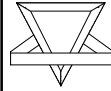


## Profile Building

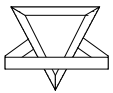
Tysen Leckie



## Profile Building

### Attacks

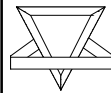
- No such thing as a false alarm. All alarms contain information.
- Different alarms require different levels of intervention. Might not be a prelude to an attack.
- Difficult to collect data representative of an attack - Can never be sure that there are no subtle attacks hiding undiscovered in the data.
- Cannot count missed detections you did not find.
- Cannot be absolutely sure that a false alarm is not in fact a true detection.



## Profile Building

### Data

- In order to get a good sample of normal behavior you might need to collect a very large amount of data.
- Collect data and separate it into:
  - Training Set - Data that does not contain malicious behavior. Normal behavior profile can then be recorded per user. Have SEADS run normally without any attack traffic for a predefined amount of time (12 hours).
  - Testing Set - Normal behavior plus malicious behavior mixed. Have SEADS run with known attack traffic for the same amount of time (12 hours).
  - Objective is to be able to distinguish the two.
- You can then calculate accurate true and false positive rates from this simulation (because you know the real attacks).



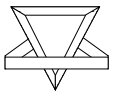
## Profile Building

### Training Set

- Run in an environment that contains no attacks.
- All activity is then defined as normal activity for that user.

### Testing Set

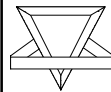
- Run in an environment that contains random behavior (normal + malicious)



## Profile Building

### Questions

- What is truth? Which packets or sessions are intrusions and which are not?
  - Can determine this if the system is run in a controlled environment.
- Is this a typical data set?
  - Partition the normal profile per time periods:
    - Weekend, Weekday, Holiday
    - Early Morning, Morning, Afternoon, Night, Late Night.
  - Interval-Based IDS: Data are collected over a period of time (above) and processed in batches.
- How long is the data going to be representative? How fast is the network or threat evolving?
  - Should updated the normal profile after a certain time period (above). Events that are above a predefined threshold are flagged as attacks. Events below a predefined threshold are added to the normal profile stats.
- Have we measured the threat? Are the attacks representative of the ones we want to detect?
  - Can determine this if the system is run in a controlled environment. Should determine the useful measurements we can track for a security protocol.

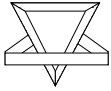


## Profile Building

### Issues

#### N-Gram Approach

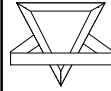
- Define the size  $n$  (6). Store the unique strings of size  $n$  in the database during normal behavior.
  - Example - A has consecutive activity with (B, C, S, B, C, D)
  - Define a window (interval) of size  $W$  (20). Tally number of sequences within each window of size  $W$  captured during the testing phase that do not match any sequence in the database. If this number is larger than some threshold, signal an alert.



## Profile Building

### ▼ Metrics

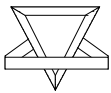
- The numbers per user, per time period of each of these metrics are stored in the profile database.
- These metrics deal with connection requests initiated by the local user (Self):
  - Successful Connection Requests by Self.
  - Failed Connection Requests by Self.
- These metrics deal with ongoing sessions that are past the initial connection request phase, initiated by the local user (Self):
  - Completed Sessions Initiated by Self
  - Failed Ongoing Sessions Initiated by Self
- These metrics deal with connection requests initiated by a remote user (Other) to a local user (Self):
  - Successful Connection Requests by Other to Self
  - Failed Connection Requests by Other to Self



## Profile Building

### ▼ Metrics

- These metrics deal with ongoing sessions that are past the initial connection request phase, initiated by a remote user (Other) to a local user (Self):
  - Completed Sessions Initiated by Other to Self
  - Failed Ongoing Sessions Initiated by Other to Self
- Other metrics:
  - Unique Session Partners
  - Repeat Session Partners
  - Concurrent Sessions with at least one Shared



## Profile Building

### ▼ Time Categories

- **Time is partitioned into the following structure:**
- Weekday Behavior = Monday, Tuesday, Wednesday, Thursday, Friday
- Weekend Behavior = Saturday, Sunday
- Holiday Behavior = Christmas, Easter, etc.
- **This is the initial structure of time and their attributes. The structure contains military time values. This helps in the storage and representation. The partition continues to refine the events by categorizing each of the above categories into:**
- Early Morning = 0001 - 0559 (12:01 A.M. - 5:59 A.M.)
- Morning = 0600 - 1059 (6:00 A.M. - 10:59 A.M.)
- Afternoon = 1100 - 1659 (11:00 A.M. - 4:59 P.M.)
- Night = 1700 - 2159 (5:00 P.M. - 9:59 P.M.)
- Late Night = 2200 - 2400 (10:00 P.M. - 12:00 Midnight)