Detecting Distributed Denial-of-Service Attacks by analyzing TCP SYN packets statistically

Yuichi Ohsita
Osaka University
Contents

- What is DDoS
- How to analyze packet
- Traffic modeling
- Method to detect SYN Flood attacks
- Evaluation
- Summary
What is DDoS?

- An attacker hacks remote hosts and installs attack tools
- The hosts attack the same server
What is DDoS?

- The number of attacks is increasing
- The number of attack nodes is very large and attack nodes are highly distributed
- The most are SYN Flood Attacks
  - Because SYN flood can put servers into denial-of-service state easily
  - More than 90% of DoS Attacks [1]

What is SYN Flood?

- Normal 3-way handshake
  - The in-progress connection requests are held in the backlog-queue
  - The queue length is limited
  
  ![Diagram](image)

Client PC

Server

- SYN
- ACK
- SYN/ACK
What is SYN Flood?

- **Mechanism of SYN Flood**

  The backlog queue is filled by malicious requests.
  Legitimate new connection requests are rejected.

  No ACK packets are replied.
  The connection requests remain in the backlog-queue till timeout.
Detection of SYN Flood

- **Problem**
  - The server cannot distinguish whether the receiving SYN packet is legitimate or malicious

- **Existing methods**
  - Detection by the mismatch of bi-directional packets
  - Detection by the mismatch between SYN and FIN

- **Remaining Issues**
  - Existing methods require long time to detect attacks
  - Existing methods may mistake high-rate normal traffic as attacks
The goal of our study

- **Objective**
  - Detecting attacks more accurately and faster
    - By using the statistical difference between normal and malicious traffic

- **What we have done**
  - Monitoring packets
  - Classification packets
  - Analyzing packets and modeling normal traffic
  - Making new mechanism to detect attacks using the model
How to monitor the traffic

- We monitored packets at the gateway of Osaka University
- We analyzed the monitored packets.
The captured traffic

- 5 days traffic from March 20, 2003 17:55
  - The number of TCP packets: 1.9 billion
  - The number of SYN packets: 21 million

![Graph showing packet arrival rate over time]

- Incoming packets
- Outgoing packets
Classification of flows

- We classified monitored packets into flow
  - Flow: a series of packets which have the same (src IP, src port, dest IP, dest port) field
- We classified the flows into groups
  - Normal traffic (85.1% of monitored traffic)
    - The flows which complete 3-way handshake
  - Incomplete traffic (14.9% of monitored)
    - The flows which do not complete 3-way handshake
Arrival rates of SYN packets

- Points where arrival rate rises sharply are due to incomplete traffic
- Arrival rate of the normal traffic changes over time
Arrival rates of SYN packets

- Points where arrival rate rises sharply are due to incomplete traffic
- Arrival rate of the normal traffic changes over time
Model of SYN arrival rate

- Arrival rate of normal traffic changes moderately
- We fit the arrival rate to a normal distribution
  - The normal distribution with the mean $\zeta$ and the variance $\sigma$ is
    \[
    F(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(y-\zeta)^2}{2\sigma^2}\right] dy
    \]
  - The arrival rate are positive value, and we only use the nonnegative part of the $F(x)$
    \[
    G(x) = \frac{F(x) - F(0)}{1 - F(0)}
    \]
The arrival rates of normal traffic can be modeled by normal distribution.
Because of many high-rate SYN packets caused by attack traffic, the distribution of SYN rates is far from a normal distribution.
How to detect attacks

- Attacks can be identified by observing the difference between SYN arrival rates and a normal distribution.

  - By calculating the average of squared difference

\[
D = \frac{\sum_{i=0}^{n} \left( G \left( r_i \right) - \frac{i}{n} \right)^2}{n}
\]

where:
- \( G(r_i) \) is the cumulative distribution function of the normal distribution,
- \( r_i \) is the sample SYN arrival rate,
- \( n \) is the number of samples.

The graph illustrates the cumulative distribution of SYN arrival rates, comparing a normal distribution (red line) with a sample distribution (green line). The difference is highlighted by arrows.
The averages of squared difference for the normal traffic are small regardless of time.

The averages of squared difference for all traffic rise rapidly at several points.
How to evaluate our method

- We wrote a program for our detecting method
- Two sample data for the trace-driven simulation
  - Monitored traffic
    - To confirm our method can detect monitored attacks
  - Traffic injected attack traffic
    - To know how small attacks can be detected
Definition of attacks

- **Attacks which must be detected**
  - Attacks which can put servers into denial-service state
  - Points where more than 1024 SYN packets are sent within 180 second
  - There are 10 attacks in our monitored traffic

- Probability of not detecting attacks = \( \frac{\text{number of attacks that cannot detect}}{\text{number of attacks satisfying the definition}} \)

- Probability of erroneous detection = \( \frac{\text{number of points erroneously detected as attacks}}{\text{number of points detected as attacks}} \)
Simulation of monitored traffic

- Smaller threshold, more erroneous detection
- Most attacks can be detected without erroneous detection

There are only two erroneous detection caused by a single client sending 20 SYN/sec.
Detectable attack rate

- We injected low-rate attack into the traced traffic
- Attacks whose rates are more than 14 SYNs/sec can be detected without erroneous detection
Time to detect attacks

- Averages of squared difference increase gradually after the beginning of attacks.
- Our proposed method can detect faster.
  - One of reasons is because our method adopts a parametric approach.

![Graph showing time to detect attacks and average of squared difference](image)
Summary and future work

- **Summary**
  - We monitored and analyzed packets at the gateway of Osaka University
  - We fit the arrival rate of SYN packets to a normal distribution
  - We can detect attacks by the difference between arrival rates of SYN packets and a normal distribution

- **Future work**
  - Setting the parameters
  - Modeling other types of traffic