# A Novel Flow-Based Network Intrusion Detection Model Based On the Inverse Potts Model

**BHANU PRAKASH DUBEY, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand, India 248002,**

**Abstract**

One of the main security issues in the modern cyber environment is intrusion detection. A sizable number of methods that are based on machine learning strategies have been created. Consequently, we have developed machine learning techniques for spotting the infiltration. Using the method, we can both detect intrusion and pinpoint the attacker's specifics. Host-based and network-based IDS are the two primary categories. A Host based Intrusion Detection System (HIDS) keeps track of each host or device and notifies the user when it notices any unusual activity, such as changing or removing a system file, making an unauthorised series of system calls, or changing the configuration. To look for intrusions in the network traffic, a Network based Intrusion Detection System (NIDS) is often installed at network points like routers and gateways. One of the main security issues in the modern cyber environment is intrusion detection. A sizable number of methods that are based on machine learning strategies have been created. CICIDS17 dataset from dataset repository was used in this system. The system is then constructed with a machine learning classifier to divide the data into attack and benign categories. The outcomes of the research will demonstrate certain capabilities, including accuracy, precision, and mistake rate.

**DOI Number: 10.48047/nq.2019.17.03.2018** **NeuroQuantology 2019; 17(03):182-187**



## 1. INTRODUCTION

The number of hacking incidences is increasing day by day as technological innovation is developed. The number of hacking incidents reported by businesses every year is high. 2007 saw the commencement of a Distributed Denial of Service (DDoS) attack targeting websites hosted at Estonia. On June 17, 2008, numerous users in one of Amazon's locations began submitting authenticated requests. The servers slowed down when the queries started to rise considerably. The Dropbox service was severely disrupted for more than 15 hours in January 2013 due to a DDoS assault, according to a report from European Network and Information Security Agency. This disruption affected all users worldwide. On September 28, 2014, Facebook was allegedly the target of a widespread denial of service assault.It has been reported that 50% of cyberattacks begin with some kind of network scanning activity. Hackers are threatening the sensitive data of users kept on workstations by conducting flooding and probing assaults and disseminating malware files as viruses, worms, and spams to take advantage of flaws contained in current software as well. According to Cisco Annual Security Report, on April 17, 2013, 40% of all spam messages sent worldwide were connected to Boston Marathon bombing. According to a 2017 Cisco report, Trojan was among the top five malware tools used to first access consumer computers and business networks. Ensuring security in such a sophisticated technical setting is difficult and requires careful consideration. Different types of assaults have been taken into account by researchers for intrusion detection. Attacks are divided into nine categories according to a recent attack dataset (UNSW-NB): Fuzzer, Analysis, Reconnaissance, ShellCode, Worm, Generic, DoS, Exploit, and Generic. Section III has a thorough discussion of each of these attacks. Middle-box security solutions including firewalls, antivirus software, Intrusion Detection Systems are currently used. Firewall regulates network traffic depending on the source or

destination address. In accordance with the firewall regulations, it modifies the traffic. The quantity of state that is available to firewalls and their understanding of the hosts receiving the material are additional limitations. An IDS is a sort of security instrument that monitors system for suspicious activity and analyses network traffic before alerting the system or network administrator. To look for intrusions in the network traffic, a Network based Intrusion Detection System is often deployed at network points like gateways and routers. These IDS employ one of three kinds of advanced detection mechanisms: hybrid detection, anomaly detection, and abuse detection. IDS contains a set of knowledge base to identify recognised types of attack as part of the abuse detection strategy. Strategies for detecting misuse may be roughly divided into two categories: knowledge-based approaches and machine learning-based approaches. In knowledge-based approach, predetermined rules or attack patterns are compared to network traffic or host audit data like system call traces. Three categories can be used to group knowledge-based approaches: Signature matching, state transition analysis, and rule-based expert systems are the first three. Our project's major goal is to successfully classify or anticipate network intrusion assaults, integrate various machine learning techniques for better results, improve the overall performance of classification algorithms and to manage and identify intrusion attacks in networks.

## 2. LITERATURE SURVEY

This project's major goal is to successfully classify or anticipate network intrusion assaults, integrate various machine learning techniques for better results, improve the overall performance of classification algorithms, and regulate and detect intrusion attacks in networks. With the advent of programmable features, network security issues may now be efficiently detected and monitored thanks to Software Defined Networking (SDN) technology. In order to safeguard computer networks and address network security challenges, machine learning techniques have recently been integrated into SDN-based Network Intrusion Detection Systems (NIDS). Deep learning technology, a stream of sophisticated machine learning techniques, starts to take shape in the setting of SDN. Researchers looked at a number of current studies on machine learning techniques that utilise SDN for constructing NIDS in the study. They specifically evaluated the deep learning methodologies used to create SDN-based NIDS. The technologies that may be used to develop NIDS models in an SDN context were examined in this survey in the interim. A discussion of current issues with implementing NIDS using ML/DL and upcoming projects concludes this survey. The complexity and high implementation costs of different feature learning techniques are their key drawbacks [1].

Users and organisations have been at danger from cyberattacks since the Internet's inception. They have developed in complexity with computer networks. To get to their target nowadays, attackers must take many intrusive stages. Multi-stage assault, multi-step attack, or attack scenario are all terms used to describe this sequence of actions. Since understanding the attack plan and identifying the danger requires the correlation of several actions, their multi-step structure makes intrusion detection difficult. The security research community has worked for developing methods for identify these types of danger and anticipate next moves ever since the year 2000. The goal of this survey is to compile all the literature outlining multi-step assault detection techniques. Researchers concentrate on techniques that attempt to disclose the entire structure of the assault and the connections between its phases rather than simply detecting a symptom. To find the pertinent material, we conduct bibliographic research in a methodical manner. They describe and categorise a corpus of 181 papers encompassing 119 methodologies as a consequence of our work. Researchers may draw some inferences about the state of study in multi-step assault detection from the examination of the papers. The benefit of this system is that it detects malicious network events using IDS signatures and tracks their progression as a series of events, matching them based on IP address or port. The collection of potential action sequences for a multi-step attack can be highly complicated since an attacker does not necessarily have to carry out each step in the exact same order [2].

The volume of real-time network data has recently increased dramatically due to the constantly expanding use of linked Internet of Things devices. At the same time, network risks are unavoidable, making it essential to spot abnormalities in real-time network data. Utilizing current techniques like K-means, isolation forest, hierarchical density-based spatial clustering of application with noise, agglomerative clustering, and spectral clustering, the assessment is carried out by doing critical comparison analysis. The evaluation's findings clearly demonstrated the effectiveness of the suggested framework, which had a far better accuracy rate of 96.51% in comparison with other





approaches. Additionally, suggested framework fared better than current techniques in terms of less memory usage and faster implementation. Analysts will ultimately be able to accurately track and find abnormalities in real time thanks to the suggested method. Large-scale machine learning algorithms may provide very efficient results thanks to the spark iterative computation architecture, and the spark.ml API for pipeline provides programmers with a wide selection of additional modules to infuse with their design. In this situation, algorithms have their own advantages with regards to memory use, anomaly detection, data processing, execution time and accuracy [3].

Anomaly detection is crucial to identify and stop security assaults because network traffic anomalies might signify a potential network incursion. Early studies in this field and Intrusion Detection Systems (IDS) that are sold commercially are mostly signature-based. The issue with signature-based methods is that they are unsuitable for real-time network anomaly detection since the database signatures must be updated when new attack signatures become available. Anomaly detection has recently become more popular, and machine learning classification methods provide the foundation. To the Kyoto 2006+ data set, researchers use seven different machine learning approaches with information entropy calculations and assess their effectiveness. The results demonstrated that the majority of machine learning approaches offer greater than 90% precision, recall, and accuracy for this particular data set. Radial Basis Function (RBF), nonetheless, outperforms the other six algorithms examined in this paper using the area under the Receiver Operating Curve (ROC) measure. This signature-based method's fundamental drawback is that the database signature must be updated when new signatures become available, making it unsuitable for real-time network anomaly detection. It might be challenging to compare the outcomes of the seven algorithms covered in this article using different performance criteria [4].

The Industrial Control System (ICS), a vital piece of the critical infrastructure, is subject to an increase in cyberthreats. This threat was amplified by the Shodan search engine's introduction. The Shodan search engine has grown into a favoured toolbox for penetration testers and attackers because it can recognise and index industrial control equipment linked to the Internet. In this study, researchers employ honeypot technology to fully explore the Shodan search engine. First, they set up six dispersed honeypot systems and gathered three months' worth of traffic information. To detect Shodan scans built on traffic characteristic and the function code, researchers create a hierarchical DFA-SVM identification model. This model is then modified to determine that Shodan and Shodan-like scanners outperform widely used approach of reverse resolving IP addresses. The influence of Shodan on industrial control systems is assessed with regards to scanning duration, frequency & port, region preferences, ICS protocol function code proportion and ICS protocol preferences. Finally, researchers undertake an extensive study for Shodan scans. As a result, they provide certain protective steps to lessen the Shodan threat. The key benefit of SVM is a machine learning model that, due to its high rate of detection for small samples and good generalisation capabilities, can handle high-dimensional and non-linear Shodan traffic from a limited number of Shodan scanners [5].

## 3. PROPOSED SYSTEM

Since NIDS are the primary focus of the current system, this article also analyses free &open-source network sniffer software in addition to NIDS implementation instruments and sample sets that are currently in use. Then, it reviews, evaluates, and contrasts cutting-edge NIDS solutions in the IoT environment with regards to architecture, detection methodology, addressed risks, validation tactics and algorithm implementations. The paper explores future prospects and works with conventional as well as machine learning NIDS approaches. We concentrate on IoT NIDS implemented by machine learning in this survey since learning algorithms have an excellent track record for performance in security and privacy. The survey, in contrast to other leading studies that focus on conventional systems, offers a thorough analysis of NIDSs using various parts of learning techniques for the Internet of Things.The study, in our opinion, will be helpful for academic and industrial research in three ways: first, in identifying IoT dangers and problems; second, in implementing NIDS; and third, in proposing novel smart approaches in IoT-context while taking IoT limits into consideration.





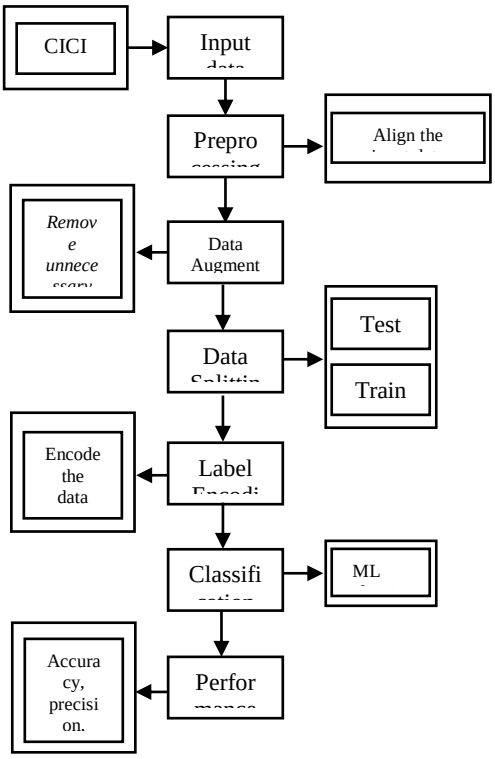


**Fig 1: System Architecture**

Additionally, the study will help security professionals distinguish between standard and IoT NIDS. When compared to what was suggested, the outcomes are poor. It is ineffective at handling vast amounts of data and theoretical constraints are also there.The CICIDS17 dataset was used as the system's input. The source data were obtained from the dataset repository. After that, the data preparation process must be put into action. To avoid inaccurate prediction at this level, we must deal with the missing data and encode the label of the source data. After that, test and train groups for the dataset must be created. Data is separated based on ratios. Most of the data will be available in train set. During the test, just a portion of the data will be available. During training phase, the model is evaluated and projections are produced during testing phase. After that, the vectorization should be employed. It implies that text must be encoded as numbers or numeric values in order to create feature vectors. Machine learning classifiers are used to categorise the various network intrusion attacks. The results of the experiment show the significance of performance metrics including recall, accuracy and precision. The suggested system has a number of advantages, including:

- It works well for a big number of datasets.
- When compared to the current system, the experimental outcome is excellent.
- Give precise and reliable forecast outcomes.
- It prevents inconsistent data and provides low time commitment.

**Fig 2: Flow Diagram**

## 4. RESULTS

Our project's primary goal is to accurately categorise or anticipate network intrusion assaults, integrate various machine learning techniques for improved performance, improve the overall effectiveness of classification algorithms, and manage and identify intrusion attacks in networks. The CICIDS17 dataset was collected from a collection of datasets for this purpose. The system is then constructed with a machine learning classifier to divide the data into attack and benign categories. Following screenshots from the experimental results exhibit several performances, including accuracy, precision, and error rate.

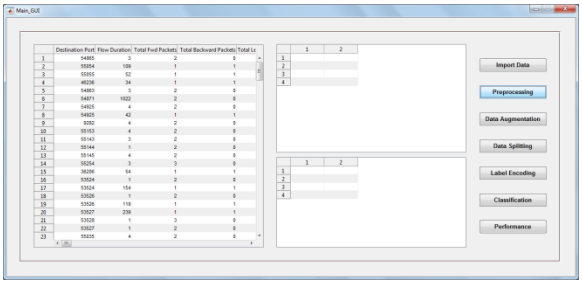


**Fig 3: Data Pre-Processing**

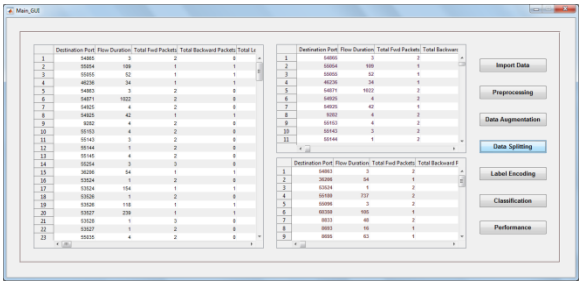


**Fig 4: Data Splitting**

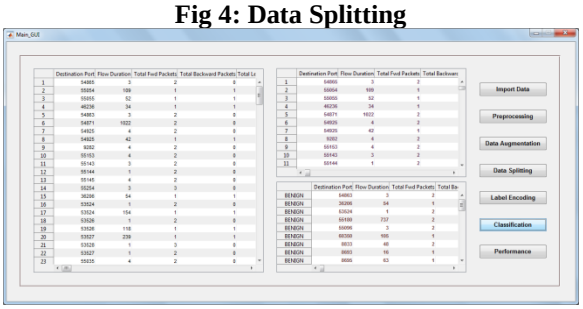


**Fig 5: Classification**

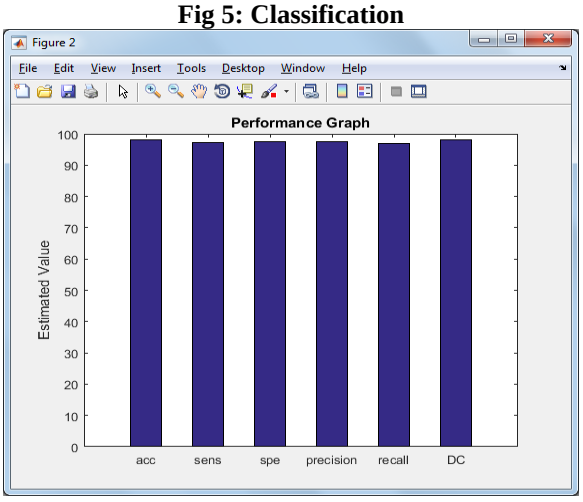


**Fig 6: Performance Analysis**

## 5. CONCLUSION

We come to the conclusion that the CICIDS17 dataset served as the input. In our study article, we stated the input dataset. We have put the machine learning and classification algorithms into practise. Finally, the outcome demonstrates that various performance parameters, including f1 score, accuracy, precision, and recall. Next, we must categorise the various intrusion assaults that have occurred in the network.

## 6. FURTHER IMPROVEMENT

In the future, we'd like to merge two deep learning algorithms or two distinct machine learning approaches. Future extensions or modifications might be made to the suggested clustering and classification methods to achieve even better results. In addition to the tried-and-true combination of data mining approaches, other clustering algorithm combos can be used to improve the detection accuracy. To increase system efficiency, a preventive system might be incorporated to the sentiment analysis detection system.

## REFERENCE

[1] Nasrin Sultana & Naveen Chilamkurti& Wei Peng&RabeiAlhadad,"Survey on SDN based network intrusion detection system using machine learning approaches" - 2017

[2] Julio Navarro, AlineDeruyver, Pierre Parrend "A systematic survey on multi-step attack detection"- 2018

[3] Jeysudha, A., Muthukutty, L., Krishnan, A., & Shivadekar, S. (2017). Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications, 162(9), 42-45.

[4] MarziaZaman, Chung-Horng Lung, "Evaluation of Machine Learning Techniques for Network Intrusion Detection"- 2018

[5]Sayed, A., Sardeshmukh, M., & Limkar, S. (2014). Improved Iris Recognition Using Eigen Values for Feature Extraction for Off Gaze Images. In ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol II: Hosted by CSI Vishakapatnam Chapter (pp. 181-189). Springer International Publishing.

[6] H. Sedjelmaci and M. Feham, "Novel Hybrid Intrusion Detection System for Clustered Wireless Sensor Network," Int. J. Netw. Secur. It's Appl. (IJNSA), Vol.3, No.4, July 2011, vol. 3, no. 4, pp. 1–14, 2011.

[7] L. Khan, M. Awad, and B. Thuraisingham, "A new intrusion detection system using support vector machines and hierarchical clustering," VLDB J., vol. 16, no. 4, pp. 507–521, 2007.

[8] S. K. Sahu, S. Sarangi, and S. K. Jena, "A detail analysis on intrusion detection datasets," Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014, pp. 1348–1353, 2014.

[9] O. Can, C. Turguner, and O. K. Sahingoz, "A Neural Network Based Intrusion Detection System For Wireless Sensor Networks," Signal Process. Commun. Appl. Conf. (SIU), 2015 23th, pp. 2302–2305, 2015.

[10] F. Lu and L. Wang, "Intrusion Detection System Based on Integration of Neural Network for Wireless Sensor Network," J. Softw. Eng. 2014.

[11] Y. Y. Li and L. E. Parker, "Intruder detection using a wireless sensor network with an intelligent mobile robot response," Southeastcon, 2008. IEEE, pp. 37–42, 2008.

[12] A. Kulakov and D. Davcev, "Tracking of unusual events in wireless sensor networks based on artificial neural-networks algorithms," Inf. Technol. Coding Comput. 2005. ITCC 2005. Int. Conf., pp. 534–539, 2005.
[13] M. Panda, "Security Threats at Each Layer of Wireless Sensor Networks," Int. J. Adv. Res. Comput.Sci. Softw. Eng., vol. 3, no. 11, pp. 61–67, 2013.
[14] Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," Proc. First IEEE Int. Work. Sens. Netw. Protoc. Appl. 2003., pp. 113–127, 2003.
[15] H. Hindy, D. Brosset, E. Bayne, A. Seeam, C. Tachtatzis, R. Atkinson, and X. Bellekens, "A Taxonomy and Survey of Intrusion Detection System Design Techniques, Network Threats and Datasets," vol. 1,no. 1, 2018.