INTERNSHIP REPORT

UNDER GUIDANCE OF

DR.ARUN RAJ KUMAR P.

AT CSED,NITC

BY MUMMANA SANJAY

BTECH
MAJOR - MECHANICAL ENGINEERING [NITK]
MINOR - COMPUTER SCIENCE [NITK]

TASK 4 ASSIGNED ON 6TH JUNE 2024

Task Overview:

Using Multi-Agent Reinforcement Learning (MARL) and Information Gain to Classify TCP SYN Flood Attacks.

Marl:

Multi-Agent Reinforcement Learning (MARL) is a subfield of reinforcement learning (RL) where multiple agents interact within an environment. Each agent learns to optimize its own policy based on the rewards it receives, which may depend on the actions of other agents. MARL can be used in cooperative, competitive, or mixed settings.

Information Gain:

Information Gain is a concept from information theory that measures the reduction in uncertainty about a random variable after observing another random variable. In the context of classification, information gain is often used to select features that provide the most significant reduction in uncertainty about the class labels.

TCP SYN Flood Attack:

A **TCP SYN Flood attack** is a type of Denial-of-Service (DoS) attack where the attacker sends a succession of SYN requests to a target's server in an attempt to consume enough server resources to make the system unresponsive to legitimate traffic.

Combining MARL and Information Gain for TCP SYN Flood Attack Classification:

Theoretical Framework:

1. **Agents and Environment**:
    - **Agents**: Multiple agents can represent different network nodes or components (e.g., routers, firewalls, intrusion detection systems).
    - **Environment**: The network under attack, where agents observe traffic patterns and make decisions.
2. **Observations and Actions**:
    - Each agent observes network traffic features such as packet arrival times, SYN packet ratios, IP addresses, etc.
    - Actions may include flagging traffic as normal or malicious, adjusting firewall rules, or reporting suspicious activity.
3. **Reward Structure**:
    - Agents receive rewards based on their accuracy in classifying traffic, reducing false positives and false negatives.
    - Penalties for misclassification to encourage accurate learning.

4. **Information Gain for Feature Selection**:
    - Agents use information gain to select features that most effectively distinguish between normal and attack traffic.
    - This helps in focusing on the most relevant features, reducing the dimensionality of the data, and improving learning efficiency.

MARL Algorithm:

1. **Initialization**: Initialize agents, each with a policy for classifying traffic.
2. **Observation**: Agents collect data from the network, observing traffic features.
3. **Feature Selection**: Calculate information gain for each feature and select the top features.
4. **Action Selection**: Based on selected features, agents choose actions (classify traffic).
5. **Reward Computation**: Calculate rewards based on the accuracy of the classification.
6. **Policy Update**: Agents update their policies using reinforcement learning algorithms (e.g., Q-learning, policy gradients) based on the rewards received.
7. **Iteration**: Repeat the observation-action-reward-update cycle.

Advantages of Using MARL and Information Gain for TCP SYN Flood Classification:

1. **Scalability**:
    - MARL can handle large-scale networks with multiple nodes working together to detect attacks.
    - Distributed nature allows for parallel processing and faster response times.
2. **Adaptive Learning**:
    - Agents continuously learn and adapt to new attack patterns.
    - MARL enables dynamic policy updates based on real-time feedback.
3. **Improved Accuracy**:
    - Information gain ensures that the most informative features are used for classification.
    - Collaborative efforts of multiple agents lead to more accurate detection.
4. **Resilience**:
    - The distributed approach increases the system's resilience to individual node failures.
    - Multiple agents can compensate for the misclassification of any single agent.
5. **Reduced False Positives/Negatives**:
    - Agents learn to optimize their actions to minimize incorrect classifications.
    - Information gain helps in focusing on the most distinguishing features, reducing errors.
6. **Enhanced Decision Making**:
    - Each agent contributes to the overall decision-making process, combining multiple perspectives.
    - Agents can share information and strategies, leading to a more comprehensive understanding of the traffic patterns.

Conclusion:

Combining MARL and information gain provides a robust theoretical framework for classifying TCP SYN Flood attacks. The agents work together to learn and adapt to new threats, leveraging the most informative features for accurate and efficient classification. This approach offers significant advantages in terms of scalability, accuracy, resilience, and adaptive learning, making it a promising solution for network security.

-----THANK YOU-----