클라우드 네트워킹의 시작과 끝 Nexus 9000 스위칭 솔루션

시스코 코리아

구 본 일 부장

Nexus 스위치 제품군

Nexus 9000 Cloud Scale

차세대 DC를 위한
Cisco Cloud Scale ASICs

- 10/25/40/100/400G를 지원하는 고성능의 Fabric
- VXLAN, Segment Routing, ACI 등 H/W 기반 Fabric 기술 지원
- H/W 기반의 텔레메트리
- 자동화를 위한 API

Nexus 9500/3600 R-Series

Core/Edge 라우팅
Broadcom Jericho ASICs

- 대용량의 멀티캐스트 지원
- MPLS, VXLAN, Segment Routing 지원
- Deep Buffers
- 코어 라우팅을 위한 대용량 라우팅 테이블

Nexus 7000 Series

DCI/Campus 코어를 위한
Cisco ASICs

- Data Center간 연결을 위한 기술 지원
- 검증 된 DC/Campus 코어 기능 제공

Nexus 3000 Series

범용적 사용을 위한
Merchant ASICs

- 상용 칩셋 기반
- 특정 칩셋 기반의 기능이 필요한 경우 (Ultra Low Latency, Data Path Programmability 등)

Nexus 스위치와 NX-OS

Nexus 9000 Cloud Scale

차세대 DC를 위한
Cisco Cloud Scale ASICs

- 10/25/40/100/400G를 지원하는 고성능의 Fabric
- VXLAN, Segment Routing, ACI 등 H/W 기반 Fabric 기술 지원
- H/W 기반의 텔레메트리
- 자동화를 위한 API

지능형 서비스 — Load Balance (ITD, PLB)
iCAM, Catena

가시성 / 분석 — Telemetry (Tetration, Netflow)
Cisco Nexus Data Broker

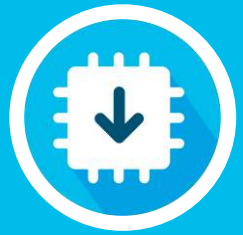
Programmability — Guestshell, NX-OS SDK
Docker 기반

Switching Infrastructure — 다양한 H/W 지원
400G 지원 계획

DCNM을 통한 관리관 향상 자동화 지원

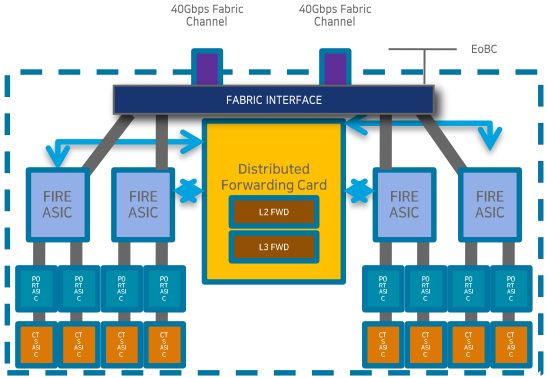
목차

- Switching Infrastructure
- Programmability
- 가시성 / 분석
- 지능형 서비스

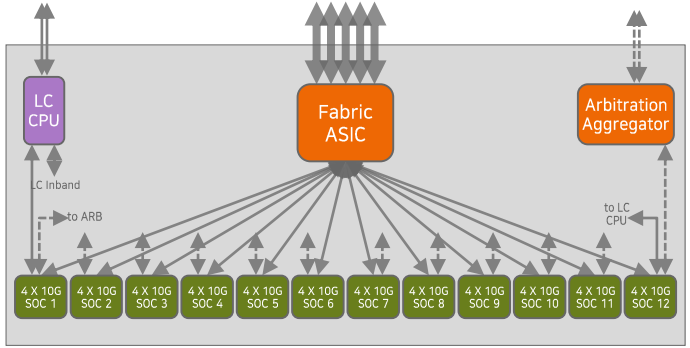


Switch Infrastructure

Switch 구조의 변화



32 x 10G Ports

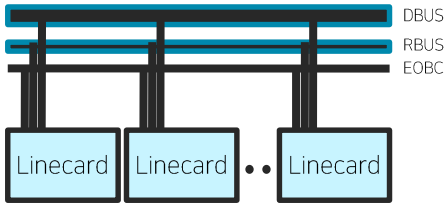


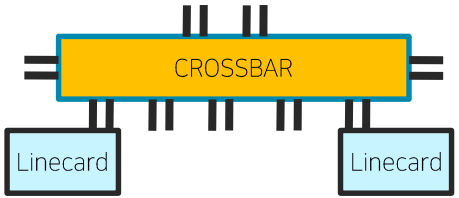
48 x 10G Ports

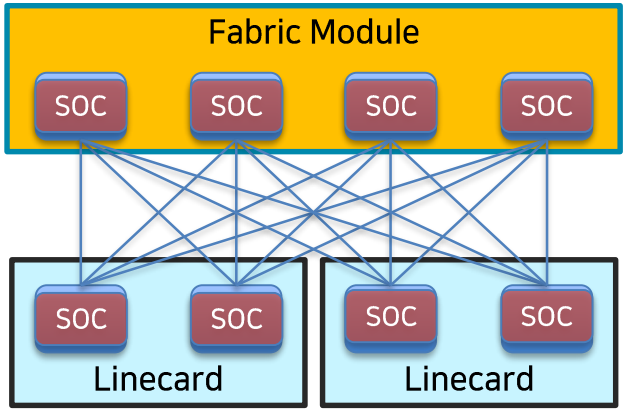
64 x 100G Ports

Bandwidth와 포트 집적도를 높이기 위한 발전

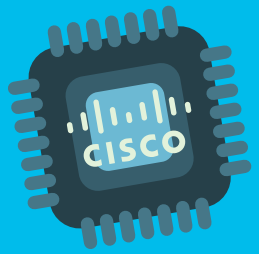
Switch 구조의 변화







Bandwidth와 포트 집적도를 높이기 위한 발전



Cisco
Cloud Scale ASICs

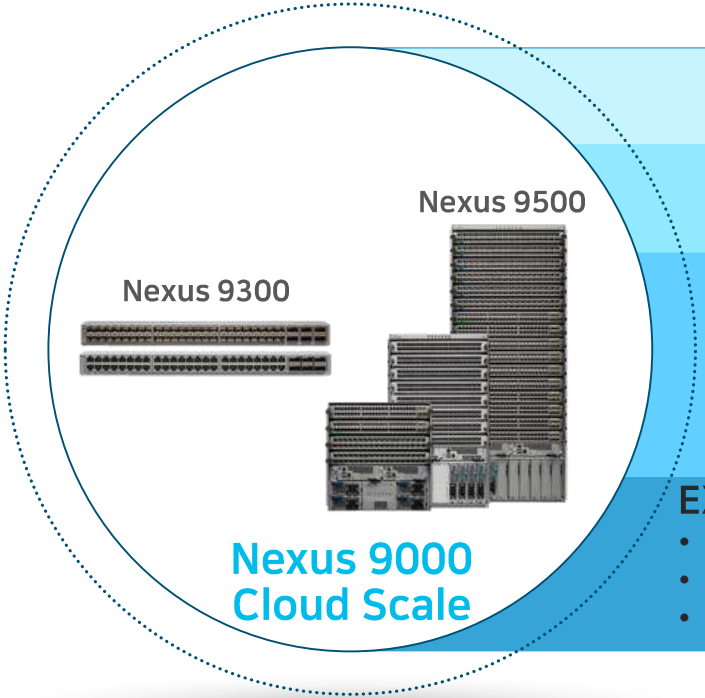
Innovations

- H/W 기반의 Telemetry
  (Flow Table, Flow Table Event, SSX)
- VXLAN Switching/Routing
- 어플리케이션 최적의 성능을 지원하는
  향상된 버퍼 관리 (AFD/DPP)
- 다양한 상황에 유연하게 대처 가능한
  Route, Policy, Host Table 등의 Scale
- Line-Rate 의 암호화
  (MACSec/CloudSec)

Better User Experience

- H/W 기반의 기능을 최대로 이용한 S/W
  지원
- 다양한 SDK의 지원으로 접근이 쉬운
  개발환경 지원

Cisco Cloud Scale ASICs 제품 군

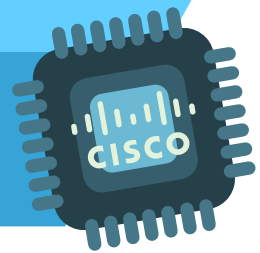


GX ASICs (12.8T per ASIC)
- 400G 지원 (16 x 400G)

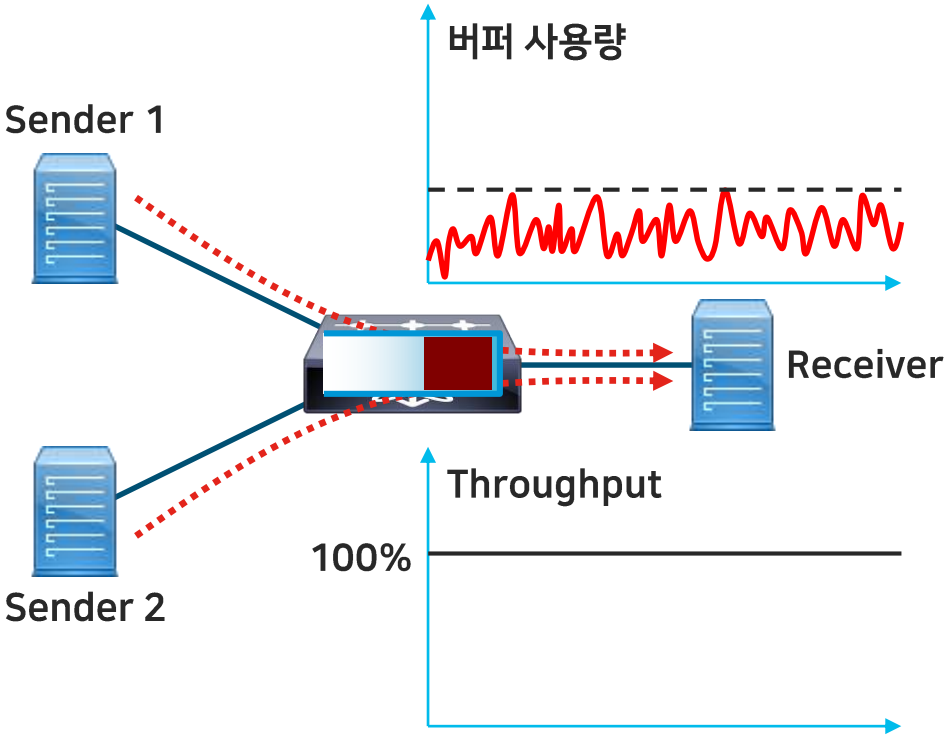
FX2/FX3 ASICs (3.6T per ASIC)
- Telemetry (Streaming Statistics eXport - SSX)

FX ASICs (1.8T per ASIC)
- Telemetry (Flow Table Event)
- Line Rate 암호화 (MACSec/CloudSec)
- 25G 서버 Optical 연결 지원 (RS-FEC)
- LAN/SAN 통합 포트 지원 (25G 이더넷/32G FC)

EX ASICs (1.8T per ASIC)
- 10/25/40/100G 등 멀티 Rate
- Smart Buffer
- Telemetry (Flow Table)

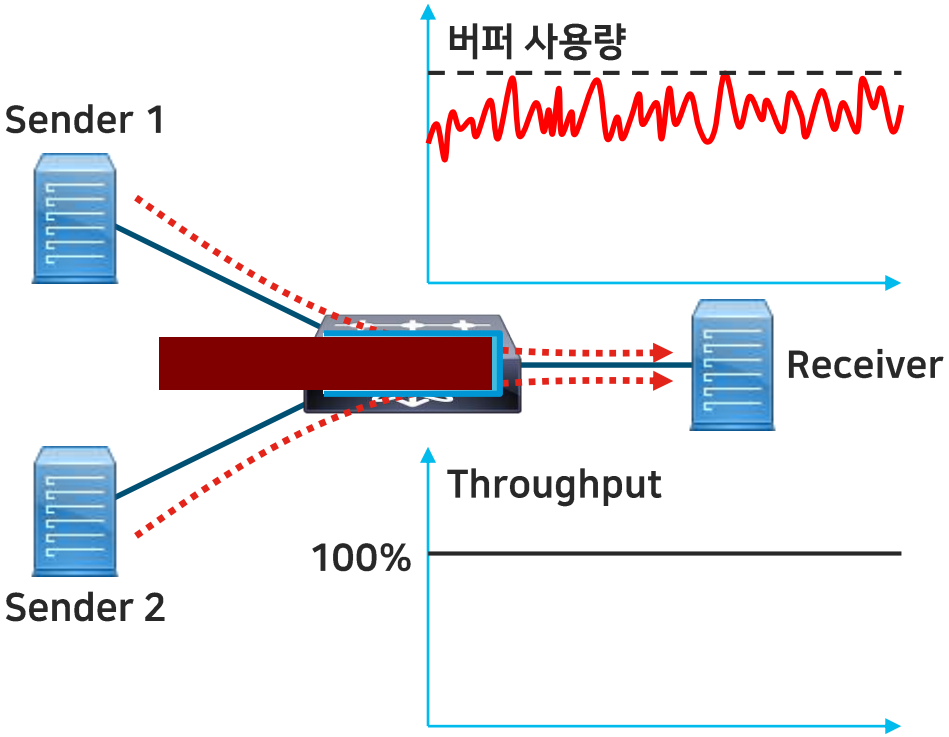


Buffer 사용의 효율적 방식 필요



Buffer 사용량 > 0
= 100% Throughput

Buffer 사용의 효율적 방식 필요



Buffer 사용량의 급격한 증가

Big Buffer
= Throughput 증가 효과 없음
+ Latency의 지연

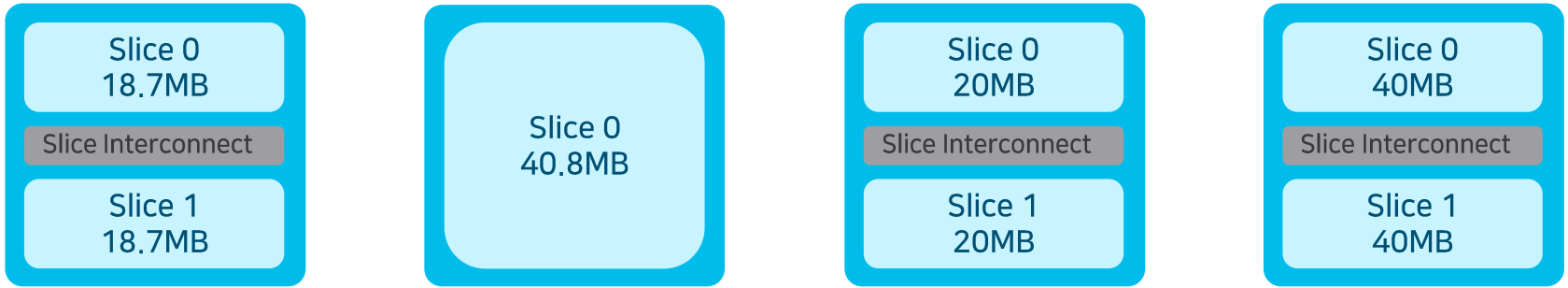
RTT = 100 usec시 권고 Buffer 용량

	10 Gbps	40 Gbps	100 Gbps
1	250KB	1MB	2.5MB
100	25KB	100KB	250KB
2500	5KB	20KB	50KB

Beheshti et al. 2008: "Experimental Study of Router Buffer Sizing"

효율적인 Buffer 관리

Dynamic Buffer Protection (DBP)	Shared-Memory 구조인 Buffer에서 능동적 Buffer 할당
Approximate Fare Drop (AFD)	Burst Traffic 으로 인한 Drop 발생 시 서비스 영향도 최소화
Dynamic Packet Prioritization (DPP)	크기가 작고 시간이 짧은 Flow를 보다 더 빨리 처리



EX
18.7MB/Slice
(Total 37.4MB)

FX
40.8MB/Slice
(Total 40.8MB)

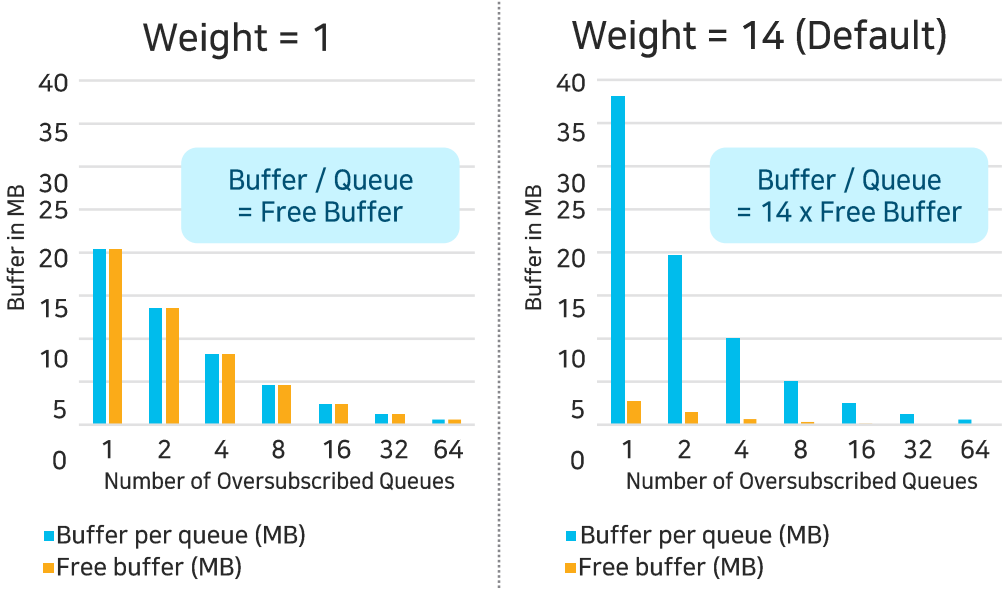
FX2
20MB/Slice
(Total 40MB)

FX3
40MB/Slice
(Total 80MB)

효율적인 Buffer 관리 - DBP

Dynamic Buffer Protection (DBP)

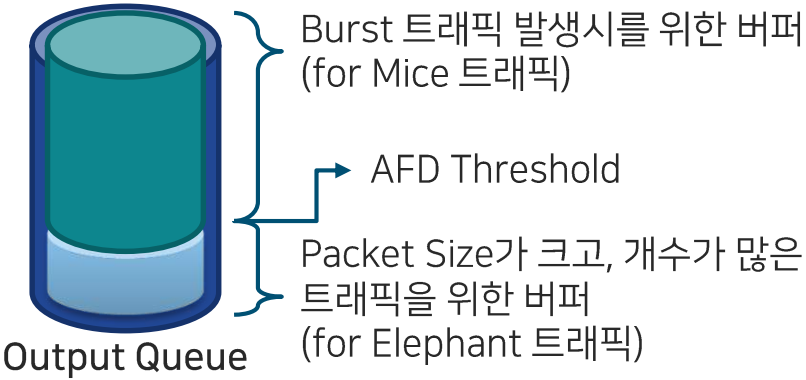
- Shared-memory 구조의 Output Queue의 Threshold를 동적으로 조절
  - 혼잡도에 따라 자동으로 Queue Limit 조정
  - 각각의 Slice에 할당 된 버퍼로 한정
- Queue Threshold는 남아있는 메모리를 기준으로 Weight 로 설정

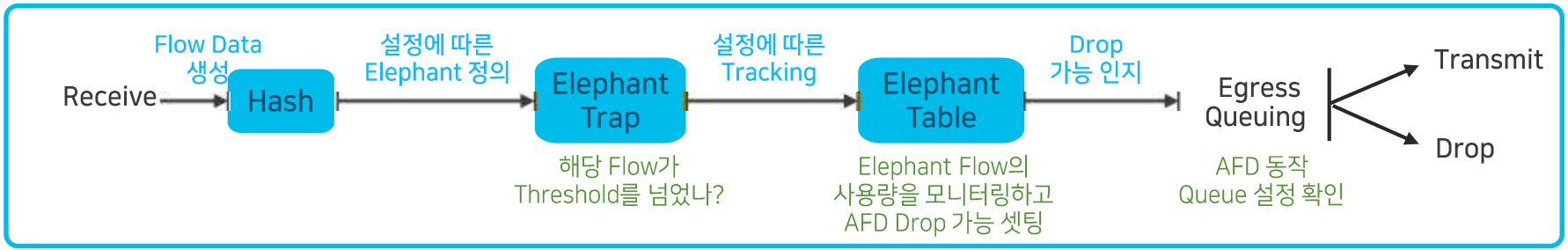


지능형 링크 혼잡 제어 – AFD

Approximate Fare Drop (AFD)

- Egress 포트의 상태를 모니터링하며 적정한(Fair) 임계치(Bytes)를 넘는 Elephant 플로우에 대하여 선제적 Drop
- Elephant 플로우 상태를 모니터링하기 위해 Ingress에서 Elephant Flow Table을 유지

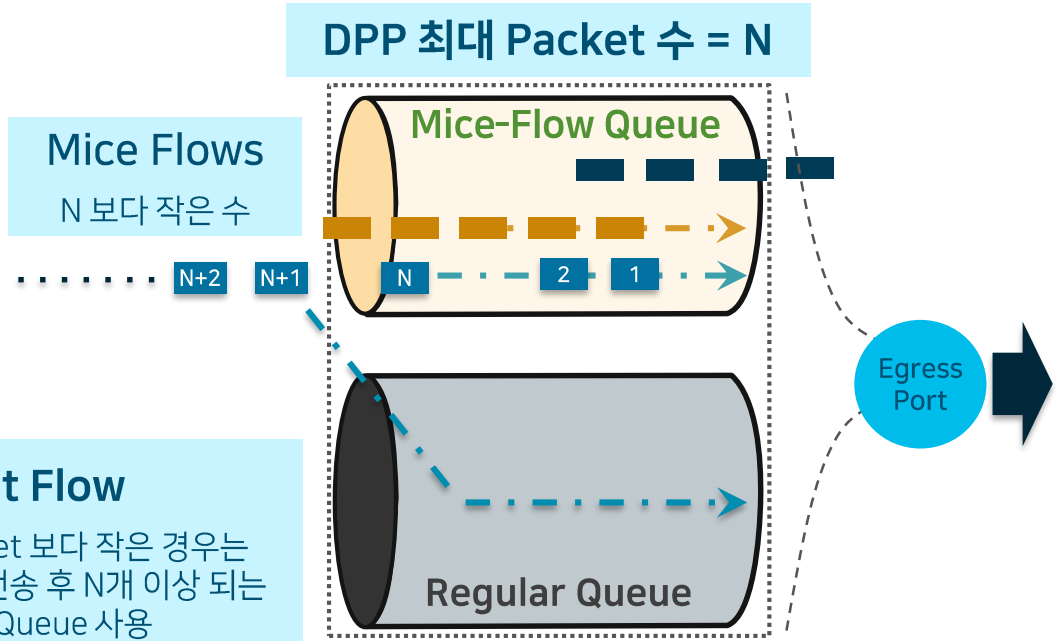




동적 트래픽 우선 처리 – DPP

Dynamic Packet Prioritization (DPP)

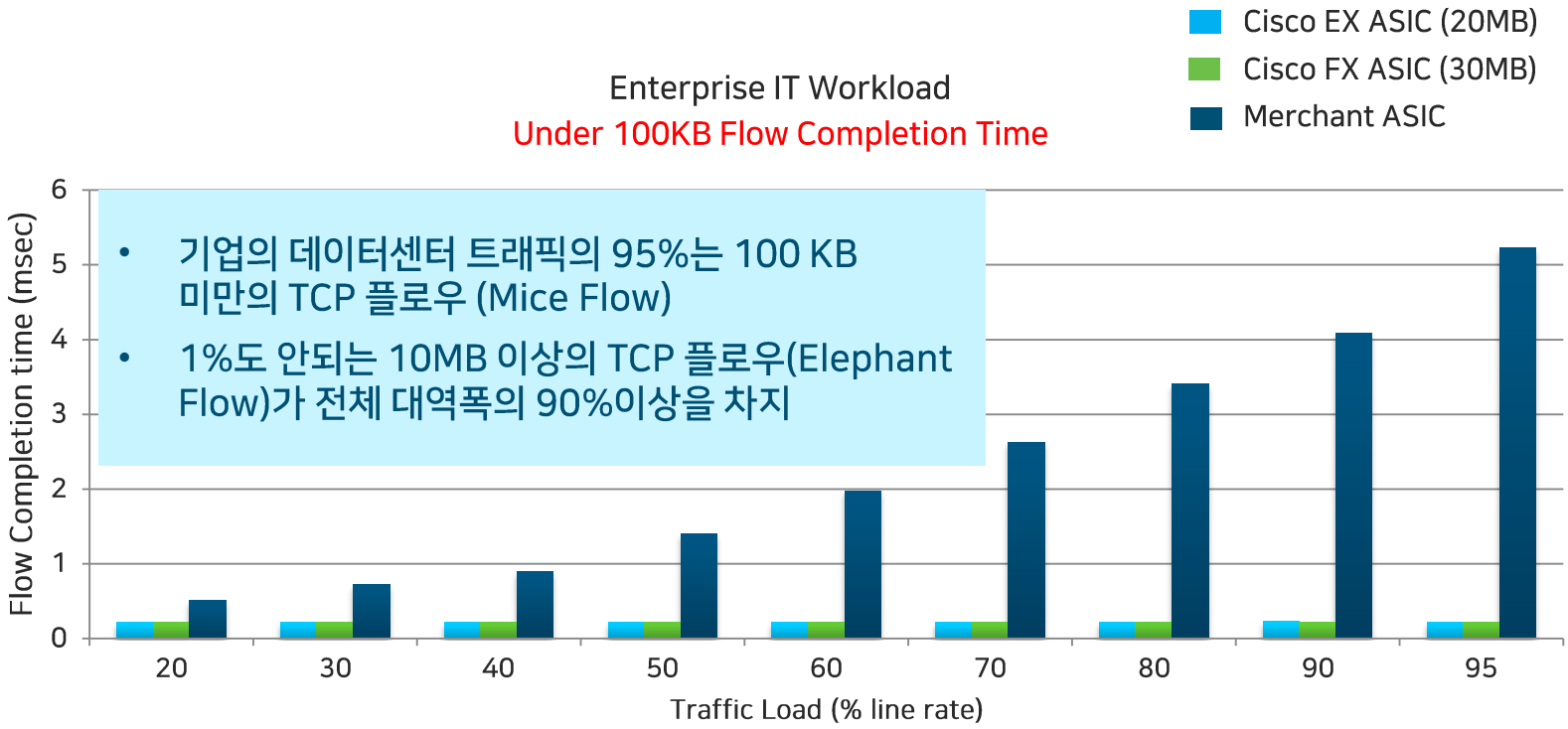
- 애플리케이션별 QoS 정책 적용 없이 작은 플로우(Mice Flow)에 대해 자동으로 높은 우선 순위의 Queue에 할당
- 패킷 수가 작은 플로우(Mice Flow)를 우선적으로 처리
- Mice Flow는 보편적으로 Control Traffic이 해당 됨



Elephant Flow

처음부터 N개의 Packet 보다 작은 경우는 Mice-Flow Queue로 전송 후 N개 이상 되는 경우 Regular Queue 사용

AFD/DPP 기대 효과



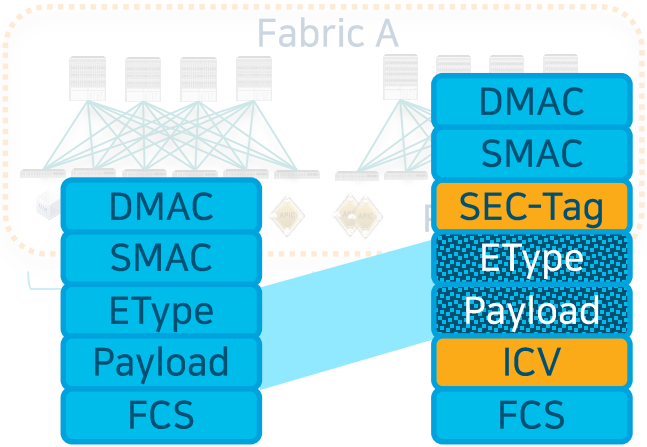
http://miercom.com/cisco-systems-speeding-applications-in-data-center-networks/

End-to-End 암호화 지원

End-to-End 암호화 지원

MACSEC

- Hop by Hop 의 L2 Link Level의 암호화
- FX 기반 이상 ASIC 지원

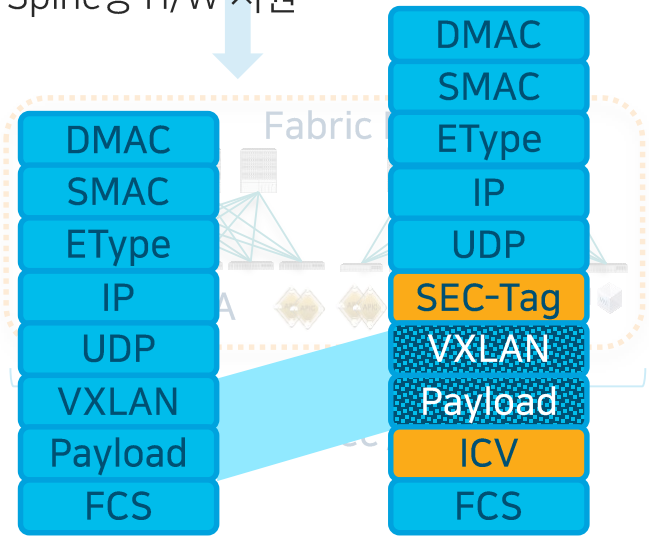


암호화 지원 범위

CloudSec
100G Line Rate의 VxLAN 통신 암호화

CloudSec

- VTEP to VTEP 트래픽의 암호화
- Spine용 H/W 지원

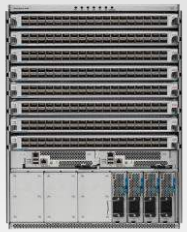


모듈형 – Nexus 9500 Switch

Cisco Nexus 9500 Chassis

Nexus 9504

7 RU
4 x I/O 슬롯



Nexus 9508

13 RU
8 x I/O 슬롯

Nexus 9516

21 RU
16 x I/O 슬롯

I/O 모듈		지원 포트	
	X9736C-FX	36p 40/100G QSFP	ACI / NX-OS
	X9732C-FX	32p 40/100G QSFP	
	X9732C-EX	32p 40/100G QSFP	
	X9736C-EX	36p 40/100G QSFP	NX-OS
	X97160YC-EX	48p 1/10/25G SFP+ + 4p 40/100G QSFP	
	X9788TC-FX	48p 1/10G T + 4p 40/100G QSFP	

고정형 – Nexus 9300 40/100G Switch

Nexus 9300
ACI Spine / NX-OS

9332C

32p 40/100G QSFP

9364C

64p 40/100G QSFP

Nexus 9300-FX2

9336C-FX2

36p 40/100G QSFP

Nexus 9300-EX

93180LC-EX

32p 40/50G
or 18p 100G

고정형 – Nexus 9300 1/10/25G Switch

Nexus 9300 -FX2/FX3

93240YC-FX2

48p 1/10/25 SFP+
+ 12p 40/100G QSFP

93360YC-FX3

96p 1/10/25 SFP+
+ 12p 40/100G QSFP

93216TC-FX3

96p 1/10GT
+ 12p 40/100G QSFP

Nexus 9300-FX

93180YC-FX

48p 1/10/25 SFP+
+ 6p 40/100G QSFP

93108TC-FX

48p 1/10G T
+ 6p 40/100G QSFP

9348GC-FX

48p 100M/1G T + 4p 1/10/25G
+ 2p 40/100G QSFP

Nexus 9300-EX

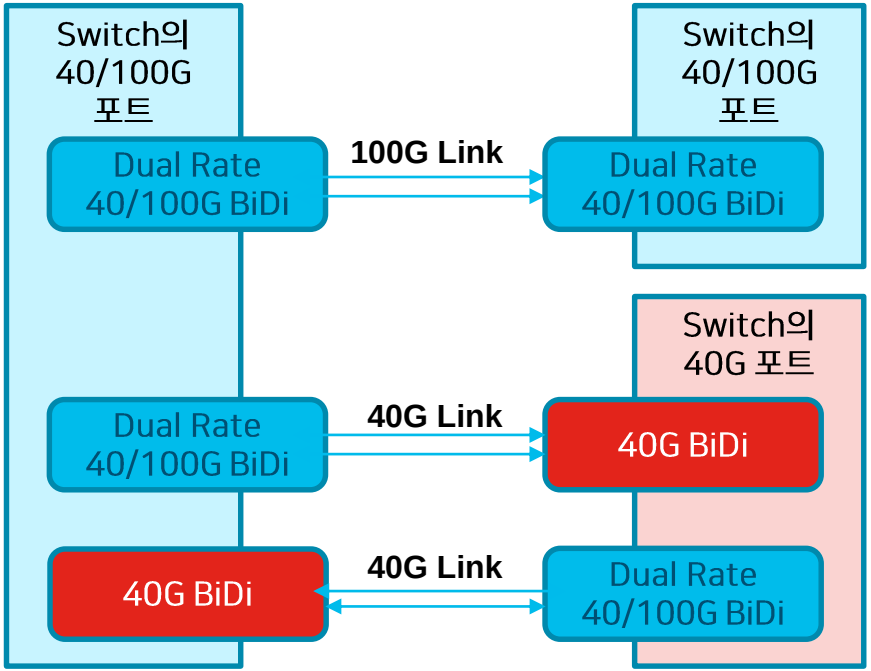
93180YC-EX

48p 1/10/25 SFP+
+ 6p 40/100G QSFP

93108TC-EX

48p 1/10G T
+ 6p 40/100G QSFP

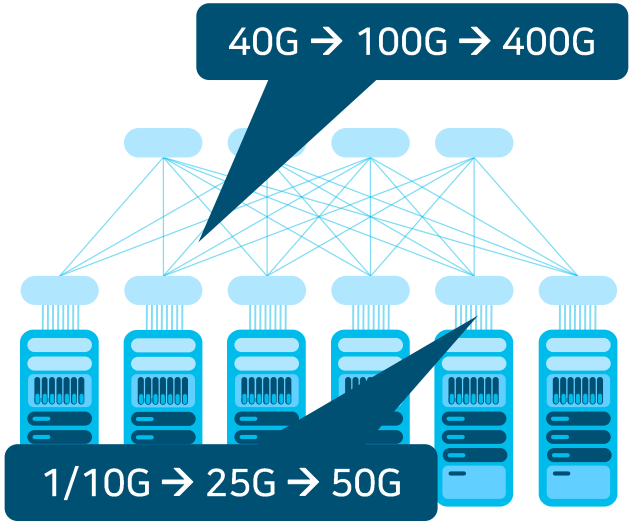
Optics – 40/100G BiDi



- 표준 방식의 QSFP Type
- 기존의 MMF 케이블 사용하여 100G 지원

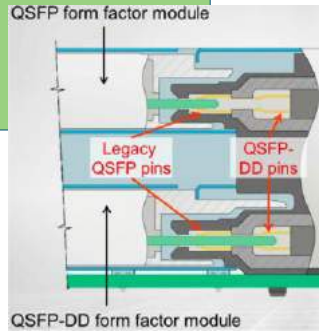
	40G Bidi	40/100G Bidi
PID	QSFP-40G-SR-BD	QSFP-40/100-SRBD
지원 거리 40G OM3 케이블 40G OM4 케이블 100G OM3 케이블 100G OM4 케이블	 100m 150m NA NA	 100m 150m 70m 100m
Form Factor	QSFP	QSFP
커넥터 타입	Duplex LC	Duplex LC

Optics – 400G



기존 QSFP와 동일 크기
1 RU에 36포트

40/100G 모듈 수용 가능
25/50/100G Breakout



포트 밀집도

하위 호환성

ASIC 기술

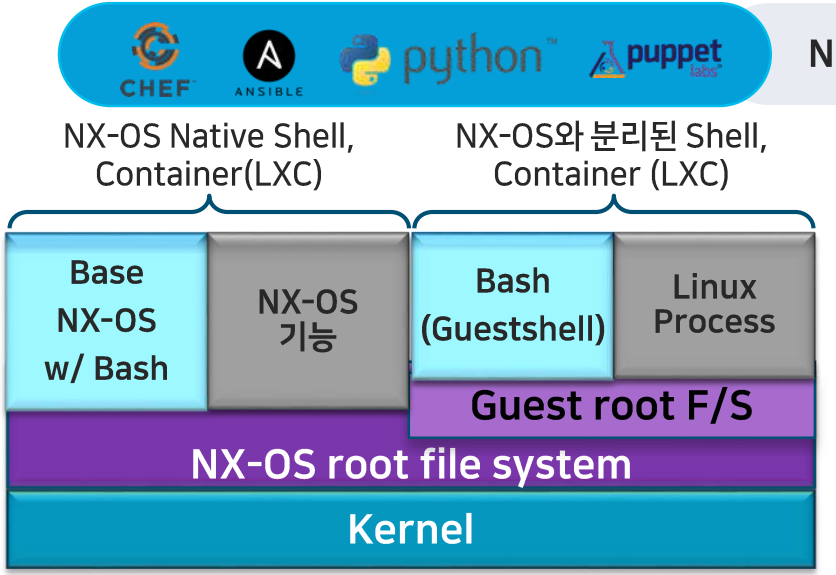
여러 케이블과 거리

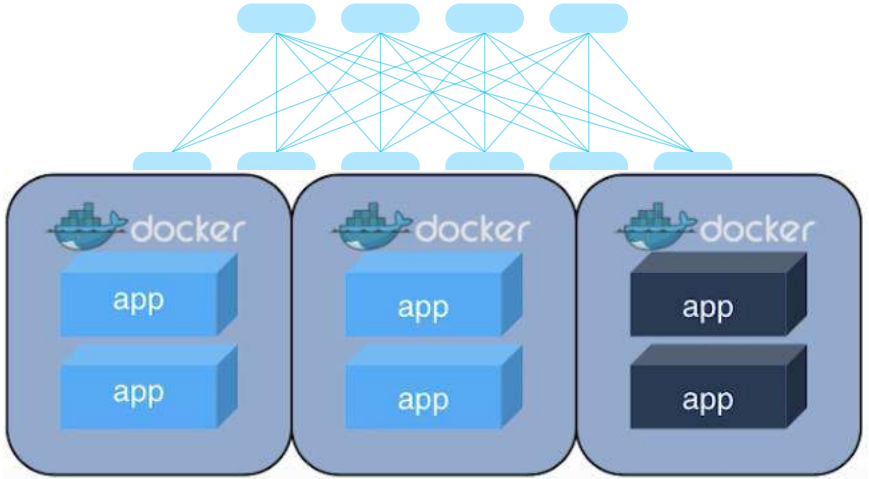
ASIC (16 x 400G)
QSFP-DD 표준화 주도

3m 에서 최장 100Km
Fiber, Copper 지원

Programmability

NX-OS 구조의 혁신

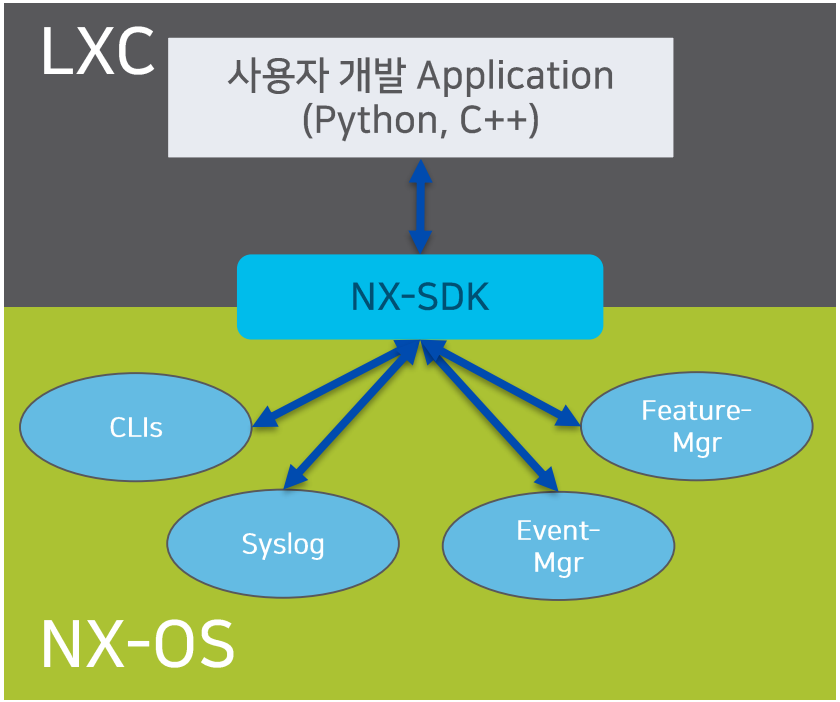




- Guestshell은 Python 설치 되어있는 64bit CentOS
- NX-OS와는 분리되어 있는 사용자 영역으로 NX-OS의 정보 호출을 위해서 NX-API 사용
- Guestshell에 CPU/Mem/FileSystem 등 자원 할당 변경

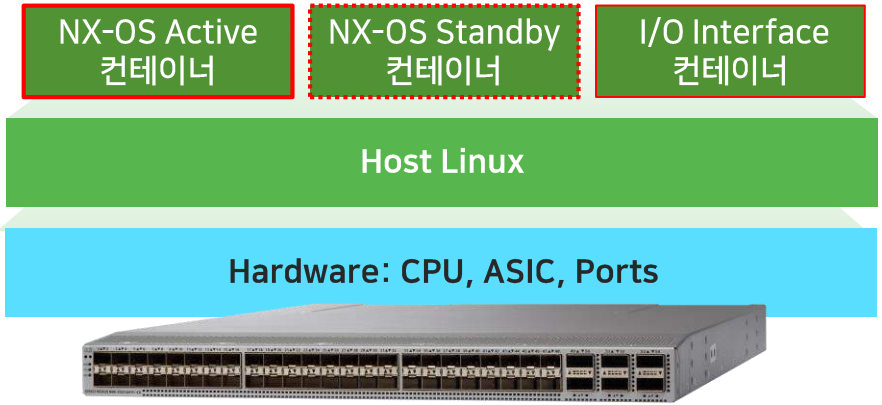
- Native Docker 지원
- 기존 개발 되어 있는 Docker 기반 Application 사용 가능

NX-SDK를 사용한 App 개발 지원



- NX-OS에 접근 가능한 사용자가 개발한 Application 개발 Tool
- 개발한 Application 은 :
  - Python 이나 C++ 로 개발
  - NX-OS의 OSPF 등과 같은 기본 기능으로 동작 가능하며 실행과 관리를 NX-OS를 통함
  - 설정이나 show 명령 등과 같은 Custom CLI 를 생성 가능
  - Custom Syslog, Event, Error 메시지 생성 가능

컨테이너 기반의 Enhanced ISSU



동작 방식

- 이중화 된 Virtual Supervisor 구조로 NX-OS 가 컨테이너로 설치
- 6초 미만의 Control Plane 다운타임
- Zero Packet Loss

특장점

- 1RU의 장비에서 In-Service-Software-Upgrade 지원
- 빠른 Upgrade 시간

가시성 / 분석

Telemetry Workflow

Nexus 9000
Software/Hardware 기반의
Telemetry

Cisco Turnkey 제품
Tetration, DCNM/ACI Fabric Insight, StealthWatch 등

Custom / 3rd Party

Telemetry를 이용하여…

가용성

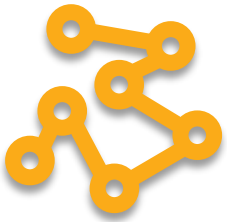
- Interface/Link 의 상태
- Data Plane 의 전반적 정보
- 장비 모니터링
- Protocol 동작 상태

네트워크 운영

- 혼잡 구간 발견 및 Congestion 모니터링
- Buffer 사용 현황
- 큐 단위의 Microburst 탐지

Flow 기반의 분석

- Flow 별 Latency 측정
- Packet 전달 경로와 이상 징후 탐지
- Flow 단위의 Microburst 탐지



Cloud Scale Telemetry Source

Data-Plane의 Traffic 정보

Flow Table (FT)

- 장비를 지나는 모든 패킷을 캡처하여 Flow정보와 Metadata를 생성하여 전달
  - 5-Tuple Flow 정보
  - 인터페이스 큐 정보
  - Flow의 시작/종료 시간
  - Flow 지연 시간

하드웨어 기반의 Export
(약 100msec 단위)

Flow Table Events (FTE)

- Flow Table로 생성 된 정보 기반으로 Threshold나 조건에 따른 Event 기반의 알림
  - 5-Tuple Flow 정보
  - 인터페이스 큐 정보
  - Buffer. FWD, ACL, Policer Drop
  - 지연/Burst Threshold 초과

Flow-Level과 설정에 따른 하드웨어 기반의 Export

ASIC 상태

Streaming Statistics Export (SSX)

- ASIC의 상태 정보를 사용자 설정에 따라 Streaming
  - 인터페이스와 큐 Statistics (packets/bytes/drops)
  - Buffer depth
  - 기타 ASIC 기반의 카운터

약 10usec 기반으로 수집하여 하드웨어 기반으로 Export

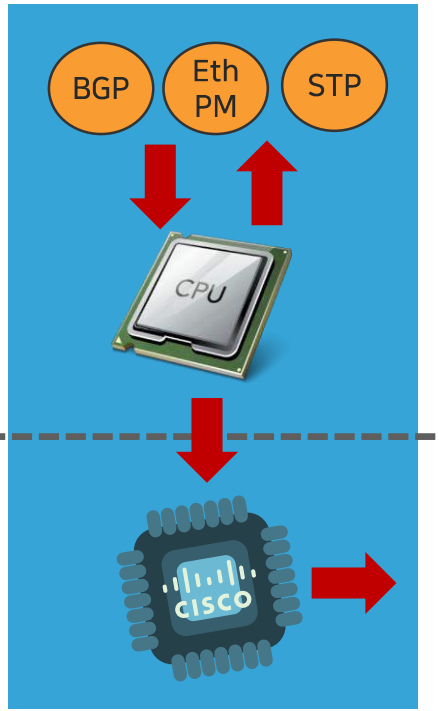
EX ASICs | **FX ASICs** | **FX2 이상 ASICs**

S/W 기반

- NX-OS 상의 CLI로 볼 수 있는 모든 정보 : 카운터, 프로토콜 상태, Control Plane 정보 등

하드웨어 기반 Telemetry 동작 방식

S/W 기반 Telemetry



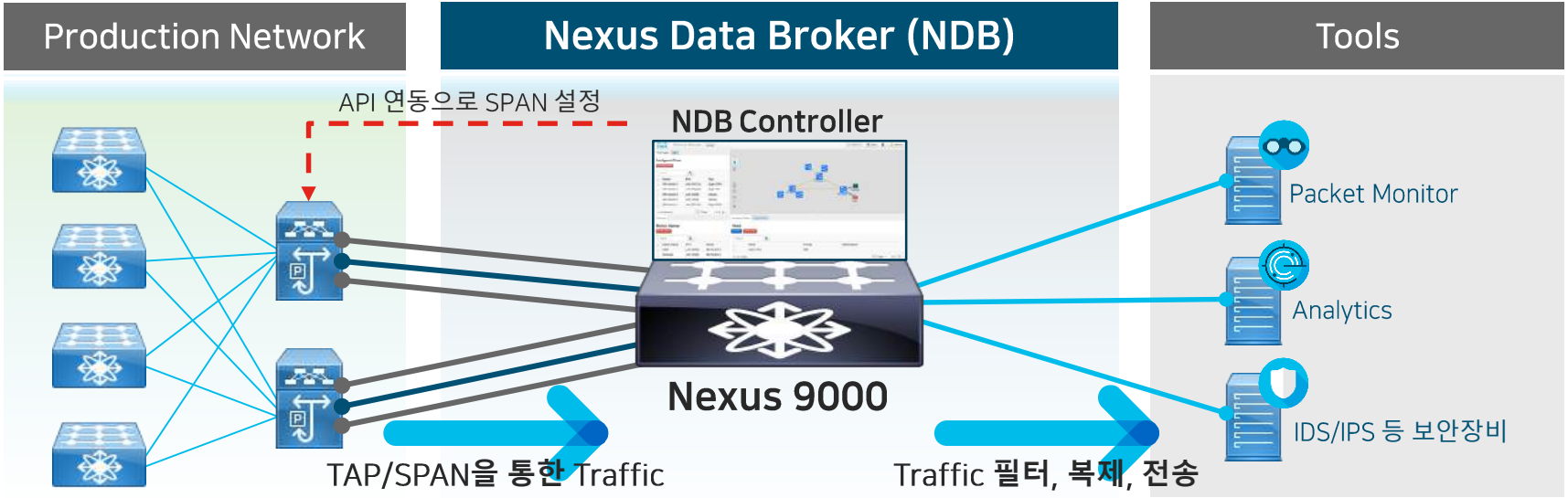
필요한 Telemetry Data 관련 설정
(S/W, H/W 기반의 모든 정보)

CPU 기반의 Telemetry
Data Export
(SNMP, Syslog, Netflow 등)

S/W 기반 Telemetry

ASIC에서 직접 H/W 기반의
Telemetry Data Export

TAP-SPAN Aggregation – Nexus Data Broker

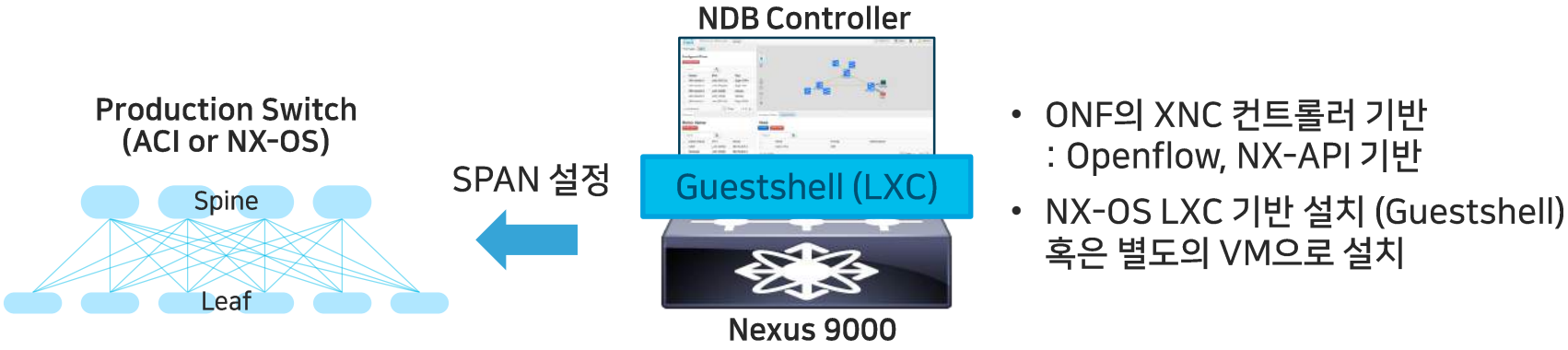


- NDB를 통한 Nexus/ACI의 SPAN 설정 중앙 관리

- Nexus 9000 하드웨어에 별도의 NDB Controller 앱 설치 후 동작
- Traffic 분배 뿐 아닌 Flow Data 생성 및 전달 지원 (Flow Generator)

- Traffic 필터링, 복제 등의 기능을 통하여 여러 Tool로 트래픽 전송

NDB 주요 기능



- ONF의 XNC 컨트롤러 기반
  : Openflow, NX-API 기반
- NX-OS LXC 기반 설치 (Guestshell)
  혹은 별도의 VM으로 설치

Packet 분류 및 분배

- Layer 1~4, HTTP, 사용자 정의 방식의 Filtering
- 1:N, N:N 등 다양한 방식으로 Packet 전송
- Symmetric / Asymmetric Hash 방식의 LB

Edit Packet

- PTP 기반의 Timestamp 생성
- Packet Truncate (Payload 제거)
- VLAN/MPLS Tag 제거
- Q-in-Q 지원

부가 서비스

- SPAN/TAP으로 전송 된 Packet을 바탕으로 sFlow, Netflow 생성 및 Export
- ERSPAN Destination 역할로 원격지 Packet 수집

지능형 서비스

NX-OS의 지능형 서비스

Load Balance

L2 기반 LB – **Smart Channel**
L3/L4 기반 LB - **ITD (Intelligent Traffic Director)**
VxLAN Fabric 분산형 LB – **PLB (Pervasive Load Balance)**

TCAM 모니터링

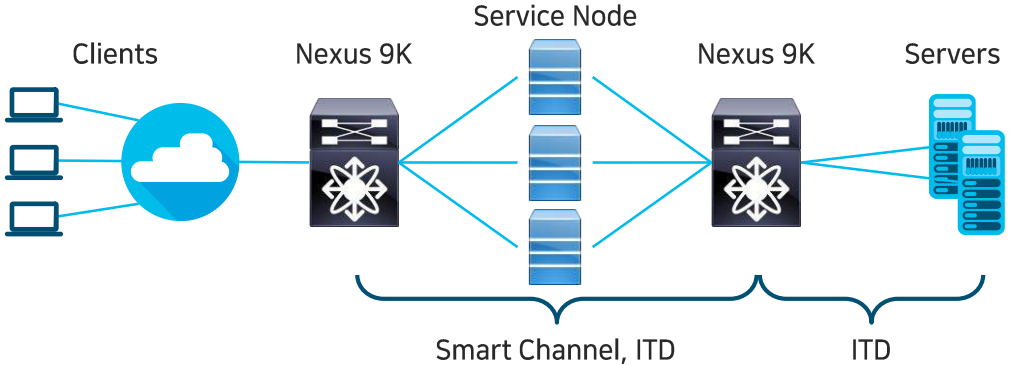
ACL, QoS, FIB 등 TCAM을 사용 현황의 모니터링 - **iCAM**

서비스 체인

트래픽 경로 상 방화벽, IPS 등 서비스 노드의 추가/삭제 - **Catena**

다양한 Load Balance 서비스

Smart Channel

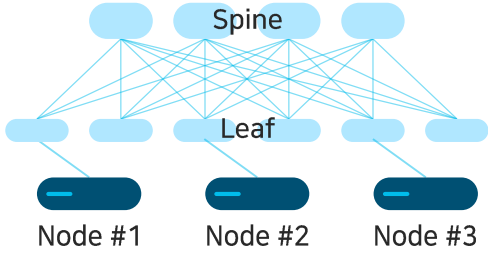


- L2 기반의 Symmetric LB
- 하드웨어 기반의 Wire-Rate 성능
- 연결 된 장비의 Health Check
- ACL 등을 통한 선별적 Redirect
- N+M Redundant

ITD (Intelligent Traffic Director)

- L3/L4 기반, VIP, NAT의 LB, SLB
- 하드웨어 기반의 Wire-Rate 성능
- 연결 된 장비의 Health Check
- ACL 등을 통한 선별적 Redirect
- N+M Redundant

PLB (Pervasive Load Balance)



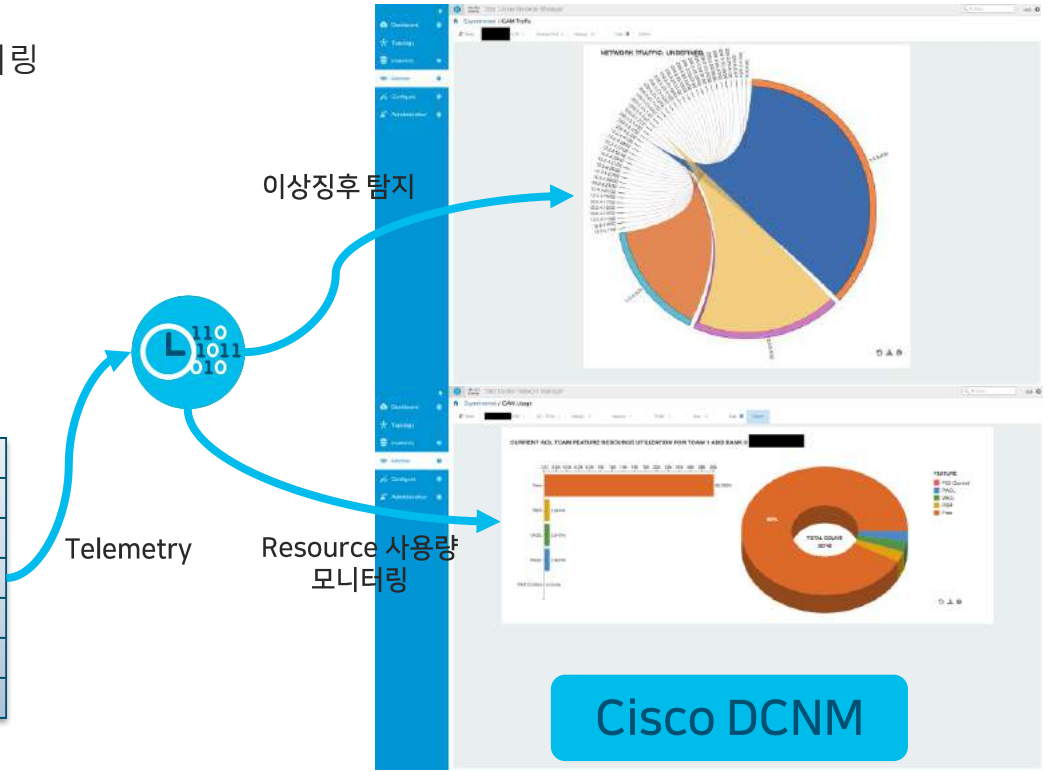
- Fabric 내 분산 배치 된 장비의 LB
- VXLAN/EVPN 지원
- 연결 된 장비의 Health Check
- DSR, Non-DSR SLB, FLB 지원

TCAM 사용량 모니터링 – iCAM

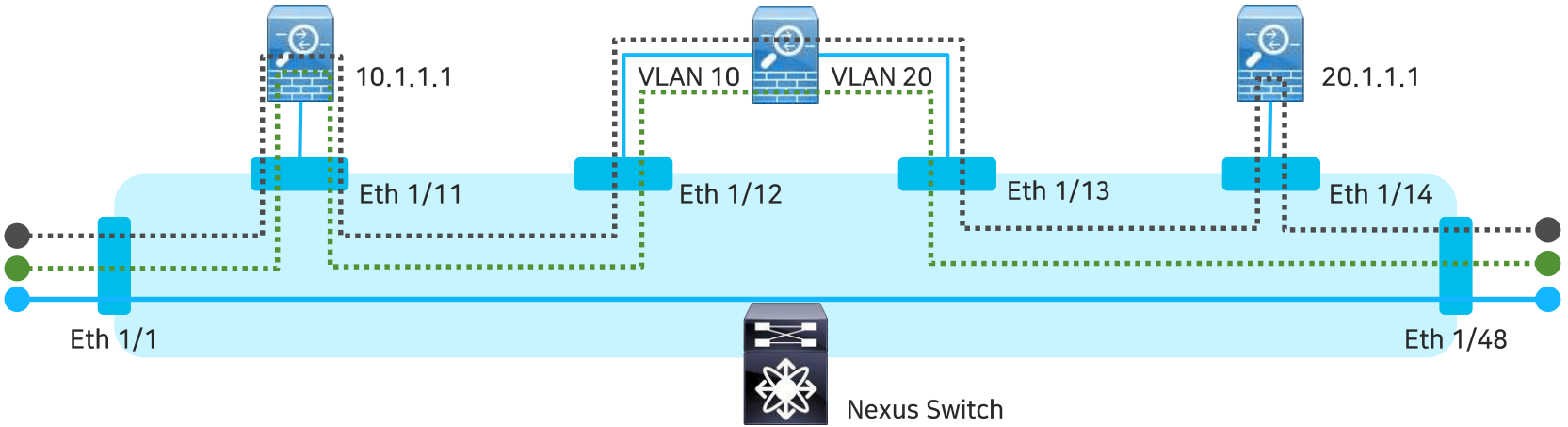
- ACL/QOS/PBR 등 TCAM 사용 현황에 대한 모니터링
  - 어떤 ACL이 현재 사용 되는 것인지 아닌지
  - 어떤 L4 기반 트래픽이 많이 사용 되는지
  - 가장 많은 통신이 일어나는 트래픽이 어떤 것인지
- 기능 별 TCAM, SRAM 자원 사용량 모니터링
- Historical Data 기반으로 미래의 사용량 예측
- CLI 및 Data Export로 모니터링 가능

TCAM Entry

Feature	ACE entry	Action	Packet #
RACL	2.2.2.0/24 → 3.3.3.0/24	Permit	1405
RACL	4.4.4.0/24 → 5.5.5.0/24	Deny	2550
PBR	2.1.1.1/32 → 3.1.1.1/32	Rewrite	2983
QOS	4.1.1.1/32 → 5.1.1.1/32	Redirect	1472
…	…	…	…
…			



서비스 체이닝 – Catena



- 하드웨어 기반의 Wire-Rate 성능을 제공하는 서비스 체이닝
- Packet 헤더의 변조 없이 동작
- Service Node의 Health Monitoring을 통한 자동화 된 Node의 추가/삭제
- ACL 기반으로 Traffic에 따라 선별적인 Service 적용

- 각각의 서비스 체인 별 Telemetry 정보 제공
- 이를 위한 별도의 추가 하드웨어 불필요
- Service Node와 연동을 위한 별도의 인증 제도, 연동, 호환성 이슈 없음

요약

Hardware + Software Innovation

Switching Infrastructure

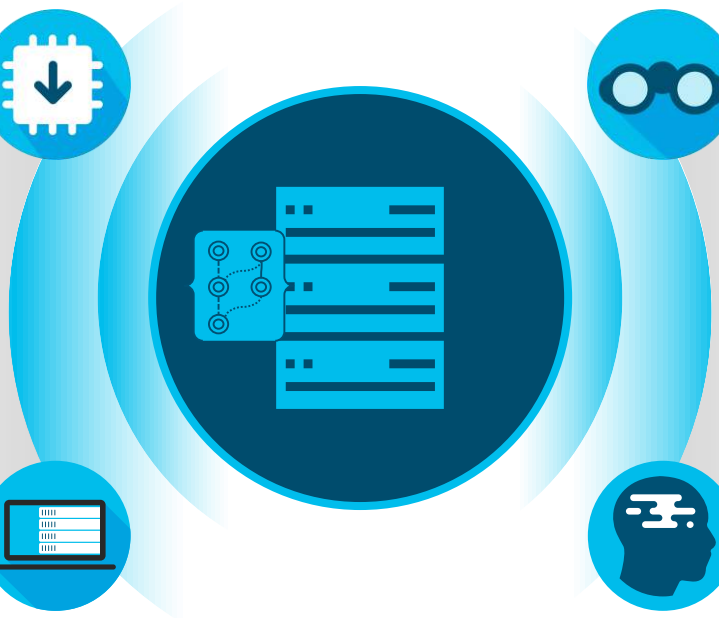
Cloud Scale ASICs

효율적인 40/100/400G

가시성 / 분석

다양한 Telemetry Source

모니터링을 위한 기능



Programmability

컨테이너 기반의 NX-OS

다양한 API/SDK 지원

지능형 서비스

다양한 방식의 L4/L7 서비스 지원

TCAM 사용 현황


CISCO