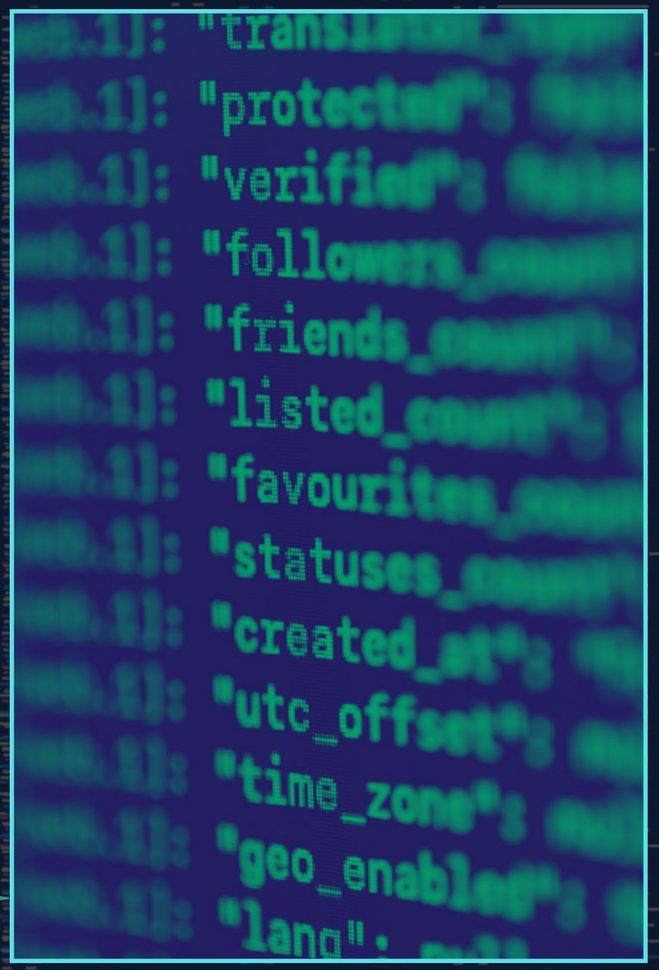
KDD CUPP - 1999

Incident Response

PRESENTATION



01. INCIDENT

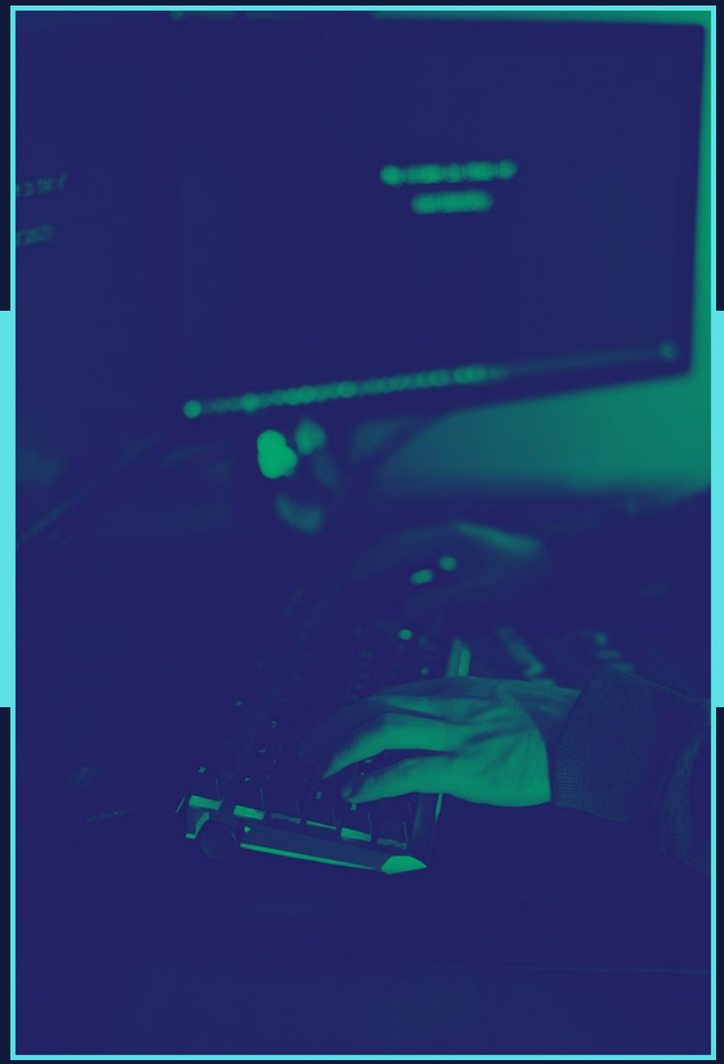
KDD CUP 1999 data : https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

Where does the data come from?
The dataset originates from the **1998 DARPA Intrusion Detection Evaluation Program** , which was prepared and managed by MIT Lincoln Labs.

What happened?
Simulated DDos attack

Who generated it?
It was generated by **MIT Lincoln Laboratory** under the sponsorship of the **Defense Advanced Research Projects Agency (DARPA)** and **Air Force Research Laboratory (AFRL)** .

What kind of devices / technologies does it target?
The dataset simulates a military network environment and targets **network-based intrusion detection systems (NIDS)** . It includes traffic data collected from a simulated network of Unix and Windows machines.



(P) Preparation

1. Patch asset vulnerabilities
2. Ensure antivirus/endpoint protection software is installed on workstations and laptops
3. Confirm that servers and workstations are logging to a central location
4. Review firewall, IDS, and IPS rules routinely and update based on the needs of the environment
5. Restrict access to critical assets as needed
6. Conduct employee security awareness training
7. Restrict users to the least privileges required

(I) Identification

1. Monitor for:
   a. Malicious tasklist commands ran in Windows environment [1]
   b. Malicious ps commands ran in Mac and Linux environments [1]
   c. Actions that could be taken to gather system and network information [1]
   d. Attempts by programs to exfiltrate process memory [1]
2. Routinely check firewall, IDS, IPS, and SIEM logs for any unusual behavior
3. Analyze web application metadata for suspicious user-agent strings and other artifacts
4. Investigate and clear ALL alerts
5. Investigate information provided by Windows Management Instrumentation and Windows API via PowerShell [1]

(E) Eradication

1. Close the attack vector by applying the Preparation steps listed above
2. Perform endpoint/AV scans on targeted systems
3. Reset any compromised passwords
4. Inspect ALL assets and user activity for IOC consistent with the attack profile
5. Inspect backups for IOC consistent with the attack profile PRIOR to system recovery
6. Patch asset vulnerabilities

(R) Recovery

1. Restore to the RPO within the RTO
2. Address any collateral damage by assessing exposed technologies
3. Resolve any related security incidents
4. Restore affected systems to their last clean backup

(C) Containment

1. Inventory (enumerate & assess) environment technologies
2. Detect | Deny | Disrupt | Degrade | Deceive | Destroy
3. Observe -> Orient -> Decide -> Act
4. Archive scanning related artifacts such as IP addresses, user agents, and requests
5. Determine the source and pathway of the attack
6. Issue a perimeter enforcement for known threat actor locations

(L) Lessons/Opportunities

1. Perform routine cyber hygiene due diligence
2. Engage external cybersecurity-as-a-service providers and response professionals
3. Implement policy changes to reduce future risk
4. Utilize newly obtained threat signatures
5. Remember that data and events should not be viewed in isolation, but as part of a chain of behavior that could lead to other activities [1]

References:

1. MITRE ATT&CK Technique 1057: https://attack.mitre.org/techniques/T1057/

Incident Response

The KDD Cup 1999 dataset is based on simulated network traffic rather than live host or process monitoring, so we have **adapted** parts of the playbook to focus on identifying attacker behaviors through **network-level indicators.**

For example, we will analyze the dataset for suspicious patterns and behaviors within the network data that align with the **Discovery techniques** outlined in the playbook (e.g., network probing or exfiltration attempts).

GSPBC-1040 - Discovery - Process Discovery:

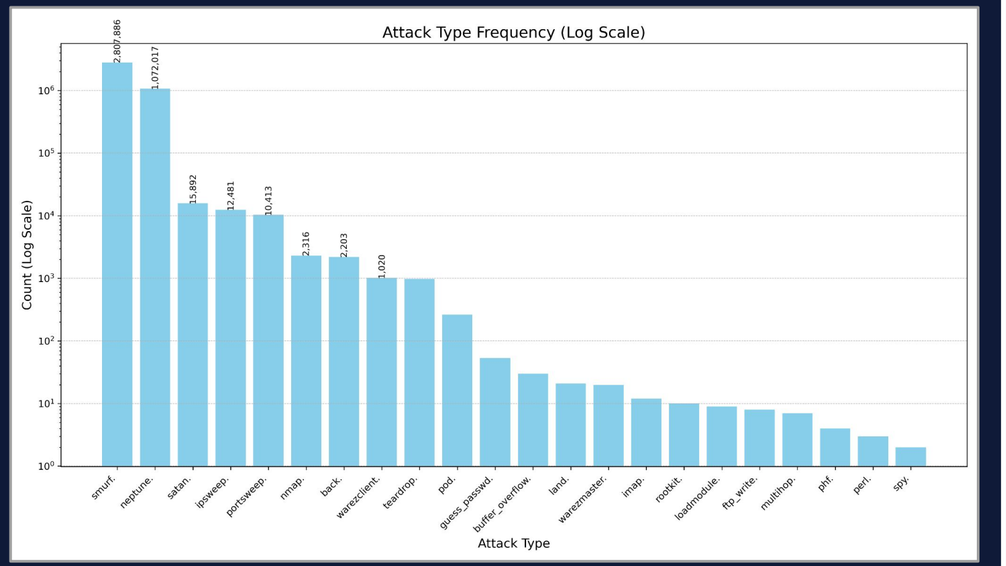
Tools Used

Wireshark
I'll use Wireshark to explore packet-level details and visualize how some of the attacks (like DoS or probing) behave in terms of traffic.

Python (Pandas, Scikit-learn)
I'll analyze and visualize the dataset using Python, focusing on patterns across features like src_bytes, duration, and count. I may also try to build simple anomaly detection models.

CIRT Playbook Battle Card: GSPBC-1040 - Discovery - Process Discovery
I plan to map the different attack types in the dataset to relevant MITRE ATT&CK Discovery techniques, helping to tie this simulated data to real-world frameworks

```python
import os
import csv

# Path to your original file
input_file = os.path.expanduser("~/Desktop/kddcup.data.corrected.txt")
# Path to save the abnormal lines
output_file = os.path.expanduser("~/Desktop/abnormal_lines.csv")

def extract_abnormal_lines(input_path, output_path):
    with open(input_path, "r") as infile, open(output_path, "w", newline="") as
        reader = csv.reader(infile)  # <-- fixed this line
        writer = csv.writer(outfile)

        for row in reader:
            if row and row[-1].strip() != "normal.":
                writer.writerow(row)

    print(f"Abnormal lines saved to: {output_path}")

extract_abnormal_lines(input_file, output_file)
```



Identify Assets Affected

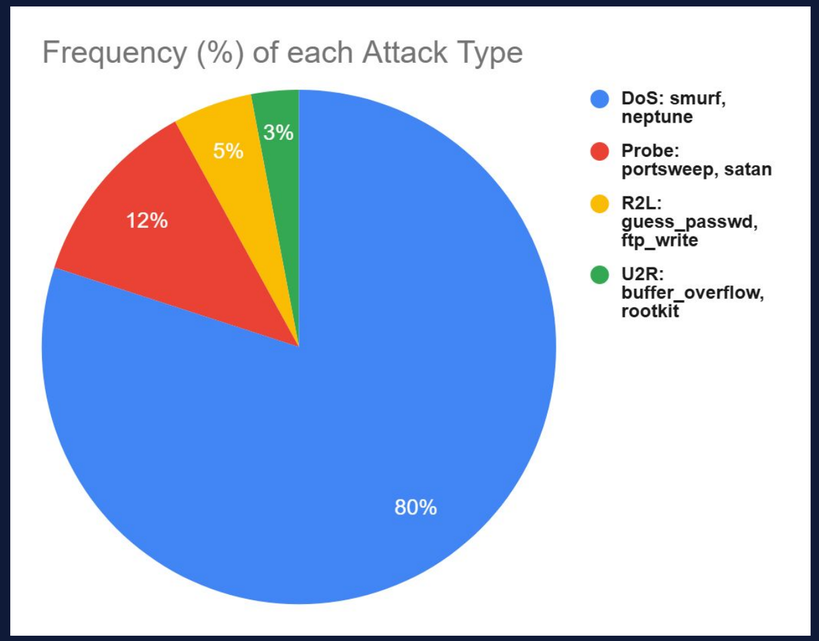
The KDD dataset represents a simulated network environment consisting of:

- Unix-based systems and Windows NT servers
- A set of workstations and servers used to host services (like HTTP, FTP, Telnet)
- User machines issuing requests or being targeted

Affected assets depended on the attack type:

- DoS attacks impacted availability of public-facing servers
- R2L (Remote to Local) and U2R (User to Root) attacks targeted system-level privileges on internal machines
- Probing attacks aimed at network devices and open ports across multiple assets

Impact Analysis and Triage



Criticality Level	Technique	Volume	System Role	Privilege Escalation Risk	Priority
Critical (Immediate)	T1057: Process Discovery	High	High-Criticality	High	Immediate
Critical (Immediate)	T1046: Network Service Scannir	High	High-Criticality	High	Immediate
High	T1057: Process Discovery	Low	Low-Criticality	Medium	High
High	DoS Attacks	High	Low-Criticality	Low	High
Medium	DoS Attacks	Low	Low-Criticality	Low	Medium
Low	T1057: Process Discovery	Low	Low-Criticality	Low	Low

As part of our impact analysis and triage process:

We first **grouped the dataset by attack type** and calculated the frequency of each

- Over 90% of attacks were either DoS or probe-based, with the “smurf” and “neptune” attacks being the most common

We assessed **criticality by volume, system role, and privilege escalation potential** :

- DoS attacks affected service availability
- U2R/R2L attacks presented a **higher risk** due to potential full system compromise

We triaged incidents by mapping them to MITRE ATT&CK techniques (e.g., T1057: Process Discovery, T1046: Network Service Scanning) to understand the tactics involved and evaluate severity.

Recommended Remediation

Strengthening Network Security

- Patch known vulnerabilities to eliminate known entry points and regularly update systems
- Rate-limit ICMP and SYN traffic to reduce impact of DoS attacks
- Update IDS/IPS rules based on observed attack traffic patterns

Enhance User Training and Policy

- Conduct security awareness training simulating DoS, R2L, and probing attacks to improve phishing and misuse detection
- Update the GSPBC-1040 playbook to reflect the most recent DDoS incident and attackers' TTPs (tactics, techniques, and procedures)

Enforce Access Control

- Implement stricter access control and enforce least-privilege policies
- Restrict the use of scripting and process activity tools such as **Powershell** , **tasklist** , and **ps** to administrator accounts

Final Thoughts

- Real-world dataset exploration
- Structured incident analysis
- Lessons learned & recommendations

Key takeaway:

- Build strong response plans over chasing perfection

THANK YOU!