Sacher-Boldewin D. [7], such models distribute functions between levels: machine learning algorithms are responsible for event prioritization, while MITRE ATT&CK taxonomies handle the contextualization of adversary behavior, ensuring a balance between analytical depth and computational efficiency.

DISCUSSION

Detection engineering as an independent direction in cybersecurity is formed at the intersection of data analytics, system monitoring, and organizational risk management. Based on an analysis of the presented research, a set of interconnected principles can be identified that define a holistic architecture for a Detection Engineering System and ensure its reproducibility in a highly critical infrastructure.

The study by Son V. N. [8] shows that the effectiveness of detection systems is directly dependent on the ability to combine low-level kernel events with high-level MITRE ATT&CK tactics. This correlation between Sysmon processes and TTPs allows for the detection of anomalies at the early stages of the attack lifecycle, ensuring the continuity of the analytical context from the system call level to the description of the attacker's objectives. An equally important area of development is the standardization of data representation. The study by Georgiadou A. [2] proposes a model for unifying risks and cultural factors based on the STIX 2.1 format, which allows for the creation of a unified compatibility space between incident analysis systems. This unification facilitates the exchange of information between participants in the response chain and makes a quantitative assessment of the maturity of monitoring centers (SOCs) possible. As shown in the study by Zipperle M. [10], the contextualization of data through provenance graphs enables a shift from isolated events to the reconstruction of complete attack scenarios. The use of directed acyclic graphs (DAGs) makes it possible to track the relationships between processes and files, minimize the false-positive rate to 0.1%, and obtain an explainable structure of the attacker's behavior.

The interaction between the analytical and computational levels of the systems is realized through feedback mechanisms. The study by Disha R. A. [1] shows that the application of the GIWRF method in combination with decision trees forms a closed-loop model update cycle, in which new data coming from the SOC automatically adjusts the weight coefficients of the features. This ensures an increase in accuracy (F1 score) by 15–20% and resilience to APT-type attacks, which are characterized by a multi-layered structure and high behavioral variability. Table 3 presents the generalized principles for building a Detection Engineering System and their implementation in specific studies.

Table 3. Fundamental principles of Detection Engineering System (Compiled by the author based on sources: [1, 2, 8, 10])

Principle	Implementation	Effect
Multilayer correlation	Sysmon ↔ MITRE TTP	Early-stage detection
Data standardization	STIX 2.1 + Cultural Matrix	SOC maturity assessment
Behavioral contextualization	Provenance DAG analysis	FPR reduction to 0.1 %
ML integration	GIWRF + DT Pipeline	F1 increase by 15–20 %

The systematization presented in Table 3 allows us to consider the detection engineering system as a multi-layered ecosystem in which telemetry sources, machine learning models, and MITRE ATT&CK taxonomies form a single analytical loop. Multi-level correlation provides cause-and-effect links between processes and attack techniques. Data standardization ensures the reproducibility and compatibility of analytical models. The graph structures of PIDS create a context for interpreting behavior, and the interaction between the SOC and ML models ensures the system's adaptability to the dynamics of threats.

Consequently, the fundamental principles of detection engineering form the methodological basis for building resilient and explainable defense systems capable of identifying incidents and predicting the development of attack scenarios. This approach sets a new standard in the organization of analytical processes in cybersecurity, where the priority shifts from reactive response to the predictive identification of threats within a unified loop of data, behavior, and context.

Modern detection engineering faces a number of fundamental limitations that hinder the transition from experimental models to industrial solutions. As shown in the study by Sacher-Boldewin D. [7], the main obstacle lies in the heterogeneity of event logs, the lack of unified formats, and the incompleteness of metadata, which reduces the reliability of provenance graphs and limits the scalability of analysis. In practice, this is evident, for example, when integrating Sysmon telemetry with corporate EDR platforms (CrowdStrike, Splunk ES): the difference in event formats and field schemas leads to a partial loss of context and a disruption of the links between processes during correlation. Even with the active use of low-level monitoring systems, the problem of overload persists—telemetry volumes reach several terabytes, and data normalization requires manual filtering and additional semantic processing. Another significant challenge is the deficit of realistic datasets. As noted by Khraisat A. [4], existing collections—DARPA TC E3 and LANL—do not reflect the current level of threats, contain incomplete descriptions of attacks, and do not account for the complex scenarios characteristic of targeted campaigns.

As a result, machine learning algorithms are trained on incomplete or unbalanced samples, which leads to overfitting and an increase in the false-positive rate.

A special role is played by the problem of threshold dependency, which is characteristic of detectors that use fixed metric values. The study by Disha R. A. [1] shows that even models with high classification accuracy, built on the GIWRF method, demonstrate a decrease in quality when the distribution of input data changes. The absence of adaptive thresholds leads to the system becoming unstable against new types of attacks, especially in a dynamically changing infrastructure.

At the same time, the prospects for the further development of detection engineering are associated with the automation and integration of heterogeneous approaches. The study by Son V. N. [8] shows that the correlation of Sysmon logs with MITRE ATT&CK techniques can be automated by constructing correlation rules formed on the basis of provenance graphs and statistical models. This approach allows for the description of attacker behavior at the level of individual events, tactics, and procedures, creating a basis for the adaptive generation of protection rules. Similarly, Disha R. A. [1] considers the possibility of forming a feedback loop between monitoring systems and machine learning models. In this case, the SOC becomes an active source of data for self-learning, and accuracy metrics serve as parameters for tuning the algorithms. In the long term, this will allow for the construction of a continuous loop for updating detection rules, where each new attack automatically strengthens the system's resilience.

Consequently, the key direction of development is the construction of a detection engineering pipeline—a holistic analytical chain that combines telemetry collection, behavior modeling, and feedback with operational centers. This structure will allow for the implementation of the adaptive threshold principle, ensuring a balance between the system's sensitivity and stability.

The practical implementation of such an approach is already emerging in the integration of open-source tools—Sigma, OpenCTI, TheHive—with industrial SIEM platforms (Elastic, Splunk, IBM QRadar), which creates a basis for the automation of event correlation and the generation of detection rules in real time. The automatic updating of ATT&CK content and the exchange of indicators via STIX/TAXII allow for the formation of self-learning defense loops, where each new attack strengthens the analytical model.

In the long term, it is this approach that will become the methodological standard for intelligent systems for detecting and neutralizing attacks, uniting machine learning, graph analytics, and MITRE ATT&CK standards in a single research and practical loop.

CONCLUSION

The conducted research has confirmed that detection engineering is not a collection of disparate analysis methods but a holistic system that forms a new architecture for ensuring cybersecurity. The synthesis of machine learning, graph-based data provenance analysis, and MITRE ATT&CK taxonomies allows for a transition from a reactive response to incidents to the proactive prevention of attacks through a contextual understanding of adversary behavior and the interconnections between events.

It has been identified that the effectiveness of a detection engineering system is determined by the degree of standardization of input data, the correctness of the correlation between the system and behavioral levels, and the presence of a feedback loop with monitoring centers. The formation of adaptive thresholds, updated based on the actual results of SOC operations, ensures the resilience of algorithms to changes in the threat structure and allows for an increase in the accuracy and explainability of decisions.

The practical significance of this approach lies in the creation of a closed-loop analytical processing cycle, where telemetry, machine learning models, and the MITRE context are combined in a single risk management system. Such integration reduces the rate of false positives, accelerates the response to attacks, and increases the maturity of detection processes across the organization.

The prospects for further research are associated with the development of a detection engineering pipeline—a holistic analytical chain that includes the automatic generation of rules, the adaptive updating of models, and integration with the response infrastructure. The implementation of such an approach will allow for a transition to the intelligent automation of defense systems, where each new attack becomes a source of learning, and the system itself becomes a self-configuring mechanism for ensuring cyber resilience.

REFERENCES

1. Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity, 5(1), 1. https://doi.org/10.1186/s42400-021-00103-8
2. Georgiadou, A., Mouzakitis, S., & Askounis, D. (2021). Assessing MITRE ATT&CK risk using a cyber-security culture framework. Sensors, 21(9), 3267. https://doi.org/10.3390/s21093267
3. González-Granadillo, G., González-Zarzosa, S., & Diaz, R. (2021). Security Information and Event Management (SIEM): Analysis, trends, and usage in critical infrastructures. Sensors, 21(14), 4759. https://doi.org/10.3390/s21144759
4. Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. Cybersecurity, 2(20). https://doi.org/10.1186/s42400-019-0038-7

5. Pahlevan, M., & Ionita, V. (2022). Secure and efficient exchange of threat information using blockchain technology. Information, 13(10), 463. https://doi.org/10.3390/info13100463

6. Rosso, M., Campobasso, M., Gankhuyag, G., & Allodi, L. (2022). SAIBERSOC: A methodology and tool for experimenting with security operation centers. Digital Threats: Research and Practice, 3(2), Article 14. https://doi.org/10.1145/3491266

7. Sacher-Boldewin, D., & Leverett, E. (2022). The intelligent process lifecycle of active cyber defenders. Digital Threats: Research and Practice, 3(3), Article 22. https://doi.org/10.1145/3499427

8. Son, V. N., Tisenko, V. N., Tuan, L. D. A., Lam, N. T., Thuong, P. T., & Anh, D. X. (2020). Detecting behavior of malware using MITRE ATT&CK. International Journal of Advanced Trends in Computer Science and Engineering, 9(5), 8285–8294. https://doi.org/10.30534/ijatcse/2020/198952020

9. Xiong, W., Legrand, E., Åberg, O., & Lagerström, R. (2022). Cyber security threat modeling based on the MITRE Enterprise ATT&CK Matrix. Software and Systems Modeling, 21, 157–177. https://doi.org/10.1007/s10270-021-00898-7

10. Zipperle, M., Gottwalt, F., Chang, E., & Dillon, T. (2023). Provenance-based intrusion detection systems: A survey. ACM Computing Surveys, 55(7), Article 135. https://doi.org/10.1145/3539605

Citation: Alina Gaifulina, "Fundamental Principles for Building an Effective Detection Engineering System", Universal Library of Innovative Research and Studies, 2023; 01-06. DOI: https://doi.org/10.70315/uloap.ulirs.2023.001.