P4IX: A Concept for P4 Programmable Data Planes at IXPs

Daniel Wagner
Matthias Wichtlhuber, Christoph Dietzel, Jeremias Blendin and Anja Feldmann
Internet Exchange Points (IXPs)

- Networking facility
- Exchange of AS traffic
Internet Exchange Points (IXPs)

- Networking facility
- Exchange of AS traffic

- 920 IXPs globally

1 According to peeringDB, accessed 2021-11-24
ASes per IXP

CDF of connected ASes/IXP

# connected ASes/IXP
(total IXP: 920, total connected ASes: 42388)
ASes per IXP

75% of the IXPs can run on a single 32/64 port switch

CDF of connected ASes/IXP

# connected ASes/IXP
(total IXPs: 920, total connected ASes: 42388)
ASes per IXP

75% of the IXPs can run on a single 32/64 port switch

50% of the IXPs have <15 member ASes
ASes per IXP

75% of the IXPs can run on a single 32/64 port switch

50% of the IXPs have <15 member ASes

5% with >200 members

CDF of connected ASes/IXP

# connected ASes/IXP
(total IXPs: 920, total connected ASes: 42388)
ASes per IXP

• Takeaways
  • Most IXPs can run on a single box
  • IXPs vary greatly in size
  • 46 ASes per IXP on average
ASes per IXP

• Takeaways
  • Most IXPs can run on a single box
  • IXPs vary greatly in size
  • 46 ASes per IXP on average
ASes per IXP

• Takeaways
  • Most IXPs can run on a single box
  • IXPs vary greatly in size
  • 46 ASes per IXP on average

Border Gateway Routers
ASes per IXP

• Takeaways
  • Most IXPs can run on a single box
  • IXPs vary greatly in size
  • 46 ASes per IXP on average

• Consequences
  • IXPs’ market share is tiny compared to that of ISPs
  • Hardware vendors tailor features for ISPs
  • IXP specific requirements not covered
IXP Setups

Single Edge IXP

Multi Edge IXP

Multi Metro IXP
IXP Setups

- BGP FlowSpec inappropriate
- Prone to Layer-2 loops

Single Edge IXP

Multi Edge IXP

Multi Metro IXP
IXP Setups

- **Single Edge IXP**
  - AS 19
  - AS 57
  - AS 6
  - AS 307
  - Prone to Layer-2 loops
  - BGP FlowSpec inappropriate

- **Multi Edge IXP**
  - Multi Metro IXP
  - AS 181
  - AS 206
  - AS 13
  - AS 4006
  - AS 27
  - AS 981
  - Very large Layer-2 domain

- **Large Layer-2 domain**
IXP Setups

Single Edge IXP
- AS 57
- AS 19
- AS 6
- AS 307

Prone to Layer-2 loops

Large Layer-2 domain

Routing required, but Layer-2 facade should be upheld

Virtualization required

BGP FlowSpec inappropriate

Multi Edge IXP

Multi Metro IXP

Very large Layer-2 domain
IXP Setups

• Takeaways
  • IXPs have specific requirements
  • They differ from ISPs’ ones
  • Complex workarounds to implement solutions
IXP Setups

• Takeaways
  • IXPs have specific requirements
  • They differ from ISPs’ ones
  • Complex workarounds to implement solutions

• Consequences
  • IXPs left to buy expensive hardware with unsuitable feature set
  • We propose P4IX
P4IX

- Scalable, P4-based implementation of an IXP

![Diagram showing Current IXP Stack and P4IX Stack](image_url)
P4IX

• Access & Optical layer untouched
P4IX

• Combine virtualization & routing layer
P4IX

• Tight integration of business logic & tailored monitoring
P4IX

- Value added services in P4 for service differentiation
P4IX Concept

Management Layer

P4 Pipeline

Parse Ethernet Headers, VLAN header(s), and IPv4/6 headers → TR2
Apply rate limit based on static port / VLAN / MAC / bandwidth table → TR5

Input from Port → Ingress Rate Limiter → Ingress MAC/IP Filter → Traffic Classification

Drop ingress packets matching certain L2/L3/L4 criteria. Source MAC is checked for VLAN admission → TR3
Classify packets into ARP / NDP, BUM, remaining traffic by matching certain Ethernet IP criteria → TR3
Rewrite packet to ARP / NDP response based on static IP-to-MAC table → TR4
Mark BUM traffic for rate limiting → TR3

Classify packets into local delivery or remote delivery based on static destination MAC table, determine group of output ports → TR1
Set output port based on static Ethernet table information, apply load balancing across multiple ports → TR1
Sample packets for statistics → TR6

Load Balancing and Fallback → Packet Sampling → Output to Port

Egress MAC/IP Filter → ARP / NDP Handler → BUM Handler → Dist. Classification

Pull in business logic from external database and convert to P4 table representation → OR1
Exposé P4 tables via P4runtime API to management tools → OR1

Database Client → Message Broker Client → SSH Access/Shell

P4runtime

Network OS

Allow for NOC management with standard Linux tools → OR2
Exposé information for monitoring & analysis → OR3
Drop egress packets matching certain L2/L3/L4 criteria. Destination MAC is checked for VLAN admission → TR3
Apply rate limit based on static port / VLAN / MAC / bandwidth table → TR5

Deparse Ethernet Headers, VLAN header(s), and IPv4 / IPv6 headers → TR2
P4IX Concept

• **Management Layer**
  • Accesses IXP customer data base and P4 device
  • Reads configuration of customer ports
  • Translates it into P4 table format
  • Populates device’s P4 tables through P4 runtime

• Reads statistical information from P4 device
• Puts it into a publish-subscribe broker system

• Offers SSH access for debugging
P4IX Concept

• Input from port
  • P4 stage to parse packet
P4IX Concept

• Ingress rate limiter
  • Shapes or drops traffic
  • According to member’s purchased service