NetAssistant: Dialogue Based Network Diagnosis in Data Center Networks

Haopei Wang, Anubhavnidhi Abhashkumar, Changyu Lin, Tianrong Zhang, Xiaoming Gu, Ning Ma, Chang Wu, Songlin Liu, Wei Zhou, Yongbin Dong, Weirong Jiang, Yi Wang

Network Engineering Team

usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

nsdi '24

ByteDance 字节跳动

Motivation

NetAssistant Design

Deployment & Gain

Lessons & Future Work

Motivation

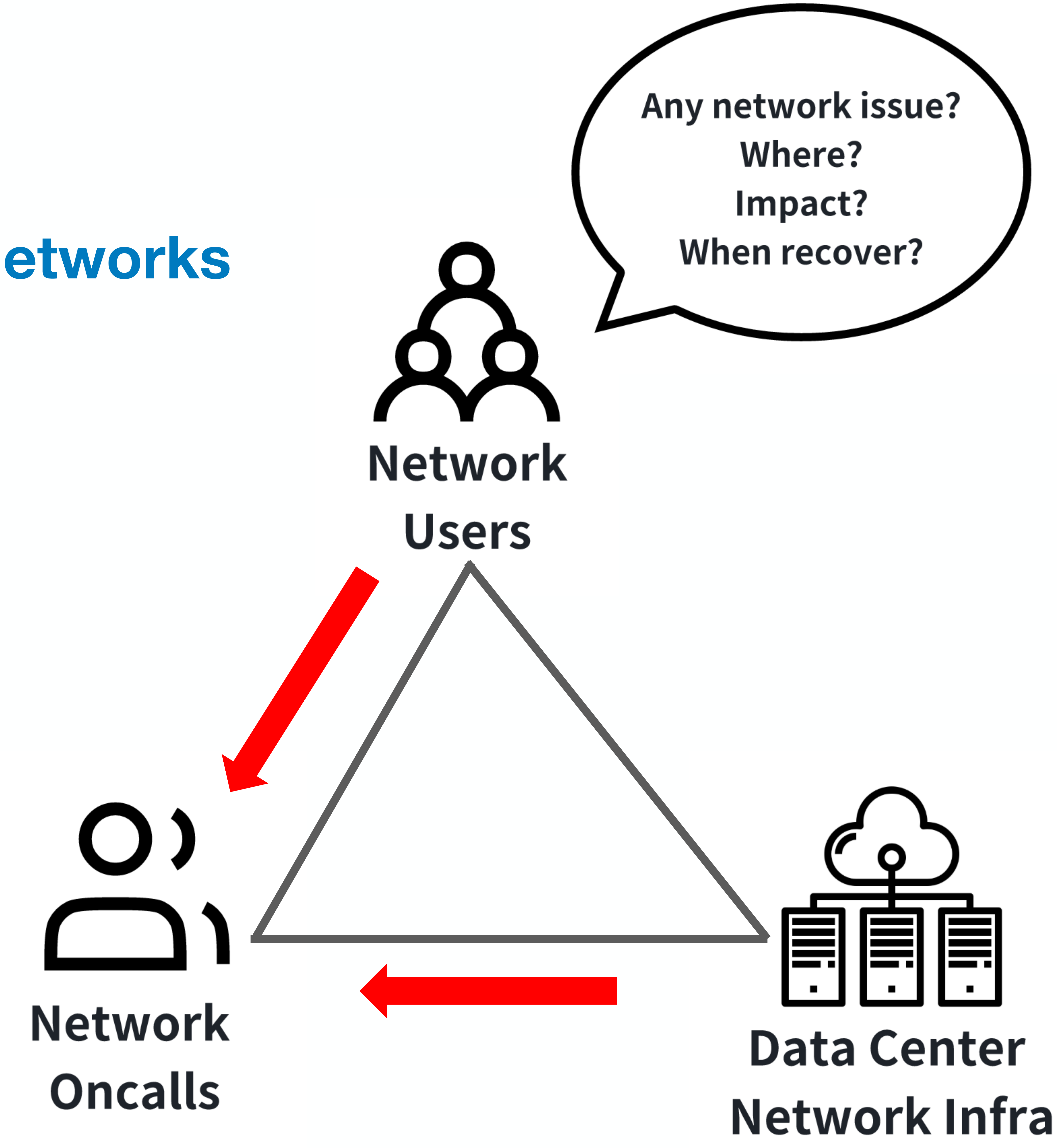
High visibility is important in datacenter networks

Especially for network users

Network oncalls handler requests/tickets

Time consuming and labor intensive
Monitoring data, incidents records

There is a gap!



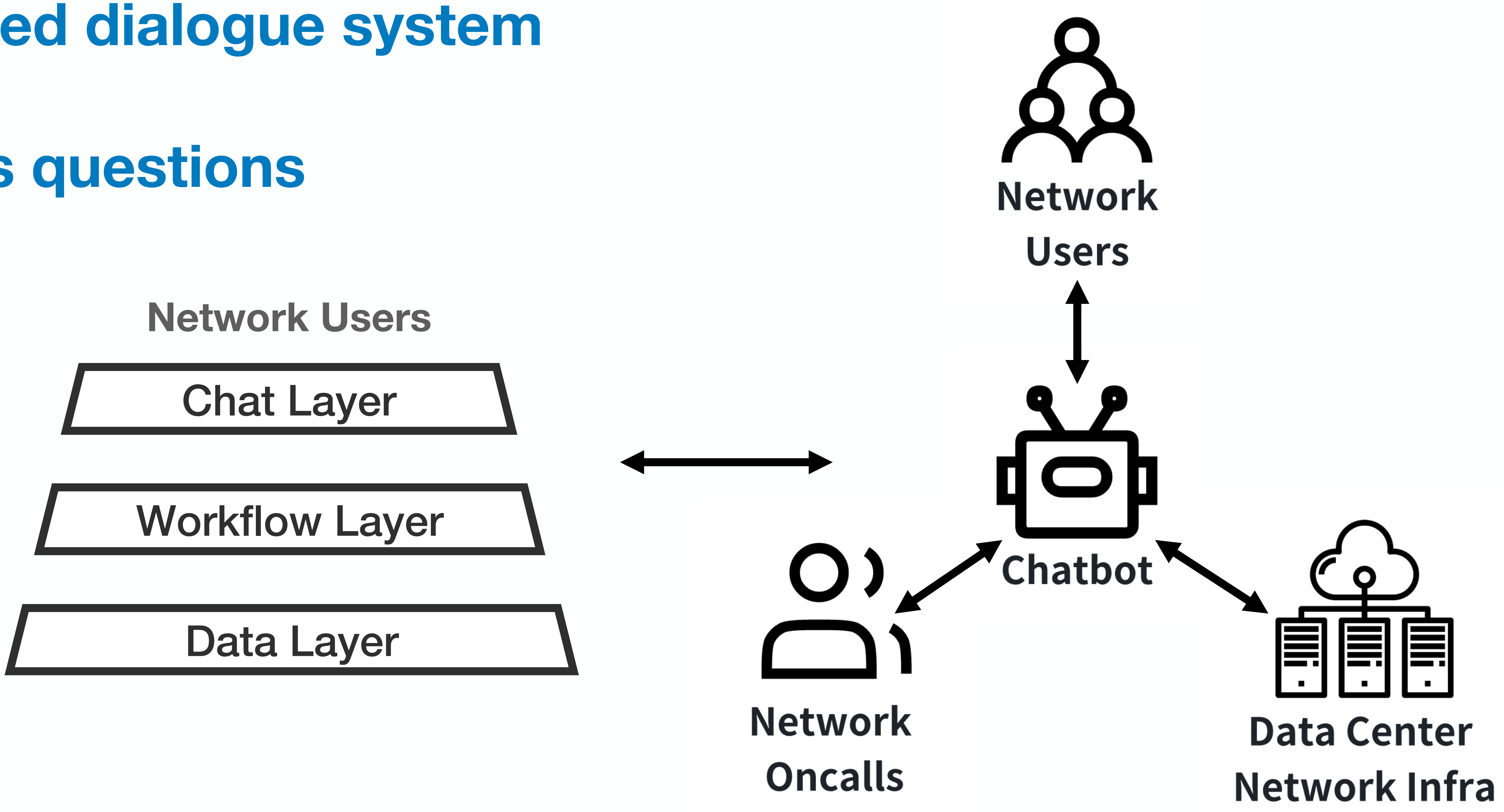
NetAssistant Design

Our Idea: Leverage a task-oriented dialogue system

Automatically answers diagnosis questions

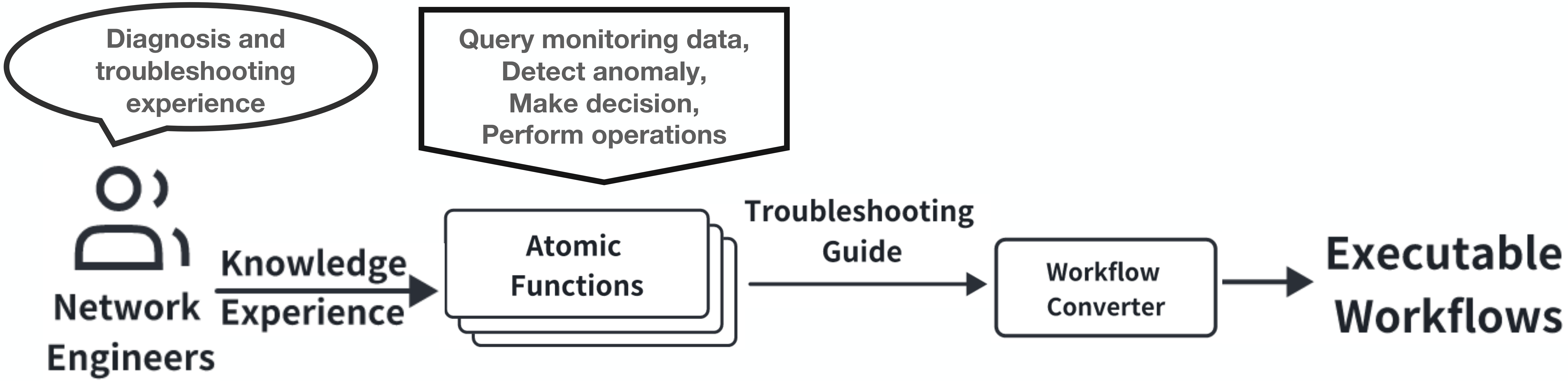
Three layers of abstraction

1. Chat service
   - Dialogue Engine
2. Diagnosis workflows
   - Workflow Engine
3. Retrieval of monitoring data
   - Data Engine



NetAssistant Design

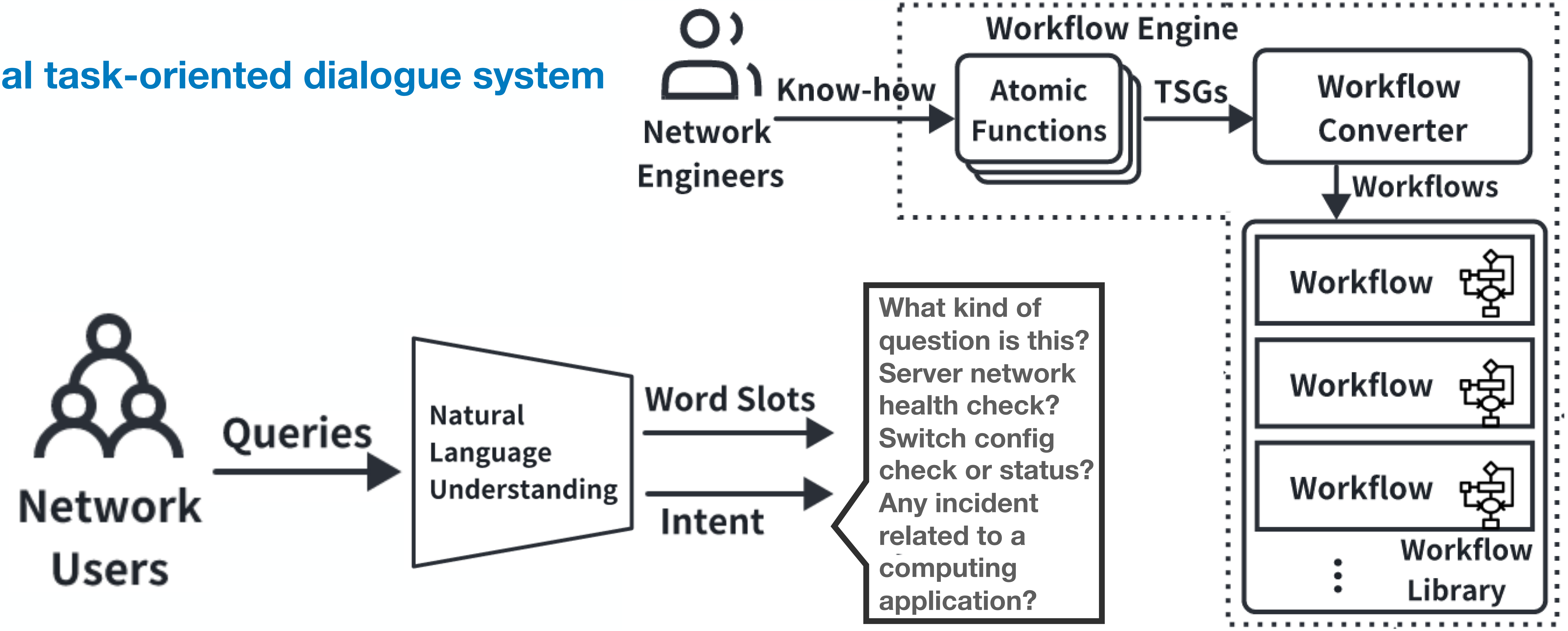
Workflow Engine:



NetAssistant Design

Dialogue Engine:

A typical task-oriented dialogue system



What kind of question is this? Server network health check? Switch config check or status? Any incident related to a computing application?

NetAssistant Design

Data Engine:

Performance bottleneck

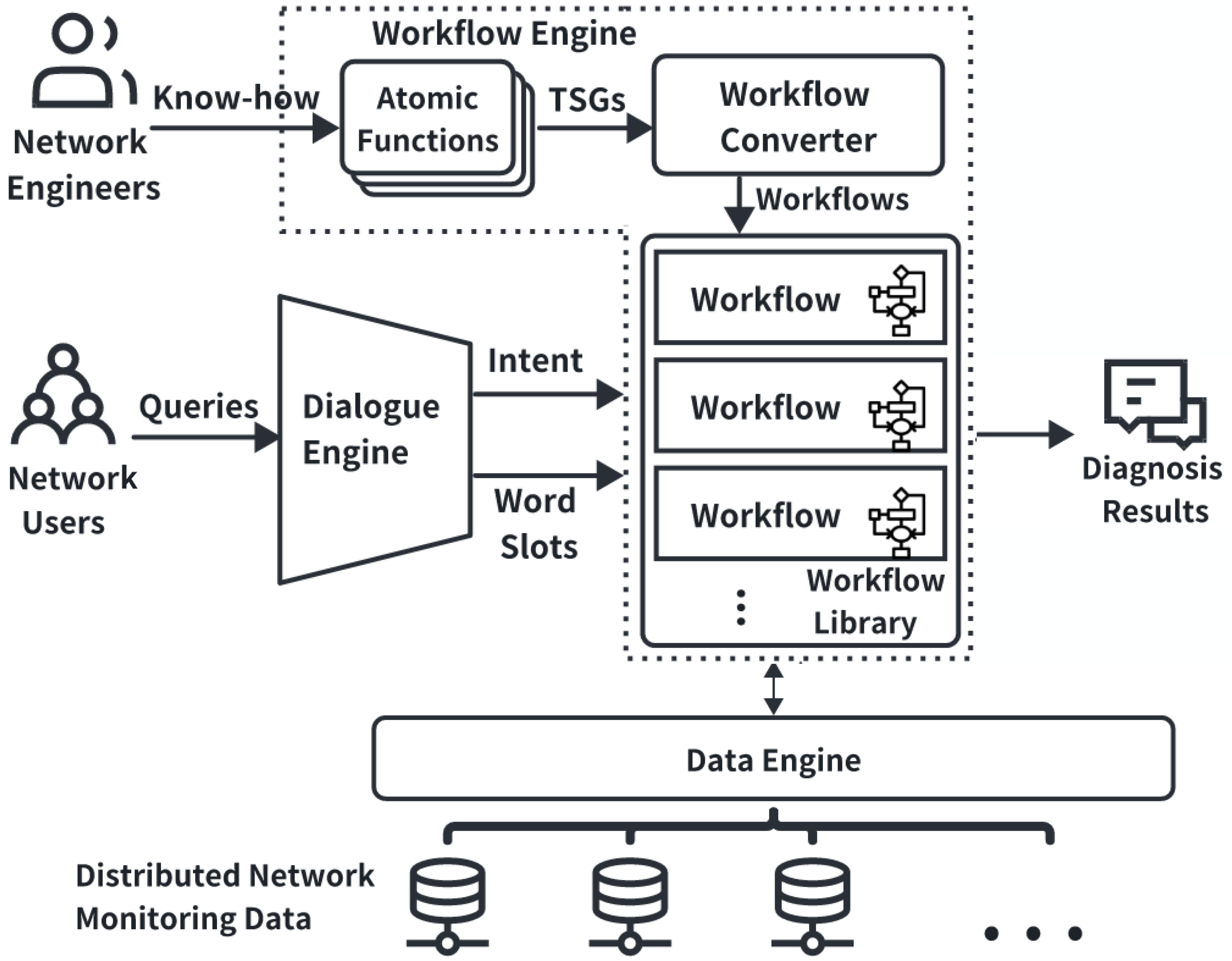
- Query monitoring data
- E.g., sFlow for 100 links
- E.g., syslog for 1000 switches

Our Idea:

Only anomalies/jitters are important

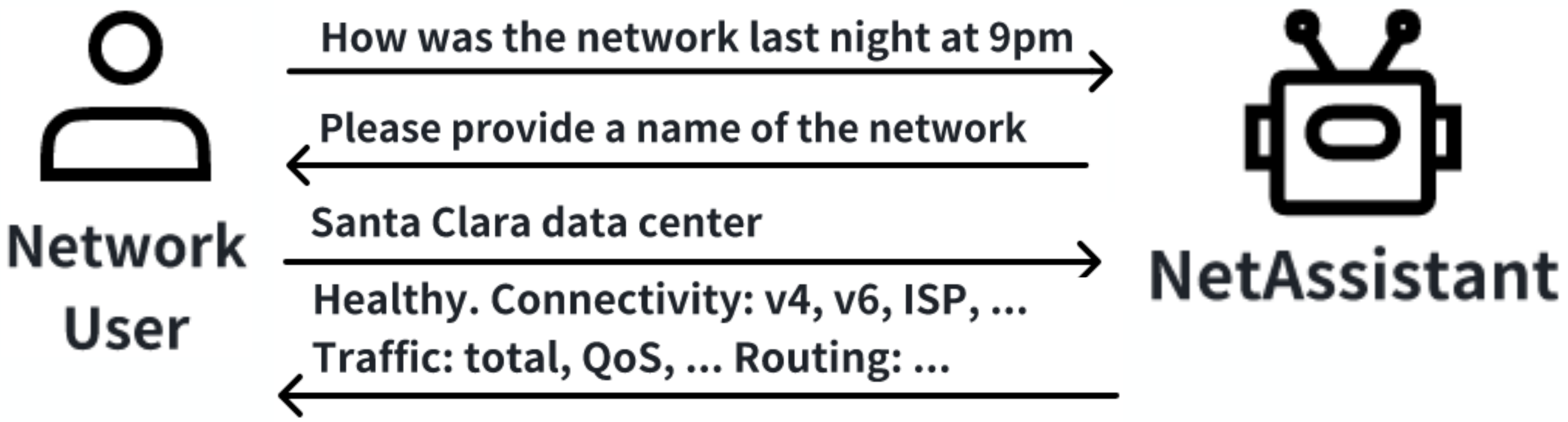
We combine:

- Reactive/on-demand querying
- Proactive alerting



NetAssistant Design

Sample dialogues between NetAssistant and the users:



Network User
How was the network last night at 9pm
Please provide a name of the network
Santa Clara data center
Healthy. Connectivity: v4, v6, ISP, ...
Traffic: total, QoS, ... Routing: ...
NetAssistant

Deployment & Gain

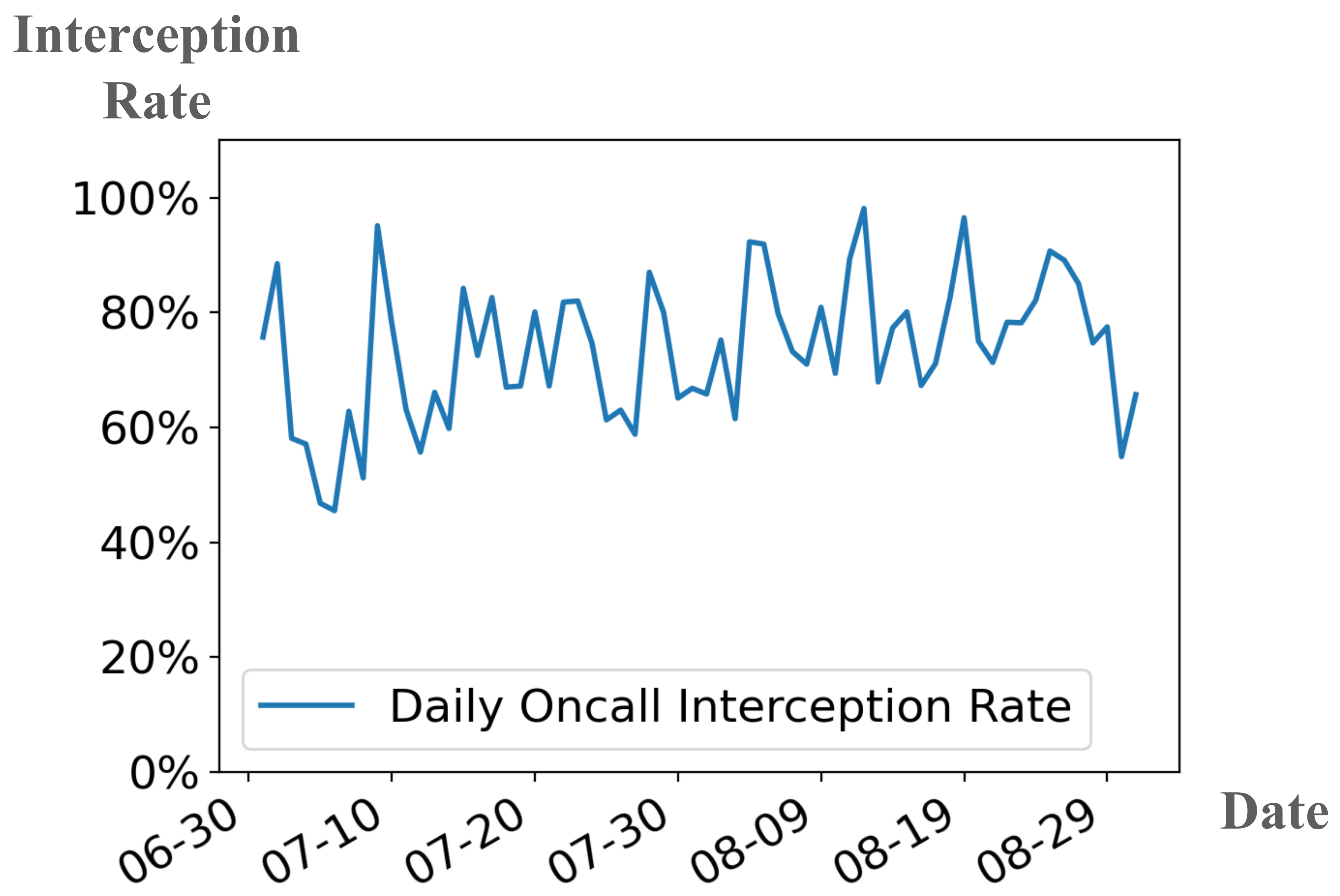
First version was launched in April 2020

- Starting with only 2 workflows, now 100+
- Iterate on technology and functionality every few weeks
- Now ~200 uses per day

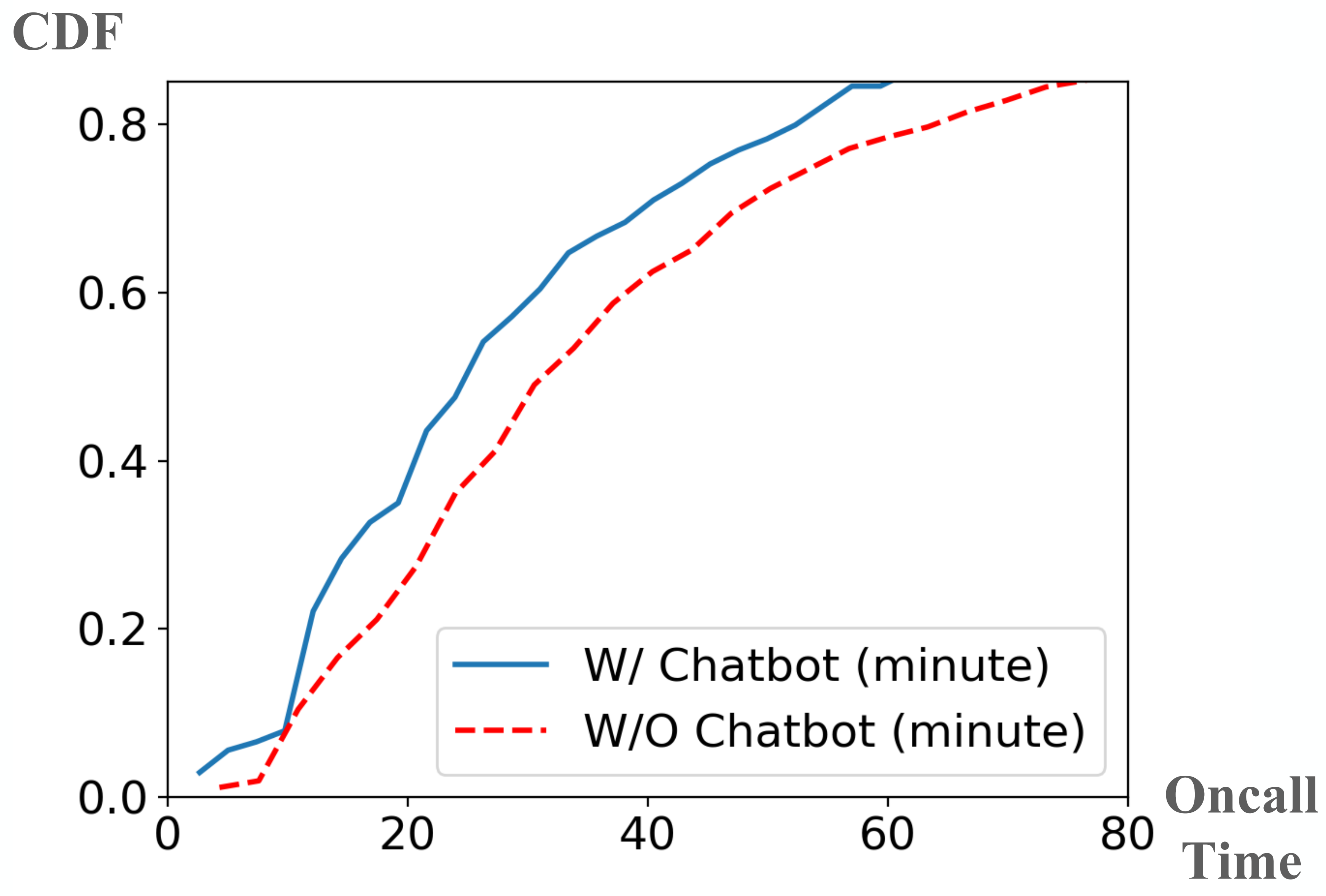
What is the gain of this project?

- Save human labor time
- Directly intercept oncalls
- Reduce oncall duration time

Deployment & Gain



Intercepted a considerable proportion of oncalls in 2023



Reduce oncall duration time

ByteDance字节跳动

Lessons & Future Work

Lessons:

- Earn user trust
- FN is more harmful than FP
- Empowering our users

Future work:

- Explore the potential of LLM in AIOps
- Challenges:

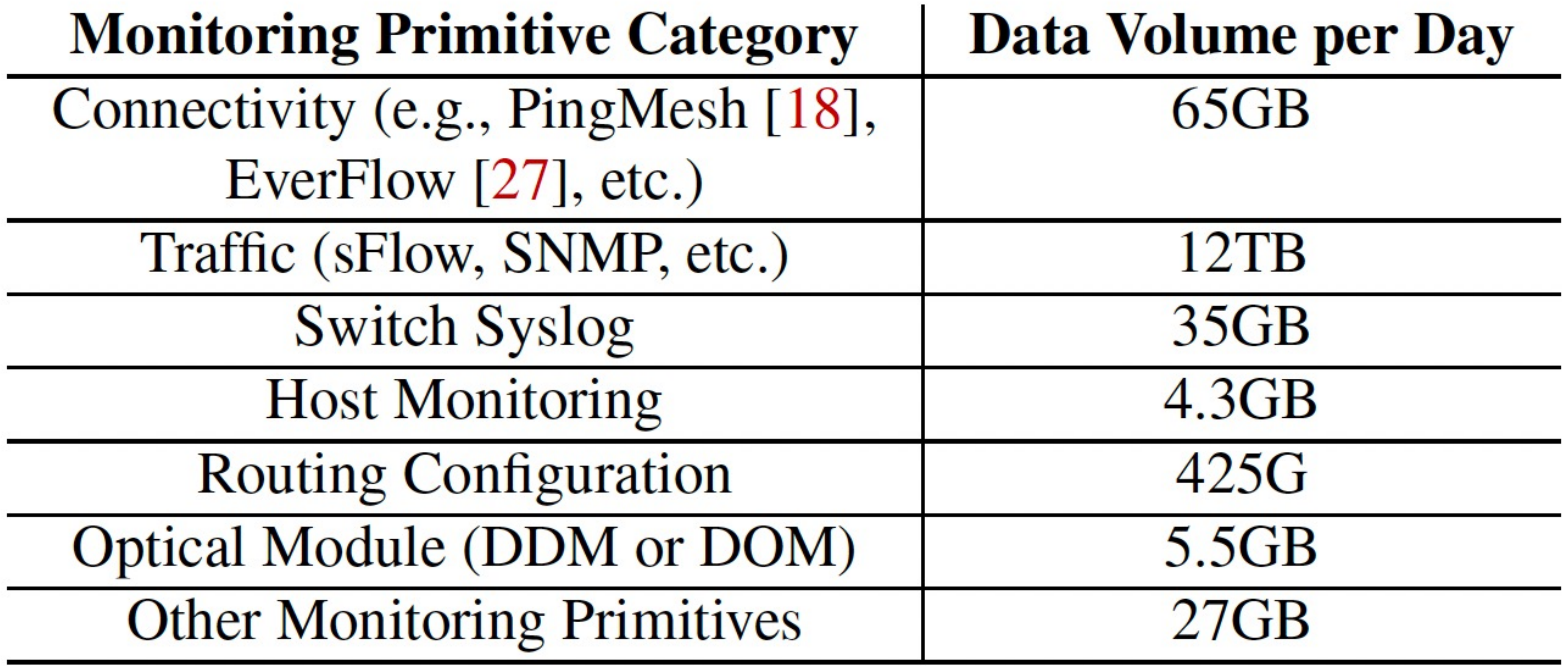
  Understanding diagnosis logic

  Processing real time monitoring data

THANKS

ByteDance 字节跳动

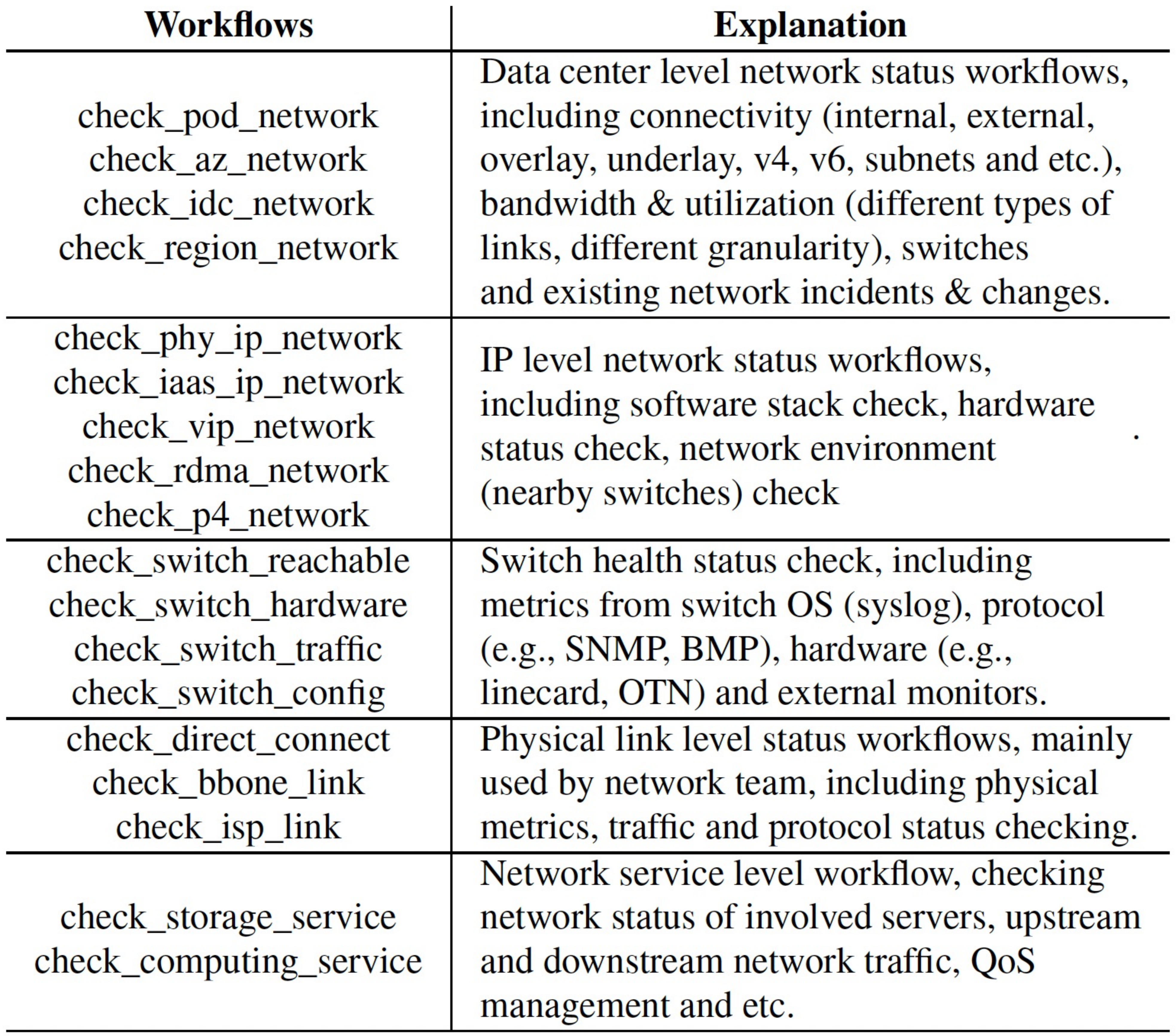
Appendix



Monitoring Primitive Category	Data Volume per Day
Connectivity (e.g., PingMesh [18], EverFlow [27], etc.)	65GB
Traffic (sFlow, SNMP, etc.)	12TB
Switch Syslog	35GB
Host Monitoring	4.3GB
Routing Configuration	425G
Optical Module (DDM or DOM)	5.5GB
Other Monitoring Primitives	27GB

Table 3: Monitoring primitives and data volume

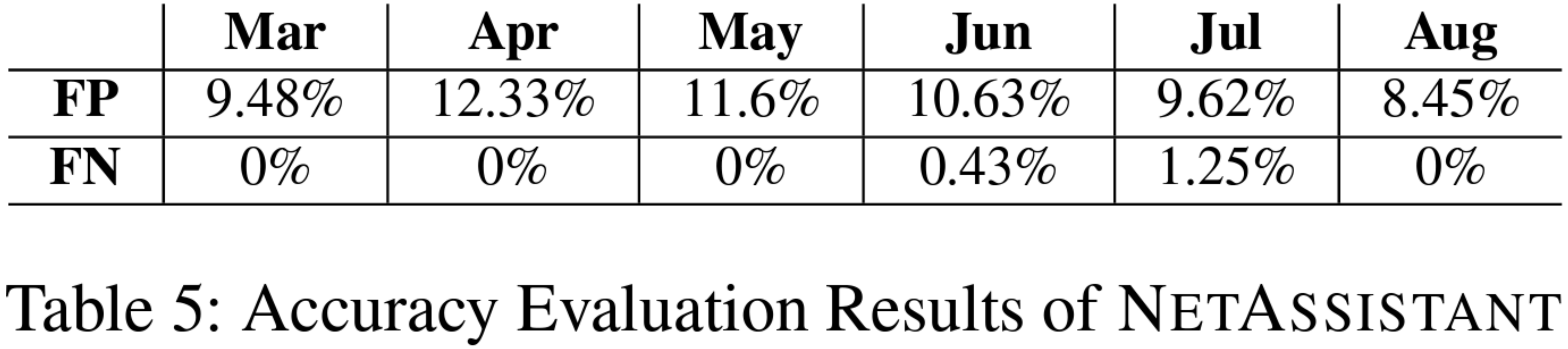
Appendix



Workflows	Explanation
check_pod_network check_az_network check_idc_network check_region_network	Data center level network status workflows, including connectivity (internal, external, overlay, underlay, v4, v6, subnets and etc.), bandwidth & utilization (different types of links, different granularity), switches and existing network incidents & changes.
check_phy_ip_network check_iaas_ip_network check_vip_network check_rdma_network check_p4_network	IP level network status workflows, including software stack check, hardware status check, network environment (nearby switches) check
check_switch_reachable check_switch_hardware check_switch_traffic check_switch_config	Switch health status check, including metrics from switch OS (syslog), protocol (e.g., SNMP, BMP), hardware (e.g., linecard, OTN) and external monitors.
check_direct_connect check_bbone_link check_isp_link	Physical link level status workflows, mainly used by network team, including physical metrics, traffic and protocol status checking.
check_storage_service check_computing_service	Network service level workflow, checking network status of involved servers, upstream and downstream network traffic, QoS management and etc.

Table 4: Commonly used workflows

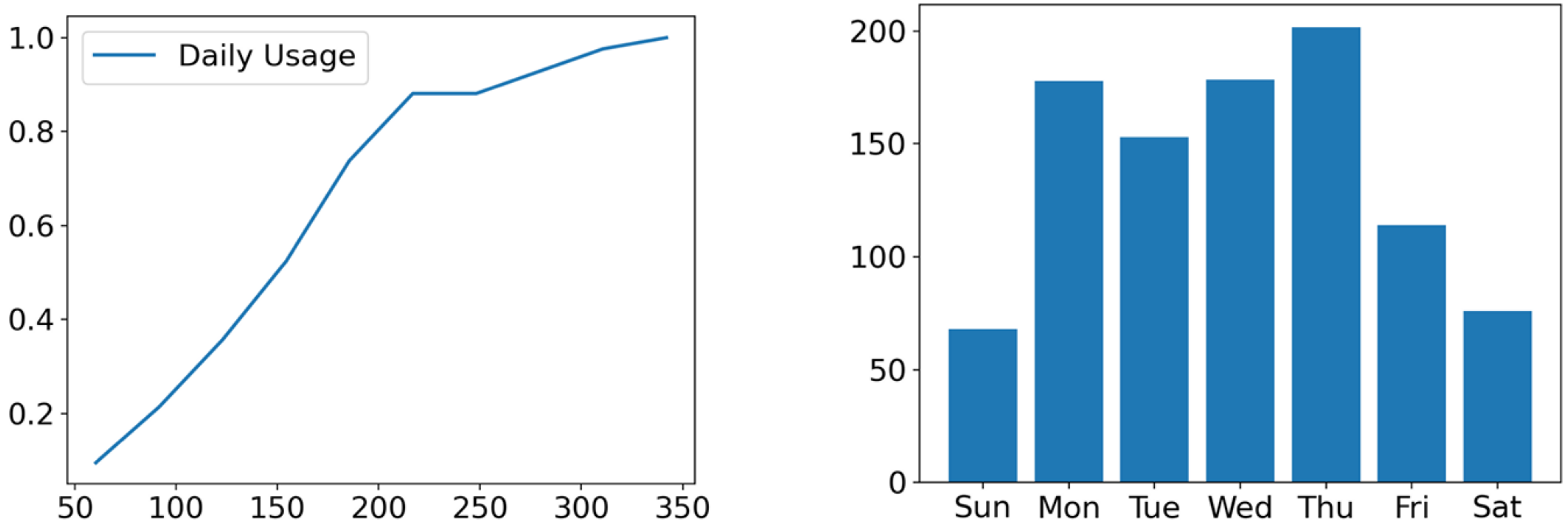
Appendix



	Mar	Apr	May	Jun	Jul	Aug
FP	9.48%	12.33%	11.6%	10.63%	9.62%	8.45%
FN	0%	0%	0%	0.43%	1.25%	0%

Table 5: Accuracy Evaluation Results of NETASSISTANT

Appendix



(a) Daily Oncall Usage for July and August in 2023 (CDF)

(b) The 50th Percentile Usage by Day of the Week

Figure 8: Daily Usage Results of NETASSISTANT