

A REVIEW OF MACHINE LEARNING VS. SIGNATURE-BASED CYBERSECURITY TOOLS: A COMPARATIVE ASSESSMENT OF DARKTRACE AND SNORT FOR NETWORK INTRUSION DETECTION

Tarannum Bano¹, Deepshikha²

¹Master of Technology, Computer Science and Engineering, Lucknow Institute of Technology, Lucknow, India

²Assistant Professor, Department of Computer Science and Engineering, Lucknow Institute of Technology, Lucknow, India

Abstract - There is need for strong to protect the computer infrastructure. Network Intrusion Detection Systems (NIDS) because of the ever increasing complexity in the cyber attacks. The review article has performed a comparative analysis of two ruling paradigms in intrusion detection domain machine based learning-based (ML)-based tools such as Dark Trace and traditional signature based systems such as Snort. The research is important since it highlights prominence of significant trade-offs associated with these approaches by detailing them on the basis of their detection capabilities, adaptability, operational performance and economic benefits. On the other hand, signature-based solutions such as Snort are extremely accurate in the case of known problems and have very low overhead in a processing operation, but the manual updating nature of such a system renders networks vulnerable to attacking as they come. On the other hand, solutions in ML such as Darktrace add more wiggle room in dealing with zero day exploits using behavioural analysis, albeit the high rate of false positives, high spending on resources and low explainability. The analysis establishes the necessity to apply contextual tools like organisational size, threat landscape, and available resources. Other than that, in the paper, there are listed systemic problems like adversarial attacks, skill gaps and dependency on a data base and it proposes the use of hybrid frameworks that advocate for the best of both worlds. The observations have practical implications for practitioners in the field of Cybersecurity and thus offer guidance to the profession, e.g., are XAI (explainable AI, and the joint, sharing of threat intelligence, to encourage adaptive and visible defenses. This review is also an aspect of an outreach concerning searching for approaches of maximizing NIDS amid the changing cyber threats.

Key Words: Network Intrusion Detection Systems (NIDS), Machine Learning, Signature-Based Detection, Darktrace, Snort, Cybersecurity, Zero-Day Attacks, Adaptive Security, Hybrid Frameworks.

1. INTRODUCTION

1.1 Background

The cyber security landscape is witnessing an alarming pace of transformation because malware authors are constantly making their malware new using sophisticated techniques and exploit towards breaching the network environment, which includes zero day exploits and polymorphic malware. In such an availability-critical environment, Network Intrusion Detection Systems (NIDS) play a major role in securing a given infrastructure by examining network traffic for suspicious activities. Signature based detection tools, such as Snort, have been the dominant tool in this field historically, and it relied upon the known pattern of the attack to detect their presence. Yet, the advent of machine learning (ML) driven paradigms such as solutions like Darktrace have heralded change in the form of an alternative that is dynamic. These ML based systems analyze the behavioral anomalies as opposed to static signatures and hence versatile to new and emerging threats. Such a change points to a deep contradiction in cybersecurity: the opposition between the precision of conventional methods and the agility of current methods which are powered by AI.

1.2 Problem Statement

Alas, signature-based tools are very proficient at identifying well-known threats, but their work is purely rule-based, and therefore they are powerless in the face of zero-day and APT's attacks. Inversely, ML-based systems such as Darktrace are faced with hurdles such as high false positives, computational complex predicaments and the "black box nature" of algorithms. The lack of such all-comprehensive comparative studies even adds another dimension of complexity towards decision making of organizations that wish to adopt or move toward these paradigms. An intensive examination of their strengths, constraints, and operational trade-offs is eminently necessary to guide cybersecurity practices in a variety of organizational context.

1.3 Research Objectives

The goal of this paper is three-dimensional, systematic comparison of Darktrace (ml-based) and Snort (signature-based): 1) reliability: 2) ease of use; 3) cost effectiveness, effectiveness in intrusion detection, flexibility to new threats, scalability in real-life deployment. In determining their technical performance, resource needs and cost effectiveness, the study is purported to outline the strength and weakness of both options. Further, it will provide actionable insights that will enable enterprises to decide in choosing NIDS that will be customized to their risk profiles, infrastructure capabilities, and threat landscapes.

2. LITERATURE REVIEW

2.1 Signature-Based Intrusion Detection

Signature based intrusion detection systems like Snort and Suricata evolved as key tools in the realm of cybersecurity for it is derived on the basis of identifying network traffic activity by comparing predefined rule-set with pattern-matching. These systems became particularly popular in the end '90s and at the start of the '00s, because of their simple logic, good ability to identify available threats, and perceivable operation that didn't require much auditing and rule tuning efforts. Although they rely on static signatures, this fact imposes a restriction on the ability to detect zero-day attacks or intelligent polymorphic malware, as the latter are not marked in signature databases beforehand. Additionally, safety demands that efficacy is constantly edited by humans in order to include new threat intelligence, an activity that creates a delay and exposes networks when there are gaps in coverage. Notwithstanding these limitations, signature-based tools are still prevalently used due to their efficacy at dealing with well-known threats and low CPU overhead on top of comparably complex systems.

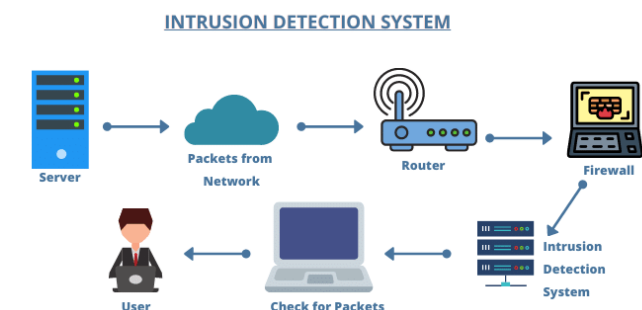


Figure-1: Intrusion Detection Systems

2.2 Machine Learning in Cybersecurity

With the advent of artificial intelligence (AI) and machine learning (ML), the anomaly detection game was changed,

and among the tools such as Darktrace, Vectra, benefit from unsupervised learning and behavioral analysis to detect anomalous fluctuations in the activity of the network. Against this, ML Models fall short of the signature based approach in detecting new or emerging threats such as zero-day exploits or insider threat as they can infer order from large tables of data without any set rules. This convenience is a signature value of ML when it is applied in a dynamic setting where threat actors keep on changing their approach. However, the ML-driven systems are not flawless because there is excessive computational cost on training and inference; its susceptibility to adversarial attacks that alter input data; and "black-box" character of the system making it impossible to trust it or put it under control. Besides, there is also a very biased choice on their part with regard to the nature and variability of training data that is applied for that purpose thus sparking a debate on speculative bias and variations advocating for different network environments.

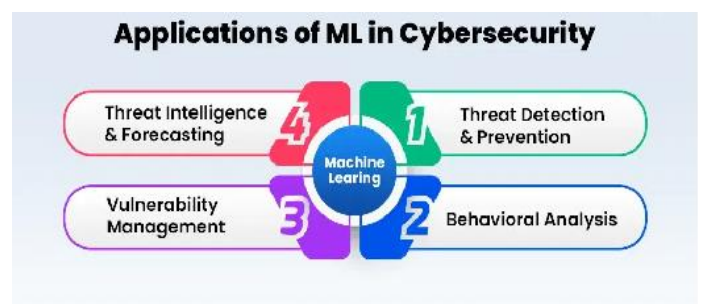


Figure-2: Machine Learning in Cybersecurity.

2.3 Prior Comparative Studies

Past studies in intrusion detection systems have examined the technical performance of ML and signature-based tools, most times, emphasizing ML's better performance over finding unknown threats and signature-based systems' greater efficiency in detection of known attacks. Comparisons among institutions of learning and industry analysts have made comparisons using measures like detection rates, false positives, and scalability, but such analyses seldom reflect such tangible factors that include deployment complexity, organizational resource limitations, and cost-benefit trade-offs. For example, despite superior accuracy in detection conditions, the real-world applicability of ML models suffers due to the necessity of their special skills and facilities. Moreover, limitations remain in the awareness of long term viability of ml systems especially in situations that involve continual retraining and realignment. By conducting a thorough analysis and review of the literature in cybersecurity, a specific need is outlined for a holistic assessment of technical, operational, and economic aspects to determine an informed decision-making process in cybersecurity practice.

3. METHODOLOGY

3.1 Framework for Comparison

The comparison between Darktrace and Snort is organized around a multi-dimensional parameter by which their performance in technical, operational, and economical domains are measured. Some of the significant criteria are detection accuracy calculated in terms of the true positive and negative rates to determine reliability when determining threats while avoiding false alarms. adaptability, which measures how effectively each of the tools could react to new or emerging vectors of attack without any human assistance, resource efficiency (both in terms of computational expenses and hardware demands); ease of deployment and maintenance (in terms of how complex the installation process is, how straightforward the update procedures are and how intuitive the tools are); and cost-effectiveness (both from the initial purchase price to the cost of long-term operations); respectively. Together, they enable a fair assessment; in addition to technical aspect of capability, those important features for actual implementation are also considered.

3.2 Tools Selection

The relevance of Darktrace and Snort as representative case studies is based on their visibility in their own paradigms, and their opposing philosophical facilities of operation. An example of traditional rule driven-detection is the open source signature-based Snort, which has been driven by out-of-source rulesets generated by the community and simple, transparent, customizable logic. The extensive use of this tool along with its development history over the decades make it a benchmark for measuring the signature-based efficacy. On the other hand, Darktrace, which is a proprietary ML-based platform, implements dishonest learning and behavioral analysis using the "Enterprise Immune System" to detect anomalous behavior in a self-sustaining manner. The choice of its voice is the reflection of the emerging industry trend toward AI-driven solutions and an entry point into the discussion on the scalability and other innovative claims of contemporary ML tools. In combination, these tools condense the essence of strengths and weaknesses of their detection measures and make a non-arbitrary comparison possible.

3.3 Data Collection

Sources of data for this study were consolidated from different sources in order to ensure that the data has some chances of being plausible and less biased. It was only from the gained inputs by the aid of the academic papers and technical documentation introduced in the two tools that the algorithms, architectures and theoretical performance was established. Case studies and

whitepapers contained practical examples of deployments and successes and challenges experienced in various contexts of organizations. Other results were supported by reviews from the audience, industry reports and direct opinions on usability, maintenance as well as cost implications. Popular metrics (detection rates- known and unknown, positive / negative ratio and scalability measure) were discussed as to be used for quantification of performance. In addition, consumption of computational resources and cost breakdowns were addressed to determine spatial feasibility. This method of triangulation of qualitative and quantitative data gains as deep evaluation, which bridges the gap between theoretical claim and empirical results.

4. COMPARATIVE ANALYSIS

4.1 Detection Capabilities

Snort is a very fast sniffer, which is able to trace not only famous attack patterns like SQL injection (SQLi) and distributed denial-of-service (DDoS) attacks but perform it effectively with the help of a massive database of signature-based rules. Such rules are customized by an international community of cyber security experts and allow to pinpointing known risks without any ambiguity. While on the contrary, unlike in Dark Gatekeeper, Darktrace does detect unsupervised machine that looks into the behavior of the network and looks for deviations from the established baseline to find anomalies such as insider threats as well as zero days exploits. While the deterministic analysis offered by Snort guarantees reliability for the detected threats, the probabilistic analysis of Darktrace reaches better performance in the detection of new attacks with no known signatures, and so, reveals a great trade-off between the discrimination and the flexibility in intrusions detection.

4.2 Adaptability and Scalability

The reliance of Snort on manual rule updates causes latency in the reaction to emerging threats as evident during the Log4j vulnerability crisis when late rules deployments left networks exposed for some time. On the other hand, Darktrace's autonomous learning architecture allows for real-time adjustment of the existing attack vectors through dynamic re-definition of its conception of normal behavior without human supervision. This self-evolving ability augments scalability in networks of a large or dynamic nature, where manual rule administration of Snort becomes all the more unwieldy. However, Darktrace's dependency on constant data ingestion is a topic that raises doubts about the capabilities of the software in wasteful or excessively controlled circumstances where the accessibility of data is limited.

Table 1: Comparative Summary of Darktrace and Snort

Criteria	Darktrace (ML-Based)	Snort (Signature-Based)
Detection Capabilities	Detects anomalies via behavioral analysis (e.g., zero-day exploits, insider threats).	Matches known attack signatures (e.g., SQLi, DDoS).
Adaptability	Autonomous learning adapts to new threats (e.g., Log4j) without manual intervention.	Requires frequent rule updates; delays in detecting emerging threats.
Operational Efficiency	High computational demands (GPU/CPU-intensive); higher false positives.	Lightweight, low resource usage; fewer false positives for known threats.
Cost & Accessibility	Proprietary licensing with high upfront/subscription costs; steep learning curve.	Open-source (free); low initial cost but higher maintenance effort for rule updates.
Ethical/Privacy	GDPR-compliant but collects extensive behavioral data; "black box" decision-making.	Transparent, auditable rules; minimal privacy concerns due to signature-based logic.

4.3 Operational Efficiency

Being lightweight and rule based Snort causes low computational overhead therefore it is suitable for resource-constrained environment or legacy systems. Darktrace, in its turn, though, requires a lot of processing power and storage to train and run its AI models, which can require its own infrastructure. This difference ignites to false positive rates: Snort’s rule specificity hinders false alarms at cost of missing the subtleties or novel threats, while Darktrace’s attention to behavioral anomalies leads to greater false positives requiring extra triage work.

Organizations, while choosing a tool, would therefore need to come to a trade-off between operational efficiency and its ability to detect thoroughly.

4.4 Cost and Accessibility

Snort’s open-source structure relegates the licenses to non-issue, providing advantages in price for cost-sensitive organisations, but costs for custom rule work, personnel and maintenance may mount up. Darktrace has proprietary pricing, whereupon initial costs and subscription charges are high that are justified by its autonomous capabilities and less manual oversight. However, the latter’s complexity might require a dedicated training, while the clear rule logic of Snort makes it possible for the in-house teams to tweak the configurations at a relatively minor cost. It depends on whether an organization is better off with affordability, control (Snort) or automation, scalability (Darktrace).

4.5 Ethical and Privacy Considerations

Darktrace’s massive data collection activities, which are crucial to training its ML models, are troubling from a privacy perspective especially with such regulations such as GDPR, which requires strict user data monitoring. Although Darktrace highlights the focus on anonymization and compliance, its enigmatic decision-making steps are no different from Snort’s viewable, audit-ready rulesets. Snort’s community-driven architecture enables organizations to audit and adjudge detection logic, thus creating trust and accountability. Such ethical specificities illustrate the balance between innovations in the AI-driven tools and the necessity of transparency in cybersecurity procedures.

5. CHALLENGES AND LIMITATIONS

5.1 Technical Limitations

Machine learning in cyber security has an inherent correlation with the availability of high-quality representative data for training because biased or insufficient data can result in incorrect models which misjudge the threat, or miss out an important anomaly. For such tools as Darktrace, this dependence provokes doubts regarding the generalization for various network environments, especially for niched industries with unique traffic patterns. On the contrary, signature based IDS’s such as Snort have built-in delays in terms of threat intelligence updates because manual rule creation and distribution processes are not well suited for agile developments in attack vectors. This gap of time results in windows of vulnerability, such as zero-day exploits that do not trigger until signatures were defined and implemented. Both paradigms therefore struggle on basic technical limitations: ML’s data-hunger and signature-based tools’ reactive properties.

5.2 Practical Barriers

Much of the organizational reluctance in embracing AI-based tools such as Darktrace may come as a lack of trust in dark processes that eventually alienate practitioners from recognizing AI-based opportunities. This is coupled with the fear of false positive disruptions. In the case of small enterprises, they may not have the infrastructure or knowledge on how to handle ML systems that require special skills in data science and cybersecurity. Although Snort's rule-based system is more transparent, it still needs trained personnel to modify and refresh rules, which are problematic for organizations lacking in IT resources. This skill gap is further compounded by the lack of professionals that are familiar with legacy systems and leading the edge AI, which in turn prevents the adoption of integrating hybrid solutions.

5.3 Emerging Threats

ML and signature-based tools are becoming more and more under the attack of advanced evasion methods. A variety of adversarial machine learning attacks that may work against Darktrace and its anomaly detection models – such as tampering with network traffic to fool them – take advantage of the brittleness of our AI systems. At the same time, polymorphic malware that modifies its code in a dynamic fashion thus subverts the Snort's dependence on static patterns. These threats point to a cybersecurity arms race involving the attackers who are constantly innovating to circumvent the detection mechanisms. Although it is versatile, ML must deal with vulnerabilities to adversarial manipulation and the overhead of real-time defense updates, an open issue requiring continued exploration in detection architectures.

6. CONCLUSION

Comparative analysis of Darktrace (machine learning-based) and Snort (signature-based) demonstrates significant differences in their advantages and shortcomings in the network intrusion detection. Snort is very competitive when it comes to recognizing piecemeal threats with high precision, however, Snort's dependence on static signatures makes it vulnerable to latent or polymorphic attacks. On the other hand, Darktrace employs unsupervised machine learning in detecting anomalies and zero-day exploits; it provides proactive defense. However, its high computational requirements, its potential for false positives and its unavoidable air of secrecy put hurdles to its use in reality. The research determines that neither the system is better than the other in all cases. Instead of that, the best choice will depend on an organization's threat picture, priorities, and capabilities.

In greater issues, both approaches are affected. Machine learning relies on good data while signature-based tools

rely on timely updates- both of which may be missing. These vulnerabilities are worsened by advanced threats such as adversarial ML and evasion malware. To deal with them, there will be a need for hybrid models that unite the strengths of both paradigms. Future improvements rest on developments in explainable AI, threat intelligence sharing, and regulatory standards as it concerns scalable, ethical, and context-driven cybersecurity solutions.

REFERENCES

1. R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," IEEE Symposium on Security and Privacy, 2010, pp. 305–316.
2. M. Roesch, "Snort: Lightweight Intrusion Detection for Networks," Proc. USENIX LISA Conf., 1999, pp. 229–238.
3. Darktrace, "The Enterprise Immune System: A Technical Overview," 2023. [Online]. Available: <https://www.darktrace.com>. Accessed: Sep. 1, 2023.
4. S. Garcia et al., "An Empirical Comparison of Botnet Detection Methods," Computers & Security, vol. 45, pp. 100–123, 2014.
5. A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cybersecurity Intrusion Detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.
6. J. W. Stokes et al., "Attack and Defense of Dynamic Analysis-Based, Adversarial Neural Malware Classification Models," IEEE S&P Workshops, 2021, pp. 298–315.
7. V. H. Pham et al., "A Comparative Study of Anomaly Detection Algorithms for Network Intrusion Detection," IEEE Access, vol. 9, pp. 106480–106496, 2021.
8. M. Tavallaee et al., "A Detailed Analysis of the KDD CUP 99 Data Set," IEEE Symposium on CISDA, 2009, pp. 1–6.
9. MITRE, "ATT&CK Matrix for Enterprise," 2023. [Online]. Available: <https://attack.mitre.org>. Accessed: Sep. 1, 2023.
10. NIST, "Guide to Intrusion Detection and Prevention Systems (IDPS)," NIST SP 800-94, 2020.
11. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

12. A. Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
13. C. Szegedy et al., "Intriguing Properties of Neural Networks," arXiv:1312.6199, 2013.
14. Snort, "Snort User Manual," 2023. [Online]. Available: <https://www.snort.org>. Accessed: Sep. 1, 2023.
15. K. Scarfone and P. Mell, "Guide to Intrusion Detection and Prevention Systems," NIST Special Publication 800-94, 2007.
16. R. Vinayakumar et al., "Deep Learning for Network Intrusion Detection Systems: A Comprehensive Analysis," *Engineering Applications of Artificial Intelligence*, vol. 96, 2020.
17. M. Husák et al., "Survey of Attack Projection, Prediction, and Forecasting in Cybersecurity," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 640–660, 2019.
18. S. M. Kasongo and Y. Sun, "A Deep Learning Method with Filter-Based Feature Engineering for Network Intrusion Detection," *IEEE Access*, vol. 7, pp. 38597–38607, 2019.
19. F. Iglesias and T. Zseby, "Analysis of Network Traffic Features for Anomaly Detection," *Machine Learning*, vol. 101, no. 1, pp. 59–84, 2015.
20. O. Depren et al., "An Intelligent Intrusion Detection System for Anomaly and Misuse Detection in Computer Networks," *Expert Systems with Applications*, vol. 29, no. 4, pp. 713–722, 2005.
21. ENISA, "Threat Landscape 2022: Emerging Cyber Threats," 2023. [Online]. Available: <https://www.enisa.europa.eu>. Accessed: Sep. 1, 2023.
22. R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems*, 3rd ed. Wiley, 2020.
23. GDPR, "General Data Protection Regulation," 2018. [Online]. Available: <https://gdpr-info.eu>. Accessed: Sep. 1, 2023.
24. M. Bishop, *Computer Security: Art and Science*, 2nd ed. Addison-Wesley, 2018.
25. S. Garcia-Fernandez et al., "A Survey on the Use of Machine Learning for Network Intrusion Detection," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–35, 2021.
26. A. Shiravi et al., "Toward Developing a Systematic Approach to Generate Benchmark Datasets for Intrusion Detection," *Computers & Security*, vol. 31, no. 3, pp. 357–374, 2012.
27. Y. Mirsky et al., "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection," *NDSS*, 2018, pp. 1–15.