International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 744-746 www.ijaem.net ISSN: 2395-5252

Examining the Impact of Data Sampling on Methods for Network Intrusion Detection Based on Machine Learning

Bhavya N Javagal[1], Shama Khanum[2], Satyam Kumar[3], Supritha V[4], Tembar Singh[5]

[1] *Asst. Professor, Department of Computer Science & Engineering, T John Institute of Technology, Bengaluru, India.*

[2345] *Students, Department of Computer Science & Engineering, T John Institute of Technology, Bengaluru, India.*

Date of Submission: 01-03-2023 Date of Acceptance: 10-03-2023

ABSTRACT - The flow features used by machine learning - based network intrusion detection systems (NIDSs) are derived from flow exporting protocols (i.e., NetFlow). The present achivement of Deep Learning and ML - based NIDS systems assumes the collection of such flow information, such as average packet size, from each packet in the flow. The evaluation strategy we suggest can be modified to accommodate various flow export step configurations.. As a result, It can still offer a reliable assessment of NIDS in the presence of sampling. According to the experiment findings, SketFlow is better than non-linear samplers. Furthermore, we discovered that SketchFlow sampling and random forest classification worked better together.

Key Words: Deep Learning, CNN, Intrusion Detection, Machine Learning, and Sampling of Network Data.

I. INTRODUCTION

Application network surveillance includes flow analysis, intrusion detection, and performance tracking. Considering the Impact of Traffic Filtering on. There is a machine learning-based network intrusion detection These methods are becoming more and more common as network traffic volume and speed continue to rise because of their efficacy, flow-based network monitoring is preferred over deep packet inspection. Network intrusion detection systems look at flow records to determine if a flow is safe or not [2]. A corpus has been demonstrated in new works of machine-learning DL NIDS [6]–[13]. Strong detection rates have been achieved with these solutions, which is promising (DRs).To our knowledge, however, the majority of inventive methods rely on records that is compiled from a sample of packets rather than which is determined from all data. It is not certain if cutting-edge techniques will work in practical uses. Taking into account real-world situations we examine the impact of sampling on NIDS (i.e., when sampling is unavoidable). distribution. Therefore, compared to colorizations created using previous techniques, those created using this technique are more vivid and perceptually realistic.

1.1 Objectives

The rapid development of technology gave us knowledge. A network connection is needed for personal use. There are certain problems that arise as technology advances. Therefore, a solution is required to stop those attacks from happening. Both internal and exterior intrusions are possible. The networking infrastructure of a business experiences an internal intrusion. They have access to the networking infrastructure. It might be an acquaintance, partner, coworker, or even a disgruntled client. Outside of the network system, there is an external attack, also known as an online assault.

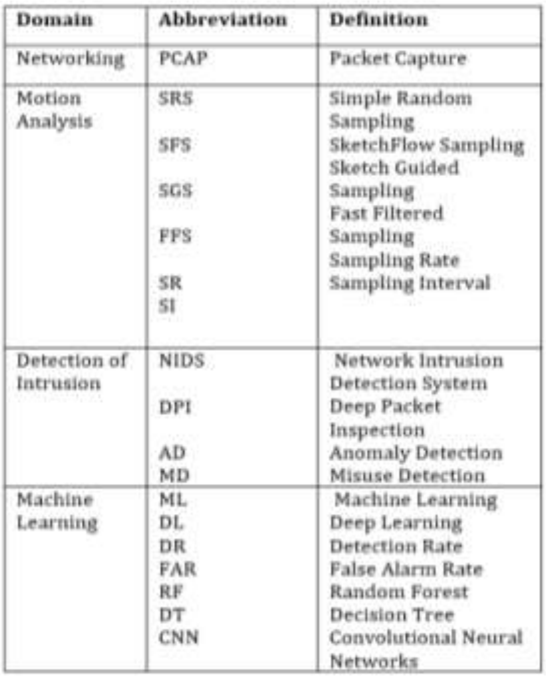
1.2 Problem Statement

The volume and speed of internet data have grown, making even record flow export for common devices difficult (i.e., switches and routers). This is due to the fact that the tracking device needs a certain amount of bandwidth, memory, and CPU time to decode each packet. As a result packet sampling reduces the overhead of the flow information tracking tool.

II. RELATED WORKS

Here, we review the pertinent literature, compare it to our results, and discuss the novelty and contributions of our work. Our research is located at the nexus of two closely linked subdomains:
(1) investigations into NIDS in relation to sampling.
(2) Comparing and discussing the freshness of our study in 2 different ways.

Table -1: Acronyms list



Domain	Abbreviation	Definition
Networking	PCAP	Packet Capture
Motion Analysis	SRS	Simple Random Sampling
	SFS	SketchFlow Sampling
	SGS	Sketch Guided Sampling
	FFS	Fast Filtered Sampling
	SR	Sampling Rate
	SI	Sampling Interval
Detection of Intrusion	NIDS	Network Intrusion Detection System
	DPI	Deep Packet Inspection
	AD	Anomaly Detection
	MD	Misuse Detection
Machine Learning	ML	Machine Learning
	DL	Deep Learning
	DR	Detection Rate
	FAR	False Alarm Rate
	RF	Random Forest
	DT	Decision Tree
	CNN	Convolutional Neural Networks

III. IMPLEMENTATION AND WORKING

3.1 Architectural Diagram

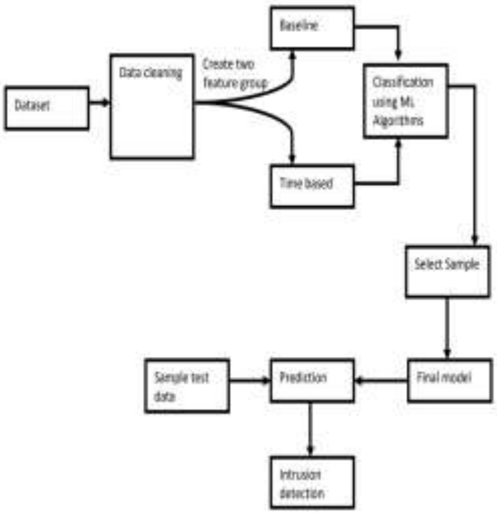
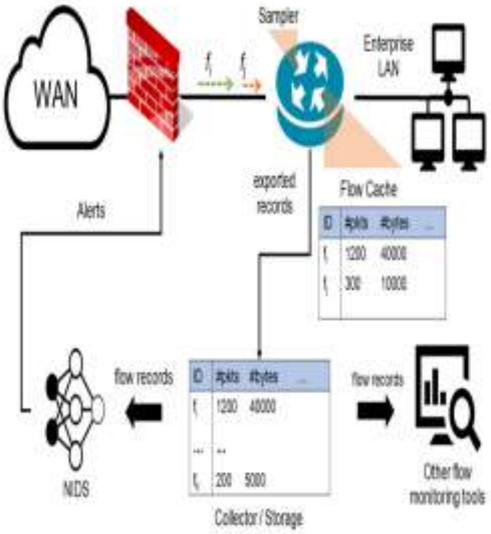


Fig -1: Diagram of the flow tracking system



3.2 SYSTEM FOR DETECTING NETWORK INTRUSION

The accuracy of the flow is verified by the NIDS. Our study's primary objective was a flow-level NIDS that classifies each record as either benign or falling under a particular malicious category. Through the observation some NIDS can identify the no.of packets passing location using signature-based deterministic rules, they are not covered in this paper. In our study, flow data that was collected using CIC Flow Meter was used to train and assess the ML-based NIDS. Flow records were taken from sampled packets for sampling studies. As a consequence, the kind and amount of sampling have a direct bearing on NIDS's effectiveness.

3.3 Flow Record

The CIC Flow Meter collects data, when one of the following circumstances happens, the system produces information, like the mean size of packet and the whole no.of bytes as a log. timeout active packet should be exported for quick analysis if it has been running for some time but has only gathered a portion of the records. Inactivity pause: The flow should be exported for quick study after it has been inactive for a set period of time. When the flow cache gets too tiny, the most recent flow data are removed.

IV. FUTURE SCOPE

Future research should focus on how sampling impacts anomaly-based NIDS. The study only took into account three sampling rates. During our flow storage tests, we could only try four different cache sizes. As a result, in-depth research on a variety of subjects may be done in the future. One method is to examine the effects of sampling using high sampling rates. Discussing different sampling techniques and ML/DL methodologies is a unique strategy. Additionally, carrying out such a study on a wide variety of diverse datasets provides more convincing justifications for how sampling affects NIDS performance.

V. CONCLUSIONS

We offer a method that is suitable for assessing ML-based NIDS. As far as we know, We are the first to propose a method that uses sampling. We were able to demonstrate that the proposed evaluation framework resulted in a notable training-data imbalance that needs to be addressed, half of the malicious flows are not exported. In the presence of sampling, our analysis of the viability of the most sophisticated ML-based NIDS reveals that sampling usually decreases the effectiveness of NIDS. However, sampling can increase efficiency the cached flow of the measuring is constrained. This research was the first to examine how sampling affected ML-based NIDS. To keep the breadth manageable, we focused only on NIDS that used misuse detection.

REFERENCES

[1] A. Sperotto, ‘‘Flow-based intrusion detection,’’ in Proc. 12th IFIP/IEEE Int. Symp. Integr. Netw. Manage. (IM) Workshops, May 2011.

[2] Aitken. P and Trammell. B, and, Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information, document RFC 7011, Internet Requests for Comments, RFC Editor, Sep. 2013.

[3] Y. Bi and Sher. M, ‘‘Flow-based intrusion detection: Techniques and challenges,’’ Computer Security,Sep. 2017

[4] B. Claise, Cisco Systems NetFlow Services Export Version 9, document RFC 3954, Internet Requests for Comments, RFC Editor, Oct. 2004.

[5] S. Venkatraman, M. Alazab, K. Soman, P. Poornachandran, Vinayakumar, A. Al-Nemrat, Deep learning method for intelligent intrusion detection system, IEEE Access, 2019.

[6] M. A. Ferrag and R. Smith, "Deep learning methods for cyber security intrusion detection: A detailed study," Proc. Electron. Workshops Computer, Sep. 2019.