MITM Detection Mechanism at client side using Machine Learning Algorithms

Mahesh Kumar Muddunuru
Lakehead University - Computer Science
email: mmuddunu@lakeheadu.ca

***Abstract*— Man-in-the-Middle attacks are talked about a lot in the technical field of computer security because they are one of the biggest concerns for professionals. MitM (Man in the Middle) attacks occur when an attacker intercepts data in the middle of a conversation by using a technique of interjecting themselves. Aside from attacking the data flow between the endpoints, attackers compromise the integrity and confidentiality of that data, as well. Through communication interception, an adversary can eavesdrop on confidential information and modify message integrity. Additionally, an adversary may intercept, modify, or destroy messages to end communication for one of the parties, thereby constituting a compromise of availability. In order to prevent the man in the middle attacks there are several methods, firewalls and intrusion detection systems, but unfortunately these methods are only applied to the administrative side of operations, i.e., Enterprise WIDS (Wireless Intrusion Detection System). The client system is not equipped to detect, prevent, and control these attacks, making it vulnerable to attack for the attacker. Our goal is to build a quality Intrusion Detection system using Anomaly Detection in Machine Learning that will detect network attacks on a client-side basis in order to prevent those attacks from happening.**

I. Introduction and Background

We are dependent on Internet and cellular networks today to conduct virtually every aspect of our lives. For example, we conduct home banking online, enjoy online entertainment and shop, make use of social networks, etc. These services contain confidential data such as the banking information, personal information and other sensitive information which becomes key target to the attacker. Aside from the individuals, cybercriminals also attack the organizations and enterprises resulting in financial losses. We sometimes read about cyber-attacks on connected things and online services on a daily basis thanks to the Internet which keeps people and things always connected. A popular attack is Man-in-the-Middle (MITM), which allows hackers to gain control of users' data as it passes through the internet.

A man-in-the-middle (MITM) attack occurs when the attacker eavesdrops on communications between two targets, then secretly relays and possibly alters messages between those parties believing they're directly communicating. Attackers might be able to intercept data and information from legitimate parties while also sending malicious links or other information via the Internet in a way that is unlikely to be detected until it is too late.

A. How MITM works

The MITM attack generally consists of two phases: interception and decryption.

1) Phase 1: Interception: The interception phase of cybercrime involves hackers gaining entry to a network from an open or poorly secured Wi-Fi router, and/or through the use of DNS servers modified to manipulate address directives (DNS). A hacker will then test the router to find potential vulnerabilities and ways of gaining access. Usually that is achieved via a weak password, regardless of the fact that cyber criminals are also capable of doing more advanced things like IP spoofing or cache poisoning.

As soon as a target is determined, an attacker usually deploys data-stealing equipment to obtain access to and gather transmitted info from the victim, redirect traffic or otherwise control the user's web experience.

2) Phase 2: Decryption: Decryption is the second phase of an MITM attack. This is when stolen data is decoded and made accessible to cybercriminals. Decrypted data is used for nefarious purposes, such as identity theft, unauthorized purchases, and other fraudulent activities. There are cases where man-in-the-middle attacks are conducted for no apparent reason other than to disrupt enterprise operations and create chaos.

The demonstration of MITM attack can be shown in the below figure 1 where the attacker acts as a legitimated access point as HOME and intercepts all the traffic between the victim and the access point.

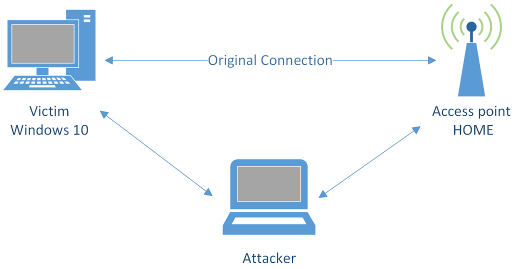


Fig. 1. MITM attack.

B. Types of MITM attacks

Depending on the target and the objective, there are a wide range of MITM techniques and outcomes. Below are some of the MITM discussed.

1) **DNS Spoofing:** Domain Name System (DNS) spoofing is a form of computer security hacking that involves the addition of false information to the DNS resolver cache and the use of altered DNS records to redirect online traffic to a fictitious website that looks like the destination website[1].
2) **IP Spoofing:** Cybercriminals use intellectual property spoofing to deceive users into thinking they are a part of a legitimate entity. The technique lets them gain access to computers within a target network, release sensitive information, launch DoS attacks, or compromise computer systems.
3) **HTTPS Spoofing:** An HTTPS spoof sends a false endorsement to the victim's browser once the request for a secured site is made. The endorsement contains a more advanced thumbprint about the bargained application that the browser confirms against an existing audit of the application's capabilities.
4) **ARP Spoofing:** Using this method, an attacker sends spoofed ARP messages over a network, which attempts to divert the traffic from its original host to the attacker's computer instead[2].
5) **Eavesdropping:** Eavesdropping can be active or passive and occurs through the interception of the traffic obtaining passwords, personal information and other sensitive information that are send over the traffic[3].
6) **Stealing Browser Cookies:** Cookies store information of login credentials, credit card information and other data that can be hijacked by the attacker using the MITM attack.

The damages and loss from these attacks can be vary depending on the motive of the attacker and the ability to deal with the data to cause mischief. With increase of the digital technology and use of Wireless communication everywhere in our daily life the consequences from these types of attacks are very serious. For example, "49 busted in Europe for Man-in-the-Middle bank attacks" where attackers were arrested on suspicion for using man-in-the-middle attacks to sniff out and intercept payment requests from email totaling fraud of €6 million[4]. It was this type of attack that hit the customers of one of South Africa's four large banks, Absa, in 2013.

II. MITM Attack Prevention

Various works have been conducted in the field of network security each with its own particular techniques. In order to detect and prevent ARP spoofing attacks[5], they conducted a comprehensive survey, analyzed existing schemes, and identified the requirements for optimal solutions, this can be used as a reference by network administrators when deciding how to secure their LANs to ARP attacks.

It was [6] who presented new defence methods, and compared them with previously proposed approaches. Most schemes can be categorized as to their implementation methods (cryptographic, voting-based, hardware), and where they are located (server/host/client). An overview of some prominent defense mechanisms against DNS spoofing by off-path spoofing adversaries was conducted by ("Antidotes for DNS poisoning by off-path adversaries") and their taxonomy of those mechanisms was described as well.

Jorge Belenguer[7] presented a low-cost embedded IDS which, when plugged into a switch or hub, is able to detect and/or prevent MitM attacks automatically and efficiently. In this paper[8] A novel application of the ID based signature scheme has been proposed to prevent IP spoofing and Man-in-the-Middle Attack across a network by implementing with key generation, packet generation with signature, and signature verification modules. There are also other MITM prevention mechanisms like strong mutual authentication at endpoints, exchanging public keys in a secure channel, using secure certificates and advanced encryption in cryptography. There are even dedicated Wi-Fi sensors in the organization networks to detect the anomalies but no standard mechanism was implemented in the client side. This specified shortcoming in the MITM prevention mechanism is focused in this paper.

III. Present Research Question

Although having various MITM defense mechanisms, heavy duty Wi-Fi sensors and Honeypots at the enterprise parameter level may prevent you from entering the organization network but the client who has no idea about the networking may vulnerable to the MITM somewhere and can compromise the whole network. Even with the very small window of opportunity the attacker can conduct various MITM attacks which was discussed above including the browser-based exploits, traffic hijacking, SSL MITM and application-level attacks. In light of these adverse effects, can Wi-Fi clients detect MITM attacks autonomously? In the enterprise side Wi-Fi sensors are deployed around the permiter and sense the Wi-Fi spectrum then it sends back to the central server/appliance, if any anomaly is detected then that device is flagged. But these infrastructure and appliances are specifically meant for the enterprise premise but this does not unfortunately extend to the client which may go out of the enterprise setup at a Hotel, Airport or any other hotspot area. In this paper we are going to use the attacker beacon frames parameters which was a broadcast network packet contains information of the access point id, physical address, authentication type, encryption mode, channel, fixed elements, variable sizes and several other information to build an intrusion detection system at the client side. These attacker beacon frames are likely similar even though using different types of attacking tools like airbase-ng[9] and Mdk3.

IV. Research Methodology

Instead of considering other network parameters we are only focusing on the attacker beacon frames because the attacker in the wild cannot know how the access point beacon looks like and its characteristics, attacker cloning capabilities also comes into picture due to the various Wi-Fi standards, hardware and software. As theses parameters from the attacker beacon frames can differ from the legitimate access point, we are going to build an intrusion detection system using machine learning anomaly detection[10] at the client side.

But there are couple of challenges while gathering the network data from the client side. Below are the few challenges to be considered:

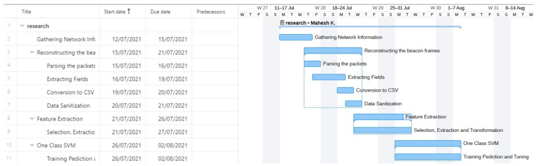
1) Client Wi-Fi Radios: The inbuilt Wi-Fi cards in the client-side device generally doesn't support the monitor mode which is a data capture where Wi-Fi network cards are able to capture all types of Wi-Fi packets. Even if they support, they don't simultaneously work in client mode.
2) Some Wi-Fi adapters may support both the modes but we even need low level driver access which are of high privileges to get out meaningful information out of the cards.
3) Roaming clients cannot collaborate with the central appliance
4) Clients can only use Wi-Fi telemetry data available by the OS, the open-source Linux provides deeper network information but whereas Windows OS limits this capability.

This paper is focused on obtaining the Wi-Fi telemetry data from Windows because if we get enough network information in a closed source OS then it is a piece of cake in the open-sourced software. We can't get enough Wi-Fi telemetry data from the default Windows configurator except the access points SSID's but there is built in feature in Windows called Windows Event Logs may get events information but not much to continue. Through the command line we can get certain network information like the mac address of the access point and network profiles. Windows also have a feature call Windows Native API which ends up giving lot of information about the Wi-Fi subsystem in Windows which includes state machine data, periodic scan results with BSS data, xml network profile data and card control such as scan, connect, disconnect lock etc. This telemetry information is derived from the network cache files. The main Information elements of the access point beacon in the network can be obtained in the form of binary blob using the Windows Native API on a per access point basis, but this information can be used and can be reconstructed again for further analyzing. Once the beacon frames are reconstructed those frames are inputted to the anomaly detection algorithm using machine learning and flag the anomalies with the fingerprinted network.

The anomaly detection happens in four stages starting with raw data which is gathering the data from the API and pcap files followed by the preprocessing method i.e., parsing the packets, extraction of network parameter fields, data sanitization. Third stage involves feature extraction and transformation of the telemetry data and then followed by the last stage which is the detection by training, predicting and tuning the anomaly using the One Class SVM.

V. Schedule Resources

Below the timelines for the project resources and tasks that had to be done before the deadlines.



	Title	Start date ↑	Due date	Predecessors
1	research			
2	Gathering Network Infc	12/07/2021	15/07/2021	
3	Reconstructing the bea	15/07/2021	21/07/2021	
4	Parsing the packets	15/07/2021	16/07/2021	
5	Extracting Fields	16/07/2021	19/07/2021	
6	Conversion to CSV	19/07/2021	20/07/2021	
7	Data Sanitization	20/07/2021	21/07/2021	
8	Feature Extraction	21/07/2021	26/07/2021	
9	Selection, Extractio	21/07/2021	27/07/2021	
10	One Class SVM	26/07/2021	02/08/2021	
11	Training Pediction ε	26/07/2021	02/08/2021	

W 27
11–17 Jul
W 28
18–24 Jul
W 29
25–31 Jul
W 30
1–7 Aug
W 31
8–14 Aug
W T F S S M T W T F S S M T W T F S S M T W T F S S M T W T F S S M T W
research • Mahesh K.
Gathering Network Information
Reconstructing the beacon frames
Parsing the packets
Extracting Fields
Conversion to CSV
Data Sanitization
Feature Extraction
Selection, Extraction and Transformation
One Class SVM
Training Pediction and Tuning

Fig. 2. Schedule.

VI. Summary

In this paper, we have studied the MITM and the different types of attacks associated with it. Also provided various MITM defense mechanisms along with their descriptions and proposed a new method to build an intrusion detection system at client side using the machine learning algorithms. In summary, there are many plug and play methods to build an intrusion-based detection based on the parameters but we are mainly focusing on providing idea of intrusion detection framework to be designed at client side.

References

[1] A. A. Maksutov, I. A. Cherepanov and M. S. Alekseev, "Detection and prevention of DNS spoofing attacks," 2017 Siberian Symposium on Data Science and Engineering (SSDSE), 2017, pp. 84-87, doi: 10.1109/SSDSE.2017.8071970.

[2] S. G. Bhirud and V. Katkar, "Light weight approach for IP-ARP spoofing detection and prevention," 2011 Second Asian Himalayas International Conference on Internet (AH-ICI), 2011, pp. 1-5, doi: 10.1109/AHICI.2011.6113951.

[3] https://www.venafi.com/blog/why-are-man-middle-attacks-so-dangerous-venafi

[4] Lisa Vaas, " 49 busted in Europe for Man-in-the-Middle bank attacks", 11 JUN 2015. https://nakedsecurity.sophos.com/2015/06/11/49-busted-in-europe-for-man-in-the-middle-bank-attacks/

[5] C. L. Abad and R. I. Bonilla, "An analysis on the schemes for detecting and preventing ARP cache poisoning attacks", Proc. 27th Int. Conf. Distrib. Comput. Syst. Workshops (ICDCSW'07), pp. 60, 2007

[6] M. Oh, Y.-G. Kim, S. Hong and S. Cha, "ASA: Agent-based secure ARP cache management", IET Commun., vol. 6, no. 7, pp. 685-693, May 2012.

[7] J. Belenguer and C. T. Calafate, "A low-cost embedded IDS to monitor and prevent Man-in-the-Middle attacks on wired LAN environments," The International Conference on Emerging Security Information, Systems, and Technologies (SECUREWARE 2007), 2007, pp. 122-127, doi: 10.1109/SECUREWARE.2007.4385321.

[8] V. Radhakishan and S. Selvakumar, "Prevention of Man-in-the-Middle Attacks Using ID Based Signatures," 2011 Second International Conference on Networking and Distributed Computing, 2011, pp. 165-169, doi: 10.1109/ICNDC.2011.40.

[9] https://www.aircrack-ng.org/doku.php?id=airbase-ng

[10] M. E. Karsligl, A. G. Yavuz, M. A. Güvensan, K. Hanifi and H. Bank, "Network intrusion detection using machine learning anomaly detection algorithms," 2017 25th Signal Processing and Communications Applications Conference (SIU), 2017, pp. 1-4, doi: 10.1109/SIU.2017.7960616.