Global Journal of Computer Science and Technology: E
Network, Web & Security
Volume 23 Issue 3 Version 1.0 Year 2023
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Mobile Adhoc Network Risk Profiles - An Overview of Existing Network Traffic Datasets to Determine Ideal Axiom Criteria

By Jedidiah Aqui & Michael Hosein

University of the West Indies St. Augustine

Abstract- A Mobile Adhoc networks also known as MANET or Wireless Adhoc Network is a network that usually has aroutable networking environment on top of a Link Layer ad hoc network. It consist of a set of mobile nodes connected wirelessly in a self-configured, self-healing network without having a fixed infrastructure. Recent studies and fieldwork have pointed in the direction of making MANETS a publicly viable option in the event of another world event/crisis such as the recent COVID-19 pandemic. As opposed to their traditional military and emergency uses, this has become a focal point due to the evident strain that was observed on mainstream Internet Service Providers as substantial adjustments had to be made to facilitate a new 'working-from-home' public. A primary aspect that must be considered before public adoption is addressing the issue of MANET risk and Security which leads into identifying and classifying risks associated with MANETS.

Index Terms: *MANET, risk profile, dataset, IDS, network, traffic.*

GJCST-E Classification: *ACM: C.2.1, C.2.3, C.4*

MOBILEADHOCNETWORKRISKPROFILESANOVERVIEWOFEXISTINGNETWORKTRAFFICDATASETSTODETERMINEIDEALAXIOMCRITERIA

Strictly as per the compliance and regulations of:

Mobile Adhoc Network Risk Profiles - An Overview of Existing Network Traffic Datasets to Determine Ideal Axiom Criteria

Jedidiah Aqui [α] & Michael Hosein [σ]

Abstract- A Mobile Adhoc networks also known as MANET or Wireless Adhoc Network is a network that usually has aroutable networking environment on top of a Link Layer ad hoc network. It consist of a set of mobile nodes connected wirelessly in a self-configured, self-healing network without having a fixed infrastructure. Recent studies and fieldwork have pointed in the direction of making MANETS a publicly viable option in the event of another world event/crisis such as the recent COVID-19 pandemic. As opposed to their traditional military and emergency uses, this has become a focal point due to the evident strain that was observed on mainstream Internet Service Providers as substantial adjustments had to be made to facilitate a new 'working-from-home' public. A primary aspect that must be considered before public adoption is addressing the issue of MANET risk and Security which leads into identifying and classifying risks associated with MANETS. This paper seeks to analyze the various existing fields and meta-data within various networking datasets, protocols as well as scenarios and subsequently establish what aspects of existing network traffic can be classified into axioms (Risk Classifying arguments) to determine Risk Profiles of MANETS. The paper also seeks to determine and propose the ideal data fields within Network traffic for classifying Risk Profiles.

Index Terms: MANET, risk profile, dataset, IDS, network, traffic.

I. Introduction

Research on the usage of wireless protocols and networks such as Bluetooth, NFC and MANETS in a public capacity has recently undergone a resurgence due to Global events such as the COVID-19 pandemic. And whilst protocols such as NFC and bluetooth has been explored in varying settings such as mentioned in, [1], [2] and further research was done in light of the Global Pandemic as per [3] and [4]. There was an evident need for greater public usage and adoption of these protocols to test the reliability and uses of them in light of the traditional reliance on mainstream Internet Service providers. To this end, advances in multiplexing connectivity for the Bluetooth protocol were made as per the work conducted in [5], [6] and [7] to allow for more simultaneous connectivity amongst mobile nodes in a network. However, the challenge of having a reliable wide-area infrastructure less network remained a challenge. Consequently, the prospects of utilizing MANETS in a public setting was explored.

As mentioned in the Abstract a MANET is a network that usually has a routable networking environment on top of a link layer ad hoc network. It consists of a set of mobile nodes connected wirelessly in a self-configured, self-healing network without having a fixed infrastructure. However, prior to delving into any discourse on public adoption, it is imperative to underscore a critical focal point: the risks and security considerations inherent to Mobile Ad-Hoc Networks (MANETs).

Based on [8] and [9] both qualitative and quantitative research has alluded to the fact that there is an evident disparity in probability-based Risk determination not only within MANETS but generally in Networking on a whole. An evident trend in Risk and Security analysis within MANETS has also shown that most Intrusion and Anomaly detection and prevention systems undertake a reactive approach to network security events which can be attributed to the dominance of 'impact-based' studies and techniques developed to address MANET and Network security.

This paper serves as an extended and in-depth analysis, aiming to substantiate the concept of Risk Profile generation introduced in [10]. Through meticulous examination, the study identifies specific domains within Network Traffic that can be readily categorized into axioms, laying the foundational groundwork for constructing an initial Risk Profile for Mobile Ad-Hoc Networks (MANETs). The research also assesses the optimal fields suitable for establishing axioms crucial to the generation of a risk profile. This analysis is integral to complementing both the passive and active phases proposed in [10] for a comprehensive solution and/or framework.

II. Literature Review

The following dataset was used in [11], [12] and [13], the work of these papers focused on developing a reference model to address the constraint of limiting user data usage in a generalized manner due to a

Author α σ: Department of Computing and Information Technology University of the West Indies St. Augustine, Trinidad and Tobago.
e-mail: aqui2_jed@yahoo.com

Global Journal of Computer Science and Technology (E) XXIII Issue III Version I Year 2023

subsection of ‘feature-rich’, ‘bandwidth-heavy’ over the top (OTT) applications. This paper focused on personalizing service degradation policies by providing guidelines for users’ OTT consumption behavior classification based on Incremental Learning (IL).

In essence, the research focused on creating a tailored framework designed to pinpoint users influencing network service quality through Over-The-Top (OTT) applications. This approach steered clear of a generalized strategy that would restrict data to all users, instead honing in on specific individuals. Notably, the dataset employed, denoted as Uninauca 141 applications, encompassed a diverse array of fields, a subset of which is exemplified in Figures 1 and 2 respectively.

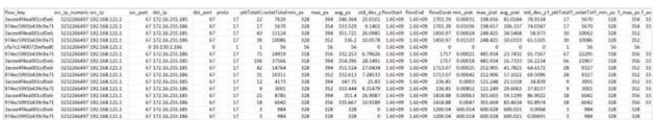


Fig. 1: Features and Fields Captured in the Uninauca Dataset

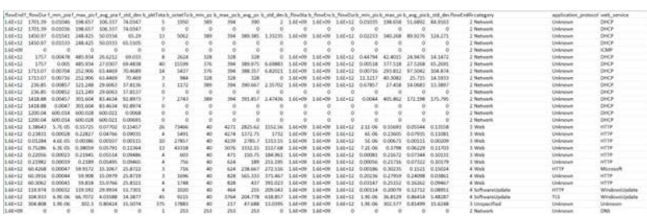


Fig. 2: Features and Fields captured in the Uninauca dataset continued

One of the immediate observations was that this network traffic focused heavily on layer 7 (application layer) information as the usage of applications by nodes connected were being monitored hence in the ‘category’ column for all Network packet captures there was a reading of ‘unkown’ for the application layer.

This dataset contains 50 features in which each instance holds the information of an IP flow generated by a network device i.e., source and destination IP addresses, ports, flow durations, interarrival times, packet sizes and layer 7 protocol (application) used on that flow as the class. For this dataset Axiom1 (Device Type) may not be easily achievable as the data suggests that of all protocols used ARP was not captured/utilized as it is a layer 3 protocol and thus, identifying MAC addresses for resolution would not be possible via this means. The only method would be to deduce the device type by patterns observed in network traffic, or type of requests made and what protocols were used.

Axiom2 (malicious node/repeat offender) is more achievable as one of the methods to identify a malicious node is to observe the identified anomalous nodes based on cached IPs or observe patterns for suspicious node behaviour, this can be derived from understanding (Source IP, Destination IP, Protocols being used).

The research conducted in [14] and [15], an in depth analysis was conducted on existing bodies of datasets to determine the accuracy of their usage in contemporary Intrusion Detection and Intrusion Prevention systems. What was found was that the 11 datasets used since the year 1998 was grossly outdated and unreliable which therefore lead to inaccurate deployment, analysis and evaluation of IDS’s and IPS’s. Additionally, it was found that some of the datasets such as ‘DARPA98’, ‘KDD99’, ‘ISC2012’, ‘ADFA13’ suffered from lack of traffic diversity and volumes, there were disparities in terms of the types of attacks the datasets covered.

Thus, the authors produced reliable datasets which contained benign and seven common attack network flows that meet real world criteria. All with the aim to evaluate performance of a comprehensive set of network traffic features and ML algorithms to give an indication of the optimum set of features for detecting certain attack categories.

The datasets (CICIDS2017dataset) provided covered network traffic for 5 days of the week (Monday through Friday) and identified different types of Network attacks such as: ‘DDos’, ‘PortScan’, ‘Infilteration’, ‘WebAttacks’, ‘Brute Force SSH’. This was a guided approach based on the attack listing provided by McAffee.

Apart from the well-known network traffic meta-data, there were several other noted meta-data fields, most notably the datasets were categorized by Network Behavioural patterns observed in traffic and subsequently labeled based on the perceived type of attack the network experienced. This label was also observed to form the basis of the “Label” field within each of the Network traffic Datasets as shown in the below figures 3, 4, 5 and 6 respectively:

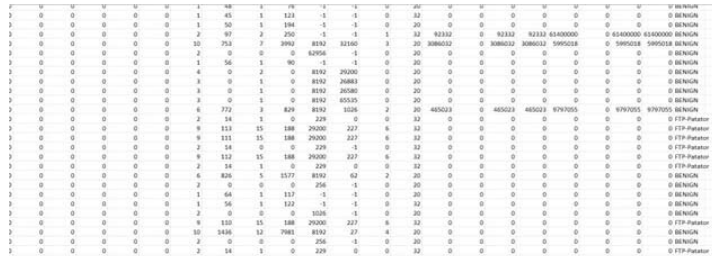


Fig. 3: Brute Force Attack

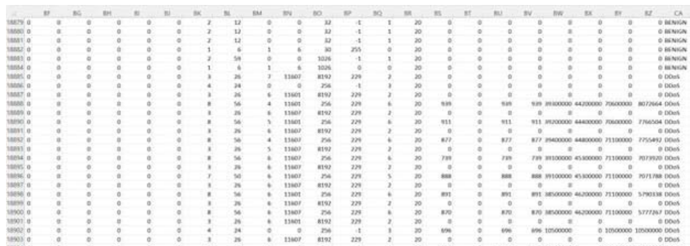


Fig. 4: Distributed Denial of Service Attack

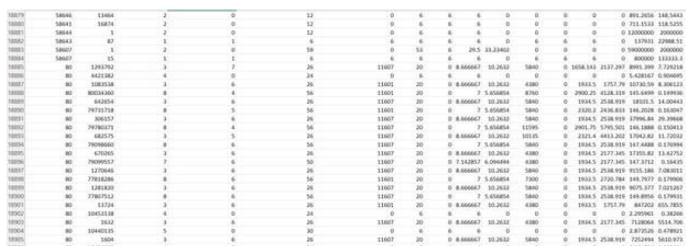


Fig. 5: Distributed Denial of Service Attack cont'd

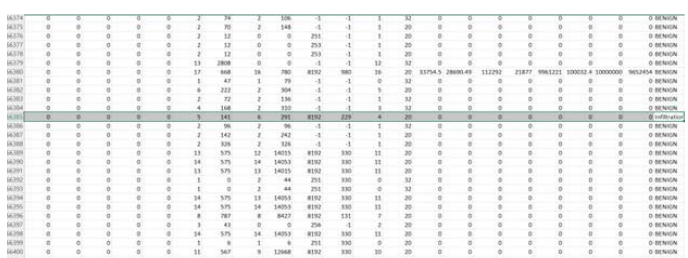


Fig. 6: Showing an Infiltration Attack

One of the most integral points to note on this paper was the method of labelling and classification utilized to further fine tune the dataset for more accurate and relevant Intrusion Detection and Prevention analysis. The most critical meta data field can be recognized by the 'Label' field as it distinguished between a benign Network activity and a specific attack. Similarly, a strain of this methodology is desired for the determination of Risk Profile for a MANET based on the identified axioms.

Another Dataset that was examined from the work conducted by [16] was captured from a Network Intrusion Detection System and captured fields such as 'Source Address Bytes', 'Destination Address bytes', 'Ip Address', 'Port Number', 'Fragmentation Bit', 'Mac ID', 'Protocol Type', 'DNS', 'TLS – transport security layer', just to name a few, However one of the most critical Meta Data fields captured in this dataset was observed to be the 'Mac ID' field. As it pertains to the current direction of the proposed solution for establishing Risk Profiles, One of the most basic Axiom defined for classifying the risk level of the MANET identifies device types. The MAC Id can be observed as an iterative step towards determining device type once it has been sourced and the device determined. This therefore, leads to a much more accurate determination of devices as opposed to observing network node behaviours which are more reliant on experience and humanistic determinations. The below figures 7 and 8 show examples of Network Meta Data fields that were captured:

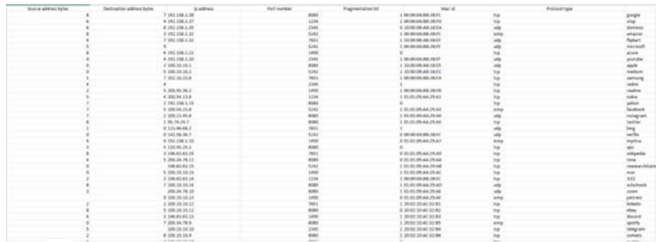
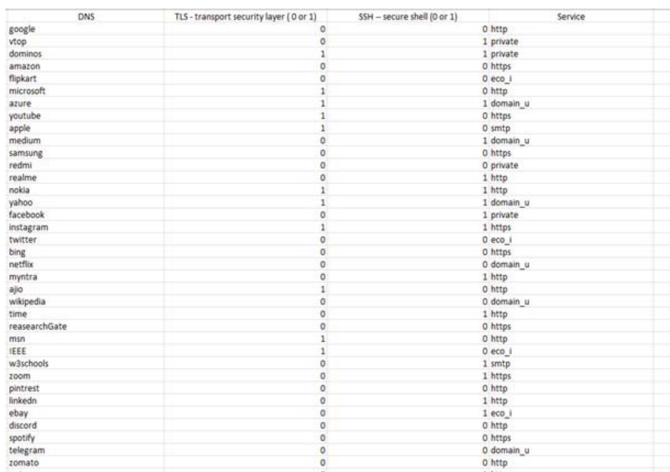


Fig. 7: Showing Packet Data Captured from an Intrusion Detection System

Another dataset (UNSWNB15) that was analyzed was derived from several papers [17], [18], [19] and [20]. These papers are namely "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)", "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 dataset and the comparison with the KDD99 dataset", "Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks", "Big data analytics for intrusion detection system: statistical decision- making using finite dirichlet mixture models."



DNS	TLS - transport security layer (0 or 1)	SSH – secure shell (0 or 1)	Service
google	0	0	http
vtop	0	1	private
dominos	1	1	private
amazon	0	0	https
flipkart	0	0	eco_i
microsoft	1	0	http
azure	1	1	domain_u
youtube	1	0	https
apple	1	0	smtp
medium	0	1	domain_u
samsung	0	0	https
redmi	0	0	private
realme	0	1	http
nokia	1	1	http
yahoo	1	1	domain_u
facebook	0	1	private
instagram	1	1	https
twitter	0	0	eco_i
bing	0	0	https
netflix	0	0	domain_u
myntra	0	1	http
ajio	1	0	http
wikipedia	0	0	domain_u
time	0	1	http
reasearchGate	0	0	https
msn	1	0	http
IEEE	1	0	eco_i
w3schools	0	1	smtp
zoom	0	1	https
pintrest	0	0	http
linkedn	0	1	http
ebay	0	1	eco_i
discord	0	0	http
spotify	0	0	https
telegram	0	0	domain_u
zomato	0	0	http

Fig. 8: Showing Packet Data Captured from an Intrusion Detection System Cont'd

The Datasets utilized in these papers can be viewed as a combination of a 100Gb network traffic (PCAP files) and the generated datasets created by the IXIA PerfectStorm tool in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS). The work done primarily focused on improving the resilience of intrusion detection within Networks specifically for the protocols: "Domain Name System (DNS)", "Hyper Text Transfer Protocol (HTTP)" and "Message Queue Telemetry Transport (MQTT)". A main focal point for these protocols was the extensive usage of same by internet of Things (IOT) devices within a network and the likelihood of cyber threats against them and the services they utilize. The Dataset contains nine distinct type of attacks namely:

- Fuzzers
- Analysis
- Backdoors
- Denial of Service
- Exploits
- Generic
- Reconnaissance
- Shellcode
- Worms

Features of the Dataset with class labels were developed by 12 algorithms. The below diagram depicts all 49 of the generated features of the Dataset:

In Addition to improving the resilience of IDS' the research conducted was also aimed at reducing the amount of 'false positives' generated by IDS in response to zero-day vulnerabilities and other type of Network threats. In [20]: Statistical Decision-Making Using Finite Dirichlet Mixture Models" focus was placed on developing a scalable framework for building an effective and lightweight anomaly detection system. The framework consisted of three (3) modules:

- Capturing and Logging – Responsible for sniffing and collecting network data.
- Pre-processing – Responsible for analyzing and filtering data to improve performance of the decision engine.
- Decision Engine – Designed based on the Dirichlet mixture model with lower upper interquartile range as decision engine.

The framework's performance was based on the two main datasets the NSL-KDD and UNSQ-NB15 with the aim of determining which technique provided a higher detection rate and lower false alarm rate than other predominant techniques.

The below are snippets of the datasets utilized as well as the 'attack' classification schema:

No.	Name	Type	Description
1	srcip	nominal	Source IP address
2	sport	integer	Source port number
3	dstip	nominal	Destination IP address
4	dsport	integer	Destination port number
5	proto	nominal	Transaction protocol
6	state	nominal	Indicates to the state and its dependent protocol, e.g. ACC, CLO, CON, ECO, ECR, FIN, INT, MAS, PAR, REQ, RST, TST, TXD, URH, URN, and (-) (if not used state)
7	dur	Float	Record total duration
8	sbytes	Integer	Source to destination transaction bytes
9	dbytes	Integer	Destination to source transaction bytes
10	sttl	Integer	Source to destination time to live value
11	dttl	Integer	Destination to source time to live value
12	sloss	Integer	Source packets retransmitted or dropped
13	dloss	Integer	Destination packets retransmitted or dropped
14	service	nominal	http, ftp, smtp, ssh, dns, ftp-data ,irc and (-) if not much used service
15	Sload	Float	Source bits per second
16	Dload	Float	Destination bits per second
17	Spkts	integer	Source to destination packet count
18	Dpkts	integer	Destination to source packet count
19	swin	integer	Source TCP window advertisement value
20	dwin	integer	Destination TCP window advertisement value
21	stcpb	integer	Source TCP base sequence number
22	dtcpb	integer	Destination TCP base sequence number
23	smeansz	integer	Mean of the ?ow packet size transmitted by the src
24	dmeansz	integer	Mean of the ?ow packet size transmitted by the dst
25	trans_depth	integer	Represents the pipelined depth into the connection of http request/response transaction
26	res_bdy_len	integer	Actual uncompressed content size of the data transferred from the server's http service.
27	Sjit	Float	Source jitter (mSec)
28	Djit	Float	Destination jitter (mSec)
29	Stime	Timestamp	record start time
30	Ltime	Timestamp	record last time
31	Sintpkt	Float	Source interpacket arrival time (mSec)
32	Dintpkt	Float	Destination interpacket arrival time (mSec)

Fig. 9: Showing features of the Dataset USNW-NB15 dataset

33	32 Dintpkt	Float	Destination interpacket arrival time (mSec)
34	33 tcprtt	Float	TCP connection setup round-trip time, the sum of 'synack' and 'ackdat'.
35	34 synack	Float	TCP connection setup time, the time between the SYN and the SYN_ACK packets.
36	35 ackdat	Float	TCP connection setup time, the time between the SYN_ACK and the ACK packets.
37	36 is_sm_ips_ports	Binary	If source (1) and destination (3)IP addresses equal and port numbers (2)(4) equal then, this variable takes value 1 else 0
38	37 ct_state_ttl	Integer	No. for each state (6) according to specific range of values for source/destination time to live (10) (11).
39	38 ct_flw_http_mthd	Integer	No. of flows that has methods such as Get and Post in http service.
40	39 is_ftp_login	Binary	If the ftp session is accessed by user and password then 1 else 0.
41	40 ct_ftp_cmd	integer	No of flows that has a command in ftp session.
42	41 ct_srv_src	integer	No. of connections that contain the same service (14) and source address (1) in 100 connections according to the last time (26).
43	42 ct_srv_dst	integer	No. of connections that contain the same service (14) and destination address (3) in 100 connections according to the last time (26).
44	43 ct_dst_ltm	integer	No. of connections of the same destination address (3) in 100 connections according to the last time (26).
45	44 ct_src_ltm	integer	No. of connections of the same source address (1) in 100 connections according to the last time (26).
46	45 ct_src_dport_ltm	integer	No of connections of the same source address (1) and the destination port (4) in 100 connections according to the last time (26).
47	46 ct_dst_sport_ltm	integer	No of connections of the same destination address (3) and the source port (2) in 100 connections according to the last time (26).
48	47 ct_dst_src_ltm	integer	No of connections of the same source (1) and the destination (3) address in in 100 connections according to the last time (26).
49	48 attack_cat	nominal	The name of each attack category. In this data set , nine categories e.g. Fuzzers, Analysis, Backdoors, DoS Exploits, Generic, Reconnaissance, Shellcode and Worms
50	49 Label	binary	0 for normal and 1 for attack records

Fig. 10: Showing features of the Dataset USNW-NB15 dataset cont'd

59.166.0.0	1390	149.171.126.6	53	udp	CON	0.001055	132	164	31	29	0	0	dns	500473.9	621800.9	2	2	0	0	0
59.166.0.0	33661	149.171.126.9	1024	udp	CON	0.036133	528	304	31	29	0	0	-	87676.09	50480.17	4	4	0	0	0
59.166.0.6	1464	149.171.126.7	53	udp	CON	0.001119	146	178	31	29	0	0	dns	521894.5	636282.4	2	2	0	0	0
59.166.0.5	3593	149.171.126.5	53	udp	CON	0.001209	132	164	31	29	0	0	dns	436724.6	542597.2	2	2	0	0	0
59.166.0.3	49664	149.171.126.0	53	udp	CON	0.001169	146	178	31	29	0	0	dns	499572.3	609067.6	2	2	0	0	0
59.166.0.0	32119	149.171.126.9	111	udp	CON	0.078339	568	312	31	29	0	0	-	43503.23	23896.14	4	4	0	0	0
59.166.0.6	2142	149.171.126.4	53	udp	CON	0.001134	132	164	31	29	0	0	dns	465608.5	578483.3	2	2	0	0	0
10.40.182.3	0	10.40.182.3	0	arp	INT	0	46	0	0	0	0	0	-	0	0	1	0	0	0	0
59.166.0.5	40726	149.171.126.6	53	udp	CON	0.001126	146	178	31	29	0	0	dns	518650.1	632326.8	2	2	0	0	0
59.166.0.7	12660	149.171.126.4	53	udp	CON	0.001167	132	164	31	29	0	0	dns	452442.2	562125.1	2	2	0	0	0
10.40.170.2	0	10.40.170.2	0	arp	INT	0	46	0	0	0	0	0	-	0	0	1	0	0	0	0
10.40.170.2	0	10.40.170.2	0	arp	INT	0	46	0	0	0	0	0	-	0	0	1	0	0	0	0
10.40.182.3	0	10.40.182.3	0	arp	INT	0	46	0	0	0	0	0	-	0	0	1	0	0	0	0
59.166.0.1	48847	149.171.126.5	53	udp	CON	0.001093	132	164	31	29	0	0	dns	483074.1	600182.9	2	2	0	0	0
59.166.0.1	24266	149.171.126.9	41049	udp	CON	0.001851	528	304	31	29	0	0	-	1711507	985413.3	4	4	0	0	0
59.166.0.1	10393	149.171.126.6	44307	udp	CON	0.001749	528	304	31	29	0	0	-	1811321	1042882	4	4	0	0	0
59.166.0.2	62539	149.171.126.4	53	udp	CON	0.001128	132	164	31	29	0	0	dns	468085.1	581560.3	2	2	0	0	0
59.166.0.1	21270	149.171.126.9	111	udp	CON	0.005153	568	312	31	29	0	0	-	661362.3	363283.5	4	4	0	0	0
59.166.0.1	8989	149.171.126.6	111	udp	CON	0.004898	568	312	31	29	0	0	-	695794.2	382196.8	4	4	0	0	0
59.166.0.4	49346	149.171.126.5	53	udp	CON	0.001111	132	164	31	29	0	0	dns	475247.5	590459.1	2	2	0	0	0
175.45.176.3	21223	149.171.126.18	32780	udp	INT	0.000021	728	0	254	0	0	0	-	1.39E+08	0	2	0	0	0	0
175.45.176.2	23357	149.171.126.16	80	tcp	FIN	0.240139	918	25552	62	252	2	10	http	28050.42	815794.2	12	24	255	255	1.71E+09
175.45.176.0	13284	149.171.126.16	80	tcp	FIN	2.39039	1362	268	254	252	6	1	http	4233.619	749.6685	14	6	255	255	3.9E+09
59.166.0.3	4192	149.171.126.9	53	udp	CON	0.001101	132	164	31	29	0	0	dns	479564	595822	2	2	0	0	0
59.166.0.2	26872	149.171.126.8	53	udp	CON	0.001082	132	164	31	29	0	0	dns	487985.2	606284.7	2	2	0	0	0
59.166.0.8	24946	149.171.126.2	53	udp	CON	0.001122	132	164	31	29	0	0	dns	470588.2	584670.2	2	2	0	0	0
59.166.0.9	5685	149.171.126.6	53	udp	CON	0.001141	146	178	31	29	0	0	dns	511831.8	624014.1	2	2	0	0	0
59.166.0.0	22848	149.171.126.1	53	udp	CON	0.001164	146	178	31	29	0	0	dns	501718.2	611683.8	2	2	0	0	0

Fig. 11: Showing the Packet Data Captured in the USNW-NB15 Dataset

0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.017	0.013	0	0	0	0	0	0	0	0	3	7	1	1
0	132	76	0	0	9.89101	10.68271	1.42E+09	1.42E+09	7.005	7.564333	0	0	0	0	0	0	0	0	2	4	2	1
0	73	89	0	0	0	0	1.42E+09	1.42E+09	0.017	0.013	0	0	0	0	0	0	0	0	12	8	1	2
0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.043	0.014	0	0	0	0	0	0	0	0	6	9	1	1
0	73	89	0	0	0	0	1.42E+09	1.42E+09	0.005	0.003	0	0	0	0	0	0	0	0	7	9	1	1
0	142	78	0	0	29.68222	34.37034	1.42E+09	1.42E+09	21.003	24.315	0	0	0	0	0	0	0	0	2	4	2	1
0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.017	0.013	0	0	0	0	0	0	0	0	12	7	1	2
0	46	0	0	0	0	0	1.42E+09	1.42E+09	0	0	0	0	0	1	2	0	0	0	2	2	2	2
0	73	89	0	0	0	0	1.42E+09	1.42E+09	0.018	0.013	0	0	0	0	0	0	0	0	6	7	1	1
0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.018	0.013	0	0	0	0	0	0	0	0	6	7	2	1
0	46	0	0	0	0	0	1.42E+09	1.42E+09	0	0	0	0	0	1	2	0	0	0	2	2	2	2
0	46	0	0	0	0	0	1.42E+09	1.42E+09	0	0	0	0	0	1	2	0	0	0	2	2	2	2
0	46	0	0	0	0	0	1.42E+09	1.42E+09	0	0	0	0	0	1	2	0	0	0	2	2	2	2
0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.018	0.013	0	0	0	0	0	0	0	0	6	9	2	5
0	132	76	0	0	0.656903	0.328339	1.42E+09	1.42E+09	0.471	0.237667	0	0	0	0	0	0	0	0	8	4	2	5
0	132	76	0	0	0.640403	0.280968	1.42E+09	1.42E+09	0.458333	0.203667	0	0	0	0	0	0	0	0	8	2	3	5
0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.017	0.015	0	0	0	0	0	0	0	0	7	7	2	1
0	142	78	0	0	1.890104	1.610554	1.42E+09	1.42E+09	1.348	1.149333	0	0	0	0	0	0	0	0	8	4	2	5
0	142	78	0	0	1.780739	1.549507	1.42E+09	1.42E+09	1.269667	1.103667	0	0	0	0	0	0	0	0	8	2	3	5
0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.018	0.013	0	0	0	0	0	0	0	0	5	9	2	1
0	364	0	0	0	0	0	1.42E+09	1.42E+09	0.021	0	0	0	0	0	2	0	0	0	1	1	1	1
1.94E+09	77	1065	1	12026	1170.482	1144.383	1.42E+09	1.42E+09	21.83082	9.570304	0.051475	0.006528	0.044947	0	1	1	0	0	3	2	2	1
2.47E+09	97	45	1	0	18786.71	941.7249	1.42E+09	1.42E+09	183.5793	474.2594	0.066088	0.017959	0.048129	0	1	1	0	0	5	2	2	1
0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.017	0.012	0	0	0	0	0	0	0	0	7	4	3	1
0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.011	0.009	0	0	0	0	0	0	0	0	7	6	1	1
0	66	82	0	0	0	0	1.42E+09	1.42E+09	0.018	0.014	0	0	0	0	0	0	0	0	8	3	1	2
0	73	89	0	0	0	0	1.42E+09	1.42E+09	0.017	0.015	0	0	0	0	0	0	0	0	5	7	1	1
0	73	89	0	0	0	0	1.42E+09	1.42E+09	0.018	0.013	0	0	0	0	0	0	0	0	3	8	2	1

Fig. 12: Showing the Packet Data Captured in the USNW-NB15 Dataset Cont'd

AS	AT	AU	AV	AW
1	1	2		0
3	2	2		0
3	1	1		0
3	2	2		0
1	1	2		0
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1	Exploits	1
2	1	1	DoS	1
1	1	1	Exploits	1
2	1	1	Exploits	1
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1		0
2	1	1		0
1	1	2		0
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1		0
2	1	1		0
1	1	2		0
1	1	1	Exploits	1
2	1	1		0

Fig. 13: Showing the Packet Data Captured in the USNW-NB15 Dataset Cont'd

The following Dataset snippets captured the various types of attack events as well as their respective number of occurrences and associated protocols/attack sub-category for the recorded period:

Attack category	Attack subcategory	Number of events
normal		2218761
Fuzzers	FTP	558
Fuzzers	HTTP	1497
Fuzzers	RIP	3550
Fuzzers	SMB	5245
Fuzzers	Syslog	1851
Fuzzers	PPTP	1583
Fuzzers	FTP	248
Fuzzers	DCERPC	164
Fuzzers	OSPF	993
Fuzzers	TFTP	193
Fuzzers	DCERPC	455
Fuzzers	OSPF	1746
Fuzzers	BGP	6163
Reconnaissance	Telnet	6
Reconnaissance	SNMP	69
Reconnaissance	SunRPC Portmapper (TCP) UDP Service	2030
Reconnaissance	SunRPC Portmapper (TCP) TCP Service	2026
Reconnaissance	SunRPC Portmapper (UDP) UDP Service	2045
Reconnaissance	NetBIOS	5
Reconnaissance	DNS	35
Reconnaissance	HTTP	1867
Reconnaissance	SunRPC Portmapper (UDP)	2028
Reconnaissance	ICMP	1739
Reconnaissance	SCTP	367
Reconnaissance	MSSQL	5
Reconnaissance	SMTP	6

Fig. 14: Showing the Attack Events of the USNW-NB15 dataset

Axiom to ascertain which fields aligned more accurately to the axiom descriptions.

After completion of substantial qualitative research and analysis. A determination of ideal fields for Risk Profiles were established based on current network traffic data. This was done to establish an idea of the accuracy of a generated Risk profile with existing datafields in MANET traffic. Additionally, proposed meta-data and nominal data fields were introduced and would be covered in the 'Discussion' section. These proposed fields would seek to establish a more accurate Risk Profile calculation.

AS	AT	AU	AV	AW
1	1	2		0
3	2	2		0
3	1	1		0
3	2	2		0
1	1	2		0
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1	Exploits	1
2	1	1	DoS	1
1	1	1	Exploits	1
2	1	1	Exploits	1
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1		0
2	1	1		0
1	1	2		0
1	1	1		0
1	1	1		0
1	1	1		0
1	1	1		0
2	1	1		0
1	1	2		0
1	1	1	Exploits	1
2	1	1		0

Fig. 15: Showing the Attack Events of the USNW-NB15 dataset cont'd

Exploits	SCCP	3
Exploits	SIP	1043
Exploits	TFTP	87
Generic	All	7
Generic	SIP	436
Generic	HTTP	1
Generic	SMTP	247
Generic	IXIA	7395
Generic	TFTP	116
Generic	IXIA	207243
Generic	Superflow	10
Generic	HTTP	5
Generic	TFTP	21
Reconnaissance	DNS	6
Reconnaissance	SMTP	1
Reconnaissance	HTTP	314
Reconnaissance	SNMP	12
Reconnaissance	SunRPC Portmapper (UDP) TCP Service	349
Reconnaissance	MSSQL	1
Reconnaissance	NetBIOS	1
Reconnaissance	SCTP	2
Reconnaissance	SunRPC	2
Reconnaissance	Telnet	1
Reconnaissance	ICMP	26
Reconnaissance	SunRPC Portmapper (TCP) TCP Service	349
Reconnaissance	SunRPC Portmapper (TCP) UDP Service	349
Reconnaissance	SunRPC Portmapper (UDP) UDP Service	346
Shellcode	FreeBSD	8

Fig. 16: Showing the Attack Events of the USNW-NB15 dataset cont'd

III. Methodology

The methodology undertaken was an iterative one which stemmed from the previously mentioned paper [3] which pro-posed an approach for identifying risk levels within MANETS. Several datasets with diverse attributes and situations such as data from:

- Network Intrusion Detection Systems
- Network Intrusion Prevention Systems
- Application layer network traffic
- MANET traffic
- Network (peer to peer, multihop, traditional) traffic
- generated Network traffic datasets from training and
- modeling data

These were subsequently sourced. This was done to gain a current perspective of the available meta-data fields that are typically captured within network traffic. Based on the identified fields within the datasets, a comparative analysis was then conducted based on the general description of each.

IV. Discussion

The analysis conducted on the datasets led to the determination of the common fields captured within typical network traffic as well as the additional fields that were captured based on the type of traffic being observed. Some realizations that were observed are as follows:

- Datasets varied based on the nature of the traffic being captured.
- Different levels of granularity were observed across the numerous datasets. In terms of what were the typical network traffic fields being captured versus more nominal value fields that were identified by the packet tracers/network monitors.

The results of the assessment conducted on current network data captures revealed that some of the most common network traffic fields identified were:

- Source IP
- Destination IP
- Protocol
- Port
- Length
- Info
- number

Some of the other datafields that were observed from the network data captures were:

- MAC Id
- application protocol
- web service (i.e. http, private, ecoi, https)
- category
- label(distinguishing type of attack experienced)
- service (i.e. http, private, ecoi, https)
- DNS
- attack cat
- label (binary value 0 = normal, 1= attack records)

Based on a general description of Axioms, they form the basis for classifying risk levels within MANETS. Axiom 1 primarily pertained to the device types that are currently on a MANET, apart from observing node behaviours to gauge what type of device they may be, some helpful fields for Axiom 1 would be: *'Source IP', 'Destination IP', 'Protocol', 'MAC Id', 'application protocol', 'label', 'attack cat', 'DNS'*

Axiom 2 would have generally pertained to whether a node is a repeat offender or not and thus, the data fields that would be most useful for determining Axiom2 would be: *'Source IP', 'Destination IP', 'Protocol', 'application protocol', 'label', 'attack cat', 'DNS',* 'category' However, these fields consist of what currently exists in typical Network traffic or IDS traffic data schemas. The addition of the following fields would improve the accuracy of the determined risk level of the given MANET as it would act as additional classification criteria to determine a malicious node, similar to the machine learning classification techniques used in [21] and [22]:

- Axiom 1 - a Binary value of (0= positive, 1= negative)
- Axiom 2 - a Binary value of (0= positive, 1= negative)
- Risk Score - Ranging from 1-5 (1= being very, 2= good, 3= fair, 4= warning, 5= critical)

V. Future Work

The prospective work outlined in this paper involves implementing the Risk Profiles methodology on datasets that align with the current spectrum of Network Traffic fields being recorded. The outcome of this implementation will unveil the present Risk Profile of a designated MANET/Network. A comparative analysis will then be conducted, juxtaposing the existing dataset schema against the proposed fields outlined in the Discussion. This comparative assessment aims to illuminate the accuracy levels in dealing with the limited data fields currently available, as opposed to the introduction of Axioms for refining the precise determination of Risk Profiles.

VI. Conclusion

In conclusion, this paper's research has revealed diverse levels of granularity in the captured data fields of Network/MANET traffic. These nuances in granularity were discerned based on the origin of the network traffic, encompassing MANETs, Mobile Networks, IDS, and IPS. Although many exhibited the conventional dataset fields, noteworthy insights emerged from integrating the previously identified fields within network traffic that readily align with classifiable Axioms. Furthermore, the study established that for an accurate assessment of the current state versus a proposed configuration concerning risk level determination, the classification methods must be applied to the existing datasets. This application involves testing against new datasets that incorporate the Axioms, thereby refining the determination of risk levels specific to MANETs.

References Références Referencias

1. M. Hosein and L. Bigram. "An educational bluetooth quizzing application in Android." International Journal of Wireless and Mobile Networks 5.6 (2013): 69.
2. M. Hosein. "Using Wireless Technology for Quick Distribution of Wireless and Mobile Course Notes and Other Resources." GSTF Journal on Computing Oct2015, Vol4 Issue3, p60-70.

3. D. Granger, and M. Hosein. "WI-FI DIRECT AS A MOBILE STUDENT QUIZZING PLATFORM–A Case Study." Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2018.
4. M Hosein and K Ramdass. "Wi-fi Direct Applications within a Post Covid Classroom–Bridging the Gap between Fully Online and Face to Face Learning." 2022 International Conference on Information Networking (ICOIN), 464-469.
5. J. Aqui and M. Hosein, "Investigating simultaneous wireless connections for a quiz management system-A case study," Global Journal of Computer Science and Technology, pp. 1–11, 2021.
6. J. Aqui and M. Hosein, "An approach to establishing simultaneous server-side connections for NFC/bluetooth enabled Quiz Management Systems," Global Journal of Computer Science and Technology, pp.1–11, 2021.
7. J. Aqui and M. Hosein, "A probabilistic determination of resilience of QMS's simultaneous server-side connections approaches/ methodologies," 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), 2021.
8. H. Michael and A. Jedidiah, "Mobile Adhoc Networks - An Overview of Risk Identification, Intrusion Detection and Machine Learning Techniques used," 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, Karnataka, India, 2022, pp. 1-5, doi: 10.1109/ICMNWC56-175.2022.10031757.
9. J. Aqui and M. Hosein, "Mobile Ad-hoc Networks Topic Modelling and Dataset Querying," 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICM-NWC), Tumkur, Karnataka, India, 2022, pp. 1-6, doi: 10.1109/ICM-NWC56175.2022.10031921.
10. A. Jedidiah and H. Michael, "Mobile Adhoc Networks - Establishing Initial Risk Profiles utilizing ML Techniques," 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, Karnataka, India, 2022, pp. 1-5, doi: 10.1109/ICMNWC561-75.2022.10031628.
11. J. S. Rojas, A. Pekar, A. Rendon, and J. C. Corrales, "Smart user consumption profiling: Incremental learning-based OTT service degradation," IEEE Access, vol. 8, pp. 207426–207442, 2020.
12. T. Sudtasan and H. Mitomo, "Effects of OTT services on consumer's willingness to pay for optical fiber broadband connection in Thailand, "in Proc. 27th Eur. Regional Conf. Int. Telecommun. Soc., 2016, pp. 1–11.
13. V. Carela-Espan˜ol, "Network traffic classification?: From theory to practice," Ph.D. dissertation, Dept. d' Arquitectura Computadors, Univ. Polite`cnica Catalunya, Barcelona, Spain, 2014.
14. I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," Proceedings of the 4th International Conference on Information Systems Security and Privacy, 2018.
15. Brown, C., Cowperthwaite, A., Hijazi, A., and Somayaji, A. (2009). Analysis of the 1999 darpa/lincoln laboratory ids evaluation data with netadhict. In 2009 IEEE SCISDA, pages 1–7.
16. S. Mishra, "Network intrusion detection system," Kaggle.com, 18-Nov-2022. [Online]. Available: https://www.kaggle.com/datasets/shalini31mishra/network-intrusion-detection-system. [Accessed: 07-Apr-2023].
17. N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia, 2015, pp. 1-6, doi: 10.1109/Mil-CIS. 2015.7348942.
18. N. Moustafa and J. Slay, "The evaluation of NETWORK ANOMALY DETECTION SYSTEMS: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," Information Security Journal: A Global Perspective, vol. 25, no. 1-3, pp. 18–31, 2016.
19. N. Moustafa, J. Slay and G. Creech, "Novel Geometric Area Analysis Technique for Anomaly Detection Using Trapezoidal Area Estimation on Large-Scale Networks," in IEEE Transactions on Big Data, vol. 5, no. 4, pp. 481-494, 1 Dec. 2019, doi: 10.1109/TBDATA.2017.2715166.
20. N. Moustafa, G. Creech, and J. Slay, "Big Data Analytics for Intrusion Detection System: Statistical decision-making using finite Dirichlet mixture models," Data Analytics and Decision Support for Cybersecurity, pp. 127–156, 2017.
21. A. Mitrokotsa and C. Dimitrakakis, "Intrusion detection in manet using classification algorithms: The effects of cost and model selection," Ad Hoc Networks, vol. 11, no. 1, pp. 226–237, 2013.
22. G. sah, S. Singh, and S. Banerjee, "Intrusion detection system using classification algorithms with feature selection mechanism over real-time data traffic," 2022.