June, 2020

DPA Focussed Testing

# Scope of the document

This document captures the methodology being planned to align towards DPA (Data Plane Abstraction) focused testing through both manual ways and automations. The focus at present is restricted to DNX based platforms only (ncs5500/5700, ncs540(L), ncs560 ..etc)

## What is DPA ?

### Data Plane Abstraction is the hardware Abstraction layer for Broadcom-DNX based Network Devices.

### DPA interfaces with upper layer EA/FIB process to program the BCM hardware through SDK APIs

### DPA maintains multiple abstraction tables which are used to program the hardware

## Intent

### To identify the touchpoints and build testing profiles which can effectively test the changes in Platform-Dependent forwarding code.

### Should be able to validate the major SDK changes & PD level commits during an XR release cycle

# diagram

# DPA MoDEL

IOS-XR PI (MA)

# 

IOS-XR PI/PD (EA/FIB) – Eng/DR

DPA SRV

DPA

BCM SDK

# touch points

Coverage in the test profile and test cases to focus on

* PD FIB
* Dependency resolution
* Different EA interactions
* Stats allocation
* OOR handling & Recovery
* Focused DPA obj tables (nhgroups, iproute, mpls\*, l2\* …etc)
* Different hardware tables like LEM, LPM/KAPS ..etc
* Forward paths (LC/NPU/Core combos)

## **Phase 1**

Componentize the MTB into the DPA areas with major code churn. Plan is to focus on these items

## **L3fib**

* + Exercise most of the PD forwarding path / CEF chain variations
  + Cover Recycle paths. Ex-   VRF fallback, SR PFP, GRE, SR-TE – Large label
  + Coverage into resource allocations (based on GRID carving)
  + Protection/fast re-route coverage
  + Load Balancing
  + Stats allocation
  + Table usages (etcam, lpm, lem) , encaps/fec
  + Heterogeneous network
  + Not the usual things Ex: QPPB, PBTS ..etc

## **L2fib**

* + Coverage on traffic types – BUM, Routed, known L2
  + Coverage on access homing (SH/MA) and active modes (SA/AA/SFA)
  + ECD interplay Ex: BUM via FRR backup tunnel, BGP, BGP-LU, IGP, SR-TE
  + L2 coverage (local, p2p, p2mp) using any of the supported L2 services
  + L2lif focus with all service tag rewrite combinations
  + Focus on Routed chains using BVI
  + Recursion over BVI
  + Recycle paths
  + Protection coverage for l2vpn AF(ex PIC) and underlay FRR/LFA
  + Control packets in punt/inject path (ex- LACP), mac learning , distribution ..etc

## **Common items**

* + Traps coverage Ex- Invalid packets
  + Interface infra inclusive of high focus on bundles
  + Feature interactions (Qos, ACL, BFD)
  + Slow path using Fragmentation
  + Multicast ->L3, L2

# Profile 1 – L3 FIB/Forwarding

Our priority focus is on l3fib as it undergoes huge churns with good share of customer issues. Never ending combinations to cover.

Coverage across the below DPA tables

* Iproute
* ip6route
* ipnh / ipnhgroup
* ip6nh / ip6nhgroup
* ipvrf
* iptunnel
* mplslabel
* mplsnh
* l3/mpls\*stats
* L3intf, lagport
* Srv6ng, srv6sid (if supported on the asic)

NPU Blocks & HW tables in Focus

* IRPP – LEM, LPM, FEC, ecmp-FEC, inlif/rif, etcam
* ERPP – encaps, outlif

# Profile 2 – L2 FIB/Forwarding

Our priority focus is on l2fib as it undergoes huge churns with good share of customer issues.

Coverage across the below DPA tables

* mhpweport
* mplspweport
* l2xc
* l2intf
* mplsnh
* ipnh
* l2vpn / l2\*stats
* L2bridge\*
* L2evpn\*
* Bviport
* L2qos
* L2acl

NPU Blocks & HW tables in Focus

* IRPP – LEM, LPM, FEC, inlif, etcam
* ERPP – encaps, outlif

# Common items

* Multi-LC coverage (inclusive of bundles)
* Streams ingress / egress out of different LCs, NPUs, cores
* Primary and bkp via different LCs, NPUs, cores
* Card interop b/w ASIC which has pipeline differences ( Ex: J+ vs J2 )
* J2 NPU mode combinations (comp vs native)
* New MDB profiles to be considered for future releases (esp on scale)

1. **PROFILE L3FIB– DETAILED COVERAGE**

|  |  |  |  |
| --- | --- | --- | --- |
| Area | Coverage | Scale | Traits |
| L3 Forwarding chains | IPv4/v6 lookup |  |  |
|  | IPv4/v6 loopkup with ecmp | 64 |  |
|  | IPv4/v6 loopkup with ucmp | 64 |  |
|  | Loose URPF |  | v4 + v6 to be enabled together |
| MPLS Label chains | Imposition |  | LDP – Separate chains for Imp and swap  SR – Uses the same fec |
|  | Swap |  |  |
|  | PHP – Implicit-null |  |  |
|  | PHP – Explicit-null |  |  |
|  | Disposition |  |  |
|  | ECMP coverage (all stages) for mpls chains | 32 | SWAP is optimized if all out-labels are same (SR) = ecmpfec will be 1  LDP will have different out-labels so more ecmpfec usage  Encap id usage will be high |
|  | ECMP b/w IP & MPLS paths |  |  |
|  | BGP LU interface peering |  |  |
|  |  |  |  |
| Hierarchical Label chains | BGP LU Loopback peering |  |  |
|  | L3VPN unipath Label\_in\_leaf |  | LWLDI EEI label push for TCAM  RHLDI EEI label push for non-TCAM |
|  | L3VPN unipath over SR/LDP Ecmp |  |  |
|  | L3VPN unipath over RSVP/TE & SR-TE |  |  |
|  | L3VPN unipath per-prefix with cef encap sharing disable |  | RSHLDI will be reused for different LWDI  No EEI push  Better convergence for TCAM cards  For non-TCAM cards, we will need to program mpls encap for service labels  7.3.1 – with 3 levels of FEC hierarchy this CLI is of no use |
|  | L3VPN Multipath default-mode |  |  |
|  | L3VPN Multipath cef encap sharing disable |  | 7.3.1 – with 3 levels of FEC hierarchy this CLI is of no use |
|  | Label in eTCAM for SE (J2 Native)  For l3vpn unipath w/o PIC |  | In 7.2.1, only FEC is shared. Grid will still allocate MPLS encapfor VPN routesIn 7.3.1, will avoid MPLS encapid allocation by GRID once 3 levels of FEC hierarchy is used |
|  | L3VPN over LU Loopback peering |  |  |
|  | 6PE chain |  |  |
|  | 6VPE – Ipv4 NH for IPv6 routes |  |  |
|  | VRF fallback |  |  |
|  | RTBH in a VRF |  |  |
|  | 3 label collapse |  |  |
|  |  |  |  |
| SR/SR-TE label chains | SRMS LDP to SR Imposition, SR to LDP Disposition |  |  |
|  | SR-TE 1HOP chain (PHP/POP) |  |  |
|  | SR-TE Multi-hop Preference base cpath |  |  |
|  | SR-TE Weighted ECMP |  |  |
|  | SR-TE using adj-sid |  |  |
|  | SR-TE with cpath ecmp | 32 |  |
|  | Static via SR-TE |  |  |
|  | SR-TE autoroute announce |  |  |
|  | SR-TE Large label stack + services label |  | For j2: With (10+2 ?) and without recycle (8+2) |
|  | SR-TE for l3vpn, LU, BGP overlay |  |  |
|  | Flex Algo coverage |  |  |
|  |  |  |  |
|  | SR-TE for EVPN overlay |  |  |
| SRv6 | SRv6 chains |  | Not supported on J2 yet |
|  |  |  |  |
| Tunnels | RSVP TE Autoroute announce |  |  |
|  | RSVP TE Forward adjacency |  |  |
|  | RSVP TE transport for l3vpn, LU, BGP overlay |  |  |
|  | LDP over RSVP-TE |  |  |
|  | Ipv6 over RSVP-TE |  |  |
|  | Heterogenous ecmp b/w TE + LDP |  |  |
|  | GRE encap/decap chains |  |  |
|  | PBTS, QPPB |  |  |
|  |  |  |  |
| Protection | BGP PIC core |  |  |
|  | BGP PIC Edge |  | “Cef encap sharing enable” required for l3vpn (per-prefix) |
|  | BGP PIC for Services [l3vpn, LU] |  | “Cef encap sharing enable” required for l3vpn (per-prefix) |
|  | BGP PIC + multipath co-existence |  |  |
|  | BGP PIC – 2 level (L3vpn over LU) |  | Cant enable on both the levels  Not supported for 2nd level |
|  | BGP PIC L2 services |  | Supported on L2 services over LU.. Need to check on L2 services |
|  | TI-LFA disjoint PQ Sids |  |  |
|  | TI-LFA for large-label-stack tunnels |  |  |
|  | TI-LFA zero,1 extra SID |  |  |
|  | TI-LFA with microloop avoidance |  |  |
|  | RSVP-TE NHOP, NNHOP |  |  |
|  |  |  |  |
|  |  |  |  |
| Resource utilization | Encap phase mapping | 192K | Different phases in the resource bank used by different consumers (SRTE, LU ..etc) |
|  | Labels in LEM |  | Non-SE card |
|  | LEM / LPM usage for Routes |  |  |
|  | Etcam usage for Routes/labels |  | SE card |
|  | ECMPFEC usage | 32K |  |
|  | RIF/LIF usage |  |  |
|  |  |  |  |
| Stats | Mplsnh, ipnh stats – interface accounting |  |  |
|  | SR-Policy stats |  |  |
|  | RSVP-TE stats |  |  |
|  | Dynamic stats allocation |  |  |
| OOR / Recovery | Ecmp FEC OOR |  | Should stop programming ecmp and program with single path once OOR condition hit  Recovery without LC reload when OOR condition turns yellow/green |
|  | LEM , LPM OOR |  | Graceful handling and auto recovery |
|  | Stats OOR |  | Graceful handling and auto recovery |
|  | FEC, Encap-id OOR |  | Graceful handling and auto recovery |
|  |  |  |  |
| Convergence | Route programming rate – IP chains,  Encap programming (nh), Recursive chains for LEM, LPM, etcam, per-prefix, Per-ce |  |  |
|  | FRR for Service traffic |  |  |
|  | PIC for Service traffic |  |  |
| Common | Hw-module profiles coverage -> LSR optimized, DLB enablement , stats (customer usecase based)  Feature interactions (ACL, Qos, BFD)  LC / NPU coverage – interop  J2 modes – native & Comp  Qos propagation/marking |  |  |

1. **PROFILE L2FIB– DETAILED COVERAGE**

|  |  |  |  |
| --- | --- | --- | --- |
| Area | Coverage | Scale<TBU> | Traits |
| L2 interface | L2 interface creation/deletion | 4k (J2:8K) |  |
|  | L2 interface bind to XC | 2k |  |
|  | L2 interface bind to BD |  |  |
|  | EFT rewrite translations |  |  |
|  |  |  |  |
| P2P chains | Local XC |  |  |
|  | P2P PW over non-Recursive underlay (SR/LDP) |  |  |
|  | P2P PW over Recursive underlay LU interface peering/Loopback peering |  |  |
|  | P2P PW preferred path via RSVP TE tunnels (cover FRR path) |  |  |
|  | P2P PW preferred path via SR\_TE tunnels ( cover large label stack, TILFA) |  |  |
|  | PWHE ( GIL variations ) |  |  |
|  | P2P PW ECMP/UCMP underlay coverage |  |  |
|  |  |  |  |
|  |  |  |  |
| P2MP Chains | BD coverage AC replications |  |  |
|  | P2MP over non-Recursive underlay (SR/LDP) |  |  |
|  | P2M over Recursive underlay LU interface peering/Loopback peering |  |  |
|  | P2MP PW preferred path via RSVP TE tunnels (cover FRR path) |  |  |
|  | P2MP PW preferred path via SR\_TE tunnels/ODN ( cover large label stack, TILFA) |  |  |
|  | Access PW into BD (H-VPLS) |  |  |
|  | EVPN – P2P, ETREE |  |  |
|  | P2MP - ECMP/UCMP underlay coverage |  |  |
|  | EVPN MH – AA, SA, SFA, SH, MH |  |  |
|  | Inter-AS PW chain for P2P/P2MP |  |  |
|  | EVPN Headend ? >To be checked |  |  |
|  | EVPNv6 - Advertising EVPN NLRI with an IPv4 NextHop |  |  |
|  |  |  |  |
| L2 to L3(BVI) or L3 to L2 chains | BVI part of default VRF |  | When BVI is configured on the resource allocation for AC interface will be taking thrice the value.. |
|  | BVI part of non-default VRF |  |  |
|  | Recursive routes over BVI |  |  |
|  | L3 routing b/w bridges |  |  |
|  | Inter-subnet stretched |  |  |
|  | Intra-subnet stretched |  |  |
|  | Routing protocols over BVI |  |  |
|  | EVPN-IRB Host-Mobility: Duplicate IP Detection |  |  |
|  | L2 GW - L2+L3 IRB |  |  |
|  |  |  |  |
| L2 features (interface/BD) | Mac limit, aging, withdrawal, mac flush (levels – Per BD/BP, bulk - ERPS) |  |  |
|  | Storm-control for BUM |  |  |
|  | Flooding disable |  |  |
|  | IGMP snooping (BVI in OIL) |  |  |
|  | L2 DHCP relay agent |  |  |
|  |  |  |  |
|  |  |  |  |
| BUM replication | AC to AC |  |  |
|  | AC to VFI PW chain with neighbour reachability via IGP |  |  |
|  | AC to VFI PW chain with neighbour reachability via BGP-LU |  |  |
|  | AC to VFI PW chain with neighbour reachability via TE and FRR paths |  |  |
|  | Replication for EVPN NH |  |  |
|  | VFI PW to AC chain |  |  |
|  | AC to access-PW vice versa |  |  |
|  | LAG pruning on BD AC |  |  |
|  |  |  |  |
| Mac Learning | MAC learning on AC | 128k |  |
|  | MAC learning over VPLS VFI PW |  |  |
|  | MAC learning over EVPN neighbour |  |  |
|  | MAC learning over VPLS access PW |  |  |
|  |  |  |  |
| Protection | BUM protection for P2MP |  |  |
|  | BUM protection for EVPN traffic |  |  |
|  | BGP PIC for l2vpn AF, underlay LU |  |  |
|  | SR-TILFA, LFA, TE-FRR |  |  |
|  | ERPS protection |  |  |
|  |  |  |  |
|  |  |  |  |
| stats | L2xc stats |  |  |
|  | Vpws stats |  |  |
|  | L2 Interface stats, L3 stats |  |  |
|  | BVI stats |  |  |
| Resource utilization | L3 Fec, LEM (mac entries) |  |  |
|  | Encap-id utilization |  |  |
|  | LIF |  |  |
|  |  |  |  |
| Convergence | VPWS PW-red |  |  |
|  | VPLS BUM convergence using FRR triggers |  |  |
|  | VPLS unicast convergence using FRR triggers |  |  |
|  | EVPN MH – node failure/ recovery |  |  |
|  | EVPN MH – link failure/recovery |  |  |
|  | EVPN BUM traffic with core triggers |  |  |
|  | EVPN unicast traffic with core triggers |  |  |
|  | EVPN traffic with access failure |  |  |
|  | CLI to prefer specific PE IP Address (Disabling ECMP) > |  |  |
|  | PRISM - PRImary Standby Mechanism (aka. L2-FRR) |  |  |
|  | Fast Recovery for EVPN DF Election upon Link / Node Insertion |  |  |
|  |  |  |  |
| Common | Hw profile – hqos, bundle-scale |  |  |
|  | Qos on L2AC, BVI |  |  |
|  | ACL on L2AC, BVI |  |  |
|  | Qos propagation/marking |  |  |
|  |  |  |  |