



# Framework for generating IDS benchmarking Data sets

Stian Skjølvik

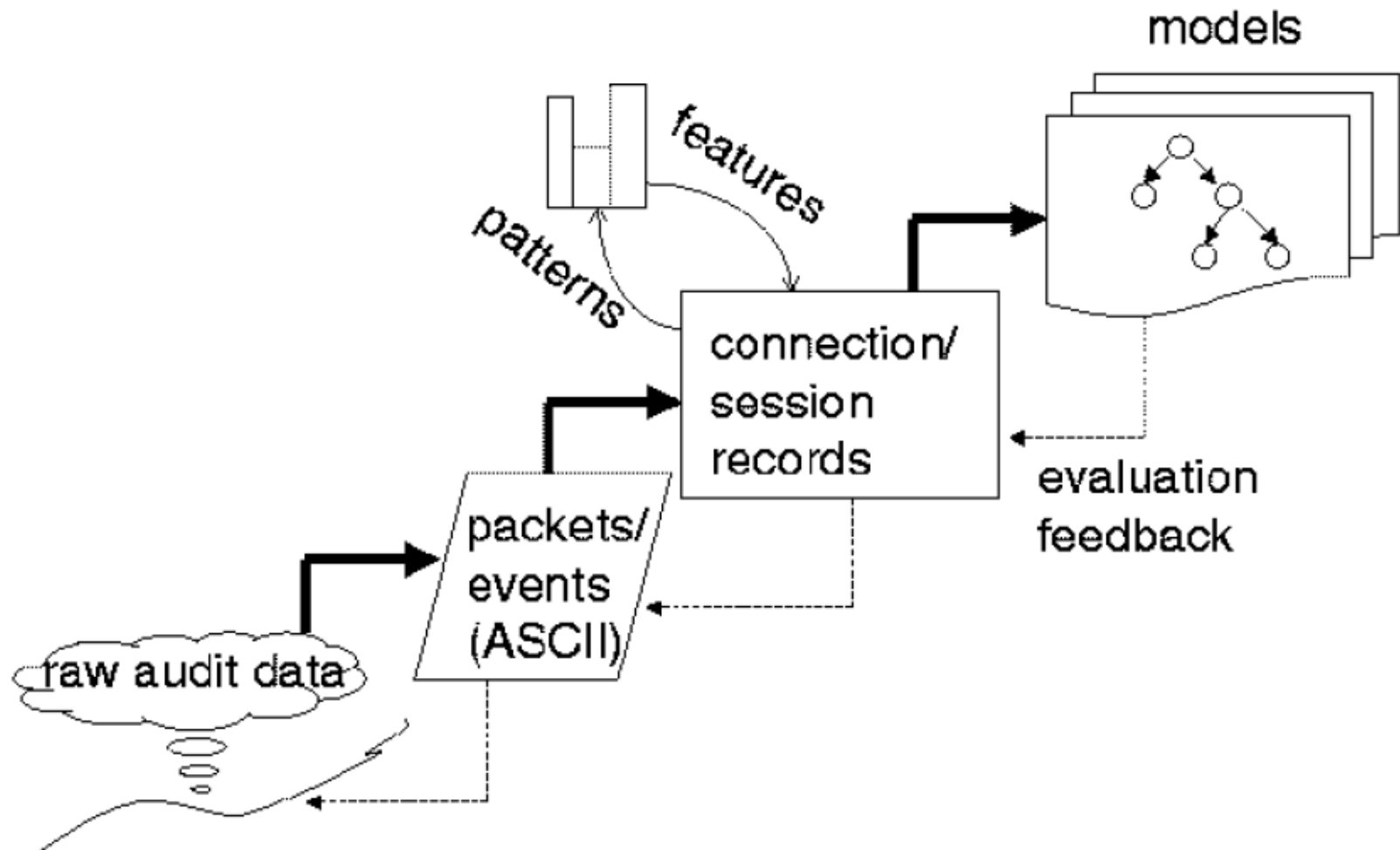
# Outline

- Introduction - KDD CUP 99
- Honeypot – capturing network traffic
- Preprocessing
- Prototype
- Experiments
- Conclusions
- Questions

# Introduction - KDD CUP 99

- KDD CUP 99 is the only publicly available labeled benchmarking data set for IDS.
  - Constructed using simulated traffic and attacks.
  - Has been criticized, mainly because the simulated traffic.
  - 4 attack class: DOS, R2L, U2R and Probing.
  - 41 features extracted: Basic, Content and Traffic features.

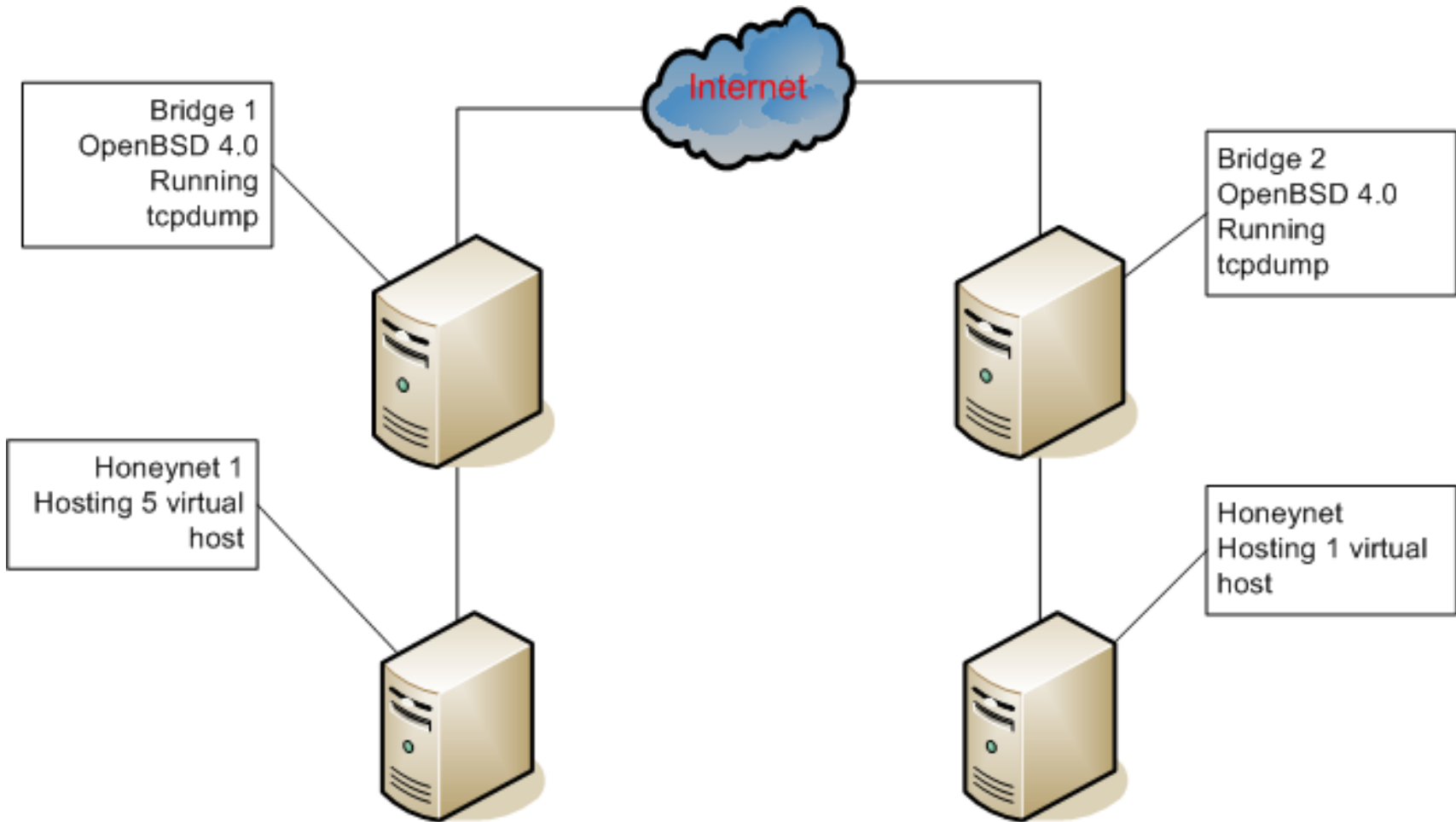
# Introduction - KDD feature extraction



# Introduction – research questions

- How can common properties of attacks be extracted from a large and evergrowing set of attacks against computers networks?
- What are the properties of an IDS benchmarking data set based on real traffic for whose generation the methods of extraction of common properties of attacks have been used?

# Honeypot – Capturing network traffic I



# Honeypot – capturing network traffic II

- tcpdump - used to capture all packets through the bridge, stored on file.
- Traffic from 23th February to 27th March.
  - Captured approximately 177 000 useful packets.
  - 8 different attack types.
- Added 6 attack types to create a modified data set.

# Preprocessing

- Export binary tcpdump file to PDML (XML).
- Remove unwanted network packets.
- Extract packet information from PDML file.
  - Generic packet information (Ethernet protocol).
  - Internet Protocol header (IP).
    - Internet Control Message Protocol (ICMP)
    - Transmission Control Protocol (TCP)
    - User Datagram Protocol (UDP)

# Prototype - tasks

- Establishing connection records from the preprocessed packets.
- Computing features for each connection.
- Label each connection, either an attack or normal traffic.
- Generating the data set and stores it to a file.

# Prototype - connection records

- Information about a connection between two host:
  - Same or opposite source and destination addresses.
  - Same or opposite source and destination ports.
  - Packets must have been sent in both directions.
  - Payload of every packet in the connection is added.

# Prototype – feature extraction

- 7 basic features: Duration, Protocol, Service, Src\_Bytes, Dst\_Bytes, Land and Urgent.
- 3 new content features: Access to passwd, Cross site scripting and Directory traversal.
- 5 traffic features: Count, Same\_srv\_rate, Diff\_srv\_rate, Srv\_count and Srv\_diff\_host\_rate.

# Prototype – new features

- Access to passwd attempted
  - Request to port 80 containing: ``/etc/passwd``.
- Cross site scripting attempted
  - Packet to port 80 containing: ``<script``.
  - False positive if web application accepts scripts.
- Directory traversal attempted
  - Packet to port 80 containing ``..\`` or ``../``.
  - False positive if web page has bad links containing these sequences.

# Prototype – labeling connections

- All occurrences of attacks are stored in a file similar to the alert file generated by Snort.
- All packets in a connection are checked:
  - Timestamp with source and destination addresses.
  - Sequence and Acknowledgement number.
  - Source and Destination ports.
  - ICMP Code and Type.
- Each connection obtains a label, either normal or the classification name of the attack.

# Prototype – data set(modified)

- 42 841 connection records – 14 attack types.

| ID  | BASIC FEATURES |          |         |           |           |       |        | CONTENT FEATURES |                      |                     |
|-----|----------------|----------|---------|-----------|-----------|-------|--------|------------------|----------------------|---------------------|
|     | Duration       | Protocol | Service | Src_Bytes | Dst_Bytes | Land  | Urgent | Access to passwd | Cross site scripting | Directory traversal |
| 1   | 1,12           | TCP      | 80      | 752       | 1474      | False | 0      | True             | False                | False               |
| 14  | 1,169          | TCP      | 80      | 750       | 1474      | False | 0      | True             | False                | True                |
| 19  | 1,197          | TCP      | 80      | 744       | 1474      | False | 0      | True             | False                | True                |
| 24  | 1,21           | TCP      | 80      | 713       | 1474      | False | 0      | True             | False                | True                |
| 333 | 1,406          | TCP      | 80      | 1208      | 3068      | False | 0      | True             | False                | True                |
| 343 | 1,342          | TCP      | 80      | 1083      | 3068      | False | 0      | True             | False                | False               |
| 370 | 1,296          | TCP      | 80      | 1031      | 3068      | False | 0      | True             | False                | False               |

| TRAFFIC FEATURES |               |               |           |                    | Status of connection    |
|------------------|---------------|---------------|-----------|--------------------|-------------------------|
| Count            | Same_srv_rate | Diff_srv_rate | Srv_count | Srv_diff_host_rate |                         |
| 1                | 1             | 0             | 1         | 0                  | Web /etc/passwd attempt |
| 1                | 1             | 0             | 1         | 0                  | Web /etc/passwd attempt |
| 2                | 1             | 0             | 1         | 0,5                | Web /etc/passwd attempt |
| 3                | 1             | 0             | 2         | 0,333333333        | Web /etc/passwd attempt |
| 1                | 1             | 0             | 1         | 0                  | Web /etc/passwd attempt |
| 2                | 1             | 0             | 1         | 0,5                | Web /etc/passwd attempt |
| 3                | 1             | 0             | 2         | 0,333333333        | Web /etc/passwd attempt |

# Experiments - methodology

- Compare the attacks found in the data sets with alarms from Snort to determine detection rates.
- Analyze the three new content features, and their abilities to identify attacks.
  - Finding a combination of features to identify each attack type.
  - Determining the relevance of each feature to be used to detect various attacks.

# Experiments – attacks in the modified data set

| Attack name                          | Attack type            | Number of occurrences | Number of attacks detected by Snort | DR    |
|--------------------------------------|------------------------|-----------------------|-------------------------------------|-------|
| Authorization basic overflow attempt | U2R, DOS               | 4                     | 4                                   | 100 % |
| Chunked-Encoding transfer attempt    | U2R                    | 1                     | 1                                   | 100 % |
| Cross site scripting attempt         | Information disclosure | 7                     | 7                                   | 100 % |
| HTTP directory traversal             | Information disclosure | 14                    | 14                                  | 100 % |
| ICMP Ping NMap                       | Probing                | 8                     | 8                                   | 100 % |
| ICMP Superscan echo                  | Probing                | 6                     | 6                                   | 100 % |
| IIS SAM attempt                      | R2L                    | 12                    | 12                                  | 100 % |
| IIS view source via translate header | Information disclosure | 15                    | 15                                  | 100 % |
| MS SQL worm propagation attempt      | U2R                    | 768                   | 765                                 | 100 % |
| SYN Scan                             | Probing                | 22295                 | 0                                   | 0 %   |
| Traceroute ICMP                      | Probing                | 1                     | 1                                   | 100 % |
| Traceroute UDP                       | Probing                | 8                     | 0                                   | 0 %   |
| Web /etc/passwd attempt              | Information disclosure | 7                     | 7                                   | 100 % |
| WebDAV search access                 | DOS                    | 1                     | 1                                   | 100 % |

# Experiments – relevance of features

| Feature              | Attack                               |                                   |                              |                          |                |                     |                 |                                      |                                 |          |                 |                |                         |                      |
|----------------------|--------------------------------------|-----------------------------------|------------------------------|--------------------------|----------------|---------------------|-----------------|--------------------------------------|---------------------------------|----------|-----------------|----------------|-------------------------|----------------------|
|                      | Authorization basic overflow attempt | Chunked-Encoding transfer attempt | Cross site scripting attempt | HTTP directory traversal | ICMP Ping NMap | ICMP Superscan echo | IIS SAM attempt | IIS view source via translate header | MS SQL Worm propagation attempt | SYN Scan | Traceroute ICMP | Traceroute UDP | Web /etc/passwd attempt | WebDAV search access |
| Duration             |                                      |                                   |                              |                          |                |                     |                 |                                      | X                               |          |                 |                |                         |                      |
| Protocol             |                                      |                                   |                              |                          | X              | X                   |                 |                                      | X                               |          |                 | X              |                         |                      |
| Service              | X                                    |                                   |                              |                          |                |                     |                 |                                      | X                               |          |                 | X              |                         |                      |
| Src_bytes            |                                      |                                   |                              |                          | X              | X                   |                 |                                      |                                 | X        |                 |                |                         |                      |
| Dst_bytes            |                                      |                                   |                              |                          | X              | X                   |                 |                                      |                                 | X        |                 |                |                         |                      |
| Land                 |                                      |                                   |                              |                          |                |                     |                 |                                      |                                 |          |                 | X              |                         |                      |
| Urgent               |                                      |                                   |                              |                          |                |                     |                 |                                      |                                 |          |                 |                |                         |                      |
| Access to passwd     |                                      |                                   |                              |                          |                |                     |                 |                                      |                                 |          |                 |                | X                       |                      |
| Cross site scripting |                                      |                                   | X                            |                          |                |                     |                 |                                      |                                 |          |                 |                |                         |                      |
| Directory traversal  |                                      |                                   |                              | X                        |                |                     |                 |                                      |                                 |          |                 |                |                         |                      |
| Count                | X                                    |                                   |                              |                          |                |                     |                 |                                      |                                 |          |                 |                |                         |                      |
| Same_srv_rate        | X                                    |                                   |                              | X                        |                |                     |                 |                                      |                                 |          |                 |                |                         |                      |
| Diff_srv_rate        |                                      |                                   |                              |                          |                |                     |                 |                                      |                                 | X        |                 |                |                         |                      |
| Srv_count            | X                                    |                                   |                              |                          |                |                     |                 |                                      |                                 |          |                 |                |                         |                      |
| Srv_diff_host_rate   |                                      |                                   |                              |                          | X              | X                   |                 |                                      |                                 |          |                 |                |                         |                      |

# Experiments – detection rates

| Attack                               | DR, TPR | FPR     | FNR     |
|--------------------------------------|---------|---------|---------|
| Authorization basic overflow attempt | 100 %   | -       | -       |
| Chunked-Encoding transfer attempt    | NA      | NA      | NA      |
| Cross site scripting attempt         | 100 %   | -       | -       |
| HTTP directory traversal             | 0,75 %  | *       | *       |
| ICMP Ping NMap                       | 38,9 %  | 61,1 %  | 12,5 %  |
| ICMP Superscan echo                  | 100 %   | -       | 16,67 % |
| IIS SAM attempt                      | 0,76 %  | 99,24 % | -       |
| IIS view source via translate header | 11,43 % | 88,57 % | 73,33 % |
| MS SQL worm propagation attempt      | 100 %   | -       | -       |
| SYN Scan                             | 99,99 % | 0,013 ‰ | -       |
| Traceroute ICMP                      | NA      | NA      | NA      |
| Traceroute UDP                       | 100 %   | -       | -       |
| Web /etc/passwd                      | 100 %   | -       | -       |
| WebDAV search access                 | NA      | NA      | NA      |

# Conclusions

- Constructed a framework for processing network packets captured by tcpdump, to extract common properties for all connections.
- Determined detection rate for Snort.
- Relevance of features.
- Detection rate, FPR and FNR for each combination of features, to uniquely identify one type of attack.

# Questions

- Opponent
- Supervisor
- Others