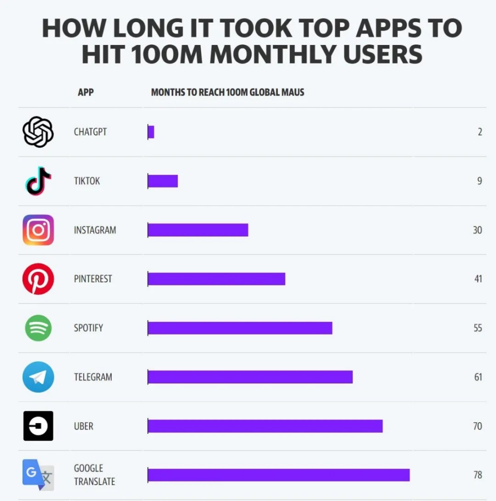
Large Language Model (LLM) for Networking: Architecture and Practice

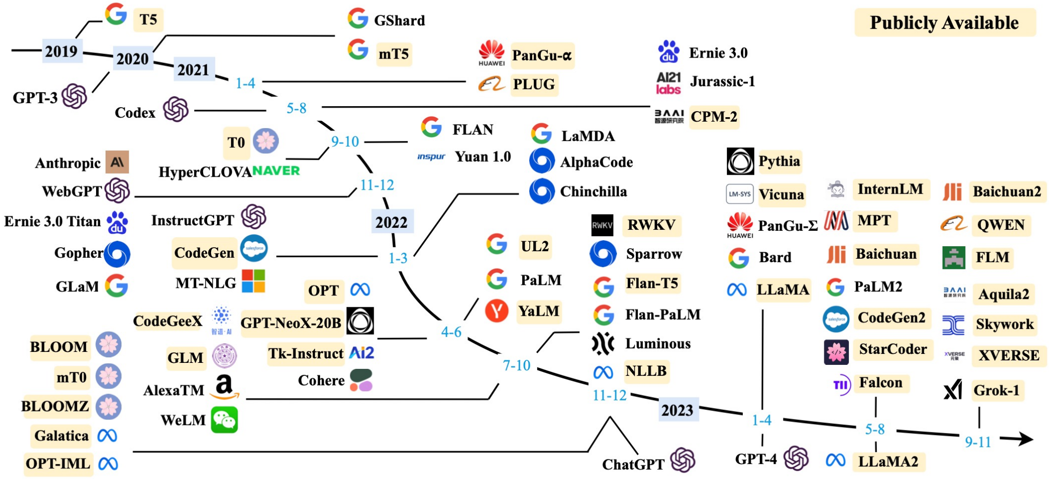
IETF 119 @ Brisbane

Xiaohui Xie
Tsinghua University, China
xiexiaohui@tsinghua.edu.cn

Background

- The emergence of ChatGPT has marked the beginning of a rapid development era for the large language model (LLM) and the generative AI



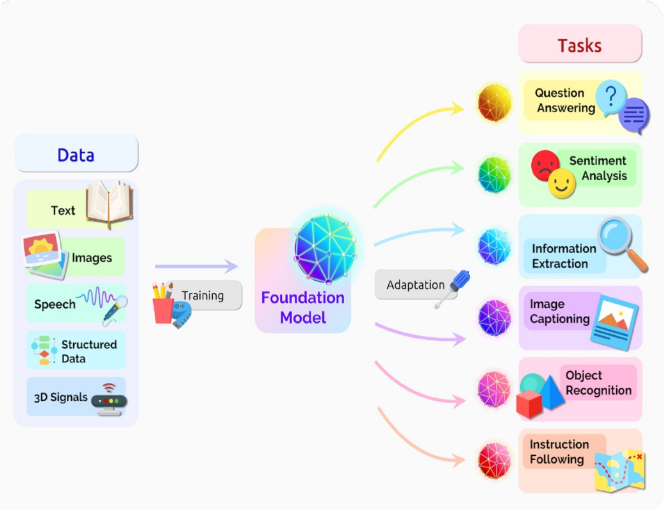


[1] finance.yahoo. ChatGPT on track to surpass 100 million users faster than TikTok or Instagram

[2] Wayne Xin Zhao et al. A Survey of Large Language Models. Arxiv 2023

Background

- LLMs show remarkable capabilities in concept understanding, mathematical reasoning, physical principle (maybe, see Sora) and tool usage

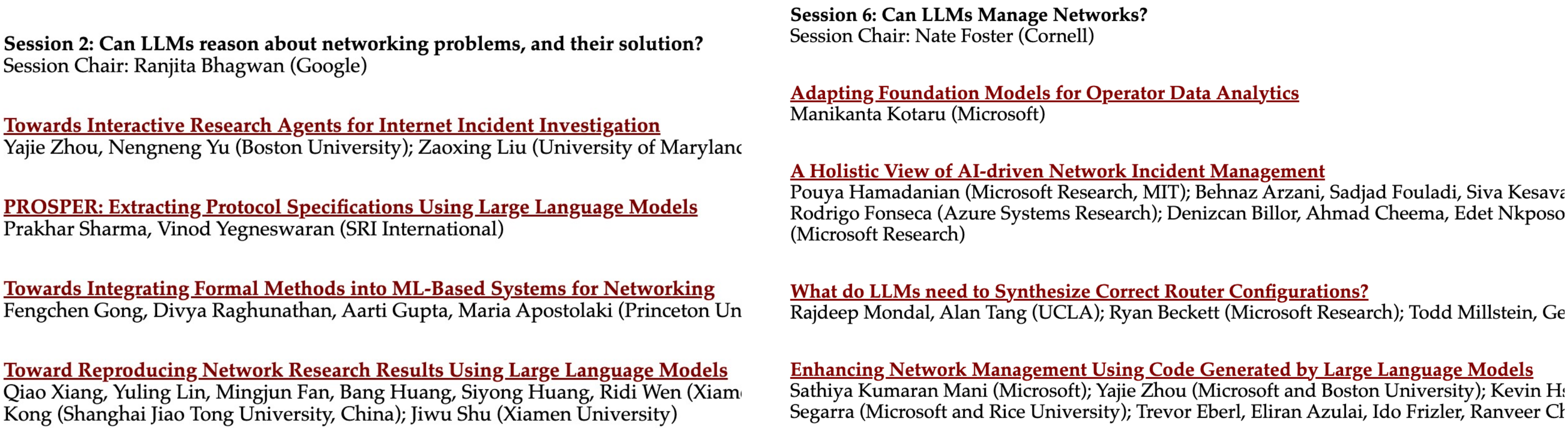


Prompt: A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket...

[1] https://openai.com/sora

Background

- The application of LLMs in the networking field is receiving increasing attention



Session 2: Can LLMs reason about networking problems, and their solution?
Session Chair: Ranjita Bhagwan (Google)

Towards Interactive Research Agents for Internet Incident Investigation
Yajie Zhou, Nengneng Yu (Boston University); Zaoxing Liu (University of Marylanc

PROSPER: Extracting Protocol Specifications Using Large Language Models
Prakhar Sharma, Vinod Yegneswaran (SRI International)

Towards Integrating Formal Methods into ML-Based Systems for Networking
Fengchen Gong, Divya Raghunathan, Aarti Gupta, Maria Apostolaki (Princeton Un

Toward Reproducing Network Research Results Using Large Language Models
Qiao Xiang, Yuling Lin, Mingjun Fan, Bang Huang, Siyong Huang, Ridi Wen (Xiam
Kong (Shanghai Jiao Tong University, China); Jiwu Shu (Xiamen University)

Session 6: Can LLMs Manage Networks?
Session Chair: Nate Foster (Cornell)

Adapting Foundation Models for Operator Data Analytics
Manikanta Kotaru (Microsoft)

A Holistic View of AI-driven Network Incident Management
Pouya Hamadanian (Microsoft Research, MIT); Behnaz Arzani, Sadjad Fouladi, Siva Kesava
Rodrigo Fonseca (Azure Systems Research); Denizcan Billor, Ahmad Cheema, Edet Nkposo
(Microsoft Research)

What do LLMs need to Synthesize Correct Router Configurations?
Rajdeep Mondal, Alan Tang (UCLA); Ryan Beckett (Microsoft Research); Todd Millstein, Ge

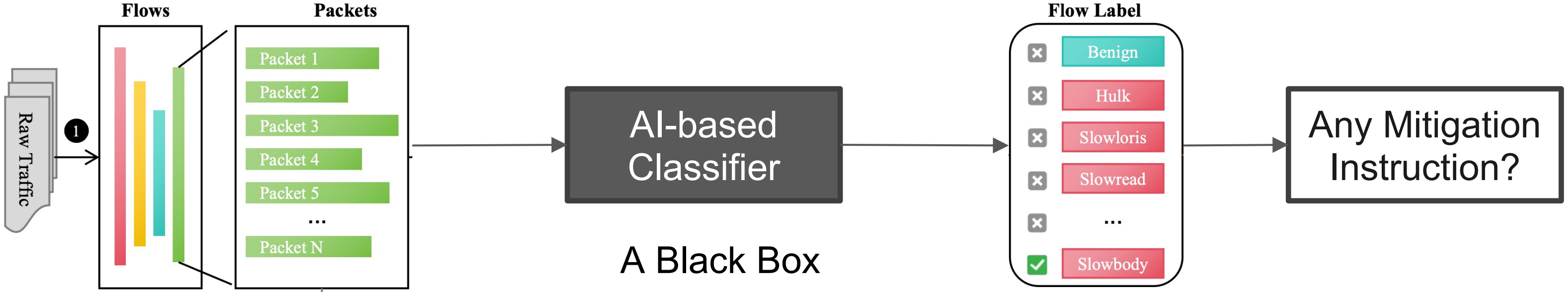
Enhancing Network Management Using Code Generated by Large Language Models
Sathiya Kumaran Mani (Microsoft); Yajie Zhou (Microsoft and Boston University); Kevin H
Segarra (Microsoft and Rice University); Trevor Eberl, Eliran Azulai, Ido Frizler, Ranveer Cl

Related Sessions @ HotNets 2023

[1] https://conferences.sigcomm.org/hotnets/2023/program.html

ShieldGPT: An LLM-based Framework for DDoS Mitigation

- The constantly evolving Distributed Denial of Service (DDoS) attacks pose a significant threat to the cyber security
- Existing AI-driven methods achieve impressive performance on DDoS detection, but two limitations hinder them from the more practical application
  - Lack of traffic-dependent explanations of detection results
  - Lack of actionable instructions for mitigation



AI-based Classifier

A Black Box

Any Mitigation Instruction?

ShieldGPT: An LLM-based Framework for DDoS Mitigation

- Challenge 1: To **represent heterogeneous information** in network scenarios, such as real-time binary traffic data and static domain-specific textual information, in a way that LLMs can understand.

- Challenge 2: To inform the LLMs of its **role for specific tasks** in preventing hallucination issues and producing the desired outcomes.

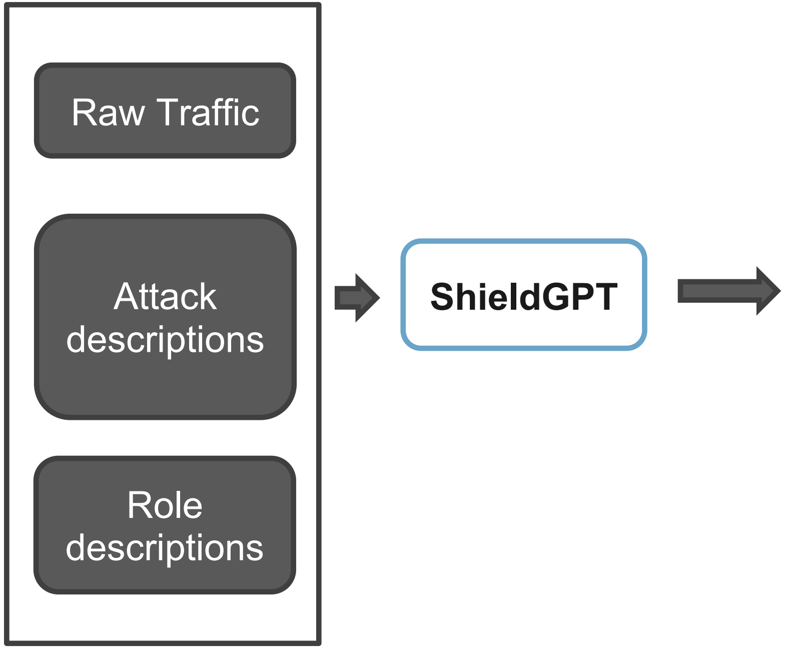
ShieldGPT: An LLM-based Framework for DDoS Mitigation

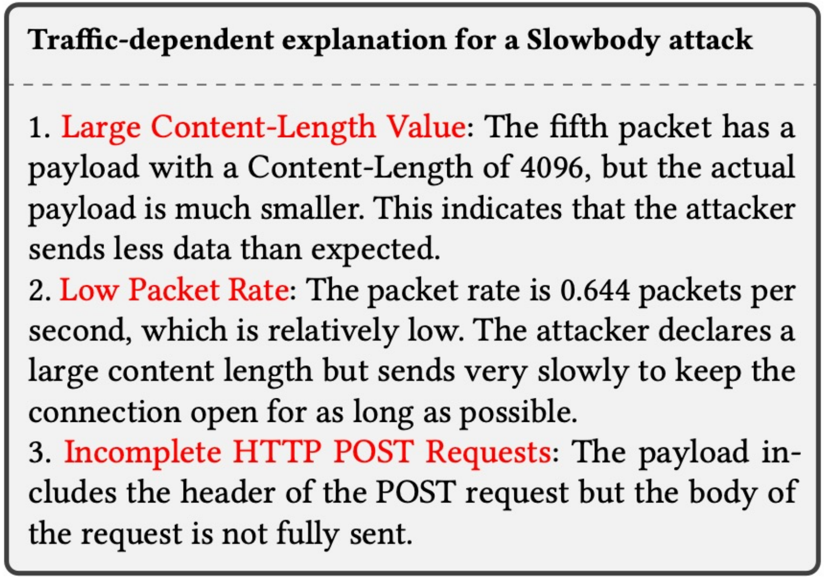
- Challenge 1: To **represent heterogeneous information** in network scenarios, such as real-time binary traffic data and static domain-specific textual information, in a way that LLMs can understand. => **Traffic representation**

- Challenge 2: To inform the LLMs of its **role for specific tasks** in preventing hallucination issues and producing the desired outcomes. => **Role representation**

ShieldGPT: An LLM-based Framework for DDoS Mitigation

- ShieldGPT can generate traffic-dependent, in-depth attack explanation



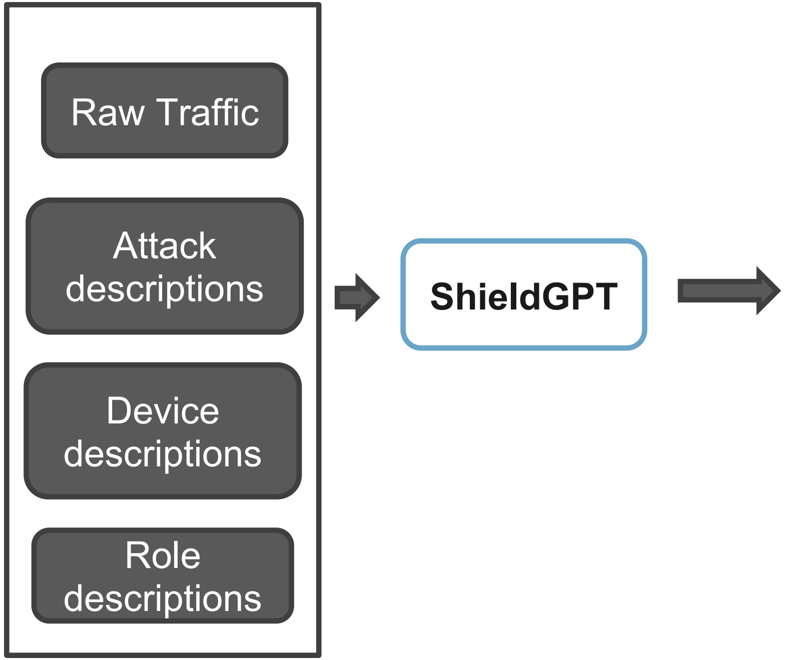


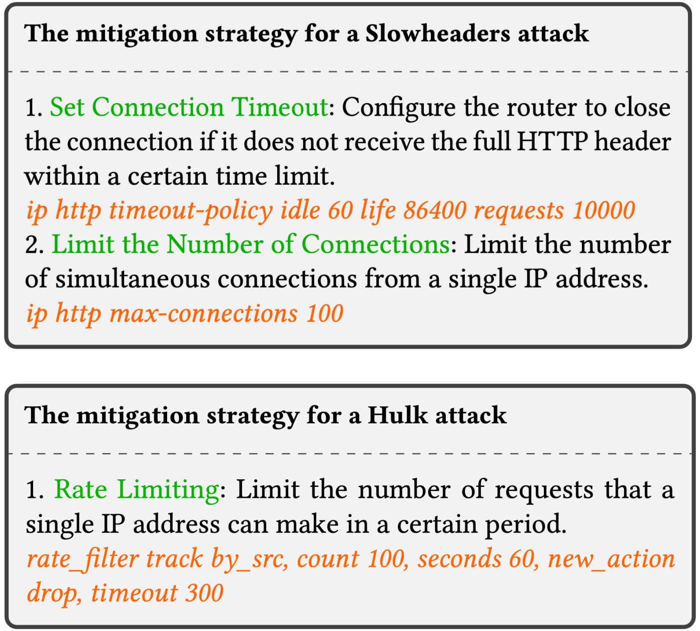
Traffic-dependent explanation for a Slowbody attack

1. Large Content-Length Value: The fifth packet has a payload with a Content-Length of 4096, but the actual payload is much smaller. This indicates that the attacker sends less data than expected.
2. Low Packet Rate: The packet rate is 0.644 packets per second, which is relatively low. The attacker declares a large content length but sends very slowly to keep the connection open for as long as possible.
3. Incomplete HTTP POST Requests: The payload includes the header of the POST request but the body of the request is not fully sent.

ShieldGPT: An LLM-based Framework for DDoS Mitigation

- ShieldGPT can generate actionable mitigation strategies





The mitigation strategy for a Slowheaders attack

1. Set Connection Timeout: Configure the router to close the connection if it does not receive the full HTTP header within a certain time limit.
ip http timeout-policy idle 60 life 86400 requests 10000
2. Limit the Number of Connections: Limit the number of simultaneous connections from a single IP address.
ip http max-connections 100

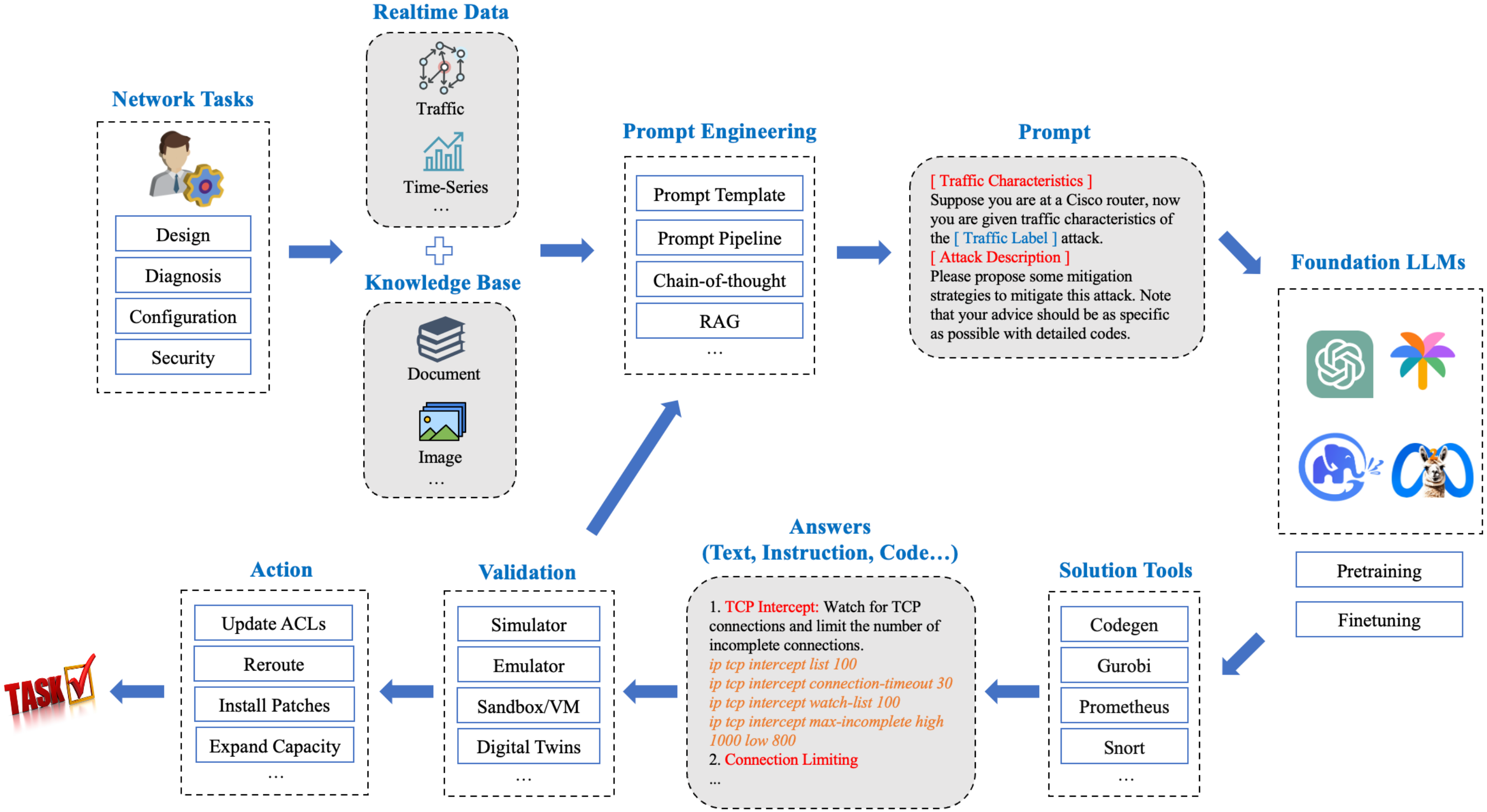
The mitigation strategy for a Hulk attack

1. Rate Limiting: Limit the number of requests that a single IP address can make in a certain period.
rate_filter track by_src, count 100, seconds 60, new_action drop, timeout 300

Future Research

- **Safety.** Establishing a robust validation mechanism is critical for future research to ensure the reliability and safety of automated mitigation strategies
- **Automatic Execution**. Enabling the generated mitigation strategies to be automatically executed. (API, mature technical stacks, support from industry)
- **Broader Applications**. Our approach can be generalized to other network tasks, such as generating diagnostic analysis in network diagnosis or generating control commands in network management

LLM-in-the-loop Architecture for Networking



Network Tasks

- Design
- Diagnosis
- Configuration
- Security

Realtime Data

Traffic

Time-Series

…

Knowledge Base

Document

Image

…

Prompt Engineering

- Prompt Template
- Prompt Pipeline
- Chain-of-thought
- RAG

…

Prompt

[Traffic Characteristics]
Suppose you are at a Cisco router, now you are given traffic characteristics of the [Traffic Label] attack.
[Attack Description]
Please propose some mitigation strategies to mitigate this attack. Note that your advice should be as specific as possible with detailed codes.

Foundation LLMs

- Pretraining
- Finetuning

Solution Tools

- Codegen
- Gurobi
- Prometheus
- Snort

…

Answers (Text, Instruction, Code…)

1. TCP Intercept: Watch for TCP connections and limit the number of incomplete connections.

```
ip tcp intercept list 100
ip tcp intercept connection-timeout 30
ip tcp intercept watch-list 100
ip tcp intercept max-incomplete high 1000 low 800
```

2. Connection Limiting

…

Validation

- Simulator
- Emulator
- Sandbox/VM
- Digital Twins

…

Action

- Update ACLs
- Reroute
- Install Patches
- Expand Capacity

…

Side Meeting @IETF 119

- Topic: Large Language Model (LLM) for Networking
- Time and Location: 16:00-17:30 (March 20, Wednesday) @ Room P6-7
- Host: Yong Cui (Tsinghua University)
- Agenda (Each talk will last 15 minutes)
  - Opening
  - Talk 1: “LLM for Networking: an overview” by Xiaohui Xie (Tsinghua University)
  - Talk 2: “Using Machine Learning and Word Embedding to Characterise the DDoS landscape with DDoS2Vec” by Marinho Barcellos (University of Waikato)
  - Talk 3: “Thinking and Practice: LLM for Cybersecurity” by Linzhe Li (Zhongguancun Lab)
  - Talk 4: “Usecases of AI for Network” by Xiaoqiu Zhang (China Mobile)
  - Free Discussion