Abbas et al LGURJCSIT 2019

ISSN: 2521-0122 (Online)
ISSN: 2519-7991 (Print)

LGU Research Jounral for
Computer Sciences & IT

Vol. 3 Issue 2, April - June 2019

Use of SPAN for Identified Network Traffic via Machine Learning

[1]Akifa Abbas, [2]Danish Ali, [3]Sidra Noureen, [4]Rehmatullah
[1]Department of Information Technology, Lahore Garrison University
[2]Department of Computer Science, Ripha International University
[3]Department of Computer Science, Lahore Garrison University
[4]Department of Software Engineering, Lahore Garrison University

[1]akifaabbas@lgu.edu.pk, [2]danish.ali6291@gmail.com
[3]sidranoureen@lgu.edu.pk , [4]Rehmatullah@lgu.edu.pk

Abstract:

Few years back the number of wireless devices and their use in our daily life has been increased a lot. All devices cell phones, laptops, tablets, camera, TVs, home appliances have become a part of network now. As the network devices are growing and getting connected to each other the security risks are getting higher. All the companies and organizations are now establishing and implanting the public and private wireless networks. Organization have to pay heavy cost to implement and integrate all devices together on a network. As wireless networks are more vulnerable to threats and in security's a huge network all the devices should be identified whenever they enter or leave a network traffic pool the experimental work in this paper will elaborate the methods to identify the network traffic identification under encryption. This paper emphases on identification of devices based on layer 2 functionality by MAC (Media Access Code). Later on, the identification was improved using labeled or tagged traffic methods by use of SPAN (Switch port analyzer technique) technology or protocol with assistance of Virtual Local Area Network. Many Supervised learning methods were examined during experiment and were referenced on data collected by real time traffic. The network traffic of multiple deceives gradually passes through network so incremental learning method is implemented as classification for streaming data.

Keywords: Network Traffic Analysis, WIFI, 802.11, Online Classification, SPAN

1. Introduction:

From past years the wireless devices are used increasingly in daily life e. g smart-phones, laptops, TVs cameras, everything is part of a network. All the organizations are trying there hard to secure networks(public/private) that provide connectivity. The first protocol of wireless introduced by IEEE as 802.11 in 1997, by 2004 WEP (Wired Equivalent Privacy), WAP converted to WPA2, Counter Cipher Mode with Block Chaining Message Authentication Code Protocol (CCMP). Temporal Key Integrity Protocol (TKIP) .NO doubt security methods help a lot but wireless mediums are vulnerable and less secure due to broadcast nature. Using supervised learning it will be easier to get information about all surrounding devices in the network. Using network traffic, analyzing per frame, it can be detected that frame belongs to which device

2. Platform:

In order to generate data collection group of physical devices was created using access point, pc, smart phones, Smart TV. Netgear N300 router used as Access point to support internet & local area network connectivity & provided authentication to devices. Rasberry Pi provides the collections capability through a wireless AP [1] USB adapter (TL-WN722N 150Mbps) & performed the data storage. Post analysis was performed on a desktop PC.

3. Data:

All data was collected in 615 hours, including 453 different transmitters.45% of network traffic was generated by top 10 transmitters, & the topmost transmitter sent 13%. There was no single chosen target within topmost 10 because all were access points. Altogether, about 787 million labelled instances, 2 well-known targets, and 9 other client devices used as targets. All instances were labelled by mac addresses.4 million instances were originated. The traffic was segmented and recorded into hour blocks by use of DUMPCAP [2].

4. Literature Review:

The common work related to research has illustrated the network traffic patterns, physical communication and physical localization for behavior analysis. In the previous work most of the research trend has observed frame relay for connection and communication identification. Frequently, The TCP layer information was used. But, in encrypted traffic this information was not found.

Desmond proposes fingerprinting by means of timing examination of 802.11 test prob request [3]. Their approach is one of a kind and related in that it utilizes aloof data as all customers disperse test demands. This research gives out passive investigation to all frames through assessment of the decoded header partitions. Foremski proposes traffic perception at the ISP level to figure out which application delivered it [4]. He utilized a SVM for traffic classification and by far most fall inside the coded (encrypted) segment of the frame. Also, the" standard" SVM isn't reasonable for classification of substantial data set. Correspondingly, Xuetal , developed a platform utilizing OpenWRT and performed essential part examination on home network system traffic [1]. Their work concentrated fundamentally on protocols and port groups. The aggregate arrangement of highlights included starting and end timestamps, source IP,DIP, source port, Destination port, protocol, packets and bytes as highlights. Which is all usually scrambled in unguided communication. Moore et al. utilized Bayesian examination to group arrange traffic by application [5]. Their work is intriguing in that they were building a classifier that was rivaling physically classified information. They uncovered just the TCP-headers for classification. Strikingly enough, they likewise found that bundle headers are regularly not sufficient for application classification. This examination demonstrates that with a little measure of data, a moderately high level of confidence can be accomplished in the matter of regardless of whether singular edges have a place with a specific gadget.

5. Implementation:

Created framework was made out of an accumulation's platform, post-handling utilities, examination instruments and classification. Post gathering manage characterization and handling. Gathering fragmented approaching information into one-hour tosses. Gathered information was prepared, translated into the Attribute-Relation File Format (ARFF) [6]. Post handling was performed on the ARFF files to change the target(s), consolidate, and control the substance.

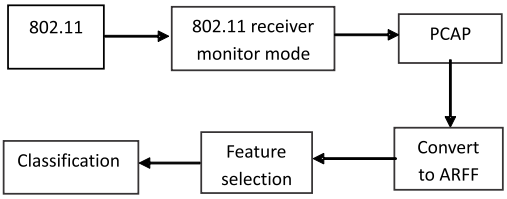

Figure 1: ARFF File Format

6. Classification:

Captured traffic will be exposed that reside within Radiotap header and MAC frame [7]. Driver would be applied and available for user space applications. Figure 1 shows the header of Radiotap, the field present in the header are configurable via multiple ways depends on the available platform. Standard Radiotap header fields can be seen in Figure 1.

Mac Frame format is demonstrated in Figure 2. All stations are liable to authenticate and interpret MAC frame's fields in order to assemble these frames by the transmitting stations. In the header of MAC frame the Address field becomes primary key to identify originating source machine.

Header Version Octet 1	Header Pad 1	HLen 2	Present Flags 4	MAC Timestamp 8	Flags 1	Data rate 1	Channel Frequency 2	Channel Type 2	SSI 1	Antenna 1	RX Flag 2

Figure 2: Radiotap Header Format

Frame Control Octet 2	Duration ID 2	Address 1 6	Address 2 6	Address 3 6	Sequence Control 2	Address 4 6	QoS control 2	HT Control 4	Frame Body 0-7951	FCS 4

Figure 3: MAC Frame

The frame control field is important in order to identify the originating source as well as nature of the frame.

Protocol Version B0-B1	Type B2-B3	Subtype B4-B7	To DS B8	From DS B9	More Fragments B10	Retry B11	Power Management B12	More Data B13	Protected Frame B14	Order B15

Figure 4: Frame Control Field Format

The reason of using MAC frame in tracking device is that it is totally unique globally as NIC [10]. Pertaining to above gathered information captured performed with a one radio operation in passive/monitoring mode. Furthermore, channel hoping did not perform to collect other frequencies i.e. single frequency will be tested. Channel 11 will be used due to its un-interruptive, wide and stable available free band. Collection based on multiple source/stations and will be stored in packet capture format. Packet capture format then translate into ARRF but with small amount of data to keep the operation smooth and bulk free. Also keep in mind that bulky data required where you received data from various sources and you have to find exact source AP via frame labeling. Conversion from PCAP to ARRF is fairly cost effective solution due to which some third party freeware are the alternatives to accomplish the statistical analysis and provision of user friendly output in charts/graphs. Weka is used in this scenario to capture the files. Massive Online Analysis is also provide you to online analysis facility, its open source framework for data streams mining [6].

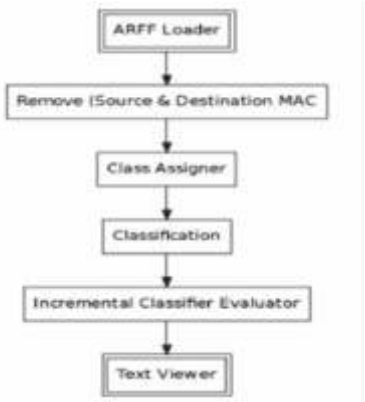

Figure 5: Knowledge Flow

Investigation explore different batch learning method. Feature selection and classification will be performed independently. These explorer have some limitation so for the rerunning test the Java code is required to accomplish this all. Despite of all that its worth mentioning here that explorer and java application only provide one hour analysis of the data so we have to move further for online data analysis that provide flexible and reliable solution.

On the basis of data set and available space, online classification is ideal [8]. Online space provide statistical analysis on intermittent basis with following six exceptions. The best method for grouping is the data nature. Despite all that no algorithm is ideal to test under different circumstances knowing all that multiple algorithms have tested to produce best output.

7. Results:

The primary test performed to identify the ability of targets. First algorithm performance measured in term of accuracy, precision and recall. A renowned target was nominated and up-sampling to yield the preciously data set.

Fig 5 and 6 show the accuracy, precision, and recall using casually increase experimented data from balanced sheet. The revealed results proved that small data set was adequate. In this case results was taken from well-known targets.

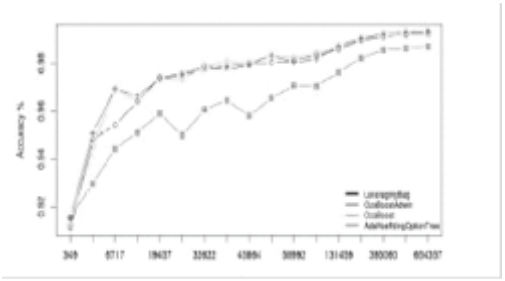

Figure 6: Accuracy

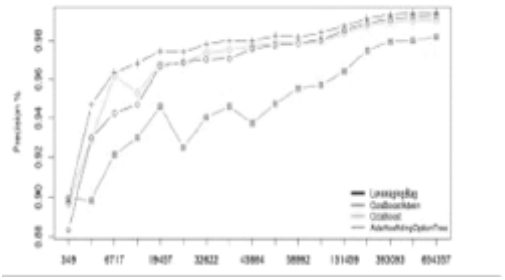

Figure 7: Precision

The proposed model failed to provide effective monitoring in multiple ways. Despite of the fact that multiple algorithms are used to manipulate the results but this model fail to provide upper level inspection based on various protocols. Transport layer as well as application level protocol line of defense mechanism vanish in this research. Proposed model only provided to identify the well know targets or end station via using wireless media broadcast nature but what if a user have valid credential who can bypass this frame identity and becomes active threat for upper layer ports/protocols. This model classify the traffic based on MAC addresses what if an object running like VRRP in which VIP have same MAC address for each connection. MAC spoofing facility available in every OS now a days can demolish all the research paper work.

TCP/UDP or ICMP traffic to inspect further to mitigate upper layer threats are not addressed in this research. Proposed model focus on only media layer protocol inspection while determining their geographical location. Being part of global village the enterprises have resource allocated according to their span rather than monitoring them globally and give intruder to chance and becomes single point of failure for the enterprise as proposed in the model.

This model only provide the target machine and login credential in case of suspected traffic so this model failed in virtualized environment i.e. the hot evolving technology in near future. This system unable to trace the end station in VAS/ VRAS environment. This model unable to give you OS level information in case of suspected target.

The forecasting mechanism implemented in this model just give you the passive stats and complex in nature. Effective forecasting mechanism like other researches give you more flexible and reliable solution which are more user friendly and easy to use. The random sampling on intermittent basis not supported in this research. Online classification itself has its own complication in it w r t security loop.

We are going to proposed solution to mitigate all shortcomings mentioned in previous section. Our wireless media traffic land on our centralized backbone switches through which providing redundancy at device level as well as media level (backup links). We will enable the Switch Port Analyzer known as SPAN which duplicates traffic from one or more CPUs, ports, Ether Channels and VLANs, network analyzer then analysis the destination of copied traffic such as a Switch Probe device or other Remote Monitoring (RMON, Wireshark) probe.

The switching of traffic on sources is not disturbed by Switch Port Analyzer. The destination for SPAN use should be defined. The copies generated by SPAN traffic strive with user traffic for switch resources.

A local SPAN is a connotation of source ports or VLANs with one or more destinations.

A local SPAN session is configured on a switch. Local SPAN does not facilitate separate source and destination sessions rather it can have ports or VLANs as sources. Both are not supported simultaneously Figure SPAN Working

It replicas traffic from one or more source ports in any VLAN or VLANs to a destination for analysis (see Fig.7). For example, as shown in Figure 7, all traffic of Ethernet port 5 (the source port) is copied to port 10. A network analyzer on port 10 receives all traffic from Ethernet port 5 which is not physically attached to port 5.

RSPAN facilitates the remote monitoring of more than one switches crossways. It supports source port, Vlans, destination on multiple switches. Figure 8 demonstrate that RSPAN practices a Layer 2 VLAN to carry SPAN traffic between switches.

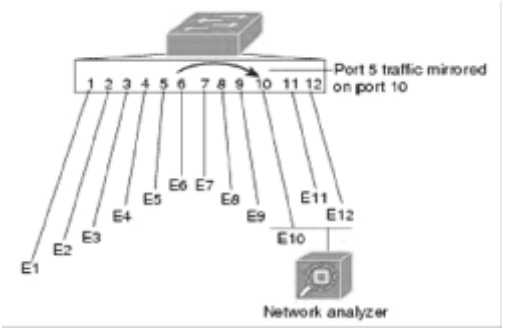

Figure 8: SPAN Working

The dedicated RSPAN session traffic is carried as Layer 2 non-routable traffic over a user-specified. Switch is connected via trunk-connection with other switch at Layer 2.

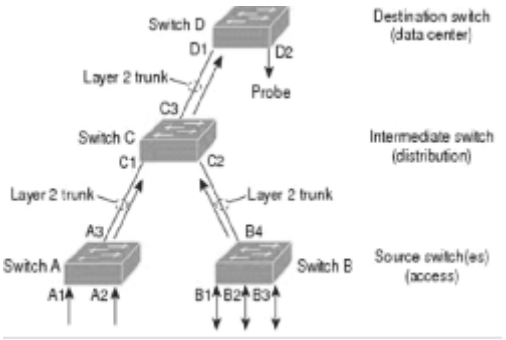

Figure 9: RSPAN

We will enable local SPAN for backbone switched while RSPAN for remote stations after that the captured traffic via SPAN and RSPAN will be monitored on Wireshark. A packet analyzer captures packets and display the maximum packet data details

A packet analyzer is used to monitor the traffic coming inside and going outside of network cable, just like a voltmeter is used by an electrician. Wireshark is perhaps one of the best open source packet analyzers available today, which can monitor the traffic covering all layers in TCP/IP stack.

Dedicated server will be managed to store and analysis of data. Strict policy to monitor the each session will be implemented and traps will be enabled to intimate about any suspected activity.

8. Conclusion:

The research has investigated capacity to differentiate a device through multiple supervised learning systems connected to scrambled remote information. Utilizing the communicated idea of the medium and inactive perception, it demonstrated that an objective can be related to a significant accuracy, review and ROC esteems utilizing an adequately substantial example informational index. The reference usage and its execution results give essential assets for further identification was done by Using SPAN.The span traafic was carried out in Layer-2. It has established the framework for distinguishing and perceiving various clients utilizing a similar gadget, examination of longer also, more various accumulations, and coordination with other frameworks and stages.

9. References:

[1] K. Xu, F. Wang, L. Gu, J. Gao, and Y. Jin, "Characterizing home network traffic: An inside view," in Wireless Algorithms, Systems, and Applications. Springer, 2012, pp. 60–71.

[2] A. Orebaugh, G. Ramirez, and J. Beale, Wireshark & Ethereal network protocol analyzer toolkit. Syngress, 2006.

[3] L. C. C. Desmond, C. C. Yuan, T. C. Pheng, and R. S. Lee, "Identifying unique devices through wireless fingerprinting," in Proceedings of the first ACM conference on Wireless network security. ACM, 2008, pp. 46–55.

[4] F. P., "Statistical, real-time classification of ip traffic in linux operating system," Master's thesis, Politechnika lska, 2011. [Online]. Available: http://www.iitis.pl/~pjf/pub/MasterThesis.pdf

[5] A. W. Moore and D. Zuev, "Internet traffic classification using bayesian analysis techniques," in ACM SIGMETRICS Performance Evaluation Review, vol. 33, no. 1. ACM, 2005, pp. 50–60.

[6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," SIGKDD Explorations, vol. 11, no. 1, 2009.

[7] "Radiotap header for 802.11 frame injection and reception," http://www.radiotap.org/.

[8] A. Bifet and R. Kirkby, "Data stream

[9] https://www.cisco.com/c/en/us/support/docs/switches/catalyst-6500-series-switches/10570-41.html

[10] Hindawi Security and Communication NetworksVolume 2017, Article ID 6235484, 21pages https://doi.org/10.1155/2017/6235484