Network Traffic Analysis with Python (PyShark) in Databricks

Setup Prerequisites (Add to Assignment Instructions)

Before starting:

1. Upload .pcap files using Databricks' UI.
2. Save the file to:
   dbfs:/tmp/pcap/http.cap
   (or use Unity Catalog volume: /Volumes/training/network/pcap/http.cap)
3. Install PySHark:

%pip install pyshark

Assignment 1: Analyze Basic Packet Structure

Title: *"Dissecting Packets: View Layers and Fields from a .pcap File"*

Objective:

- Understand packet structure (layers: Ethernet, IP, TCP, etc.)

Tasks:

1. Load a .pcap file using PyShark.
2. Print out the first 5 packets.
3. For each packet, print:
   - packet.number
   - packet.length
   - packet.highest_layer
   - packet.transport_layer

Expected Outcome:
Students will learn how packet structure is represented and identify key headers.

Assignment 2: Extract and Count IP Traffic

Title: *"Top Talkers: Who's Talking the Most?"*

Objective:

- Count most frequent IP source addresses in a capture file.

Tasks:

1. Load a .pcap file.
2. Iterate through all packets.
3. Extract IP source addresses.
4. Count and rank top 5 source IPs.

Expected Outcome:
Learn basic analysis logic + Python dictionary usage.

Assignment 3: HTTP Traffic Analysis

Title: *"Web Tracker: List Visited URLs"*

Objective:

- Extract HTTP request info from packet capture.

Tasks:

1. Use a .pcap file with HTTP traffic (e.g., http.cap).
2. Use display_filter="http" in PyShark.
3. For each HTTP request, print:
    - Host
    - URI
    - Full URL

Expected Outcome:
Understand HTTP requests at the packet level.

Assignment 4: DNS Query Tracker

Title: *"Who’s Asking? Analyzing DNS Queries"*

Objective:

- Detect DNS queries and group by domain.

Tasks:

1. Use display filter "dns".

2. For each DNS query packet:
    - Extract packet.dns.qry_name
    - Count frequency
3. Output top 10 most queried domains.

Assignment 5: Real-Time Packet Capture (for Local Use)

Note: Only applicable outside Databricks in a local machine with TShark.

Title: *"Sniff It Live: Capture and Analyze Real-Time Traffic"*

Tasks:

- Use LiveCapture on Wi-Fi/ethernet
- Filter on TCP.port == 80
- Print first 10 HTTP packets
