White Paper

Communication Service Providers
AI

ZTE's High Performance 5G UPF DPI Implementation Based on Intel® TADK

Deep Packet Inspection (DPI) is a key component of the 5G core network user plane function (UPF) used to monitor clear-text, network traffic characteristics. DPI identifies protocols and applications associated with customer data to facilitate accurate billing and quality of service.

Communication service providers are encrypting more network traffic which makes traditional DPI methods more difficult. Artificial intelligence and machine learning offer a new way to inspect encrypted data in-flight by learning topological characteristics of network traffic and building models to classify it.

ZTE implemented DPI with AI parsing and reasoning capabilities to inspect and analyze encrypted packets. The solution leveraged the latest 3rd Gen Intel® Xeon® Scalable processors with built-in AI acceleration, instructions, and optimized AI frameworks plus the Intel® traffic analysis development kit (TADK).

Abstract

Data centers and networks must evolve to keep pace with the exponential growth of data, rapid expansion of cloud-scale computing and 5G networks, and convergence of big data processing and artificial intelligence (AI). These demands are driving a new architecture for the modern data center and network of the future, enabling it to grow rapidly and scale.

This paper describes the functional evaluation and performance optimization exploration of ZTE's 5G UPF DPI module based on the Intel Traffic Analysis Development Kit (TADK) using 3rd Gen Intel® Xeon® Scalable processors.

This test was completed by ZTE on 2022-02-25. For specific test configuration, please see the chapter on system test environment. For more complete information on performance and testing, visit www.intel.com/benchmarks, and refer to http://software.intel.com/en-us/articles/optimization-notice for more information on Intel software product performance and Information on optimization solutions.

Table of Contents

1. ZTE 5GC DPI solution

1.1. 5G Core DPI Overview

Deep Packet Inspection (DPI) is an important component in the user plane function (UPF) network element of the 5G core network. The UPF uses the embedded DPI module to identify and classify the exact services and application classes accessed by users and support dynamic policy control and charging control among other function modules like security. ZTE DPI module is a scalable component, when paired with the UPF, and delivers critical traffic identification logic without impacting UPF performance and stability. This traffic analytics and Deep Packet Inspection engine fits seamlessly into a cloud-native 5G service-based architecture (SBA).

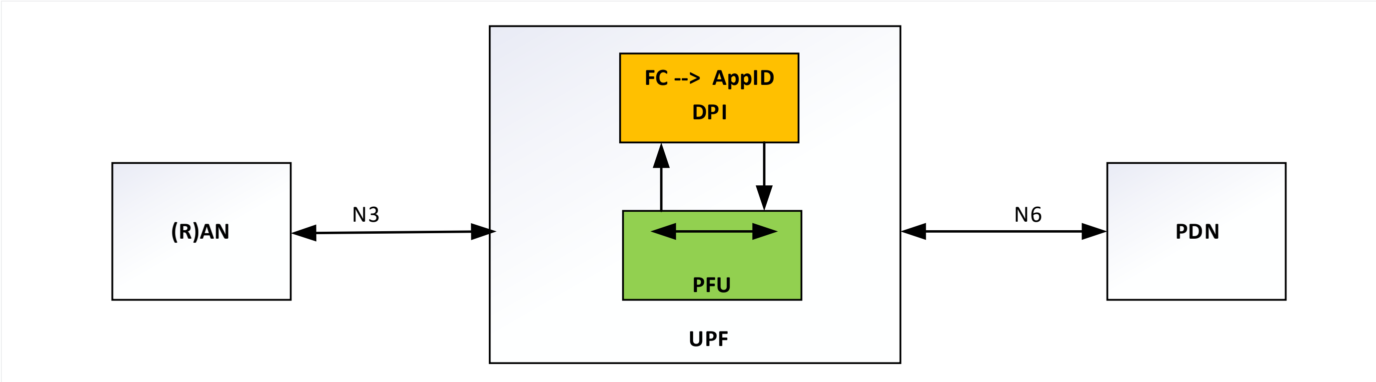

Figure 1. DPI module in 5GC network

The 5G UPF supports user data processing, application detection, packet routing and data forwarding under heavy throughput conditions. Flow classification and AppID generation can be offloaded to the DPI module. The overall deployment scheme is shown in Figure 1.

1.2. Challenge for Traditional DPI

DPI typically relies on pattern matching to analyze the payload content and identify the application related to each packet. However, increased use of encryption with network traffic makes it difficult to classify traffic using packet payload characteristics. This emphasis on network security means the industry must develop new ways for the DPI to complete protocol classification.

Advanced AI capabilities promise to improve analysis using statistical features to identify the correct traffic type without decrypting the traffic. However, this new approach requires establishing a new processing model for encrypted traffic, selecting appropriate features and machine learning algorithms, and providing high forwarding performance with guaranteed availability.

1.3. Enhanced ZTE DPI with Intel AI

ZTE and Intel collaborated to address changing requirements with ZTE's 5G DPI capabilities. New AI technologies and methods will address the dual challenge of classifying applications under encrypted traffic while supporting high-performance, packet forwarding in the UPF.

3rd generation Intel Xeon Scalable processors are one type of data-center CPU with built-in AI acceleration, end-to-end data science tools, and an ecosystem that's driving real-world applications. As a result, Intel lowered the technical threshold for network application manufacturers to enter the new field of AI.

ZTE and Intel will work to accelerate big data analysis through the highly optimized algorithm and modules provided by Intel® oneAPI Data Analytics Library (oneDAL), Intel® oneAPI Deep Neural Network Library (oneDNN), and others.

ZTE DPI leverages Intel AI modules to build a software framework and Intel® Advanced Vector Extensions 512 (Intel® AVX-512) instruction set acceleration to find the most relevant features and algorithms. This combination supports inspection of encrypted packets through advanced AI inferencing and machine learning while also promoting high-performance on Intel CPUs.

2. ZTE DPI enhanced by Intel TADK

2.1. Intel TADK

Intel TADK offers optimized libraries and tools for network traffic analysis and classification and enables developers to build their own AI/ML models. The TADK also provides plugins for integration with key open-source projects, such as the widely used vector packet processing (VPP) from the Linux Foundation's Fast Data Project (FD.io) that serves as a fundamental network framework for vBNG and5G UPF. In addition, TADK can extract network flow features and classify the flow with AI/ML Model or Deep Packet Inspection, with one example being Application Identification (AppID).

The folowing diagram shows the key modules based on Hyperscan, oneDAL, and reference solutions.

Key modules:

- **Flow Feature Extraction Library (FFEL):**

Configurable and extendable library to obtain data and information within a network flow, including Packet feature, Protocol feature, Bag-of-words (BOW) feature.

- **Flow Classifiers:**

Bi-direction flow classification on 5-tuple and flow table management and time-wheel based flow aging mechanism.

- **Protocol Detection:**

Protocol parser can detect and parse protocols including: IPv4, UDP, TCP, HTTP, TLS, QUIC, and DNS.

- **AI Engine:**

TADK wrapped oneDAL library uses its random forest algorithm to build the AI/ML classifier.

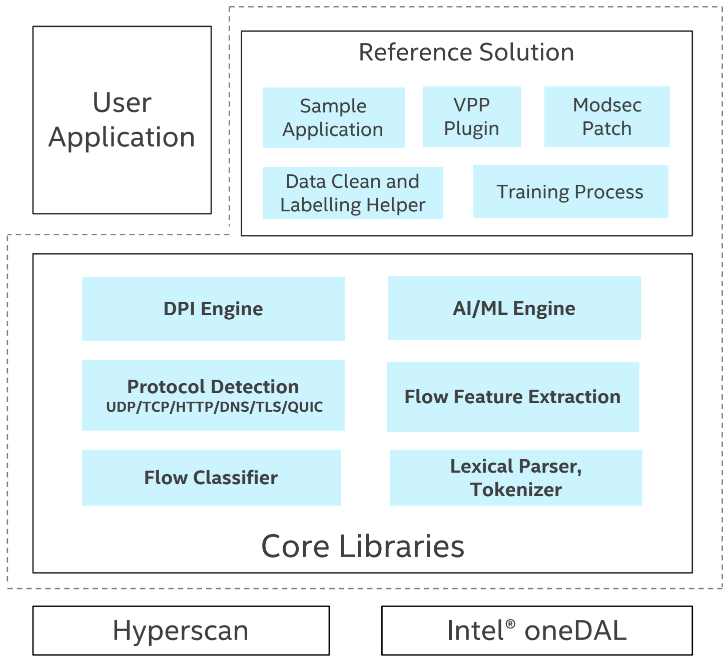

Figure 2. TADK software architecture and building modules

2.2. Integration with Intel TADK

ZTE will use solutions from the Intel TADK to identify protocols or applications with weak or even non-existent seven-layer, load characteristics. The solution collects the packets of these protocols or applications offline and labels them as specific applications. ZTE will select appropriate algorithms from the set of feature extraction algorithms in Intel TADK for encrypted traffic processing to perform AI modeling. The DPI module then loads the learned model onto the live network to classify traffic in real time. The overall DPI scheme with TADK integration is shown in Figure 3.

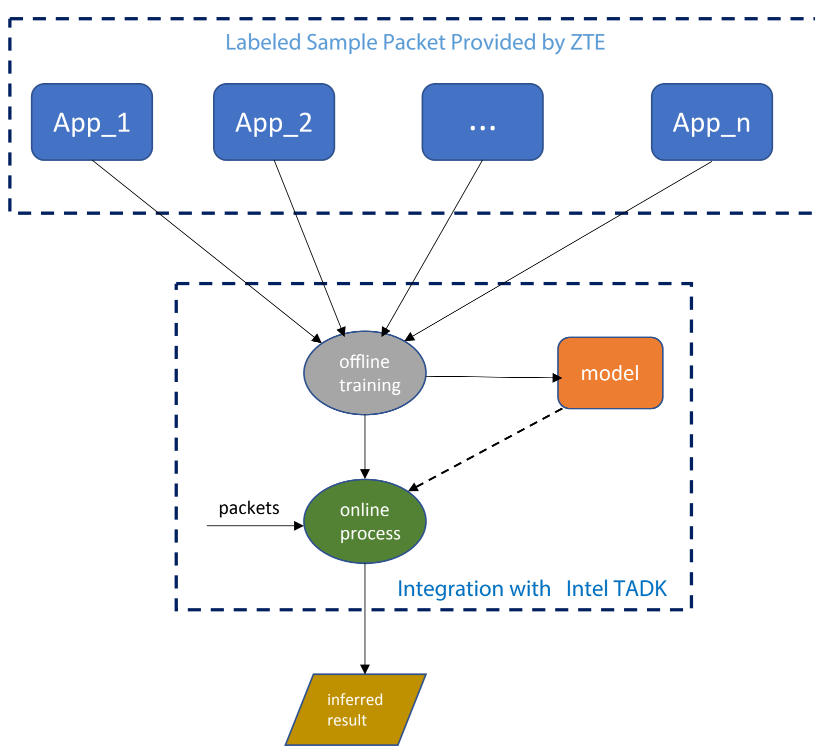

Figure 3. Overall ZTE DPI architecture diagram

Intel TADK supports flow-based traffic classification for offline training as well. After offline sampling of the target application, ZTE tags each flow with a service type label, and then uses the offline training tool provided by Intel TADK to generate an AI model. The FFE-L module can extract a set of rich features using these and corresponding labels as input to Intel's machine learning framework, oneDAL. The outputs from TADK include a business classification model file and selected final model feature. The training method is shown in Figure 4.

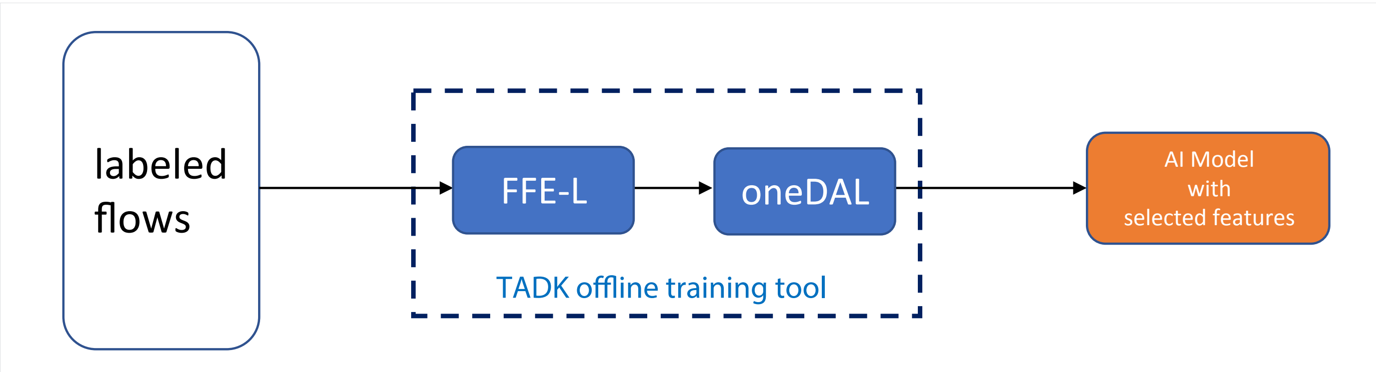

Figure 4. Schematic diagram of TADK offline training

In the online inference stage, the packet passes through ZTE DPI where the flow filter performs triage processing and then steers the network flows to either the ZTE legacy DPI classification engine or the Intel TADK module.

Intel TADK loads the model files obtained during offline training and uses the FFE-L module to manage the flow of received packets and extract flow features.

The online inference module processes each stream and outputs the inference results by flow. The online inference method is shown in Figure 5.

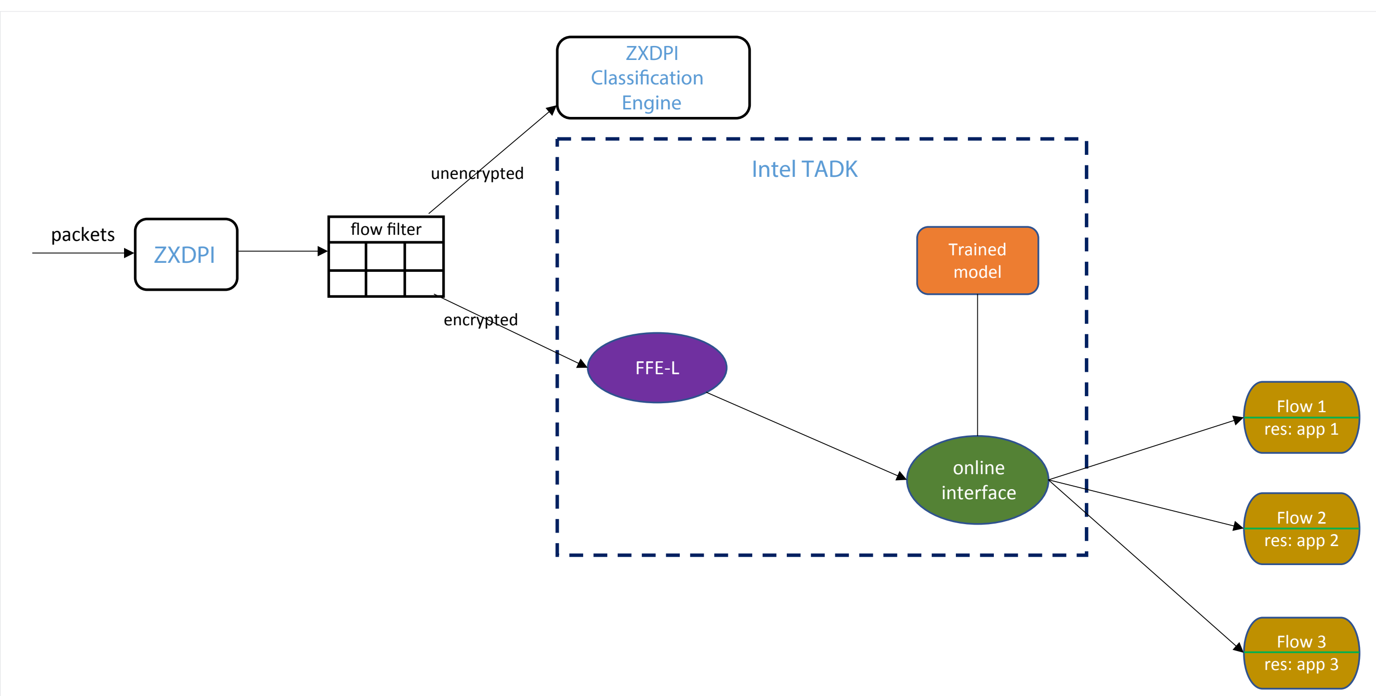

Figure 5. Schematic design of online inference with TADK

3. System Test Environment

3.1. Hardware configuration

The servers were ZTE's self-developed 5300-G4X, powered by 2-socket 3rd Generation Intel Xeon Scalable processors with each socket being connected to two Intel® E810-CQDA2 NICs.

Server	ZTE 5300G4X
uCode	0d000280
CPU	Intel® Xeon® Gold 6330N CPU@2.20GHz
Number of CPU	2
Memory	512GB
Network Adapter	Intel® E810-CQDA2

3.2. Software configuration

OS	ZTE CGSL 4.18.0-193.14.2.el8.x86_64
TCF Platform	TECS OpenPalette V7.21.20.06POS
UPF CNF	ZXUN-xGW(GUL) v7.22.10.b3
DPI	v7.22.10.b3
TADK	21.08

3.3. Networking Topology

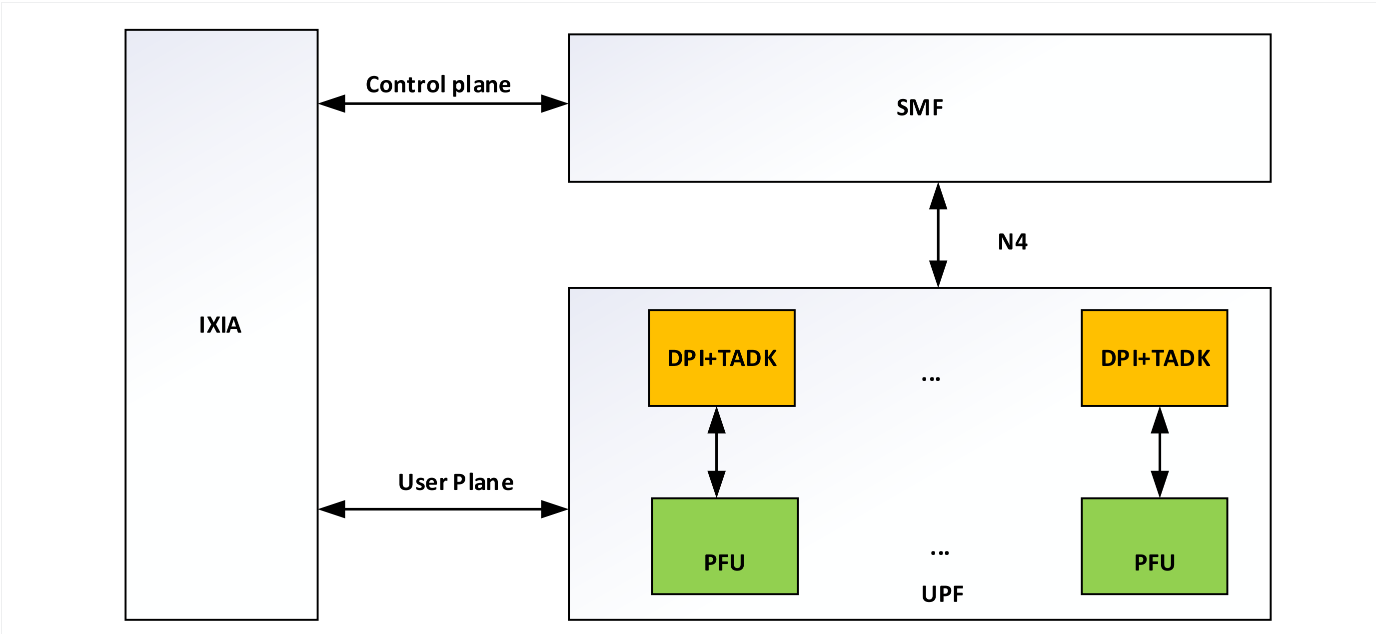

Figure 6. Test topology for the Intel® Xeon® based server

3.4. BIOS configuration

MENU	PATH TO BIOS SETTING	BIOS SETTING	REQUIRED SETTINGS
CPU CONFIGURATION	ADVANCED -> PROCESSOR CONFIGURATION	INTEL® HYPER THREADING TECH	ENABLED
		INTEL® VIRTUALIZATION TECHNOLOGY	ENABLED
POWER CONFIGRATION	ADVANCED -> POWER&PERFORMANCE	CPU POWER & PERFORMANCE POLICY	PERFORMANCE
	ADVANCED -> POWER&PERFORMANCE -> CPU P STATE CONTROL	ENHANCED INTEL® SPEEDSTEP TECH	ENABLED
		INTEL® TURBO BOOST TECHNOLOGY	ENABLED
	ADVANCED -> POWER & PERFORMANCE -> HARDWARE P STATES	HARDWARE P-STATES	DISABLED
	ADVANCED -> POWER & PERFORMANCE -> CPU C STATE CONTROL	PACKAGE C-STATE	C0/C1 STATE
		CIE	DISABLED
		PROCESSOR C6	DISABLED
IO CONFIGRATION	ADVANCED -> INTEGRATED IO CONFIGURATION	INTEL® VT FOR DIRECTED I/O	ENABLED

4. DPI test result

4.1. Test Design

ZTE chose several popular apps all with traffic encrypted for validation. Identification results, based on indicators such as recall rate and precision rate, will be counted to evaluate the classification ability of the model.

To verify the classification capabilities of Intel TADK, use cases need to be carefully designed. These flows are subject to the constraints to satisfy requirements shown in the following table.

Protocol type	Support application classification based on typical protocols: HTTP, HTTPS, QUIC etc.
Least packet number for flow Identification	Capability to identify flow type with maximum 48 packets of same flow.
Classification capability	Achieve consistent classification performance in three-category, five-category, nine-category and even more complex scenarios
Sample packet number	Sample traffic of each application have at least 500 packets.
Identification Criteria	In any classification scenario, the precision and recall rate of each application detection result are required to reach more than 90%

The specific testing, software tuning process is shown in Figure 7

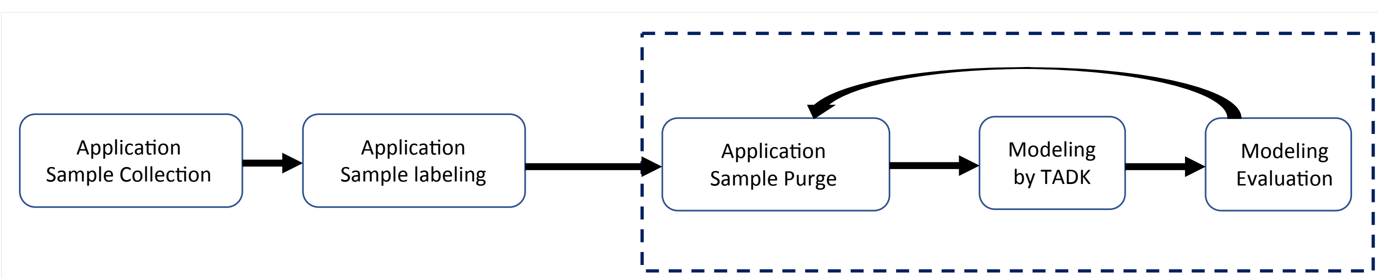

Figure 7. TADK function validation and integration pipeline

4.2. Function test

ZTE selected QQ*, Toutiao*, Baidu*, Tencent*, and other popular applications to test the accuracy of the classification algorithm.

We summarized the flow prediction results and accuracy of each classification model in the confusion matrix table in Figure 8. As the main diagonal line indicates, all the flows of Baidu, Tmall*, and Bilibili* were correctly classified; three (3) flows of Kuaishou* were mistakenly identified as Toutiao; two (2) messages of Tencent News* were mistakenly recognized as QQ. The F1 index of the overall classification results reached more than 98%. In the use cases, TADK achieves more than 90% precision and recall for each application in various scenarios, which can meet the requirements in the production environment. Accuracy of the models will improve over time through optimization of the AI algorithm.

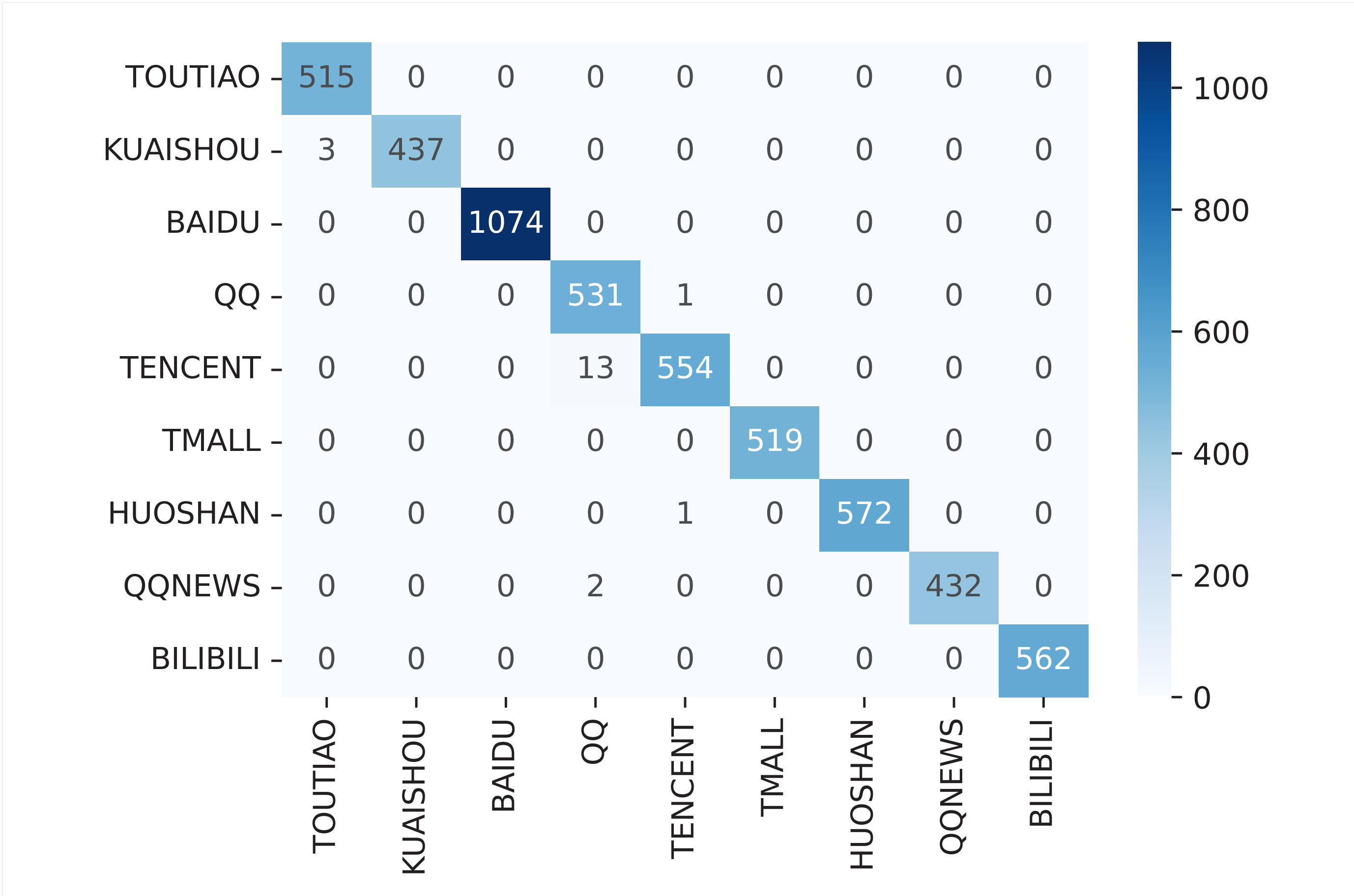

Figure 8. The confusion matrix of 9 applications

4.3. Performance test

We also measured UPF throughput with models that can classify 9 applications, also evaluated TADK traffic processing capability per core in Figure 9. The maximum throughput is 3.58Gbps (599kpps) with 9 applications.

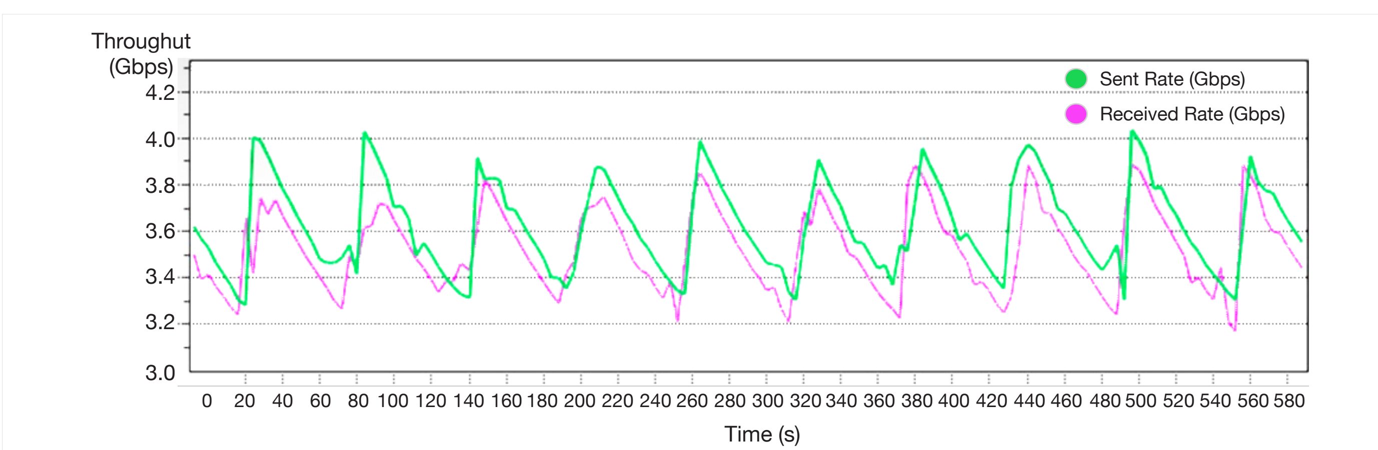

Figure 9. BAIDU, TMALL, BILIBILI, TENCENT, TOUTIAO, KUAISHOU, QQ, HUOSHAN, QQNEWS

5. Conclusions

One of the main functions of DPI is to use network traffic characteristics to identify the protocols and applications used by customers. There are fewer clear-text traffic characteristics in the network because of higher levels of encrypted traffic. As a result, the traditional methods are challenged.

Artificial intelligence can learn the topological characteristics of network traffic and build models to classify network traffic. The Intel TADK solution is built on 3rd Gen Intel® Xeon® Scalable processors with built-in AI acceleration, integrates the Intel oneDAL machine learning library, and supports the Intel AVX-512 instruction set. This combination provides feature extraction, learning modeling, and online inference for encrypted traffic, and high-performance packet processing.

6. Abbreviations

TADK	Traffic Analytics Development Kits
OneDAL	Intel® oneAPI Data Analytics Library
OneDNN	Intel® oneAPI Deep Neural Network Library
DPI	Deep Packet Inspection
FC	Flow Classification
AppID	Application Identification
UPF	User Plane Function
GTP	GPRC Tunneling Protocol
IP	Internet Protocol
OEMs	Original Equipment Manufacturers
OMU	Operations Manager (Unix)
PDN	Packet Data Network
PFU	Packet Forwarding Unit
PPPOE	Point-to-Point Protocol Over Ethernet
SR-IOV	Single Root I/O Virtualization
TCP	Transmission Control Protocol
UE	User Equipment

Authors

Intel

Yong Li

Jianwei Ma

Jack Ma

Rujun Cai

Guoliang Gu

Xiaodong Ying

NPG TADK RD team

ZTE

Liu Xiaobo

Dai Shuo