ICS 33.080
CCS L78

YD

中 华 人 民 共 和 国 通 信 行 业 标 准

YD/T 4263—2023

基于 SDN 的数据中心 OpenFlow 交换机的 TTP 模型

TTP model for openflow switch based SDN data center

2023-04-21 发布　　2023-08-01 实施

中华人民共和国工业和信息化部　发布

目　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由中国通信标准化协会提出并归口。

本文件起草单位：中国移动通信集团公司、中兴通讯股份有限公司、新华三技术有限公司、华为技术有限公司。

本文件主要起草人：王瑞雪、杨红伟、李志强、顾戎、翁思俊、罗鉴、王宇、汪军、张继江、陈博、万晓兰。

基于 SDN 的数据中心 OpenFlow 交换机的 TTP 模型

1 范围

本文件包括的主要内容有 SDN 数据中心 OpenFlow 交换机 TTP 模型的使用场景、流表定义、动作描述及流表转发过程描述等，供数据中心 Overlay 网络使用。

本文件适用于 SDN 数据中心硬件接入交换机。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

本文件没有需要界定的术语和定义。

4 缩略语

TTP	流表典型模型	Table Type Pattern
ONF	开放网络基金会	Open Networking Foundation
SDN	软件定义网络	Software Defined Networking
VxLAN	虚拟可扩展局域网	Virtual Extensible LAN

5 背景描述

TTP 是 ONF 提出的一种 OpenFlow 交换机抽象模型，使用 OpenFlow 流水线处理方式，实现 OpenFlow 控制器对 OpenFlow 交换机的转发控制。目前，OpenFlow 标准在转发面依然存在一些不足，尤其是 OpenFlow 芯片的研发与生产滞后等，制约了 OpenFlow 协议的使用。

SDN 数据中心 OpenFlow 交换机 TTP 模型的目标是提炼出 OpenFlow 芯片转发过程中的共性因素，通过简化模型加上传统芯片，推动了 OpenFlow 协议的广泛应用。一方面，TTP 模型利用传统芯片处理逻辑和表项组合出 OpenFlow 多级流表的基本功能，满足数据中心应用需求，快速进行产品开发。另一方面，通过 TTP 模型标准化，实现 SDN 控制器和 OpenFlow 交换机南向解耦，推动网络开放和产业链

发展。

根据实际需求，SDN数据中心OpenFlow交换机TTP模型适用于数据中心Overlay网络，并且Overlay网络采用VxLAN隧道封装技术；TTP模型不包括交换机Underlay网络的转发功能，不包括QinQ、MPLS等数据中心非常用场景。

6 基于SDN的数据中心OpenFlow交换机TTP模型概述

6.1 场景描述

本规范适用于数据中心Overlay网络，特征如下：

a） Overlay网络隧道实现技术选用VxLAN；

b） 不包括Underlay网络的转发功能；

c） 不包括QinQ、MPLS等场景。

6.2 TTP流水线示意图

数据中心OpenFlow交换机TTP模型流水线如图1所示。

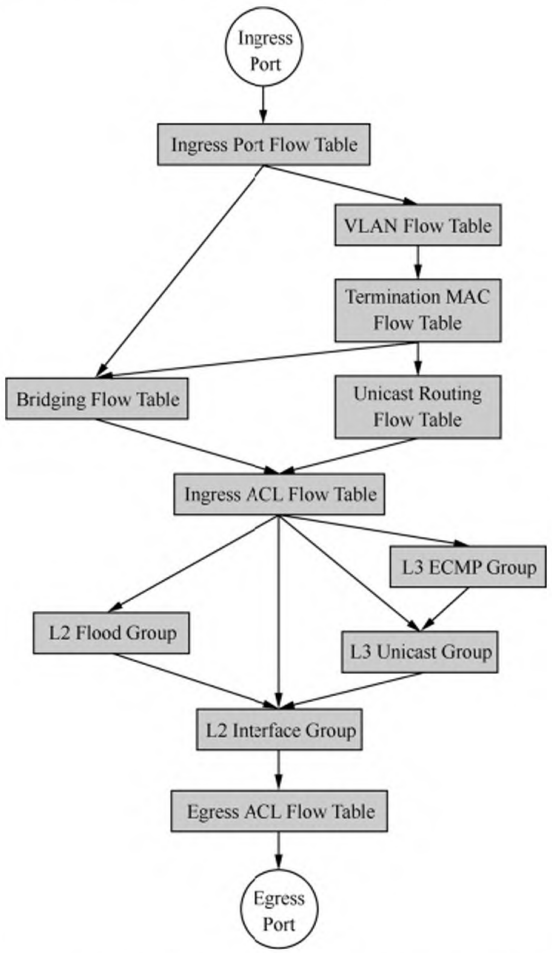

图1 数据中心OpenFlow交换机TTP模型流水线

7 流表定义及动作描述

7.1 Flow Table

7.1.1 Ingress Port Flow Table

识别报文入端口类型：交换机物理端口、VxLAN 逻辑端口。对于物理接口输入报文，转 table 10 处理；对于 VxLAN 逻辑端口输入报文，设置匹配的 VNI 值，转 table 50 处理。

流表编号为 table 0，table 0 表项内容见表 1。

表 1 table 0 表项内容

Case	Match	Len/bit	Action
物理端口输入	In_port:port number	32	Goto table 10
	Tunnel_ID:无	32	
VxLAN 逻辑端口输入	In_port: port number	32	Set metadata(VNI) Goto table 50
	Tunnel_ID:VNI	32	
其他	Table-miss		Drop

关键字段说明如下。

a） In_Port：端口包含物理端口和 VxLAN 逻辑端口两种，字段为 32bit，前 2 个 byte 表示端口类型，物理端口（编码：0x0000 xxxx）是指交换机面向主机/虚拟机的接入端口，可以携带 vlan tag 或 untagged。

b） VxLAN 逻辑端口（编码：0x0001 xxxx）：是指 VxLAN 隧道对应的逻辑端口，连接的是 SDN Overlay 网络。

c） Tunnel_ID：在 Overlay 网络中指 VxLAN VNI。

d） Metadata：在流水线中，通过 Metadata 来标识和传递 VNI、VRF 等信息，并作为匹配域跟随数据包一起参与流量的匹配过程；Metadata 为 64 位数据，定义低 32 位为 VNI，高 32 位为 VRF。

7.1.2 VLAN Flow Table

对于物理端口输入报文，根据携带的 vlan tag，判断是否进行下一步操作或丢弃。

流表编号为 table 10，table 10 表项见表 2。

表 2 table 10 表项

Case	Match	Len/bit	Action
物理端口输入 untagged 报文	In_port:port number	32	Set metadata(VNI) Goto table 20
	Vlan_ID:无	16	
物理端口输入 tagged 报文	In_port:port number	32	Set metadata(VNI) Goto table 20
	Vlan_ID:VID	16	

表 2　table 10 表项（续）

Case	Match	Len/bit	Action
其他	Table-miss	/	Drop

7.1.3　Termination MAC Flow Tables

根据物理端口入口报文 DMAC 是否是本机网关 MAC，识别做 L2 或 L3 处理，L2 处理转 table 50，L3 处理转 table 30。

流表编号为 table 20，table 20 表项见表 3。

表 3　table 20 表项

Case	Match	Len/bit	Action
IPv4 unicast MAC	ETH_TYPE:0x0800	16	Set metadata(VRF) Goto table 30
	metadata:VNI	64	
	ETH_DST:GW_MAC（网关 MAC）	48	
其他（L2 处理）	Table-miss	/	Goto table 50

7.1.4　Unicast Routing Flow Table

实现 L3 层单播报文转发，根据 DIP 对应网关，修改报文 DMAC 和 SMAC，修改 TTL，更新 VNI。

流表编号为 table 30，table 30 表项见表 4。

表 4　table 30 表项

Case	Match	Len/bit	Action
IPv4 unicast	ETH_TYPE:0x0800	16	write action: Set L3_Unicast_Group_ID 或 Set L3_ECMP_Group_ID Set TTL(TTL 减 1) Goto table 60
	metadata:VRF	64	
	IP_DST：DIP	32	
通向 GW 的目的 IP 未知流量	ETH_TYPE:0x0800	16	write action: Set L2_Unicast_Group_ID Goto table 60
	metadata:VRF	64	
	IP_DST：0.0.0.0/0	64	
特定网段上送	ETH_TYPE:0x0800	16	Send to controller
	metadata:VRF	64	
	IP_DST：DIP(网段)	32	
目的 IP 为本地地址	ETH_TYPE:0x0800	16	Send to controller
	metadata:VRF	64	

表 4　table 30 表项（续）

Case	Match	Len/bit	Action
目的 IP 为本地地址	IP_DST：DIP	32	Send to controller
其他	Table-miss	/	Drop

7.1.5　Bridging Flow Table

实现 L2 报文处理，根据 Dst_MAC 识别出端口是本地物理端口还是 VxLAN 逻辑端口，转 L2 Interface Group 处理；如果出口是本地物理端口，根据端口设置是否需要封装 VLAN Tag；如果出口为 VxLAN 逻辑端口，设置 VNI，剥离 vlan。

对于未知 Dst_MAC，上送控制器学习。

流表编号为 table 50，table 50 表项见表 5。

表 5　table 50 表项

Case	Match	Len/bit	Action
L2 转发处理	ETH_DST：Dst_MAC	48	write action: Set L2_Interface_Group_ID 或 L2_Flood_Group Goto table 60
	metadata:VNI	64	
其他	Table-miss	/	Send to controller

7.1.6　Ingress ACL Flow Table

实现入口 ACL，对识别的报文进行限速、镜像或者丢弃处理：

a）　此表不下发缺省表项，没有命中的报文执行 actionset 中的动作；

b）　业务链（引流），指定报文输出端口；

c）　镜像（复制报文，不影响原有转发），指定报文输出端口；

d）　安全组，先下发一条低优先级的通配表项（黑底），在此基础上下发允许通过的规则。

流表编号为table 60，table 60表项见表6。

表 6　table 60 表项

Case	Match	Len/bit	Action
LLDP	ETH_TYPE	16	Send to controller
ARP	ETH_TYPE	16	Send to controller
DHCP	ETH_TYPE	16	Send to controller
	IP_PROTO	8	
	UDP_SRC	16	

表 6　table 60 表项（续）

Case	Match	Len/bit	Action
DHCP	UDP_DST	16	Send to controller
Match 全集	IN_PORT	32	丢弃： Clear-actions 或 业务链： Clear-actions output port number 或 重定向： Clear-actions output port number 或 镜像： output port number 或 限速： write meter-id
	Metadata(VNI/VRF)	64	
	IPv4SRC	32	
	IPv4DST	32	
	IP_Proto	8	
	TCP_SRC/UDP_SRC	16	
	TCP_DST/UDP_DST	16	

7.1.7　Egress ACL Flow Table

实现出口 ACL，对匹配的报文进行丢弃、镜像或限速：

a） 所有从 Ingress ACL Flow Table 的报文默认进入表 7；

b） 表 7 不下发缺省表项，没有命中的报文透传。

流表编号为 table 61，table 61 表项见表 7。

表 7　table 61 表项

Case	Match	Len/bit	Action
出口 ACL	Metadata(VNI)	64	丢弃： Clear-actions 或 镜像： output port number 或 限速： write meter-id
	IPv4SRC	32	
	IPv4DST	32	
	IP_PROTO	8	
	TCP_SRC/UDP_SRC	16	
	TCP_DST/UDP_DST	16	
	VLAN	16	

7.2 Group Table

Group Table 格式如下。

Group ID	Group Type	Counters	Action Buckets

Group Table 表项见表 8。

表 8 Group Table 表项

Index	bits	描述
ID	[27:0]	组表 ID
Type	[31:28]	区分组表类型 0000：L2 Interface Group 0010：L3 Unicast Group 0100：L2 Flood Group 0111：L3 ECMP Group

参数定义如下。

a） Group ID：长度为 32 位无符号整数。

b） Group Type：组表类型，分为 all、select、indirect 等。

c） Counter：统计通过组表处理的报文数量。

d） Action Buckets：动作指令集。

7.2.1 L3 ECMP Group

实现报文分发，将报文发送到多个 L3 unicast group。

组表类型：select，即 Group 中有多个 Bucket，但根据交换机内部调度算法，仅会执行其中的某一个 Action Bucket。

L3 ECMP Group 表项见表 9。

表 9 L3 ECMP Group 表项

Case	Bucket（多个）
L3 ECMP 多路径	Bucket 1 {Set L3_Unicast_Group_id} Bucket 2 {Set L3_Unicast_Group_id} Bucket 3 {Set L3_Unicast_Group_id} …… Bucket n {Set L3_Unicast_Group_id}

7.2.2 L3 Unicast Group

修改 DMAC 和 SMAC，进入 L2 interface group。

组表类型：indirect，即组表中仅有一个 Action Bucket。

L3 Unicast Group 表项见 10。

表 10 L3 Unicast Group 表项

Case	Bucket
L3 routing	Set ETH-DST Set ETH-SRC Set L2_Interface_Group_id

7.2.3 L2 Interface Group

对出接口实现 VLAN 标签设置和弹出。

组表类型：indirect，即组表中仅有一个 Action Bucket。

L2 Internet Group 表项见表 11。

表 11 L2 Internet Group 表项

Case	Bucket
VxLAN 逻辑口	Set Tunnel_ID POP VLAN（可选，对于 untagged 报文无此操作） Out VxLAN 逻辑口
Untagged 物理口	POP VLAN Out port number
tagged 物理口	PUSH VLAN Set VLAN Out port number

7.2.4 L2 Flood Group

主要用于组播或者广播场景。

组表类型：all，即表中所有的 Action Buckets 都会被执行，数据包会被克隆为多份。

L2 Flood Group 表项见表 12。