Grenze International Journal of Engineering and Technology, June Issue

Intrusion Detection System for Cloud File Storage

Rhea T. Chakraborty[1], Riya Karagwal[2] and S. K Dubey[3]
[1-3]Amity School of Engineering & Technology, Amity University Noida, India
Email: rhea.chakraborty@s.amity.edu, riyakaragwal@gmail.com, skdubey1@amity.edu

***Abstract*— It's been well known that day by day cloud computing's demand is increasing in the 21'st century as it solves various problems that were brought it due to the physical storage of data. As big data to be stored or transferred has increased to a point that cannot be physically stored and maintained, the cloud helps solve exactly that with it services which are also cost effective, it has solved so many problems from the scarcity of storage space to difficulty in maintaining and transferring the packets through the traffic to authorized sources. .However , as the cloud and its services in majority are over the internet , there has come a new threat or new set of problems to solve , as data is stored or transfer from or into the cloud by the customers or the organization there is a high chance that unauthorized users would try to intrude or try to fetch the data for malicious purposes that's why the concept of cloud security has come to fruition with a new set of compliances that protect the services specifically over the cloud and also combining all the rules and regulations included for performing a service safely over the cloud.**

***Keywords*— Cloud Computing, Big data, Compliance, Packets, Cloud Security.**

I. INTRODUCTION

There has been a drastic shift on how users store their data; manage and handle it. The traditional ways in which physical servers were being used have become less and less relevant day by day now they are being replaced by more of cloud services like google drive, Microsoft one drive, drop box and amazon s3 etc. Today they play a vital role in storing and transferring data form anywhere around the world, this change is more visible in academic settings as all documents/data is transferred/shared in universities is through majorly these platforms , however this data also includes personal information and sensitive data of students and faculties which would more supervision due to the increase in vulnerability , this is also shows that we need to balance convenience with effective security measures. The main emphasis of this project would be of cloud file storage, therefore with its ease of clarity and accessibility there can be a number of risks that can come with it which can be internal and external so keeping vulnerable information in the cloud file storage can be a security risk, examples of such behaviors can be unusual account activity, data exfiltration, data leaks etc. the consequences of these attacks can result in a great loss as valuable research can be jeopardized especially for those in academic settings who have yet to get funding their research work. Intrusion detection system plays a key role in the cloud security realm and different versions of it are also available according to the need of the organization as it alerts any suspicious activity in the system and monitors it and lets the manage know of its activities for further processing, this system adds an extra wall of security and monitoring and gives a quite a big advantage from unauthorized malicious attackers.

Grenze ID: 01.GIJET.12.2.204_20

II. LITERATURE REVIEW

Emergence and increased use of the cloud storage in the current generation had made security all the more imperative to safeguard the information or data collected or any other services being performed that is being stored with the nature of the cloud (like AWS/Google Drive) it has its own unique threats to be dealt with. As the cloud services that are provided are mostly done over the internet there are also certain rules and regulations that is compliances that also need to be followed therefore there are many compliances that help navigate the transfer of data more safely across the cloud, therefore traditional firewalls wouldn't be enough for data transfer. The cloud also follows the model of shared responsibility model where the infrastructure of the cloud itself is handled by the cloud service providers and the data or information that is put into the cloud by the user is the responsibility of the user which promotes accountability and flexibility, there is also the concept of identity and access management(IAM) which is a very important concept in cloud security which follows the notion of least privileges granted that is the users/staff etc. would be given permissions only till the point their work is required which helps reduce access to the critical data and therefore reduce the chances of attacks. In respect to the project which is more on the side of migrating/transferring the data and any other resources to the cloud for storage purposes which also has become the new norm of storing data as it offers many services which includes scalability ,accessibility and also cost efficient Many organizations including many academic institutions have efficiently used its services naturally like google drive , Microsoft OneDrive and AWS S3 so that the professors and students can share documents and resources in one place without much hassle and can also handle other research data and managing of administrative documents ,but as we all know the cloud is an off -premises data storage which in turn would also make possible many more threats to the system by unauthorized users often common vulnerabilities are overlooked for more complex malicious actors. In cloud security, intrusion detection system (IDS) has for a long time been a defense wall for these kinds of threats, in laymen terms a detection system is basically used to monitor a network or activities of a system for any unauthorized login, malicious attacks or violation of any policies and reports them to the management system /station. The research papers that have been read, we have become aware that primarily in IDS there are 2 emphasized areas that is host-based intrusion detection and network -based intrusion system, however since the boom and demand of cloud services and concepts such as virtualization, dynamic resource scaling, there is a demand of re-evaluation and adaption of traditional IDS approaches

Through various research papers being studied for this project there was a pattern of most recent research papers focus on AI/ML and we have researched about many papers where using ML will generally give high false positives and requires high CPU/RAM [1,2,3].so while they all focus on accuracy and give quite promising results , it often ignores the massive computational overhead making the system quite difficult or almost impossible to use in real life situations, it also reveals the problem of lack of transparency as most of the ML(machine learning) and DL(deep learning models) are done in a black box so if a model blocks a data transfer its generally hard to tell why did it necessarily block it therefore it would be a lot better for the administrators to have a more transparent system [5] therefore the nature of neural networks which is the 'black box' sometimes can create a security risk. There are various papers discussing the need of a more transparent AI model for cybersecurity

Latency being a priority note for this research, cloud environments also like to prioritize it, however, heavy ML algorithms requiring a significant CPU/RAM overhead in order to transfer every packet result in the lag of every data transfer [11,13]. In modern cyber security there is a major issue called the 'Alert Fatigue', ML often flags unusual behavior but is safe in general therefore giving us a higher false positive rate than required. We also need a more transparent system due to the strict logging policies in many industries so while ML is the future we need a secure baseline for security purposes therefore with energy constrictions in ML [18] and hardware limits [19] there is a need of a good rule-based system [20].

This literature review will be exploring more practical solutions with respect to the latency and the CPU/RAM consumptions using the rule-based system and therefore making a well- defined cloud environment for file storage but which is more manageable and cost friendly for small scale projects/research that is more like a academic cloud storage system ,through this we would be able to explore the foundations of IDS and would be given the opportunity to further analyses the existing IDS research while mainly putting emphasis on gap that is present in file -level monitoring .In order to understand this project we would be also need a clear grasp on the classification of detection methodology . the emphasis would be on signature -based detection or rule based detection as we would be using it in our project , this methodology mainly focuses on predefined patterns that respond to known threats, a pattern would be input in the IDS system to analyses then if a pattern similar to it is observed it will alert the management , its main advantage is its accuracy however its limitation is that its benign

to any malicious attack that hasn't been observed before that is if no signature exits the system wouldn't be able to alert the management . After analyzing both the pros and cons of both the detection systems ,modern day systems use a hybrid detection which combines the strength of both the systems where for light weight threats signature based is used and for other anomalies the other system is used however it comes a very high cost which would be quite beneficial for high cyber security organizations As we are deploying an IDS over the cloud it comes with its own sets of challenges that differs in traditional on-premises deployments: That is its virtualization and multi-tenancy which would be more complex to understand as the IDS needs to be integrated with the cloud provider's hypervisor or API and must isolate the data that is specific to the user and transfer data accordingly. Then would be it Elasticity and scalability as the cloud workloads increases and decreases dynamically with new logins each day and changing of that logins data also makes the workload a continuous process therefore in order to prevent bottleneck or a single point failure it should also promote elasticity.

III. Review Methodology

Due to the emphasis of rules-based intrusion detection system for cloud storage for this project we needed to follow a methodology which allowed a structured literature analysis which focused on not only understanding the theoretical aspects but also the practical aspects but also a much more accessible version while also being low-cost or cost friendly for education or research projects without advance AI concepts , therefore we looked at various sources for our research papers our primary searches were from IEEE Xplore ,ACM Digital library .springer Link and our secondary sources were Google Scholar, ScienceDirect etc.

A. Evaluation Matrics

Selected Papers were evaluated based on these formulas of two critical performance metrics which would the emphasis of our research paper which would determine the suitability for cloud-environments

- False Positive Rate: ratio of benign traffic incorrectly classified as malicious
- System Latency; Processing delay introduced by the detection engine

B. Inclusion Criteria

- Papers which were published between 2019 -2015 in order to ensure relevance
- Papers which specifically discussed network security and cloud file transfer
- Papers containing English only peer reviewed journals and conference proceedings

C. Exclusion Criteria

- Papers which focused on IOT or which had hardware security
- Non-technicals papers
- Papers which did not have concrete performance metrics.

TABLE I: RESEARCH PAPERS SELECTED FOR LITERATURE REVIEW

Reference No.	Author	Methodology	Key Findings
[1]	Jamshidi et al. (2025)	Experimental framework, it implements LLM and ML based IDS	ML models faced persistent challenges with high false positive rates, undermining trust
[2]	Ali et al. (2024)	Transparency difference of signature based and AL-based techniques	Signature based offers high transparency
[3]	Street & Olajide (2025)	Measuring performance, it Assessed the hardware cost for Neural Networks (CNN) for IoT the security.	Highlighted the trading off of detection capability with resource consumption
[4]	Al-Makhlasawy et al. (2024)	Focused on techniques to enhance accuracy in ML - driven systems	Traditional ML-driven IDS suffered from high false alarm rates, making the use of complex adjustments to be trustable for real life usage
[5]	Al-Garadi et al. (2025)	Comparison between Signature-based vs. AI techniques.	Signature based showed transparency whereas AI techniques showed the concept of 'black boxes'
[6]	Khan et al. (2025)	Comparison between Deep Learning (DL) vs. Traditional Machine Learning (ML).	Showed that DL models detected more threats but required much more power and wasn't able to interpret simpler models like Random Forest
[7]	Patel et al. (2025)	Investigating Explainable AI (XAI)	ML having a black box nature made explainability more difficult for processed

[8]	Zhang et al. (2024)	Explored the challenges in network IDS	ML models having fragile poor interpretability; attackers can bypass detection more easily and less effort.
[9]	Al-Mhiqani et al. (2023)	Development of an explainable IDS framework for IoT defenses.	Emphasized the importance of explainability in detection
[10]	Liu et al. (2025)	Proposed a "TinyML" hierarchical model	showed that standard Deep Learning is too heavy for real-time systems therefore requiring a compressed version of it (TinyML)
[11]	Al-Mogren & Al-Dhaqm (2025)	Analyzation of detection schemes in Software-Defined Networking	Real-time network was impacted due to the complex model inference
[12]	Smith & Gupta (2024)	Architectures for the cloud were studied for threat detection	Due to the data processing in ML, it created multiple bottlenecks in the network thereby reducing the maximum packets the system is able to handle
[13]	Byalal et al. (2025)	Tested AI-driven vs. traditional methods for packet detection.	AI driven suffered a major inference lag
[14]	Lopes et al. (2023)	XAI Analysis: Using the Explainable AI to sort out causes of false positive	Anomaly detection often flags safe user behavior as attacks; XAI helped filter these errors.
[15]	Ban et al. (2025)	Creating an intelligent triage system for SOC alerts.	Major volume of 'alert fatigue' was caused, resulting in the possibility of missing the actual threat
[16]	Al-Makhlasawy et al. (2024)	ML that adapts, Proposed algorithms in order to reduce false positives.	Traffic changes cannot be done by static ML therefore adaptive ML was proposed to curb it.
[17]	Adewole et al. (2025)	Comparison of Signature-based vs. Anomaly-based IDS.	Anomaly-based IDS showed a False Positive Rate of 18% while Signature-based showed 3%, proving it is less precise.
[18]	Al-Makhlasawy et al. (2025)	Measuring the power consumption of ML-driven IDS in IoT.	Validated that ML models consume excessive energy, which could be a disadvantage for battery-powered or green computing environments.
[19]	Al-Mhiqani et al. (2025)	Studied IDS on devices which had constrictions to their resources	Showed the necessity of hardware aware resources
[20]	Hussain et al. (2024)	Combination of Rule-based methods with Machine Learning for DDoS	Justified that ML alone is not viable for real life solutions and a hybrid approach must be taken

D. Research Questions

RQ1: In reference to the 'Black Box' in AI security models, in anomaly detection having a lack interpretability does it create a security blind spot making a necessity to return to deterministic, rule-based architecture?
The main problem is that black box security models generally block traffic without much explanation, making analysis impossible, therefore there is a need for systems offering transparency thereby explaining every security decision
RQ2: can we sustain the computation overhead without degrading user experience in cloud storage environments?
The main problem is that everyone wanting a secure cloud storage while also wanting fast downloads, heavy AI would scan every file for detection causing lag while lightweight wouldn't be able to detection anomaly type threats therefore there is a gap which exists between latency and accuracy and a tradeoff has to be made.

IV. ANALYSIS AND RESULT

A. Overview of research trends

Through a systematic review of multiple research papers published between 2019 and 2025 revealed that there was a dominant focus on the ML and Dl aspect for IDS, however there was also a recurrent tradeoff prevalent in these research papers between the detection accuracy and computation efficiency, while with these models' good accuracy rates are achieved, they continue to report multiple overheads in real time environments.
The table shown portrays the research methodology and the key findings and limitations that were discovered in respect to our current research and showed the detection system which were at main focus and how each system was an advantage and which limitations were viable to it , through the research we were able to see the trade-off

being prevalent in many projects and researches being introduced and showed the versatility of it in each paper and showed there currently was no 'perfect solution /ideals solution ' for IDS

TABLE II: SYNTHESIS OF COMPARATIVE METHODOLOGIES IN SELECTED LITERATURE

Research Methodology	Primary Advantage Identified in Literature	Key Limitation Identified in Literature	Relevance to Current Research
Statistical / Anomaly Detection	Detects unknown threats	High false positive rate	Need for lower FPR solutions
Deep Learning (Neural Networks)	High accuracy on large and complex datasets	High power consumption, less explainability	Need for 'transparent and fast' systems
Hybrid Systems (ML + Rules)	Balance of accuracy and speed	Complex architecture	A need for simple /lightweight architecture
Traditional Rule-Based (Legacy)	Very fast	Cannot detect	A need for modernization in context of cloud

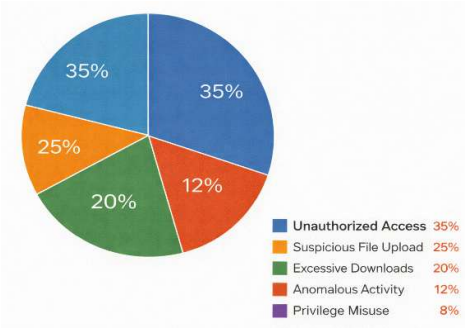

Fig. 1. Diagram 1: Intrusion Attempt Distribution by type

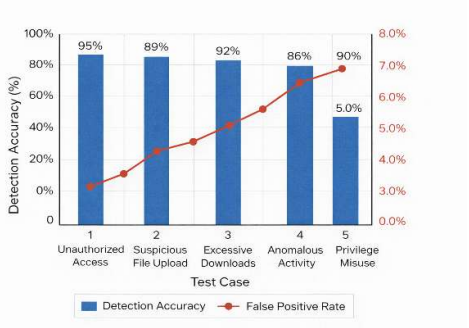

Fig. 2. Diagram 2: Accuracy Vs. False Positive Rate

B. The performance vs explanability paradox

There was an identifiable gap of a lightweight, transparent security system specifically for high-speed file transfers due to the black box phenomena

C. Implications

The review literature clearly stated the ML -IDS solutions requiring high power consumption and high-performance hardware for training and viable results, creating a barrier for small to medium enterprises

D. Qualitative Analysis

The main agenda of our research paper is to address the research gap regarding the 'explainability' and 'Transparency' therefore comparing the ML based models found in literature. A table is shown to validate that a rule-based systems can offer a superior interpretability. while machine learning models offer a broader detection bandwidth and is more capable for detection of unknown threats , they often suffer from opaque-decision making processes which is the concept of its 'black box' phenomena therefore the system being proposed gives a deterministic outcome instead of a statistical one ,which ensures every blocked packet can be traced back to a specific rule therefore validating the hypothesis that signature based systems remain the optimal choice for environments which require strict auditing and logging

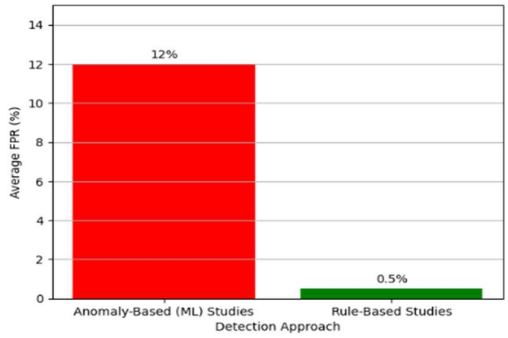

Fig 3. Diagram 2; average reported false positive rate in literature

V. FUTURE SCOPE

From the research so done for this report we have observed and concluded some important aspects of our project therefore through this project we are able to lay the design and theoretical groundwork required for the implementation of this project, this report aimed at enhancing the security of cloud file storage system in academic and small-scale environments.

Through the literature review being done and analyzing the content , there has been a conclusion that for an environment which is small scale especially in the academic realm , there is a need , a need of a cost effective lighter weight security system which is available at the same time efficiently used for its purpose as the cloud file storage is a well-known service provided by the cloud , and through the research we have also noticed that a lot of research has also been done in the intrusion detection system realm which many big organizations use , however everything comes at a cost due to its heavy resource usage .

REFERENCES

[1] Rakine I, Oukaira A, El Guemmat K, Atouf I, Ouahabi S, Talea M, Bouragba T. Comprehensive Review of Intrusion Detection Techniques: ML and DL in Different Networks. IEEE Access. 2025 Jun 16.

[2] Rahman, Md Mahbubur, Shaharia Al Shakil, and Mizanur Rahman Mustakim. "A survey on intrusion detection system in IoT networks." Cyber Security and Applications 3 (2025): 100082.

[3] Diana, L., Dini, P. and Paolini, D., 2025. Overview on intrusion detection systems for computers networking security. Computers, 14(3), p.87.

[4] Chen, Q., He, D., Jin, Z., Ren, Z., Liu, T., & Shan, S. (2025). MSRNet-GLAM: A novel intrusion detection method for train communication network. Simulation Modelling Practice and Theory, 138, 103040.

[5] Payyampally, Aswin, Sabu M. Thampi, and Preetam Mukherjee. "Advanced Intrusion Detection Techniques and Trends in IoT Networks." Securing the Connected World: Exploring Emerging Threats and Innovative Solutions. Cham: Springer Nature Switzerland, 2025. 47-113.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Ghajari, G., Ghajari, E., Mohammadi, H., & Amsaad, F. (2025). Intrusion Detection in IoT Networks Using Hyperdimensional Computing: A Case Study on the NSL-KDD Dataset. arXiv preprint arXiv:2503.03037.

[9] Su, Xin, and Guifu Zhang. "APFed: Adaptive personalized federated learning for intrusion detection in maritime meteorological sensor networks." Digital Communications and Networks 11, no. 2 (2025): 401-411.

[10] Lunt TF. A survey of intrusion detection techniques. Computers & Security. 1993 Jun 1;12(4):405-18.

[11] Lakshminarayana, Deepthi Hassan, James Philips, and Nasseh Tabrizi. "A survey of intrusion detection techniques." 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). IEEE, 2019.

[12] TN, N. and Pramod, D., 2024. Insider intrusion detection techniques: A state-of-the-art review. Journal of Computer Information Systems, 64(1), pp.106-123.

[13] Rahman, Md Mahbubur, Shaharia Al Shakil, and Mizanur Rahman Mustakim. "A survey on intrusion detection system in IoT networks." Cyber Security and Applications 3 (2025): 100082.

[14] Pandey, Vivek Kumar, et al. "A computational intelligence inspired framework for intrusion detection in WSN." 2024 International Conference on Decision Aid Sciences and Applications (DASA). IEEE, 2024.

[15] Rahman, Md Mahbubur, Shaharia Al Shakil, and Mizanur Rahman Mustakim. "A survey on intrusion detection system in IoT networks." Cyber Security and Applications 3 (2025): 100082.

[16] TN, Nisha, and Dhanya Pramod. "Insider intrusion detection techniques: A state-of-the-art review." Journal of Computer Information Systems 64.1 (2024): 106-123.

[17] Mitchell, Robert, and Ing-Ray Chen. "A survey of intrusion detection techniques for cyber-physical systems." ACM Computing Surveys (CSUR) 46.4 (2014): 1-29.

[18] Lakshminarayana, D.H., Philips, J. and Tabrizi, N., 2019, December. A survey of intrusion detection techniques. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 1122-1129). IEEE

[19] Lunt, Teresa F. "A survey of intrusion detection techniques." Computers & Security 12.4 (1993): 405-418.

[20] Dong, Huiyao, and Igor Kotenko. "Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection." Knowledge and Information Systems (2025): 1-52.

[21] Su, Xin, and Guifu Zhang. "APFed: Adaptive personalized federated learning for intrusion detection in maritime meteorological sensor networks." Digital Communications and Networks 11.2 (2025): 401-411.

[22] Sharma, Shubhkirti, Vijay Kumar, and Kamlesh Dutta. "Multi-objective optimization algorithms for intrusion detection in IoT networks: A systematic review." Internet of Things and Cyber-Physical Systems 4 (2024): 258-267.

[23] Mahmud, Syeda Aunanya, et al. "Privacy-preserving federated learning-based intrusion detection technique for cyber-physical systems." Mathematics 12.20 (2024): 3194.

[24] Samantaray, M., Barik, R. C., & Biswal, A. K. (2024). A comparative assessment of machine learning algorithms in the IoT-based network intrusion detection systems. Decision Analytics Journal, 11, 100478.
[25] Saheed, Yakub Kayode, Oluwadamilare Harazeem Abdulganiyu, and Taha Ait Tchakoucht. "Modified genetic algorithm and fine-tuned long short-term memory network for intrusion detection in the internet of things networks with edge capabilities." Applied Soft Computing 155 (2024): 111434.