📚 Complete Network Security Terminology Guide

This guide provides a comprehensive overview of essential network security terminology, core network concepts, dataset features for analysis, common attack types, enterprise terms, and interview-ready explanations.

🌐 Core Network Concepts

Network Protocols:

- **TCP (Transmission Control Protocol):** A reliable, connection-based communication protocol. It ensures that data packets are delivered in order and without errors, making it suitable for applications where data integrity is crucial.
- **UDP (User Datagram Protocol):** A fast, connectionless communication protocol. It prioritizes speed over reliability and does not guarantee delivery or order of packets, making it ideal for applications like video streaming or online gaming where slight data loss is acceptable for continuous flow.
- **ICMP (Internet Control Message Protocol):** Used by network devices to send error messages and operational information, such as whether a requested service is unavailable or a host or router cannot be reached. It's fundamental for network diagnostics.
- **IP (Internet Protocol):** The primary communication protocol for relaying datagrams across network boundaries. It defines how data is addressed and routed between devices on a network.

Connection Basics:

- **Source IP:** The unique numerical label identifying the originating device in a network connection.
- **Destination IP:** The unique numerical label identifying the target device where a network connection is directed.
- **Port:** A specific, numbered endpoint within a host that identifies a particular application or service. Think of it like a specific "door number" for a service on a computer.
- **Packet:** The fundamental unit of data transmitted over a network. Each packet contains a portion of the data along with control information like source and destination addresses.
- **Session:** A complete, interactive communication exchange or conversation between two networked systems, often involving multiple packets and requests.

🔧 Your Dataset Features Explained

These features are commonly found in network intrusion detection datasets and are

crucial for identifying malicious activities.

Basic Connection Metrics:

- **duration:** The total time, in seconds, that a network connection remained active.
- **src_bytes:** The total number of data bytes transmitted from the source (originator) of the connection to the destination.
- **dst_bytes:** The total number of data bytes transmitted from the destination back to the source of the connection.
- **land:** A binary indicator (1 or 0) signifying whether the source IP address and destination IP address, along with their respective ports, are identical. This is a classic indicator of a loopback attack.
- **wrong_fragment:** The count of fragmented packets that were received incorrectly (e.g., out of sequence or incomplete). A high count can suggest an attempt to evade detection or a malformed attack.
- **urgent:** The count of urgent packets sent. This flag is rarely used legitimately and can be a strong indicator of suspicious activity if present.

Connection Flags (Very Important!):

These flags represent the state and outcome of a network connection, often indicating normal behavior or potential attack patterns.

- **SF (Normal):** Indicates a "Successful, normal FINISH" connection. The connection was established, data was exchanged, and then properly terminated by both sides.
- **S0 (Suspicious):** Signifies a "SYN, no reply" state. The initiator sent a connection request (SYN packet), but received no response. This is a common indicator of a port scan or a DoS attempt targeting closed ports.
- **REJ (Rejected):** Indicates that a connection attempt was explicitly "REJECTED" by the destination host, often due to a firewall or security policy blocking the connection.
- **RSTO:** Stands for "Reset by ORIGINATOR." The connection was reset by the initiating host, often due to an error or an abnormal termination from the client side.
- **RSTR:** Stands for "Reset by RESPONDER." The connection was reset by the responding host, typically indicating an error or abnormal termination from the server side.

Services (What Was Being Accessed):

These represent common application layer protocols and services that might be accessed over a network.

- **HTTP:** Hypertext Transfer Protocol, primarily used for web browsing (default port 80).
- **HTTPS:** Hypertext Transfer Protocol Secure, the secure version of HTTP, encrypted using SSL/TLS (default port 443).
- **FTP:** File Transfer Protocol, used for transferring files between computers on a network (default port 21).
- **SSH:** Secure Shell, a cryptographic network protocol for secure remote access to computers (default port 22).
- **Telnet:** An older, insecure network protocol used for command-line access to remote computers (default port 23).
- **SMTP:** Simple Mail Transfer Protocol, used for sending email (default port 25).
- **DNS:** Domain Name System, used to translate human-readable domain names into IP addresses (default port 53).
- **private:** Refers to services operating on non-standard or internal-only ports, often used for custom applications or internal network communication.

Error Rate Metrics:

These features quantify the presence of errors during connection establishment or data transfer, which can be indicative of malicious activity.

- **serror_rate:** The percentage of connections to the current host that resulted in SYN errors (S0 flag). A high rate suggests port scanning or DoS attacks.
- **rerror_rate:** The percentage of connections to the current host that resulted in rejection errors (REJ flag). A high rate suggests a firewall blocking attempts or a DoS attack against a protected service.
- **srv_serror_rate:** The percentage of connections to the same *service* that resulted in SYN errors. Focuses on errors specific to a particular service.
- **srv_rerror_rate:** The percentage of connections to the same *service* that resulted in rejection errors. Focuses on rejections specific to a particular service.

Host Behavior Analysis:

These features capture aggregated behavior patterns related to a specific host or service within a recent time window, helping to identify anomalous activity.

- **count:** The number of connections to the same destination host as the current connection within the past 2 seconds.
- **srv_count:** The number of connections to the same service (on the same destination host) as the current connection within the past 2 seconds.
- **same_srv_rate:** The percentage of connections to the same service out of all connections to the same host within the past 2 seconds. A high rate indicates consistent use of one service.

- **diff_srv_rate:** The percentage of connections to different services out of all connections to the same host within the past 2 seconds. A high rate might suggest exploration or scanning.
- **dst_host_count:** The number of connections made to the overall destination host from *any* source within a defined time window (often 2 seconds).
- **dst_host_srv_count:** The number of connections made to the same service on the destination host from *any* source within a defined time window (often 2 seconds).

🚨 Attack Types & Security Terms

Understanding various attack types and security infrastructure is crucial for network data scientists.

Common Network Attacks:

- **Port Scan:** A reconnaissance technique where an attacker systematically tries different ports on a target system to discover open ports and services, indicating potential vulnerabilities.
- **DoS (Denial of Service):** An attack designed to overwhelm a system, server, or network resource with traffic or requests, making it unavailable to legitimate users.
- **DDoS (Distributed DoS):** A more potent version of a DoS attack, where the malicious traffic originates from multiple compromised systems (botnet), making it harder to mitigate.
- **Brute Force:** An attack method involving systematic, repeated attempts to guess credentials (like usernames and passwords) or encryption keys until the correct one is found.
- **Buffer Overflow:** A vulnerability that occurs when a program attempts to write data to a fixed-size memory buffer, but the data exceeds the buffer's capacity, overwriting adjacent memory locations. This can lead to system crashes or arbitrary code execution.
- **Root Kit:** A stealthy type of malicious software designed to hide the existence of other malware or unauthorized access on a computer. It grants attackers administrator-level access.
- **Probe:** A general term for reconnaissance activities aimed at gathering information about a target network or system, often a precursor to more targeted attacks.

Security Monitoring:

- **SOC (Security Operations Center):** A centralized unit within an organization

that is responsible for continuously monitoring and improving an organization's security posture, preventing, detecting, analyzing, and responding to cybersecurity incidents.
- **SIEM (Security Information Event Management):** A software solution that aggregates and analyzes security alerts generated by network hardware and applications. It provides real-time analysis of security alerts generated by network hardware and applications.
- **IDS (Intrusion Detection System):** A system that monitors network traffic or system activities for suspicious activity or policy violations and alerts when such activity is found. (What you're building!)
- **IPS (Intrusion Prevention System):** Similar to an IDS, but with the added capability to actively block or prevent detected intrusions.
- **False Positive:** An alert generated by a security system (like an IDS) that identifies legitimate, normal traffic or activity as an attack. This is a common challenge in security monitoring.
- **False Negative:** The failure of a security system to detect an actual attack or malicious activity. This is generally more dangerous than a false positive.

🏢 Enterprise/Cisco Terms

These terms are relevant to the business and technical environment of a large enterprise like Cisco.

Business Terms:

- **CISO (Chief Information Security Officer):** A senior-level executive who is responsible for an organization's information and data security.
- **SLA (Service Level Agreement):** A contract between a service provider and a customer that specifies the level of service expected from the provider.
- **ROI (Return on Investment):** A performance measure used to evaluate the efficiency or profitability of an investment. In security, it can refer to the business value gained from security investments.
- **KPI (Key Performance Indicator):** A measurable value that demonstrates how effectively a company is achieving key business objectives. In security, this could include incident response time or number of vulnerabilities patched.
- **Compliance:** Adherence to a set of rules, standards, laws, or regulations. In cybersecurity, this involves meeting requirements like GDPR, HIPAA, or industry-specific security standards.

Technical Infrastructure:

- **LAN (Local Area Network):** A computer network that interconnects computers

within a limited area such as a residence, school, laboratory, or office building.
- **WAN (Wide Area Network):** A telecommunications network that extends over a large geographical distance for the primary purpose of computer networking. The internet is the largest WAN.
- **Firewall:** A network security device that monitors incoming and outgoing network traffic and permits or blocks data packets based on a set of security rules.
- **Router:** A networking device that forwards data packets between computer networks. Routers perform the traffic directing functions on the Internet.
- **Switch:** A network device that connects devices in a computer network by using packet switching to receive, process, and forward data to the destination device.

Data Science Terms:

- **ETL (Extract, Transform, Load):** A process that extracts data from various sources, transforms it into a structured format, and loads it into a data warehouse or database for analysis.
- **Feature Engineering:** The process of using domain knowledge of the data to create new input features for a machine learning model from raw data, improving model performance.
- **Binary Classification:** A type of supervised learning problem where the goal is to predict one of two possible outcomes (e.g., normal vs. anomaly, spam vs. not spam).
- **Supervised Learning:** A type of machine learning where an algorithm learns from a labeled dataset, meaning the input data is already tagged with the correct output.
- **Model Deployment:** The process of integrating a trained machine learning model into an existing production environment where it can receive new data and make predictions in real-time.

💡 Interview-Ready Explanations

Practice these concise explanations to confidently articulate complex concepts in interviews.

- When asked "What's TCP vs UDP?"
  "TCP is like registered mail - reliable, confirms delivery, used for web browsing. UDP is like regular mail - fast but no delivery confirmation, used for video streaming where speed matters more than perfection."
- When asked "What's a port scan?"
  "It's like someone walking through an office building trying every door handle to see which offices are unlocked. In network terms, attackers probe different ports to find vulnerable services."

- When asked "What makes this dataset realistic?"
  "It simulates a real Air Force LAN with actual attack patterns - port scans showing high error rates, DoS attacks with unusual byte patterns, and normal traffic with predictable service usage."

Practice saying these terms out loud so you sound confident in interviews! 🎯