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 a routable 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

Strictly as per the compliance and regulations of:

© 2023. Jedidiah Aqui & Michael Hosein. This research/review article is distributed under the terms of the Attribution-Non Commercial-NoDerivatives 4.0 International (CC BY-NCND 4.0). You must give appropriate credit to authors and reference this article if parts of the article are reproduced in any manner. Applicable licensing terms are at https://creativecommons.org/licenses/by-nc-nd/4.0/.
Mobile Adhoc Network Risk Profiles - An Overview of Existing Network Traffic Datasets to Determine Ideal Axiom Criteria

Jedidiah Aqui\textsuperscript{a} & Michael Hosein\textsuperscript{a}

\textbf{Abstract} - A Mobile Adhoc network 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. 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.

\textbf{Index Terms}: MANET, risk profile, dataset, IDS, network, traffic.

\section{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, \cite{1}, \cite{2} and further research was done in light of the Global Pandemic as per \cite{3} and \cite{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 \cite{5}, \cite{6} and \cite{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 \cite{8} and \cite{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 \cite{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 \cite{10} for a comprehensive solution and/or framework.

\section{Literature Review}

The following dataset was used in \cite{11}, \cite{12} and \cite{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...
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.

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 ‘unknow’ for the application layer.

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', 'Infiltration', '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:
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:
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.”

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 types 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:
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:

<table>
<thead>
<tr>
<th>Attack category</th>
<th>Attack subcategory</th>
<th>Number of events</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>FTP</td>
<td>2285/15</td>
</tr>
<tr>
<td>Attacks</td>
<td>HTTP</td>
<td>1597</td>
</tr>
<tr>
<td>Attacks</td>
<td>RIP</td>
<td>2051</td>
</tr>
<tr>
<td>Attacks</td>
<td>SYLIC</td>
<td>1911</td>
</tr>
<tr>
<td>Attacks</td>
<td>PPTF</td>
<td>1583</td>
</tr>
<tr>
<td>Attacks</td>
<td>TFTP</td>
<td>248</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>IMAP</td>
<td>184</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>OSPI</td>
<td>953</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>SFOP</td>
<td>1748</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>RUP</td>
<td>6153</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>Telnet</td>
<td>6</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>SMTP</td>
<td>99</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>SunRPC Portmapper (TCP) UDP Service</td>
<td>2080</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>SunRPC Portmapper (TCP) UDP Service</td>
<td>2086</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>SunRPC Portmapper (TCP) UDP Service</td>
<td>2084</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>SunRPC Portmapper (TCP) UDP Service</td>
<td>2085</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>SunRPC Portmapper (UDP) UDP Service</td>
<td>2086</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>SunRPC Portmapper (UDP) UDP Service</td>
<td>2087</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>ICMP</td>
<td>1439</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>SCTP</td>
<td>90</td>
</tr>
<tr>
<td>Reconnaissance</td>
<td>MSSQL</td>
<td>25</td>
</tr>
</tbody>
</table>

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.

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
- Length
- Info
- number

Some of the other data fields 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 Axiom 2 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


© 2023 Global Journals


