CYBR 520

Module 7: Network Traffic Analysis Classification

Chapter 5: Network Traffic Analysis***

PLEASE REFER TO MODULE 7 ON GITHUB TO OBTAIN THE CODES OR REFER TO THE BOOK'S GITHUB REPO

Topics

1. Theory of Network Defense
2. Machine Learning and Network Security
3. Building A Predictive Model to Classify Network Attacks (on Github)

Introduction

- The most common point of entry for attackers is through computer networks.
- Network security is a comprehensive practice that involves safeguarding computer networks and the devices connected to them from malicious activities, misuse, and denial of service attacks.

Network Security

- Firewalls, well-known in network security, serve as essential defense systems.
- They enforce access policies and filter unauthorized traffic between network devices.
  - However, network security extends beyond firewalls.

Note

- In this chapter, we explore methods for categorizing network traffic. We'll establish a model of network defense as the foundation for our discussions.
- Additionally, we'll delve into specific areas of network security that have seen advancements in artificial intelligence and machine learning.

Note

- In the latter part of this chapter, we'll walk through an example of using machine learning to detect patterns and correlations in network data.
- We'll leverage data science techniques to apply classification to complex datasets, aiding in identifying potential attackers within the network.

The focus of this chapter

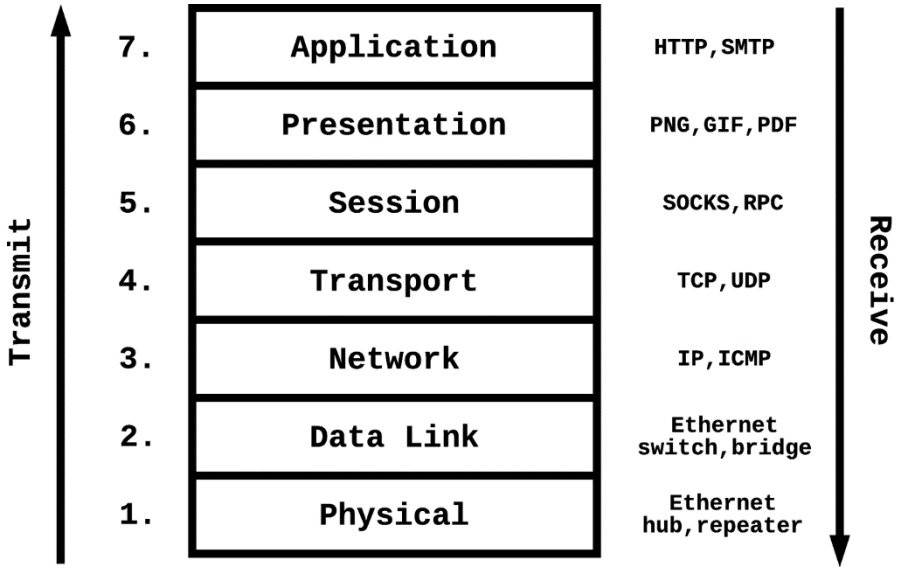
- Our focus in this chapter is on packet-based information transmission. In this approach, data streams are divided into smaller units, each containing metadata about origin, destination, and content.
- These packets are transmitted at the network layer and formatted using appropriate protocols at the transport layer. Information reconstruction from individual packets occurs at the session layer or higher.

The focus of this chapter

- The security of the network, transport, and session layers (layers 3, 4, and 5 of the OSI model) is the central theme of this chapter.

The OSI Model

- Throughout this chapter, we'll refer to different components of a typical networking stack using the Open Systems Interconnection (OSI) model.
- The OSI model consists of seven layers.



Layer 1: Physical Layer

- Converts digital data to electrical or mechanical signals for network transmission.
- Converts signals received over the network back into digital data.

Layer 2: Data Link Layer

- Facilitates data transfer between adjacent nodes in a physical network.

Layer 3: Network Layer

- Routes packets and manages flow control between two network points.

Layer 4: Transport Layer

- Provides end-to-end communication and determines data transmission quality and reliability.

Layer 5: Session Layer

- Initiates, maintains, and terminates sessions between application processes.

Layer 6: Presentation Layer

- Translates binary data into formats that applications can understand.

Layer 7: Application Layer

- Displays data received from the network to users.

Network Security

- Networks have a complicated defense model because of the broad range of attack surfaces and threat vectors.
- As is the case when defending any complicated system, administrators need to engage with attackers on multiple fronts and must make no assumptions about the reliability of any one component of the solution.

Theory of Network Defense

- The Theory of Network Defense is a framework or set of principles used in cybersecurity to protect computer networks from unauthorized access, misuse, malfunction, modification, destruction, or improper disclosure.
- This theory encompasses various strategies, techniques, and tools to secure a network and its resources.

Access Control and Authentication

- A client's interaction with a network begins with the access control layer.
- Access control is a form of authorization by which you can control which users, roles, or hosts in the organization can access each segment of the network.
  - Firewalls

Firewalls

- In Linux, built-in firewall, iptables, which enforces an IP-level ruleset that dictates the ingress and egress capabilities of a host, configurable on the command line.
- In Windows, the Windows Firewall with Advanced Security WFAS
  - Configure how others can connect to your SSH

Iptables exmple

- An attacker who gains control of a server in the 192.169.100.0/24 subnet can access the server because this passive authentication method relies on only a single property—the connection's origin—to grant or deny access.

```
# ACCEPT inbound TCP connections from 192.168.100.0/24 to port 22
iptables --append INPUT --protocol tcp --source 192.168.100.0/24
         --dport 22 --jump ACCEPT

# DROP all other inbound TCP connections to port 22
iptables --append INPUT --protocol tcp --dport 22 --jump DROP
```

Access Control and Authentication

- Access Control Methods that relay on singular data checkers are not very effective.
- *Active authentication methods* gather more information about connecting clients
  - Diversify authentication means
  - Minimize single point of failures

Active authentication methods

- For instance, in addition to using the connection origin as a signal, the system administrator might require an SSH key and/or authentication credentials to allow a connection request.
- In some cases, multifactor authentication (MFA) can be a suitable method for raising the bar for attackers wanting to break in.
- MFA breaks up the authentication into multiple parts, causing attackers to have to exploit multiple schemes or devices to get both parts of the key to gain the desired access.

Detecting In-Network Attackers

- In many cases –and in this chapter- the security measures are implemented at different levels (outside and inside).
- In-networks attackers do exist
  - Due to undetected traffic
  - Just protecting the perimeter is not sufficient, given that an attacker who spends enough time and resources on breaching the perimeter will often be successful.

Security segmentation

- Proper segmentation of a network can help limit the damage caused by in-network attackers.
  - Separate public from sensitive info
- Microsegmentation is the practice of segmenting a network into various sections based on each element's logical function

Network segmentation

- Nevertheless, network segmentation allows administrators an opportunity to enforce strict control on the number of possible paths between node A and node B on a network, and also provides added visibility to enable the use of data science to detect attackers

Data-Centric Security

- When the perimeter is compromised, any data stored within the network is at risk of being stolen.
  - Employing data-centric view limits possible damage of data loss when a network is compromised.
- Data-centric security emphasizes the security of the data itself, meaning that even if a database is breached, the data might not be of much value to an attacker.
  - Encryption?

Encryption of data in databases

- You may do that however, when could this be unsuitable?
- Actively used databases for data analysis makes this very pricy.
- Can we perform data analysis on encrypted data?

Homomorphic encryption

- A technique that allows data analysis using encrypted data.
- The *Brakerski-Gentry-Vaikuntanathan* (BGV) scheme is widely famous in performing computation without decrypting encrypted data.
  - This allows different data processing services that work on the same piece of data to pass an encrypted version of the data to each other potentially without ever having to decrypt the data.
  - Not so great with large scale data

Honeypots

- Honeypots are decoys intended for gathering information about attackers.
- Honeypots strategically placed in environments that experience a sizable volume of attacks can be useful for collecting labeled training data for research and improving attack detection models.

Machine Learning and Network Security

- Pattern mining is one of the primary strengths of machine learning, and there are many inherent patterns to be discovered in network traffic data.

No.	Time	Source	Protocol	Destination	Length	Info
1	0.000…	192.168.1.104	TCP	216.18.166.136	74	49859 → 80 [SYN] Seq=0 Win=8192 Len…
2	0.307…	216.18.166.136	TCP	192.168.1.104	74	80 → 49859 [SYN, ACK] Seq=0 Ack=1 W…
3	0.307…	192.168.1.104	TCP	216.18.166.136	66	49859 → 80 [ACK] Seq=1 Ack=1 Win=17…

Figure 5-2. TCP three-way handshake
(source: Wireshark screen capture)

From Captures to Features

- Capturing network traffic is similar to video surveillance.
- Packet analysis/ network / protocol sniffers
- This data is overwhelming
  - How can we extract features?
  - See examples on GitHub:
    - PacketsToFeatures.py

Note about features

- While we do focus on manually created featured, many features can be extracted using unsupervised feature learning algorithms
- These accept raw data and try to come up with their own features.
  - This is NOT Unsupervised Machine Learning

Threats in the Network

- We will only focus on OSI layers 3,4, and 5.
- Threats are:
  - Passive attacks
  - Active attacks
- Attacks:
  - Breaches
  - Spoofing
  - Pivoting
  - Denial of Service (DoS)

Passive Attacks

- Do not communication with any terminals in the network
- Info gathering and reconnaissance
- Ports scan are examples

Active attacks

- Active Attacks: Aggressive, deliberate actions to compromise network security.
- Categories Explained:
  - **Breaches**: Unauthorized network access, exploiting perimeter vulnerabilities.
  - **Spoofing**: Identity falsification to misdirect or intercept communications.
  - **Pivoting**: Navigating through networks post-access breach.
  - **Denial of Service (DoS)**: Overloading systems to deny user access.
  - **Botnets**: Utilizing networks of hijacked computers for coordinated attacks.
- Network Vulnerability: Emphasis on the need for robust security measures to counter these threats.

Network Breaches

- Definition: Unauthorized access to private systems, exploiting network perimeter weaknesses.
- Network Nodes as Perimeter: Role of routers, proxies, firewalls, switches, load balancers in defense.
- Intrusion Detection Systems: Mechanisms for detecting attacks on network perimeters, utilizing anomaly detection techniques.
- Server Infrastructures: Vulnerability analysis in the context of server setups and network configurations.

Techniques in Network Breaches

- **Information Gathering**: Passive reconnaissance to identify vulnerabilities. (see InforGathering.py on GitHub Repo)
- **Application Vulnerability Exploitation**: Methods attackers use to gain shell or root access via public endpoints.
- **Traffic Inspection:** Detecting remote application attacks by monitoring server communications.
- **Polymorphic Attacks**: Challenges in detecting evolving threats and the role of machine learning in addressing these challenges.

Insider Threats and Data Breaches

- Insider Threats: Risks posed by legitimate users with system access, such as corrupt employees.
- Access Control and Auditing: Implementing role-based access control policies and auditing internal security systems.
- Data Encryption: Securing stored data to reduce attack surface.
- Anomaly Detection: Using data science to detect inconsistencies in access patterns indicative of insider compromises.

Spoofing Attacks

- Spoofing Definition: Sending falsified data to mislead or intercept communications.
- DNS and ARP Spoofing: Methods of cache poisoning to misdirect client communications.
- Mitigation: Implementing DNSSEC for authenticating DNS resolutions and preventing most DNS spoofing attacks.
- Impact on Confidential Communications: How spoofing enables passive wiretapping and information exfiltration.

Pivoting in Cyber Attacks

- Pivoting: Technique of moving between servers within a network post initial access.
- Vulnerability of Infrastructures: Risks in networks with poorly designed or configured service boundaries.
- Attack Techniques: Switch spoofing, VLAN hopping, and their role in facilitating network movement.
- Penetration Testing: Use of tools like Meterpreter for identifying and exploiting pivoting vulnerabilities.

Denial-of-Service (DoS) Attacks

- DoS Attacks Defined: Targeting system availability to disrupt user access.
- Techniques: SYN Flooding and its impact on TCP handshake process.
- Botnet Role in DDoS: Use of botnets for amplifying the impact of DoS attacks.
- Variants of DoS: Various methods to drain server resources and interrupt service availability.

Botnets and Their Impact

- Botnet Definition: Networks of compromised computers used for malicious activities.
- Bot Traffic: Analysis of bot-generated web traffic, distinguishing between benign and malicious bots.
- Botnet Detection: Employing machine learning and statistical methods for detecting bot activities.
- Botnet Topology and Control: Understanding the structure and command mechanisms of botnets for effective countermeasures.

Understanding and Countering Botnets

- Importance of Botnet Knowledge: Understanding botnets for better network security.
- Botnet Zombies: The role of compromised machines in botnet operations.
- Machine Learning in Botnet Detection: Application of DNS query analysis and pattern mining for bot behavior identification.
- Command and Control (C&C) Mechanisms: Analysis of botnet control structures, including star/centralized, multileader, hierarchical, and P2P networks.

Building a Predictive Model to Classify Network Attacks

Building a Predictive Model to Classify Network Attacks

- Next, we move on to building a network attack classifier from scratch using Machine Leaning.
- Code and dataset are given on GitHub.