JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Anomaly-based Detection in Cyber Threat Intelligence for Real-Time Threat Identification & Mitigation

[1]Sathyaprakash M. Sahoo, [2]Dr. Amol Joglekar

[1]Dept. of Computer Science, [2]Asst. Professor Dept. of Computer Science,
[1]Mithibai College of Arts, Chauhan Institute of Science & Amrutben
Jivanlal College of Commerce and Economics, Mumbai, India

Abstract : With the escalating sophistication of cyber threats, there is an imperative need for advanced detection mechanisms in Cyber Threat Intelligence (CTI) to enable real-time identification and mitigation of threats. This research paper delves into the realm of anomaly-based detection, a real-time proactive approach that aims to identify deviations from expected behaviour within a network or system. The study explores the significance of anomaly-based detection in enhancing CTI capabilities, providing a nuanced understanding of its application for real-time threat identification and mitigation. The research investigates various anomaly detection techniques, ranging from statistical models to machine learning algorithms, and assesses their efficacy in discerning malicious activities within network traffic and system behaviours. Special emphasis is placed on the integration of anomaly-based detection into a broader CTI framework, emphasizing the synergy between automated detection mechanisms and human analysts. Furthermore, the paper addresses the challenges associated with anomaly-based detection, such as false positives and the dynamic nature of cyber threats. It proposes potential solutions and optimisations to mitigate these challenges, ensuring a more accurate and reliable threat detection system. Through an in-depth analysis of case studies and empirical evidence, the research demonstrates the practical implications of incorporating anomaly-based detection into CTI practices. The results showcase the effectiveness of this approach in identifying emerging threats and enabling swift, targeted mitigation strategies. Additionally, the paper discusses the potential impact of anomaly-based detection on reducing the dwell time of cyber threats and minimizing the overall risk to organizations.

IndexTerms **- Anomaly detection, cyber threat intelligence, cyber security, machine learning, intrusion detection, enterprise network.**

I. INTRODUCTION

In the dynamic landscape of the digital era, cybersecurity stands as an imperative bastion against an ever-growing number of sophisticated cyber threats. The integrity of systems and the confidentiality of data are constantly under siege, demanding a proactive approach to real-time threat identification and mitigation. This research endeavours to address this pressing need by delving into the realm of anomaly-based detection within the domain of Cyber Threat Intelligence (CTI). The primary objective of this research is to develop and propose a solution that enhances real-time threat detection with high accuracy while minimizing false positives and negatives. By leveraging anomaly detection techniques, the research aims to bridge the gap in current cyber threat intelligence practices, providing a robust and adaptive approach to identify and mitigate emerging threats.

A. Cyber Threat Intelligence

Cyber Threat Intelligence (CTI) constitutes a comprehensive process involving the systematic collection, analysis, and interpretation of data pertinent to potential cybersecurity threats and adversarial activities. This multifaceted approach mirrors the investigative role in the digital domain, wherein practitioners scrutinize and contextualize diverse information sources to discern cyber threat landscapes. Sources range from surface-level platforms to the more clandestine dark web, encompassing a spectrum of online forums and security logs. The synthesized intelligence serves as a foundation for the strategic formulation of defence mechanisms, enabling organizations to preemptively identify and thwart cyber attacks. This paradigm aligns with a proactive stance, akin to the foresight of a detective, as the derived insights empower analysts to comprehend evolving threat vectors. Through the discernment of patterns and analysis of adversarial methodologies, CTI professionals contribute to the fortification of digital defences, thereby safeguarding systems and data against potential cyber threats.

B. Anomaly Detection

Anomaly detection, at its core, involves the meticulous process of identifying patterns or behaviours that deviate from established norms within a given context. In the context of our research, we adopt this methodology as a versatile tool with applications beyond traditional cybersecurity measures. The essence of anomaly detection lies in its ability to discern irregularities, be it in data, system operations, or user behaviours. This methodological approach serves as a proactive line of defence beyond conventional signature-based strategies. Employing diverse techniques, such as statistical models and machine learning algorithms, anomaly detection scrutinizes network traffic, user behaviours, and system activities for irregularities. This sophisticated analytical framework, rooted

in continuous learning and adaptability, addresses the dynamic nature of the cyber environment. By virtue of its early and real-time threat identification capabilities, anomaly detection is instrumental in minimizing false negatives and diminishing reliance on pre-existing knowledge.

Real-life Scenario: In a scenario where a sophisticated zero-day exploit is utilized, anomaly detection can identify the anomalous patterns of behaviour associated with the exploit, even in the absence of specific signatures or prior knowledge. This adaptability and flexibility are crucial in staying ahead of emerging threats.

C. Context and Justification

Cyber threats have become increasingly pervasive, necessitating the evolution of cybersecurity measures to keep pace with the rapidly changing threat landscape. In this context, the paramount importance of real-time threat identification and mitigation cannot be overstated. The traditional approach of relying solely on predefined signatures and known patterns has proven insufficient in dealing with the onslaught of the novel and adaptive cyber threats. The significance of this research lies in its exploration of anomaly-based detection techniques to fortify cybersecurity measures. Anomaly detection offers a proactive and adaptive strategy, allowing for the identification of deviations from expected behaviour within a network or system. This is particularly crucial in addressing a critical gap in CTI.

Example: Consider a scenario where a malicious actor gains unauthorized access to a corporate network. Traditional signature-based detection systems may fail to identify the threat if the attacker employs novel techniques or modifications to known attack vectors. Anomaly detection, on the other hand, has the potential to discern unusual patterns of behaviour, alerting cybersecurity professionals to the anomaly and enabling swift response.

D. Importance of Rapid Detection

Swift detection is vital in cybersecurity. Identifying threats quickly limits their stay (dwell time), minimizing their impact. Real-time threat identification isn't just helpful; it's crucial for preventing breaches. For instance, if a financial institution faces a new malware targeting user passwords and sensitive data, an anomaly-based detection system can spot its unusual behaviour. It triggers alerts immediately, allowing the cybersecurity team to react swiftly, preventing financial losses and data breaches.

E. Current Gaps in Cyber Threat Intelligence

Anomaly-based detection for CTI faces challenges that limit its effectiveness: Excessive false alarms or missed detections, Difficulty detecting insider threats, and Inability to adjust to natural variations in network usage. For example, changes in a user's behaviour due to job duties can trigger false alarms. Overcoming these challenges is crucial to fully utilize anomaly-based detection in CTI.

II. METHODOLOGY

The methodology underpinning the development of the proposed solution is systematically presented in Table 1, where the focus lies on the purpose and respective anomaly types. The adopted approach entails an ensemble method in anomaly detection, strategically incorporating diverse techniques to amplify overall efficacy. This strategy is deliberate, aiming to elevate accuracy levels and curtail instances of false positives. Within this ensemble method, Table 2 elucidates the sub-methods corresponding to distinct anomaly detection types. The amalgamated anomaly detection techniques encompass statistical anomaly detection, machine learning-based anomaly detection, behavioural anomaly detection, and signature-based anomaly detection. Table 2 serves to delineate the nuanced purpose behind incorporating each sub-method, providing a comprehensive understanding of the rationale guiding the ensemble methodology. This meticulous combination not only introduces a spectrum of anomaly detection models but also contributes to heightened performance, concurrently addressing concerns related to both false positives and negatives in the proposed solution.

Table 1: Purpose of various anomaly detection types

Anomaly Type	Purpose
Statistical	To check deviations from established statistical norms
Machine Learning	To check if observed data deviates significantly from learned patterns
Behavioral	To check unusual user activity within a network
Signature-based	To check known malware or attack pattern based on predefined signatures

Table 2 meticulously outlines the diverse anomaly detection types integrated into the research framework, categorized under statistical, machine learning, behavioural, and signature-based. The research adopts a strategic approach of harnessing the specific strengths within each category and cross-verifying their outputs, with the overarching goal of augmenting accuracy through a collaborative and complementary methodology.

In the statistical anomaly detection category, the research encompasses both point anomalies and collective anomalies. Point anomalies involve detecting instances that significantly differ from the norm, while collective anomalies identify patterns deviating

from expected group behaviour. This dual approach broadens the scope of anomaly detection, capturing various types of irregularities within the data.

Within the machine learning paradigm, the research opts for semi-supervised techniques. Semi-supervised learning strikes a balance by leveraging a limited set of labelled data and a larger pool of unlabeled data. This choice is deliberate, acknowledging the challenges of obtaining fully labelled datasets in real-world scenarios. By incorporating semi-supervised learning, the research aims to benefit from the strengths of both supervised and unsupervised learning, enhancing adaptability and generalization.

Behavioural anomaly detection involves User and Entity Behavior Analytics (UBA, EBA). UBA focuses on recognizing anomalies in individual user behaviours, while EBA extends this analysis to entities, such as devices or applications. This approach is vital in dynamic environments where threats may evolve rapidly.

In the signature-based anomaly detection category, the research employs pattern matching and YARA (Yet Another Recursive Acronym) rules. Pattern matching involves identifying predefined patterns associated with known threats, while YARA rules enable the creation of custom rules to detect specific malicious activities. This combination ensures a comprehensive approach to identifying both well-known and emerging threats.

By integrating these varied anomaly detection methods, the research aims to create a robust and adaptive system. UBA and EBA enhance the understanding of user and entity behaviours, crucial in identifying insider threats or compromised entities. The choice of semi-supervised learning acknowledges the practical challenges of obtaining labelled data while still maintaining the benefits of guided learning. Overall, the ensemble approach, encompassing statistical, machine learning, behavioural, and signature-based methods, facilitates a comprehensive analysis, cross-verifying outputs, and contributing to the development of a sophisticated anomaly detection system with broader applicability in diverse scenarios.

Table 2: Methodology used

Statistical	Point Anomalies	Collective Anomalies
Machine Learning	Semi-Supervised	
Behavioral	User Behaviour Analytics (UBA)	Entity Behaviour Analytics (EBA)
Signature-based	Pattern Matching with YARA Rules	

A. Architecture

The architectural framework of this research's anomaly detection model is meticulously crafted to accommodate diverse data inputs from key sources, including network logs, firewall logs Security Information and Event Management (SIEM) systems, and behavioural monitoring systems. The initial phase involves the collection of raw data from these varied sources, capturing network activities, security events, and user behaviours. Following data acquisition, a preprocessing stage is implemented to cleanse, normalize, and structure the information, ensuring uniformity and compatibility for subsequent analysis.

The processed data is then seamlessly channelled into the ensemble anomaly detection model, encompassing statistical, machine learning, behavioural, and signature-based approaches. This architecture is adept at integrating and orchestrating these diverse methods to conduct a comprehensive examination of anomalies. The model's outputs are cross-verified to identify anomalies flagged by multiple approaches, enhancing the reliability of the results. The culmination of this architectural design is an adaptive system that effectively harnesses data from various sources to provide holistic anomaly detection insights, supporting cybersecurity professionals in their efforts to fortify digital defences.

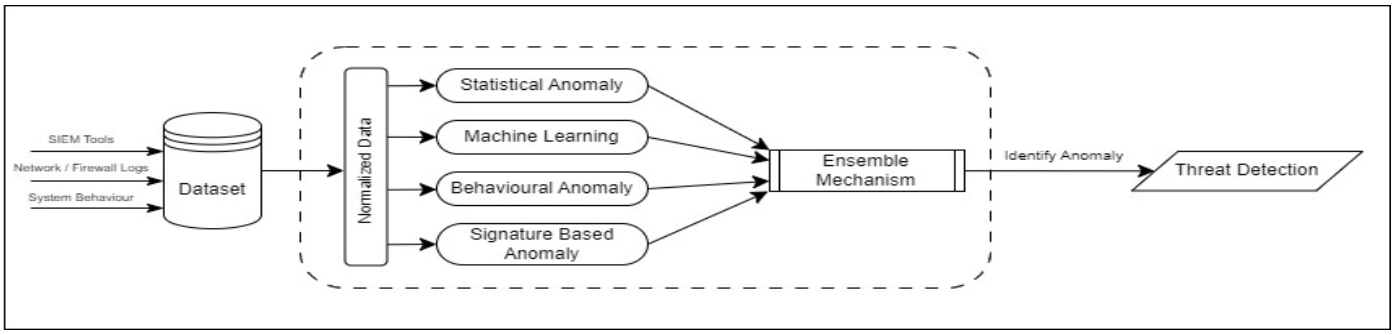

Fig 1 - Architectural Diagram of Proposed Methodology

The initial phase of our architecture involves gathering data from various sources such as SIEM tools, network/firewall logs, and system behaviours. This diverse dataset is consolidated into a standardized JSON log, which then undergoes normalization and preprocessing. This step ensures that only pertinent features are retained, setting the stage for effective anomaly detection.

Statistical Anomaly - Under statistical anomaly, we are using point anomaly and collective anomaly detection

a. **Point Anomaly Detection**: Employing the Isolation Forest method, known for its scalability, outlier resilience, and interpretability. This method excels in identifying individual anomalies in large-scale cybersecurity datasets.
b. **Contextual Anomaly Detection**: Utilizing DBSCAN, a method adept at identifying clusters of anomalies based on data density. DBSCAN's adaptability to density variations and noise handling enhances its efficacy in contextual anomaly identification.

Machine learning - Leveraging a semi-supervised model, specifically One-Class SVM, due to its ability to identify anomalies by modelling normal data patterns. Well-suited for high-dimensional feature spaces, it accommodates the numerous parameters in our input data. Notably, One-Class SVM's resilience against overfitting ensures robust generalization to dynamic cybersecurity data.

Behavioural Anomaly- Establishing parameters for potential malicious activity within User and Entity Behavior Analytics (UBA and EBA). Rule sets include identifying multiple failed logins, unusual activity timestamps (e.g., midnight logins), and abnormal browsing patterns to enhance threat identification.

Signature-based Anomaly - Implementing YARA rules to define patterns indicative of malicious activities. For instance, rules can be set to identify malicious logins, brute force attacks, or match specific malware signatures within the network. A YARA rule example is provided for clarity.

```
rule malicious_activity_rule {
   strings:
      $malicious_login = "malicious_user"
      $brute_force = "attacker"
      $malware_signature = "Malware Signature"

   condition:
      any of ($malicious_login, $brute_force, $malware_signature)
}
```

This approach enriches the system's ability to identify potential threats by recognizing specific activities aligned with known malicious patterns.

Ensemble Mechanism - The outputs and anomalies identified by the individual anomaly detection methods are collectively processed through an ensemble mechanism. This ensemble mechanism seeks the convergence of anomalies identified across all methods, and those anomalies with the highest consensus are regarded as potential threats. In essence, this ensemble approach ensures a rigorous verification process, requiring a unanimous agreement from at least three out of the four anomaly detection methods to classify an event as suspicious or indicative of malicious behaviour. By establishing such a stringent criterion, the ensemble mechanism enhances the reliability and accuracy of the threat identification process, minimizing false positives and fortifying the overall anomaly detection framework. This collaborative verification strategy significantly contributes to the robustness of the cybersecurity system, ensuring a comprehensive and validated response to potential threats.

1. Input Phase

The architecture of the anomaly detection system unfolds in three integral phases, commencing with the Input Phase. In this initial stage, the system acquires data from diverse sources crucial to the anomaly detection process. Notably, data is sourced from network logs, Security Information and Event Management (SIEM) systems, and behavioural monitoring systems. This heterogeneous input pool captures a comprehensive spectrum of activities, ranging from network behaviours to user interactions, forming the foundational dataset for subsequent analysis. Example of sample input data:


```
{
  "timestamp": "2024-03-02T12:30:45.123Z",
  "event_id": 123456,
  "source_ip": "192.168.1.100",
  "source_port": 4829,
  "destination_ip": "203.0.113.5",
  "destination_port": 80,
  "protocol": "TCP",
  "user": "john.doe",
  "user_role": "Employee",
  "department": "IT",
  "user_privileges": ["Read", "Write"],
  "event_category": "Authentication",
  "event_type": "Login",
  "event_outcome": "Success",
  "event_description": "User successfully logged in",
  "application": "Windows Security",
  "device": "Domain Controller",
```

```
        "bytes_sent": 589,
        "bytes_received": 1532,
        "source_location": "CityA, CountryX",
        "destination_location": "CityB, CountryY",
        "session_id": "ABC123",
        "session_duration": "00:15:30",
        "file_name": "important_document.docx",
        "file_path": "/docs/confidential/",
        "file_type": "Word Document",
        "file_size": 2048,
        "network_protocol_details": {
          "url": "http://example.com/page",
          "status_code": 200,
          "user_agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/58.0.3029.110 Safari/537.3"
        },
        "event_severity": "High",
        "vlan": "VLAN1",
        "subnet": "192.168.1.0/24",
        "additional_metadata": {
          "custom_field": "value",
          "organization_specific_info": {
            "compliance_status": "Compliant",
            "last_audit_date": "2024-02-15",
            "regulatory_requirements": ["GDPR", "HIPAA"]
          }
        },
        "historical_data": {
          "last_login_timestamp": "2024-03-01T18:45:30.567Z",
          "login_count_last_24h": 3
        },
        "device_logs": {
          "firewall_logs": {
            "timestamp": "2024-03-02T12:35:00.567Z",
            "log_data": {
              "action": "ALLOW",
              "source_ip": "192.168.1.100",
              "destination_ip": "203.0.113.5",
              "protocol": "TCP",
              "port": 80,
              "message": "Allowed traffic from source IP to destination IP on port 80"
            }
          },
          "antivirus_logs": {
            "timestamp": "2024-03-02T12:36:30.567Z",
            "log_data": {
              "file_path": "/docs/confidential/important_document.docx",
              "scan_result": "Clean",
              "message": "File scanned and found to be clean"
            }
          },
          "intrusion_detection_logs": {
            "timestamp": "2024-03-02T12:38:15.567Z",
            "log_data": {
              "event_type": "Brute Force Attack",
              "source_ip": "192.168.1.100",
              "destination_ip": "203.0.113.5",
              "protocol": "SSH",
              "message": "Detected a potential brute force attack on SSH"
            }
          }
        }
}
```


This exemplary input data captures a snapshot of an authentication event, encompassing details such as timestamp, source and destination IP addresses, user information, event outcome, and other relevant parameters. A distinctive feature of this dataset is the

integration of device logs, encompassing firewall, antivirus, and intrusion detection logs. These logs provide granular insights into related activities, such as allowed traffic, scan results for files, and potential security threats like brute force attacks. The inclusion of such diverse information ensures a comprehensive dataset, allowing the anomaly detection system to draw insights from various facets of network and system activities during subsequent processing phases.

It is imperative to note that the sample data has been intentionally generated to rigorously test the proposed methodology. Spanning a 24-hour timeframe, the dataset incorporates both malicious activities and intentional anomalies, ensuring a diverse and challenging dataset for the robust evaluation and adaptation of the anomaly detection system. This intentional approach underscores the methodology's efficacy in handling real-world scenarios and its resilience in the face of sophisticated cyber threats.

2. Processing Phase

The Preprocessing Phase plays a crucial role in refining and structuring the raw input data to prepare it for subsequent analysis. In this phase, various steps are taken to cleanse, normalize, and organize the data. During the preprocessing phase, the input data may undergo various transformations such as data type conversions, removal of irrelevant features, handling missing values, and normalization of numerical values. Additionally, categorical variables may be encoded, and timestamps may be standardized for consistency. The sample preprocessed data retains the essential features while being primed for subsequent analysis in the anomaly detection system. The specific preprocessing steps may vary based on the characteristics of the data and the requirements of the anomaly detection model.

3. Output Phase

In the Output Phase, the results from the ensemble method, combining all four types of anomaly detection methods, are distilled into actionable insights. The ensemble method leverages the collective intelligence of various techniques to enhance accuracy and mitigate false positives and negatives. Below is a sample representation of the output data in the context of the ensemble method


```
    {
     "timestamp": "2024-03-02T12:30:45.123Z",
     "user": "john doe",
     "source_ip": "116.25.98.11"
     "ensemble_decision": "Threat Detected",
     "anomaly_methods": {
      "statistical": "No Threat",
      "machine_learning": "Threat Detected",
      "behavioural": "Threat Detected",
      "signature_based": " Threat Detected"
     },
     "confidence_scores": {
      "statistical": 0.15,
      "machine_learning": 0.85,
      "behavioural": 0.90,
      "signature_based": 0.95
     }
}
```


In this sample output:

a. **timestamp**: the time when the anomaly detection output is generated.
b. user: denotes which user's activity was flagged or found malicious
c. **source_ip**: provides the Internet Protocol address of the user whose activity was flagged malicious
d. **ensemble_decision**: provides the final decision made by the ensemble method based on a majority vote from the individual anomaly detection methods. In this case, the ensemble method concludes "Threat Detected" as it aligns with the majority.
e. **anomaly_methods**: showcase the individual decisions made by each anomaly detection method. The ensemble method combines these to arrive at the final decision.
f. **confidence_scores**: represent the confidence or strength of each anomaly detection method's decision. This information can be useful for understanding the reliability of each method and contributes to the ensemble's decision-making process.

III. RESULT AND DISCUSSION

In this paper, four anomaly-based detection methods were employed: statistical anomaly involving point anomaly and collective anomaly using isolation forest and DBSCAN, one-class SVM for machine learning, EBA and UBA with various parameters for behavioural analysis, and pattern matching and YARA rules for signature-based detection to identify malicious activities in the network. Figure 2 illustrates the results of the calculated output for threat detection using these four methods. As depicted, three out of the four methods exhibit high confidence in detecting anomalies, leading to the final outcome of "threat detected."

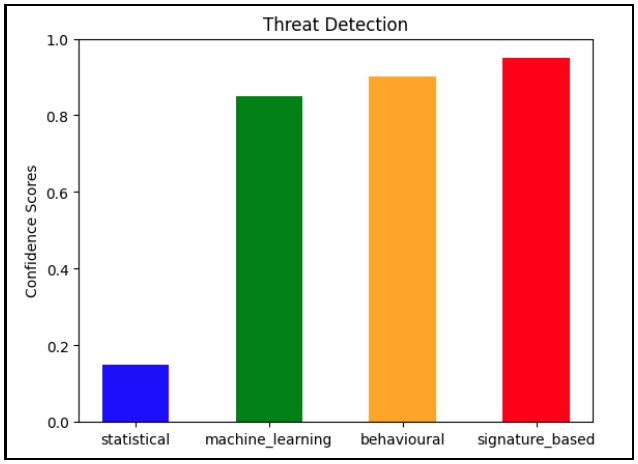

Fig 2 - Result of Ensemble Mechanism

Fig 3 illustrates the anomalies detected by the statistical-based anomaly method. Both point anomaly and collective anomaly are calculated independently, and the common anomalies are considered for higher confidence.

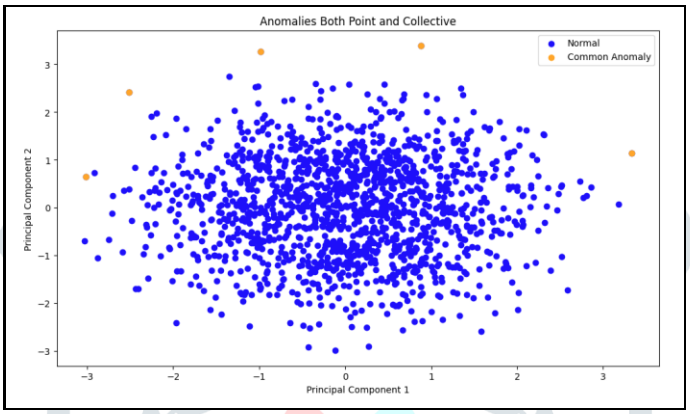

Fig 3 - Anomalies detected using Statistical Anomaly

Fig 4 shows the anomalies found using one-class SVM under machine learning technique.

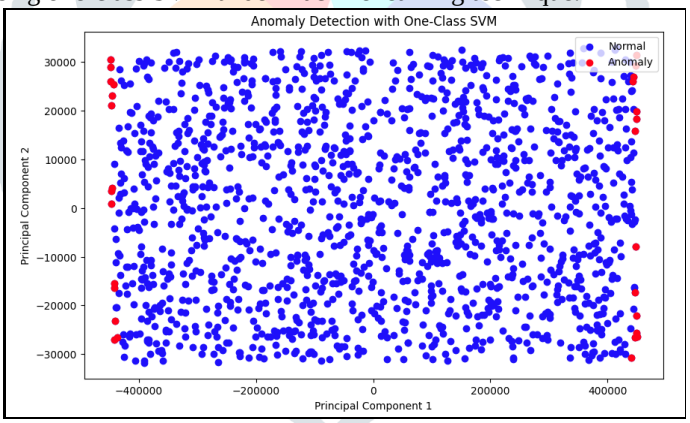

Fig 4 - Anomalies detected using Machine Learning

Fig 5 below provides insights from behavioural anomaly analysis, highlighting various types of events that occurred in the network.

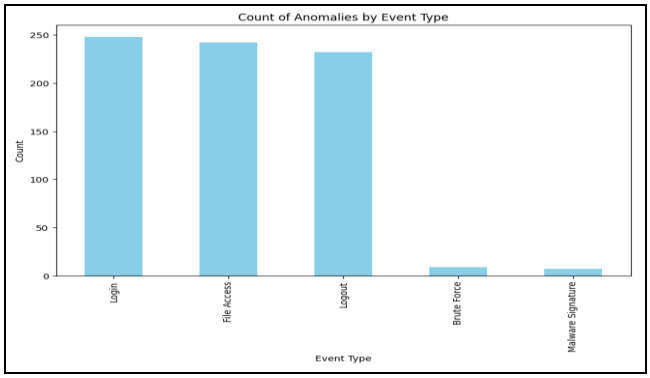

Fig 5 - Behavioural Anomaly by Event Type

Fig 6 shows the malicious activities found by pattern matching & YARA rules using Signature- methods.

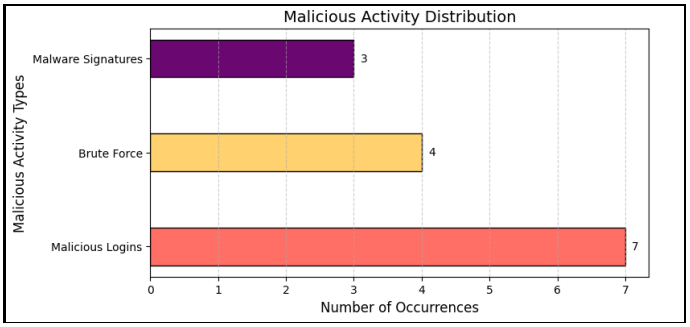

Fig 6 - Malicious Activity found using Signature-based Anomaly

IV. Challenges & Opportunity

A. Limited Training Data and Signatures

The anomaly detection approach encounters a substantial challenge due to the restricted scope of training data and signatures. Expanding the diversity and volume of training data is crucial to unravel a nuanced understanding of potential threats and enhance the model's adaptability to a broader spectrum of scenarios. Addressing this challenge presents an opportunity to curate an extensive and diverse training dataset, fostering a more comprehensive understanding of anomalous patterns and refining the model's capabilities.

B. Live Stream Data Refinement

The application of the approach to a larger live stream data set introduces both challenges and opportunities. While unlocking valuable insights and improving real-time threat identification is a potential benefit, the sheer volume of data poses computational challenges. Efficient refinement strategies are essential to optimize the model's performance in handling large-scale, dynamic datasets in real-time. This challenge provides an opportunity to develop sophisticated refinement strategies, ensuring the model's efficiency in processing live stream data. Strategies aimed at real-time data refinement could enhance the model's accuracy and responsiveness.

C. Resource Intensiveness

The intricate computations involved in the approach, especially in a live network setting, pose a challenge in terms of resource intensiveness. Efficient resource management and optimization strategies become imperative to prevent potential bottlenecks and ensure the scalability of the model. Exploring distributed computing approaches emerges as an opportunity to mitigate the resource burden associated with extensive computations. Implementing distributed systems and parallel computing architectures can optimize the computational load, enhancing efficiency and scalability.

D. Heterogeneity of applications & Defining Behavioural Boundaries

The intrinsic diversity among applications poses a significant challenge in anomaly detection. Healthcare data, for instance, may exhibit distinctive patterns compared to supply chain data. The varied characteristics and behaviours across different application domains require a nuanced approach to anomaly detection. Developing methodologies capable of accommodating this heterogeneity becomes imperative for ensuring accurate and context-aware threat identification. Establishing clear boundaries between normal and abnormal behaviour represents a formidable challenge in anomaly detection. The dynamic nature of cyber threats and the evolving landscape of network activities contribute to the complexity of this task. Crafting effective algorithms and models that can discern subtle deviations without generating false positives is essential.

V. Conclusion

Anomaly-based detection in Cyber Threat Intelligence (CTI) is a proactive approach that aims to identify deviations from expected behaviour within a network or system. This research paper explores the significance of anomaly-based detection in enhancing CTI capabilities and provides a nuanced understanding of its application for real-time threat identification and mitigation. The study investigates various anomaly detection techniques, ranging from statistical models to machine learning algorithms, and assesses their efficacy in discerning malicious activities within network traffic and system behaviours. Special emphasis is placed on the integration of anomaly-based detection into a broader CTI framework, emphasizing the synergy between automated detection mechanisms and human analysts. The paper addresses challenges associated with anomaly-based detection, such as false positives and the dynamic nature of cyber threats. It proposes potential solutions and optimisations to mitigate these challenges, ensuring a more accurate and reliable threat detection system. The results showcase the effectiveness of anomaly-based detection in identifying emerging threats and enabling swift, targeted mitigation strategies.

Anomaly detection is a methodological approach that identifies patterns or behaviours that deviate from established norms within a given context. It serves as a proactive line of defence beyond traditional cybersecurity measures, employing techniques such as statistical models and machine learning algorithms. This approach addresses the dynamic nature of the cyber environment, minimizing false negatives and diminishing reliance on pre-existing knowledge. Anomaly detection can identify anomalous patterns of behaviour in real-life scenarios, even in the absence of specific signatures or prior knowledge.

Real-time threat identification is crucial in cybersecurity, as it allows for the identification of deviations from expected behaviour within a network or system. Rapid detection can help minimize the impact on systems and data, making real-time threat identification a pivotal aspect of cybersecurity. However, current gaps in cyber threat intelligence include an excessive number of false positives or negatives, limitations in insider threat identification, and challenges in adapting to normal changes within network behaviour.

The proposed solution employs an ensemble method in anomaly detection, strategically incorporating diverse techniques to amplify overall efficacy and reduce instances of false positives.

VI. REFERENCES

[1.] Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” ACM computing surveys (CSUR), vol. 41, no. 3, p. 15, 2009.

[2.] H. Schulze, “Insider threat report: 2018.” https://www.ca.com/content/dam/ca/us/files/ebook/insider-threat-report.pdf. [3] Experian, “2019 global identity and fraud report.”https://www.experian.com/decision-analytics/global-fraud-report.html.

[3.] J. Jiang, J. Chen, K.-K. R. Choo, K. Liu, C. Liu, M. Yu, and P. Mohapatra, “Prediction and detection of malicious insiders’ motivation based on sentiment profile on webpages and emails”, in MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM), pp. 1–6, IEEE, 2018.

[4.] Abdallah, M. A. Maarof, and A. Zainal, “Fraud detection system: A survey”, Journal of Network and Computer Applications, vol. 68, pp. 90–113, 2016.

[5.] Adam Khalid, Anazida Zainal, Mohd Aizaini Maarof, Faud A. Ghaleb, “Advanced Persistent Threat Detection: A Survey”, in 3rd d International Cyber Resilience Conference, IEEE, 2021.

[6.] S. Ghosh and S. Sampalli, "A survey of security in SCADA networks: Current issues and future challenges," IEEE Access, vol. 7, pp. 135812-135831, 2019.

[7.] M. Jouini and L. B. A. Rabai, "A security framework for secure cloud computing environments," in Cloud Security: Concepts, methodologies, tools, and applications: IGI Global, 2019, pp. 249-263.

[8.] X. Wang, K. Zheng, X. Niu, B. Wu, and C. Wu, "Detection of command and control in advanced persistent threat based on independent access," in 2016 IEEE International Conference on Communications (ICC), 2016, pp. 1-6: IEEE

[9.] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network Aanomaly detection techniques," Journal of Network and Computer Applications, vol. 60, pp. 19-31, 2016.

[10.] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," IEEE Communications Surveys & tutorials, vol. 16, no. 1, pp. 303-336,2013.

[11.] P. Chen, L. Desmet, and C. Huygens, "A study on advanced persistent threats," in IFIP International Conference on Communications and Multimedia Security, 2014, pp. 63-72: Springer

[12.] M. A. Kabir and X. Luo, ‘‘Unsupervised learning for network flow based anomaly detection in the era of deep learning,’’ in Proc. IEEE 6th Int. Conf. Big Data Comput. Service Appl. (BigDataService), Aug. 2020, pp. 165–168

[13.] G. Andresini, A. Appice, N. D. Mauro, C. Loglisci, and D. Malerba, ‘‘Multi-channel deep feature learning for intrusion detection,’’ IEEE Access, vol. 8, pp. 53346–53359, 2020.

[14.] Y. Fu, F. Lou, F. Meng, Z. Tian, H. Zhang, and F. Jiang, ‘‘An intelligent network attack detection method based on RNN,’’ in Proc. IEEE 3rd Int. Conf. Data Sci. Cyberspace (DSC), Jun. 2018, pp. 483–489.

[15.] Kwon, K. Natarajan, S. C. Suh, H. Kim, and J. Kim, ‘‘An empirical study on network anomaly detection using convolutional neural networks,’’ in Proc. IEEE 38th Int. Conf. Distrib. Comput. Syst. (ICDCS), Jul. 2018, pp. 1595–1598.

[16.] Creech and J. Hu, ‘‘A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns,’’ IEEE Trans. Comput., vol. 63, no. 4, pp. 807–819, Apr. 2014.

[17.] M. Al-Zewairi, S. Almajali, and A. Awajan, ‘‘Experimental evaluation of a multi-layer feed-forward artificial neural network classifier for network intrusion detection system,’’ in Proc. Int. Conf. New Trends Comput. Sci. (ICTCS), Oct. 2017, pp. 167–172.

[18.] M. Xie, J. Hu, and J. Slay, ‘‘Evaluating host-based anomaly detection systems: Application of the one-class SVM algorithm to ADFA-LD,’’ in Proc. 11th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD). Xiamen, China: Springer, Aug. 2014, pp. 542–549.

[19.] Matthew Mahoney, ‘‘Network traffic anomaly detection based on packet bytes,’’ in Proc. ACM Symp. Appl. Comput. (SAC). New York, NY, USA: Association for Computing Machinery, 2003, pp. 346–350.

[20.] J. McHugh, ‘‘Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln laboratory,’’ ACM Trans. Inf. Syst. Secur., vol. 3, no. 4, pp. 262–294, Nov. 2000.

[21.] J. Liu, L. Yin, Y. Hu, S. Lv, and L. Sun, “A Novel Intrusion Detection Algorithm for Industrial Control Systems Based on CNN and Process State Transition,” 2018 IEEE 37th International Performance Computing and Communications Conference, IPCCC 2018, 2018.