
AWS
re:Invent
DECEMBER 2 – 6, 2024 | LAS VEGAS, NV

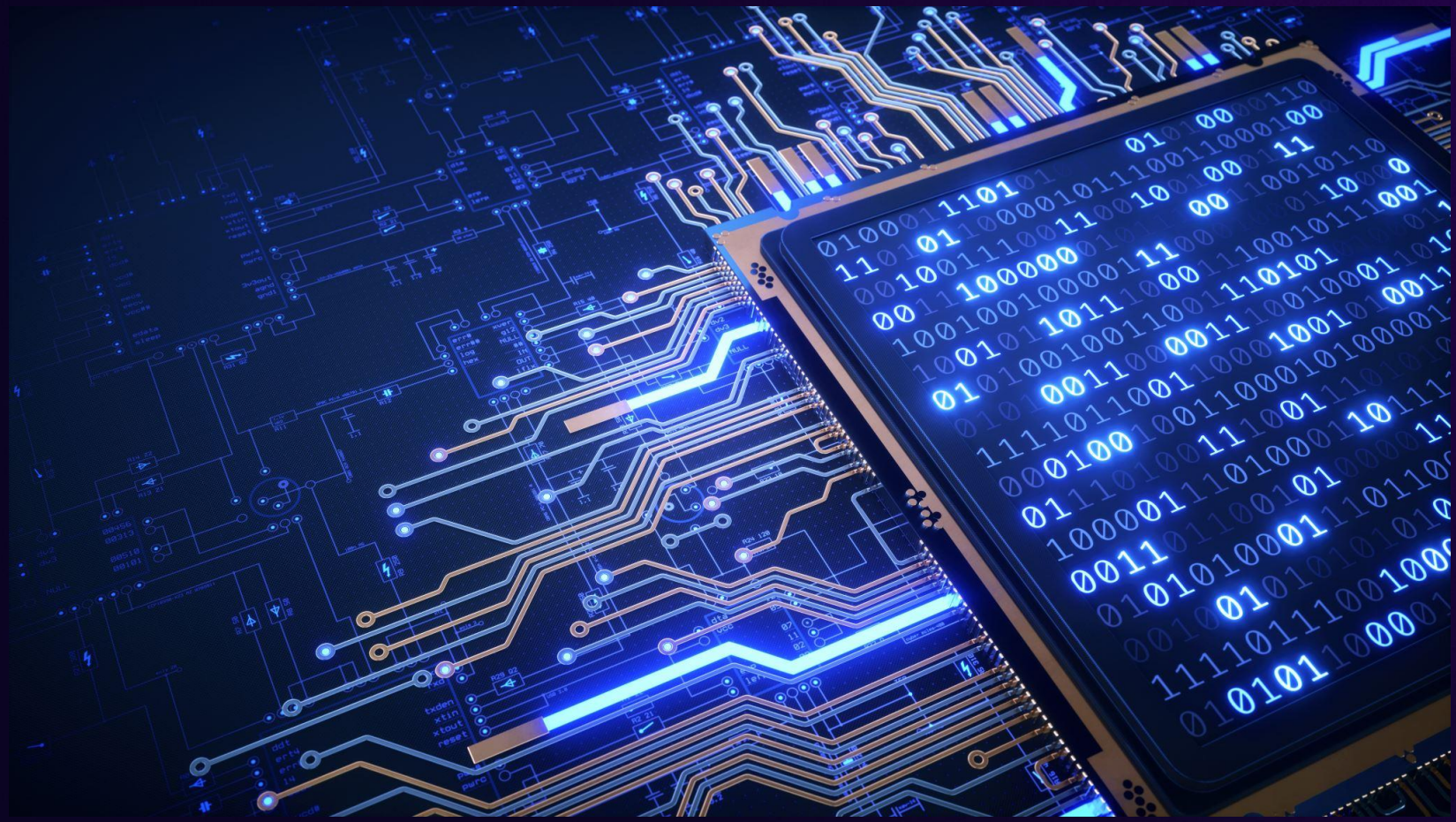
NET402

EC2 Nitro networking under the hood

John Pangle
(he/him)
Senior Product Manager Technical
AWS

Scott Wainner
(he/him)
Principal Solution Architect
AWS

Why are you here?



Why are we here?

VPC Nitro Networking Session Agenda

Packet Analysis

Flow Analysis

Multi-Flow Analysis

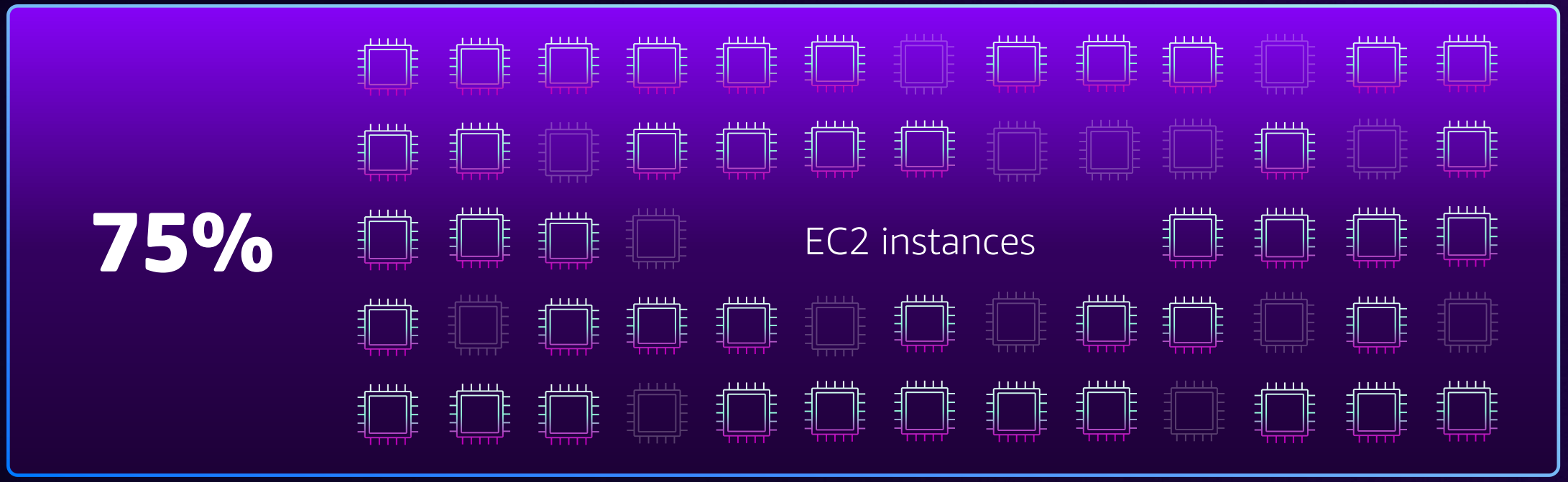
Tools

Action Plan

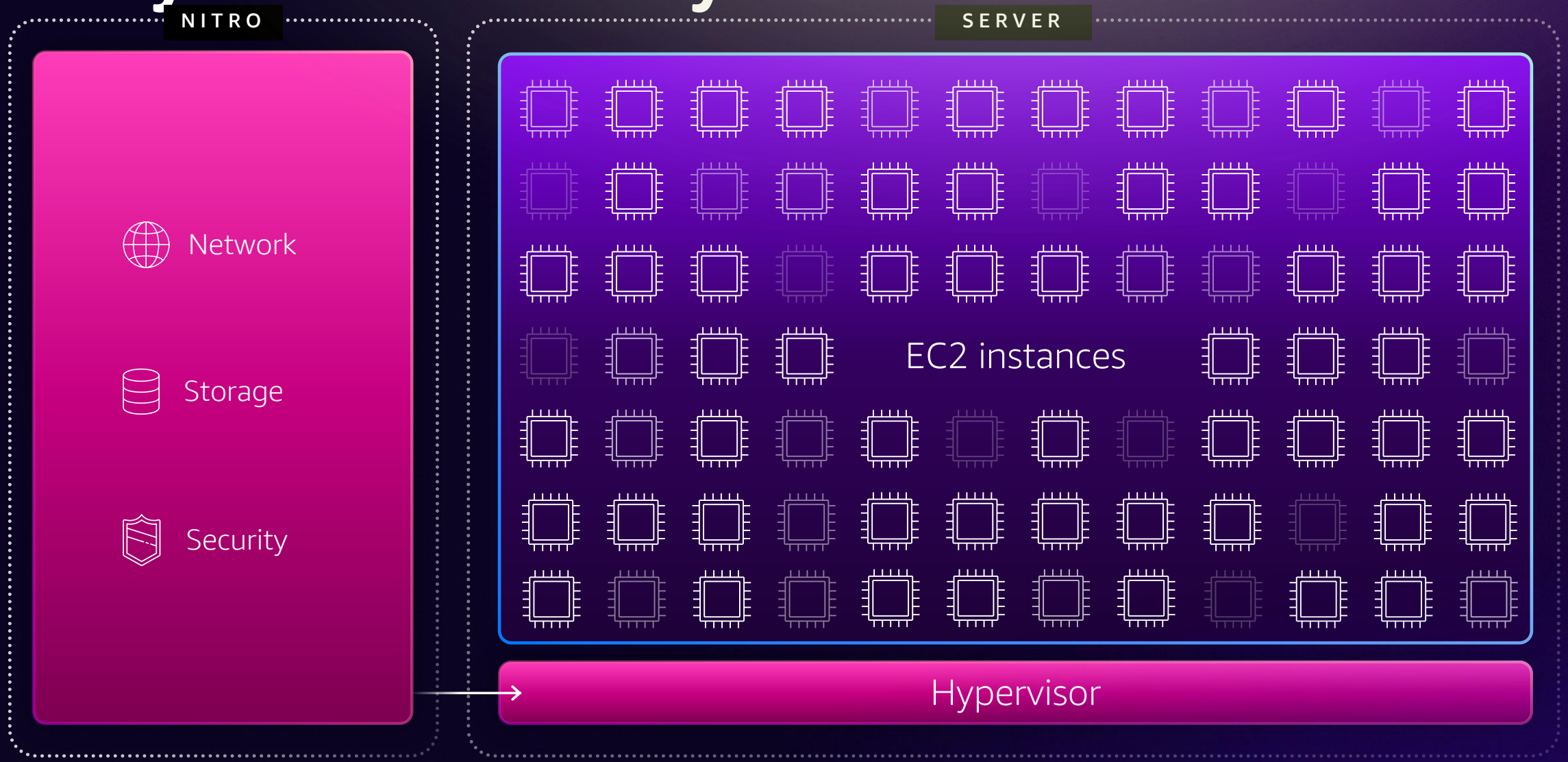
Introducing Nitro

History of AWS Nitro System

SERVER



History of AWS Nitro System



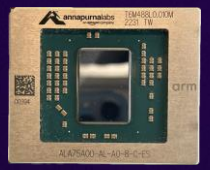
Nitro History

Nitro History

Constantly improving performance and security







Nitro Generation	Nitro v1	Nitro v2	Nitro v3	Nitro v4	Nitro v5
EC2 Instance	C4	C5	C5n	C6in	C7gn
Bandwidth	10 Gbps	25 Gbps	100 Gbps	200 Gbps	200 Gbps
Packet Rate		2x	3x	2x	1.5x
Encryption in Transit			X	X	X

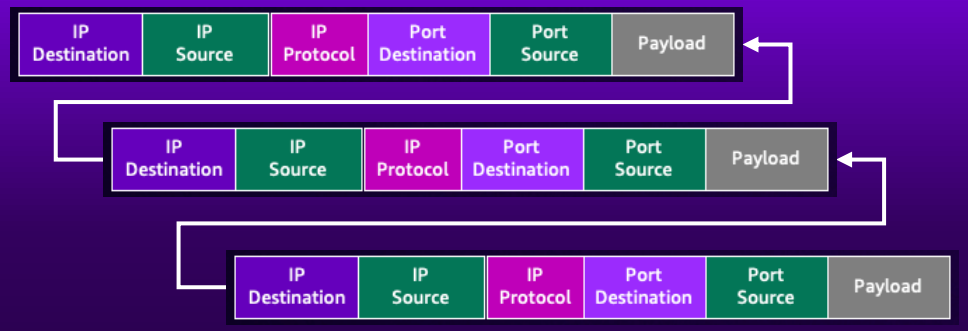
What are we going to cover today?

Packet Analysis

What are we going to cover today?

Packet Analysis

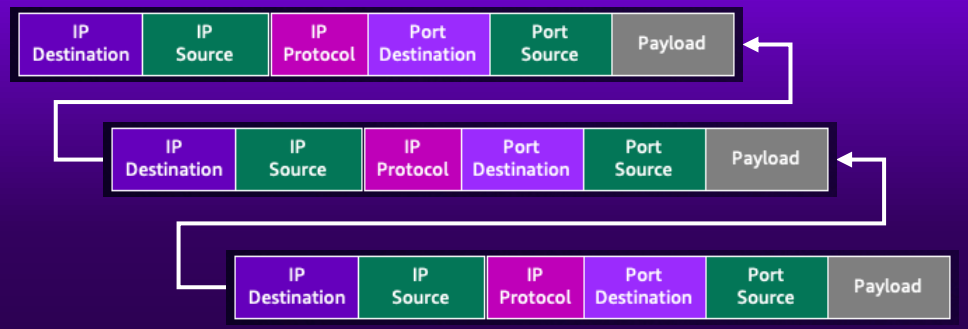
Flow Analysis



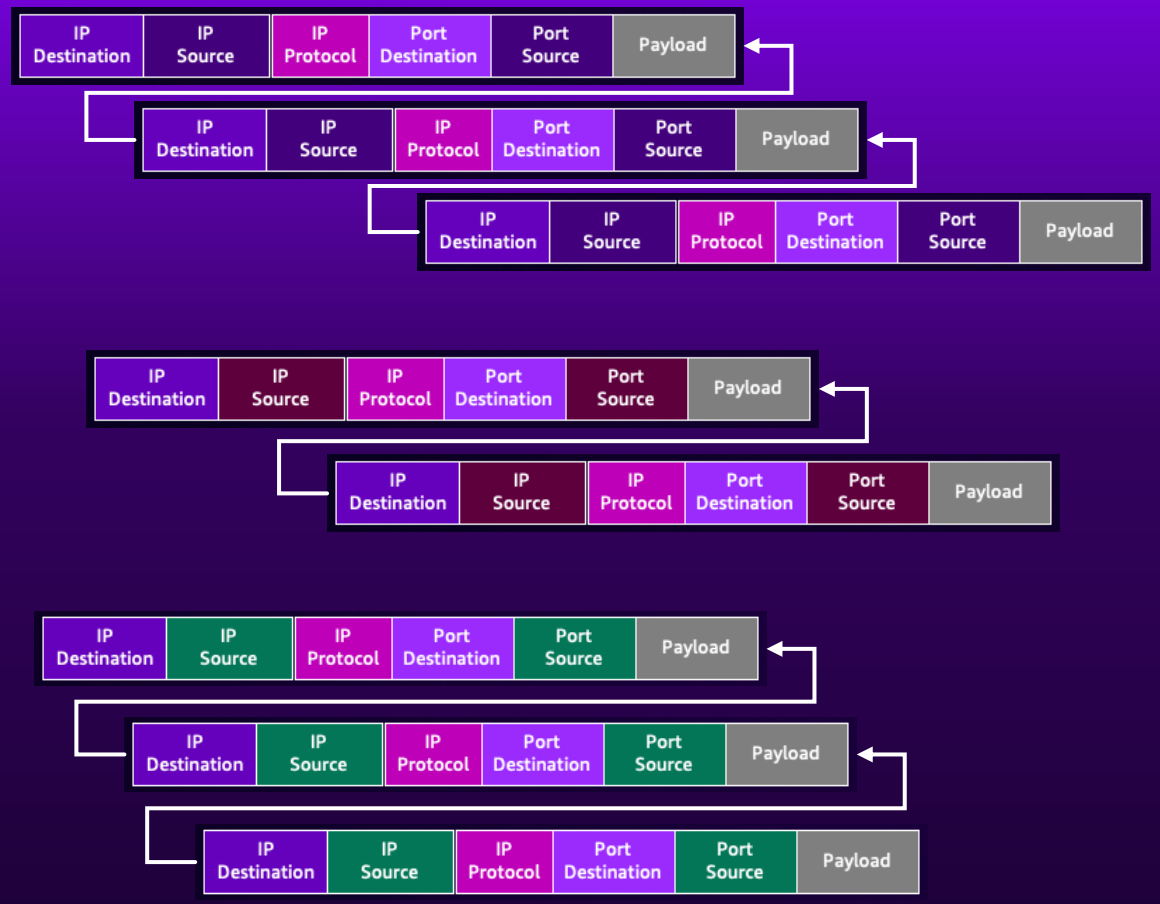
What are we going to cover today?

Packet Analysis

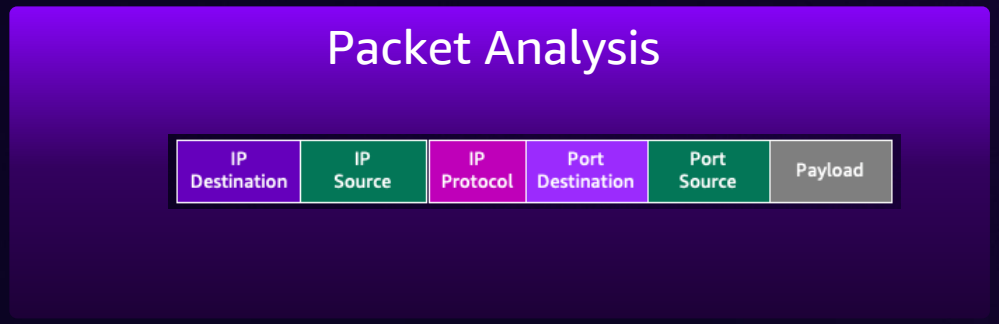
Flow Analysis



Multiple Flows

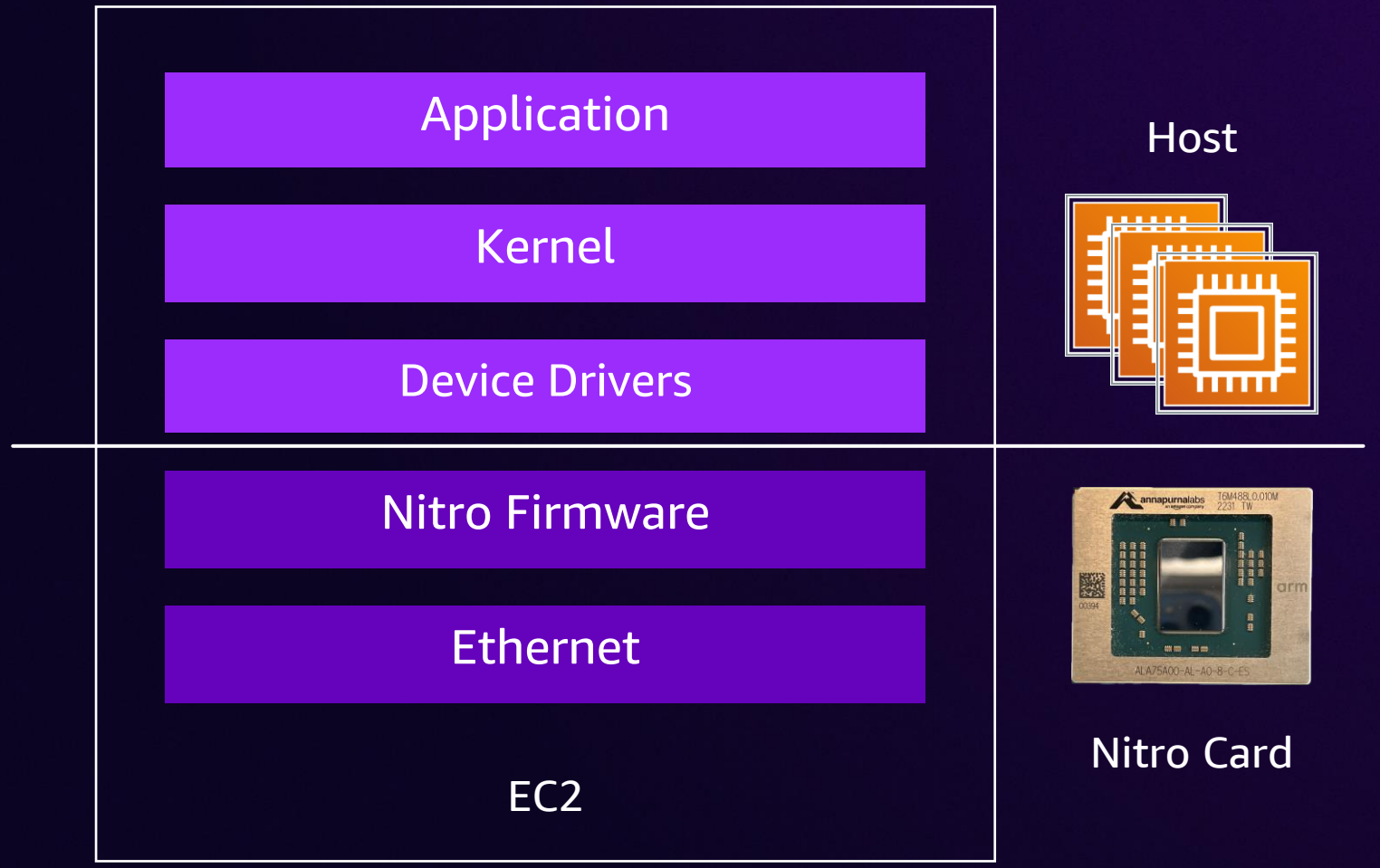


Starting with Packet Analysis

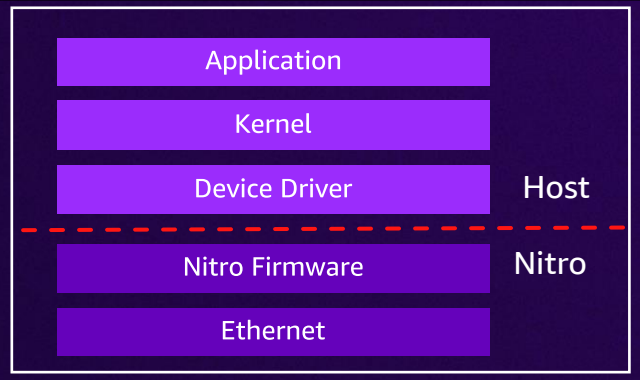


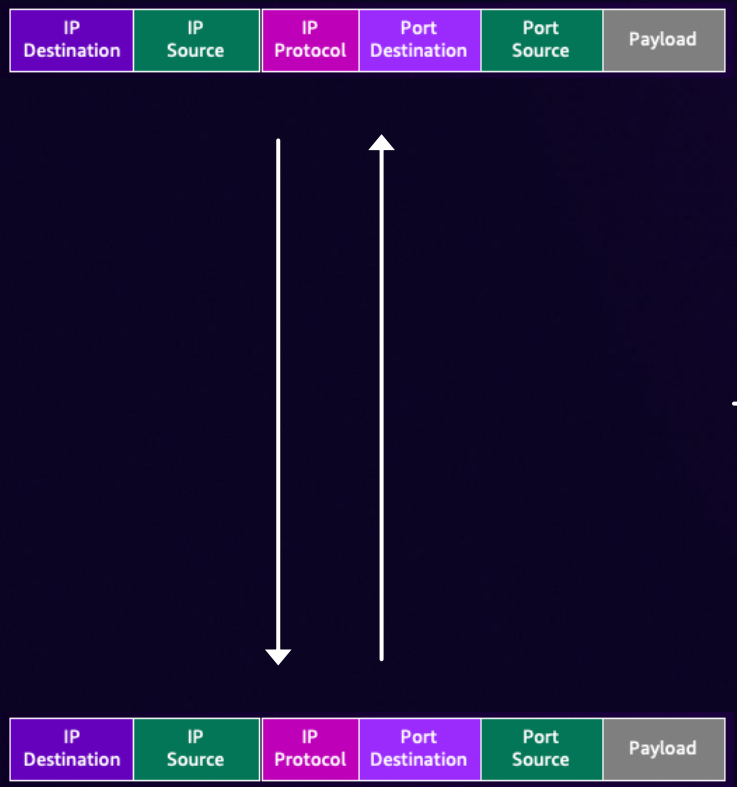
- Processing Stack
- VPC Overview Flows
- Packet Processing
- Nitro State Machine

Processing Stack

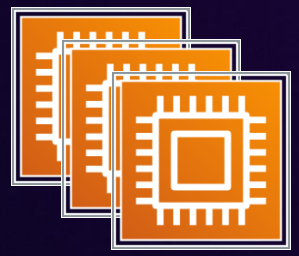


Processing Stack



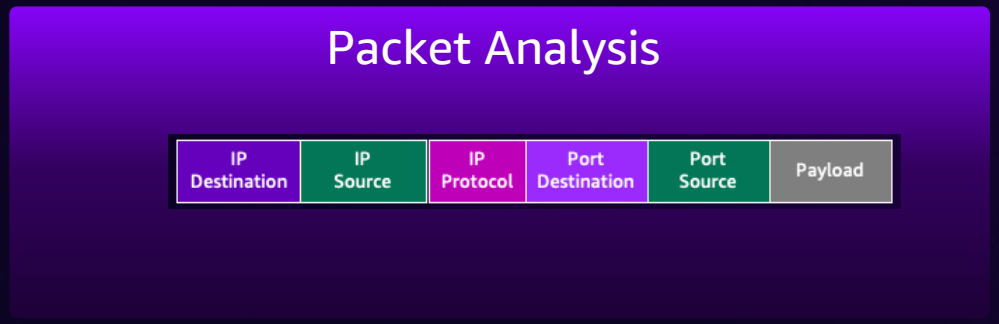


Host



Nitro Card

VPC Overview



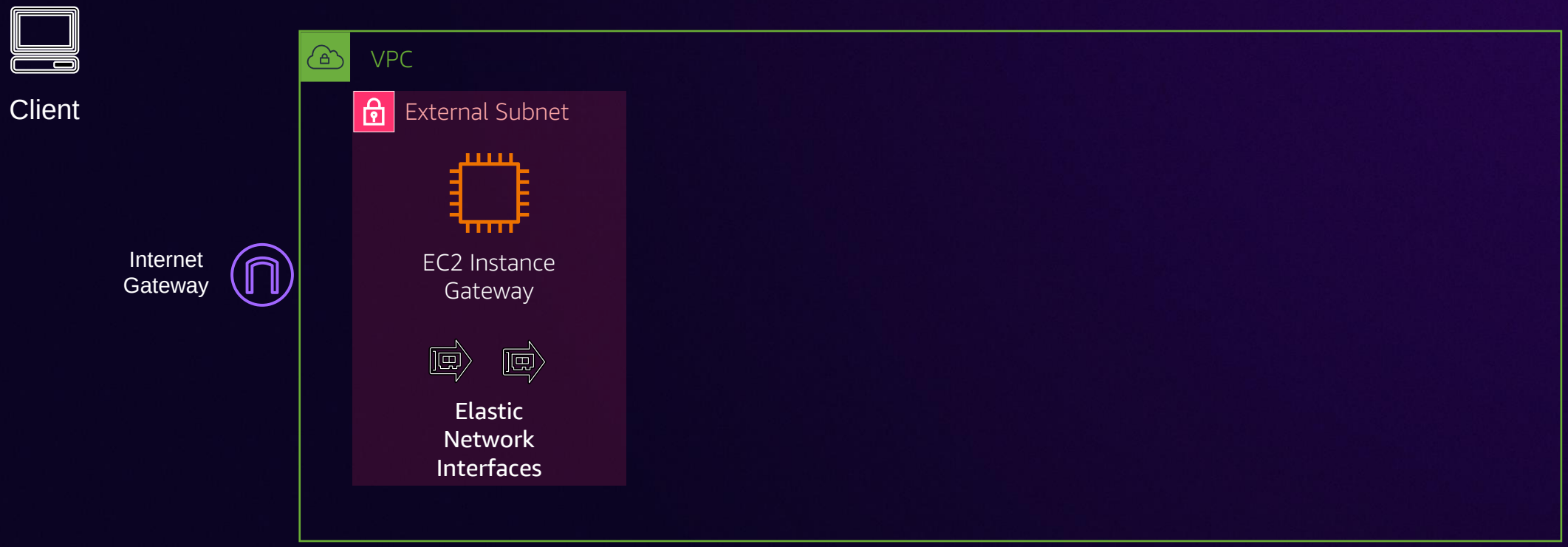
- Processing Stack
- VPC Overview Flows
- Packet Processing
- Nitro State Machine

VPC Overview

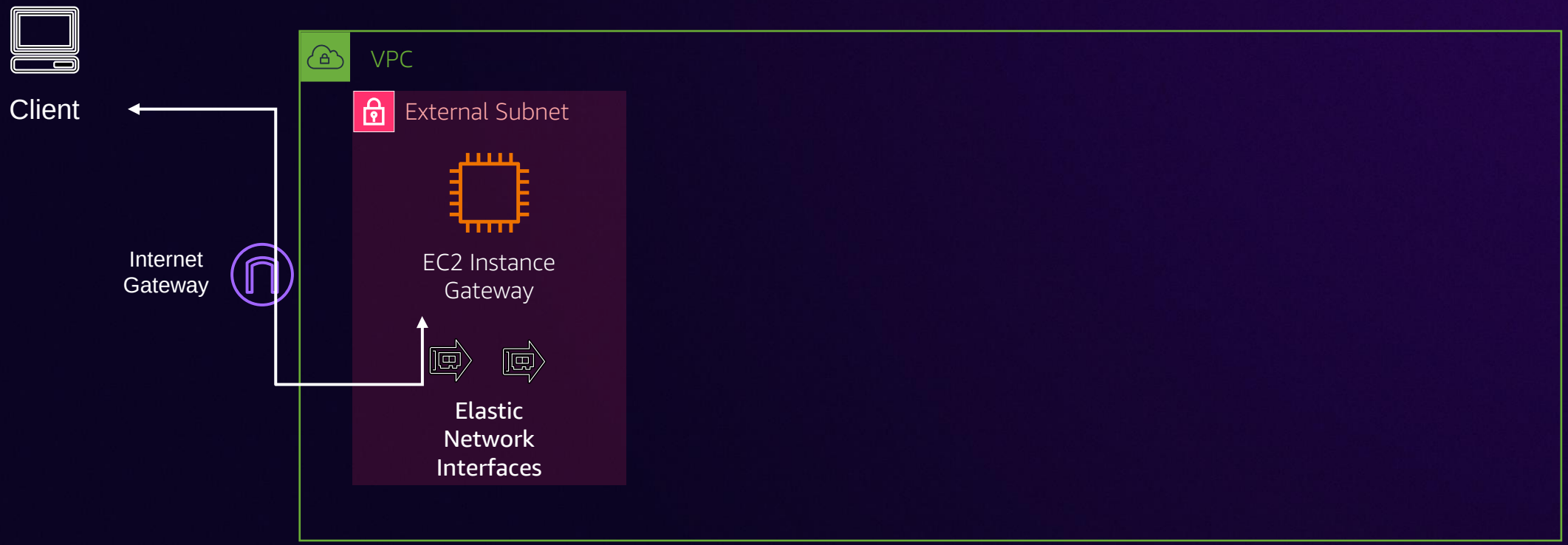
Client

Internet Gateway

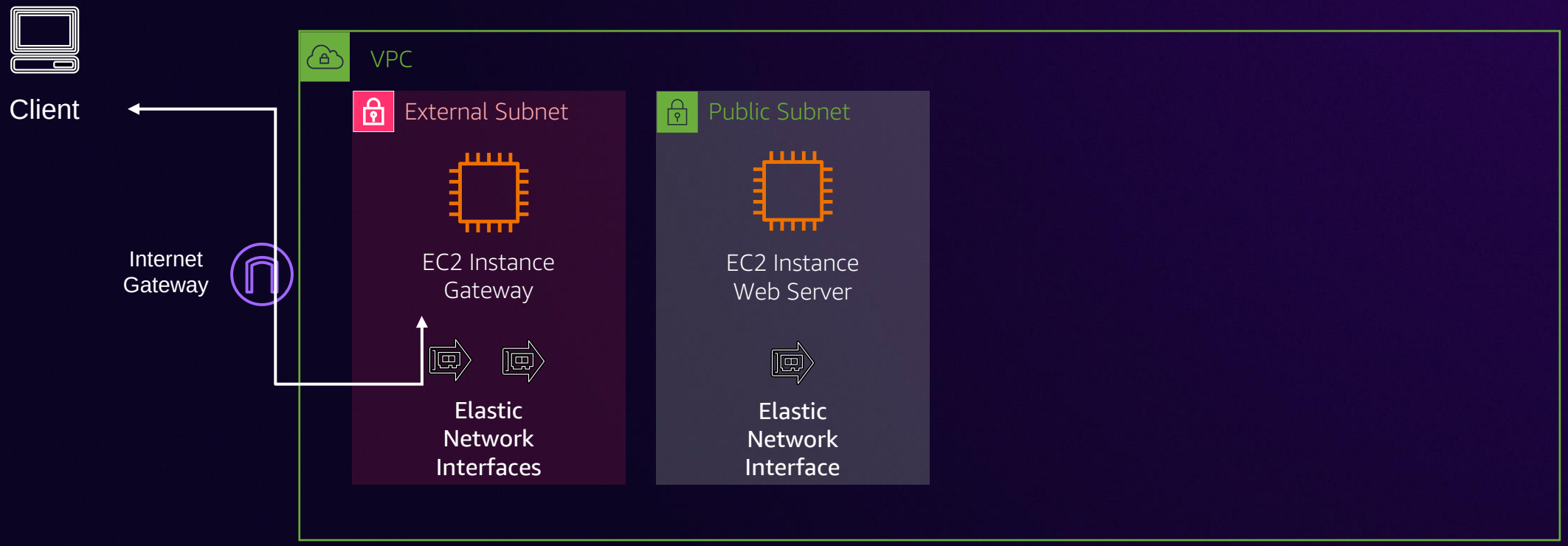
VPC Overview



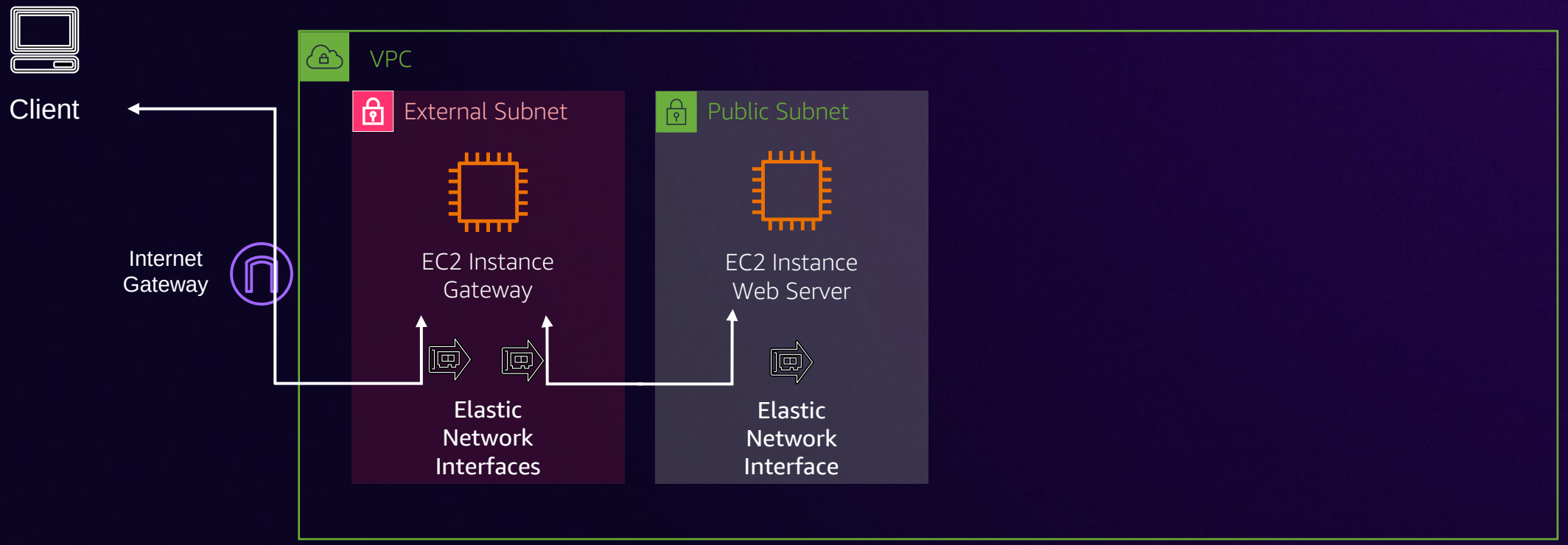
VPC Overview



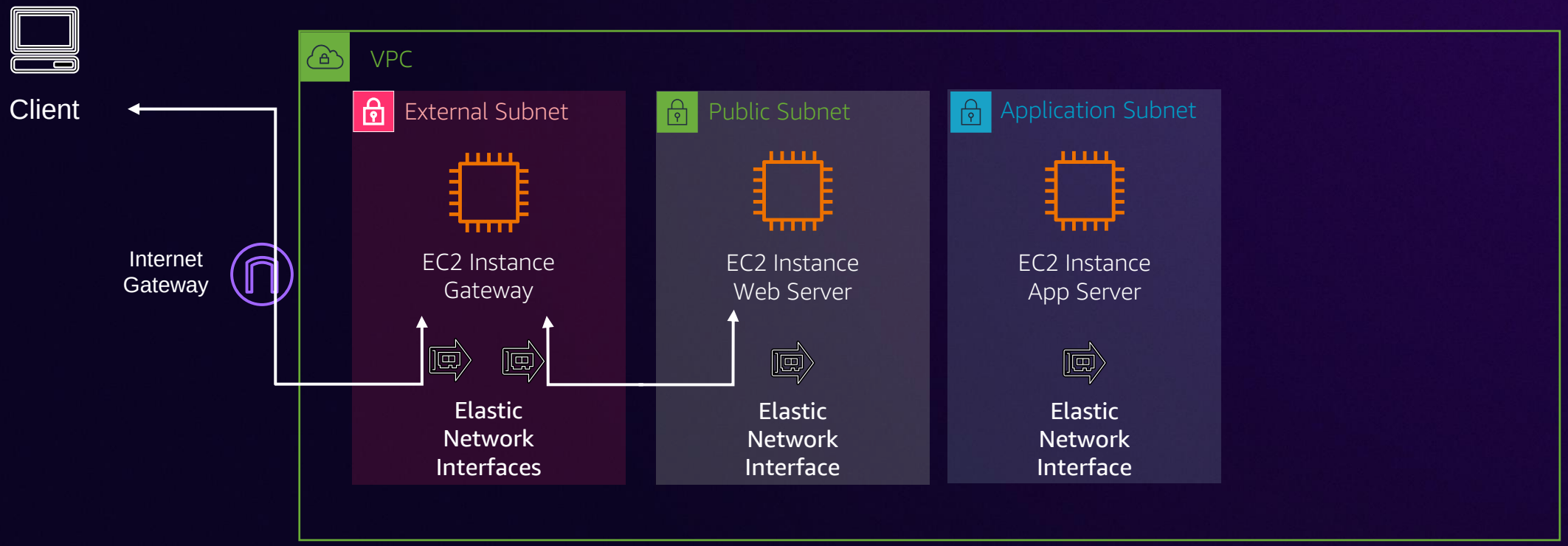
VPC Overview



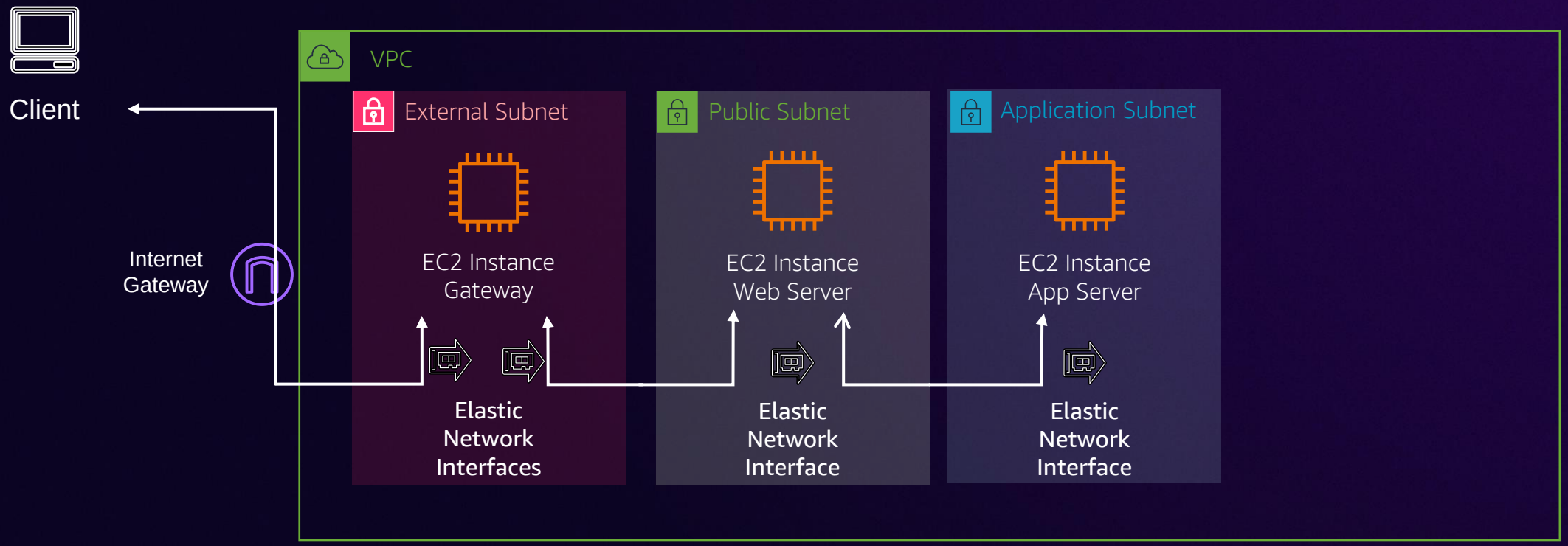
VPC Overview



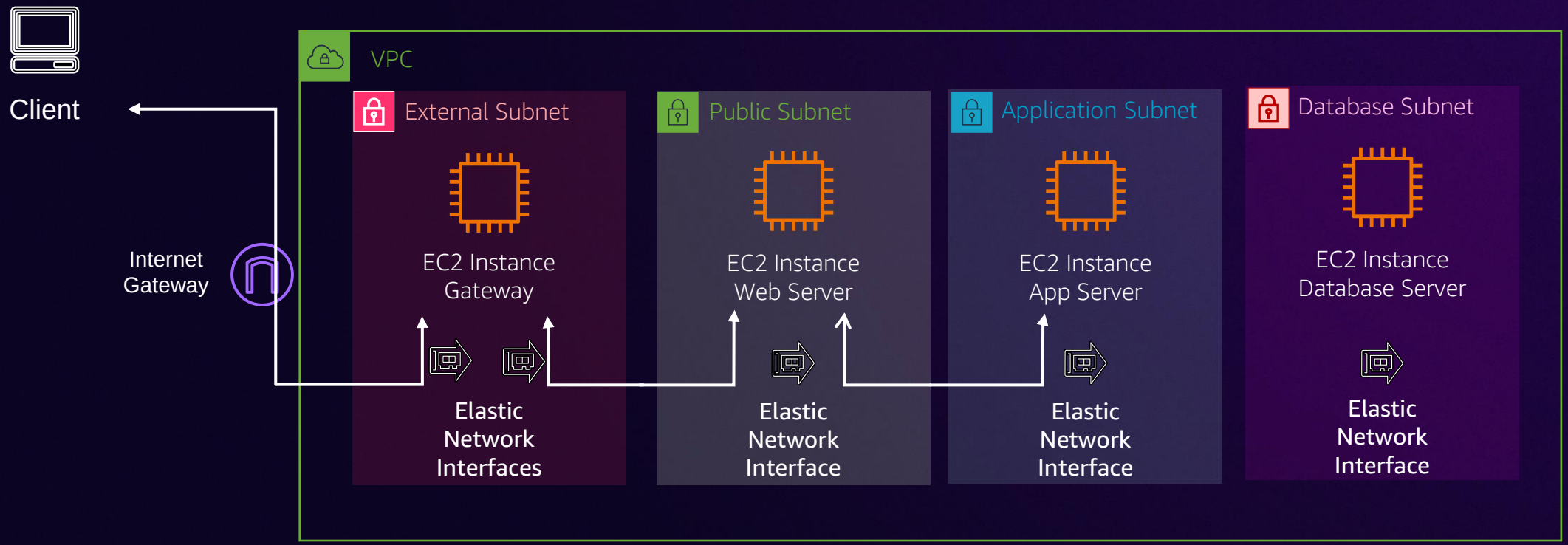
VPC Overview



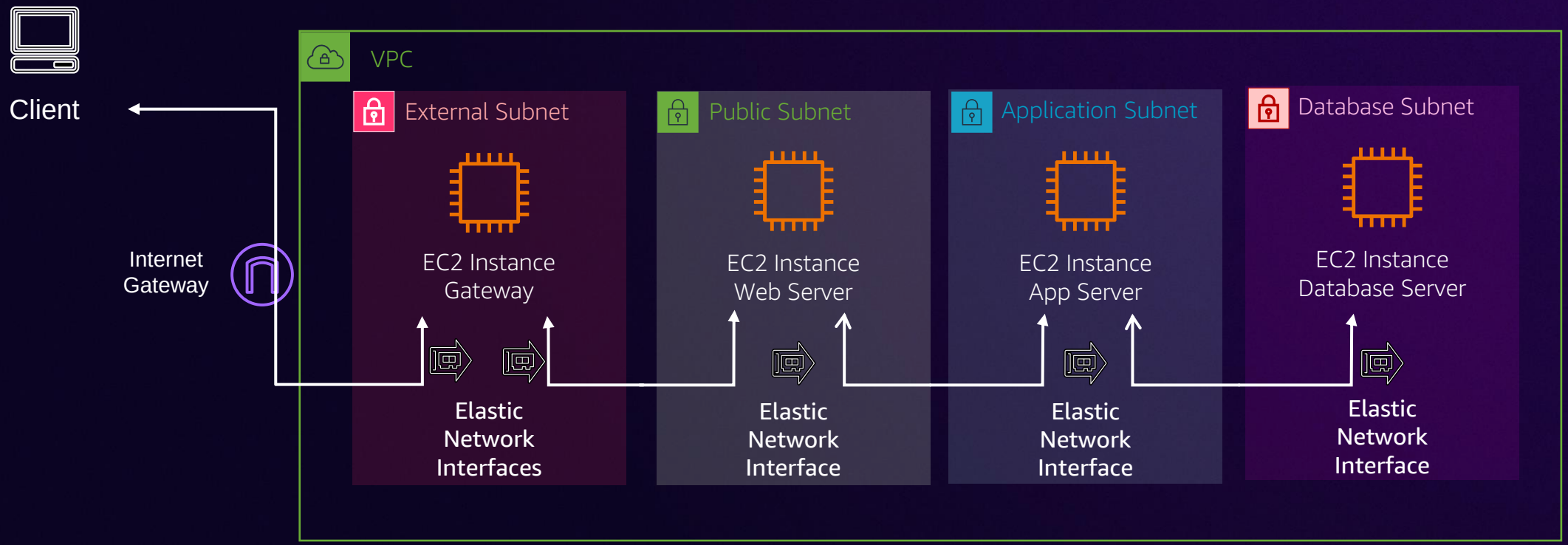
VPC Overview



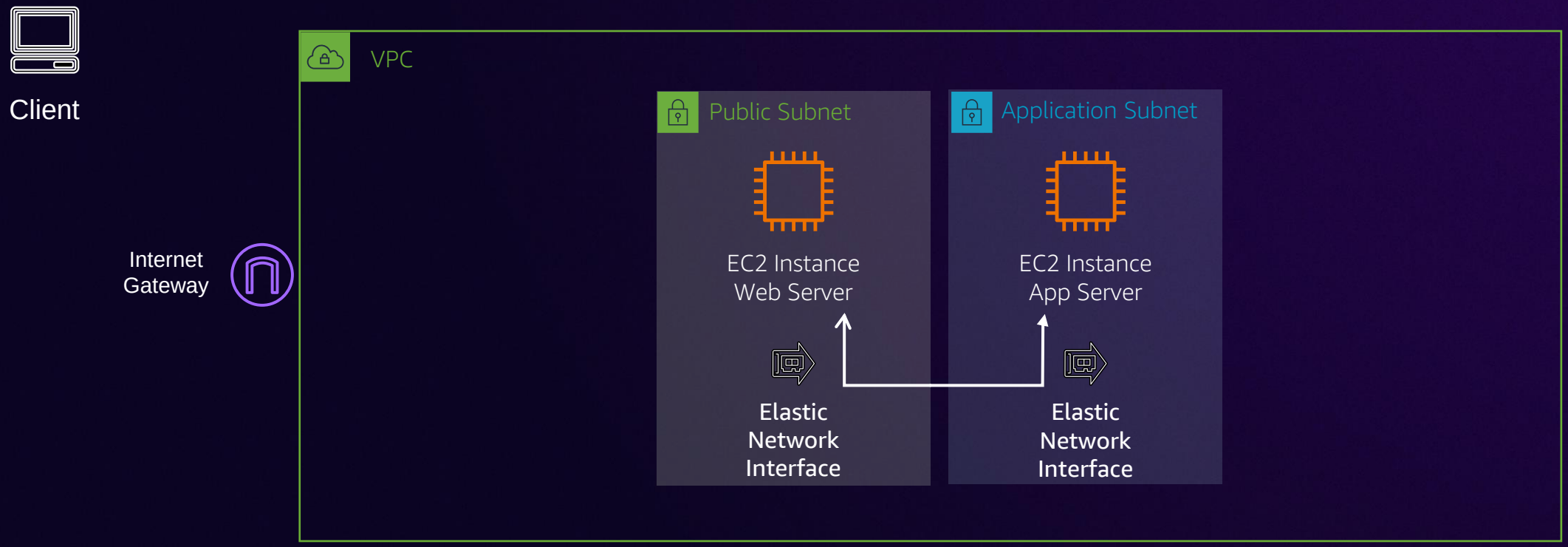
VPC Overview



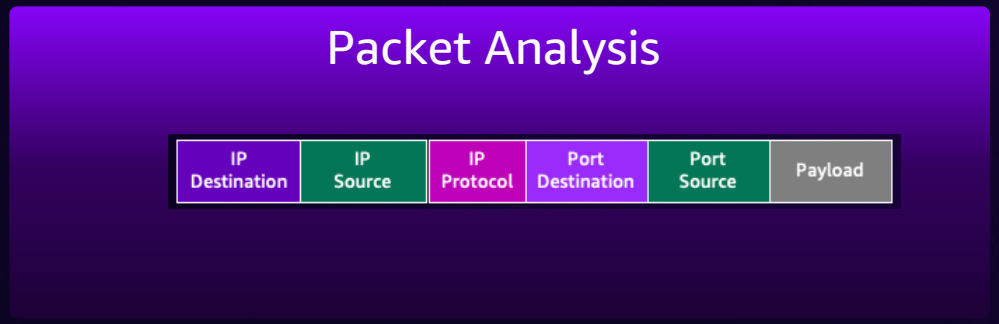
VPC Overview



VPC Flow Scenario

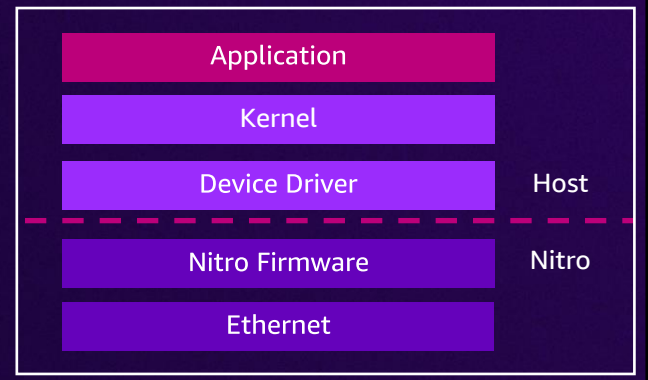


Packet Processing



- Processing Stack
- VPC Overview Flows
- Packet Processing
- Nitro State Machine

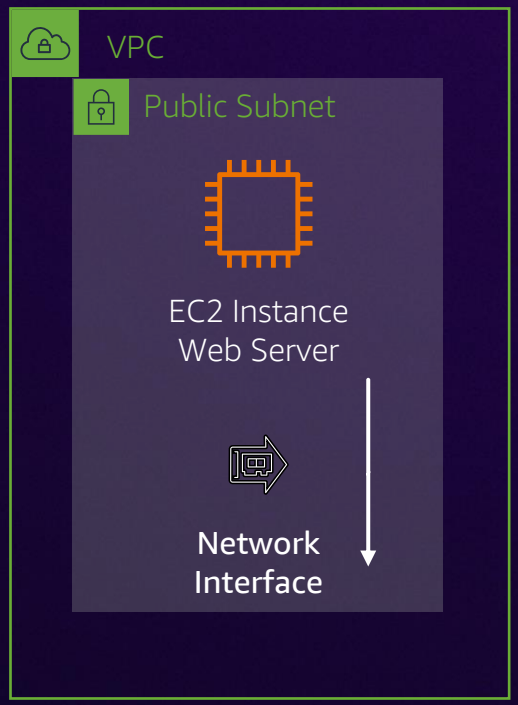
Target Host: App Server



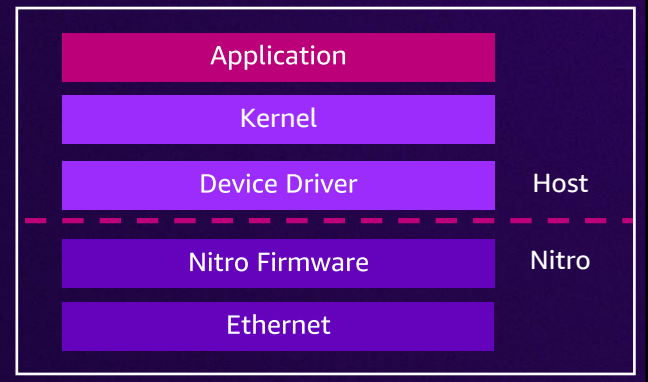
IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload

Web_Server> curl http://App_Server/a_file

DNS: Resolution of App_Server to IP Destination



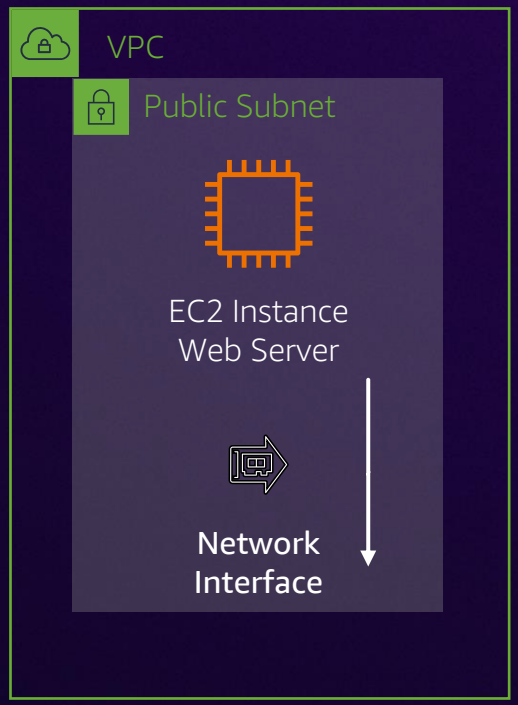
Target Host: App Server



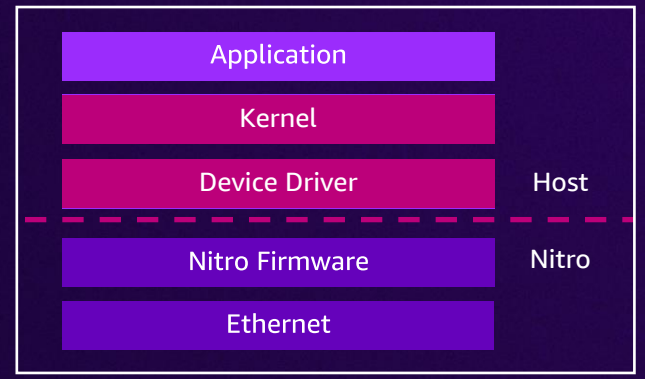
IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload

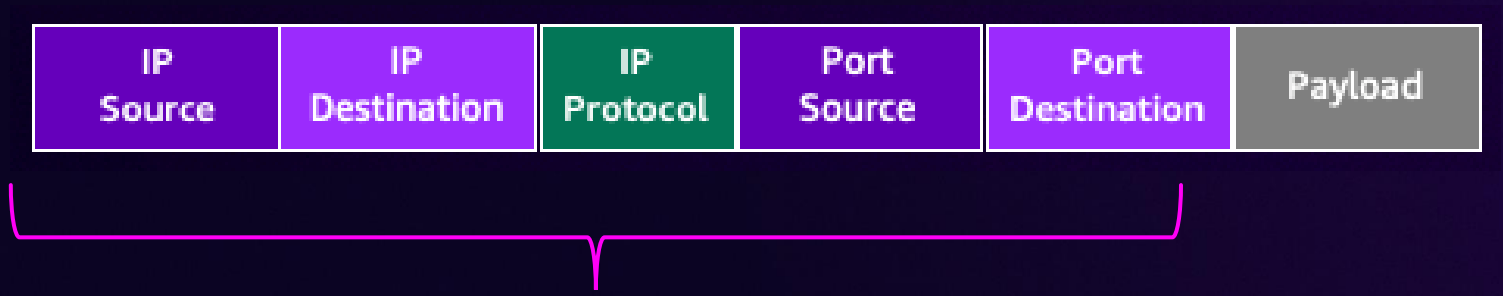
Web_Server> curl http://App_Server/a_file

DNS: Resolution of App_Server to IP Destination



Kernel Routing to Interface



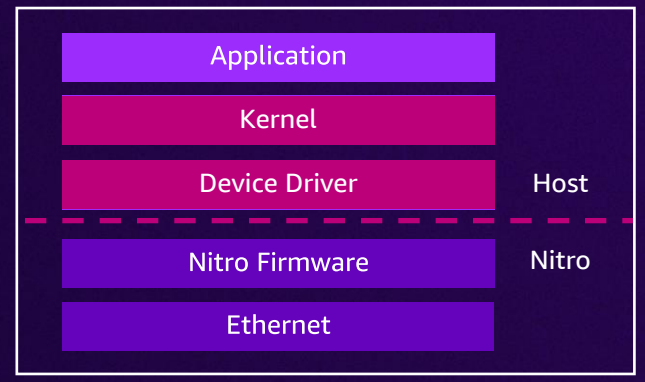


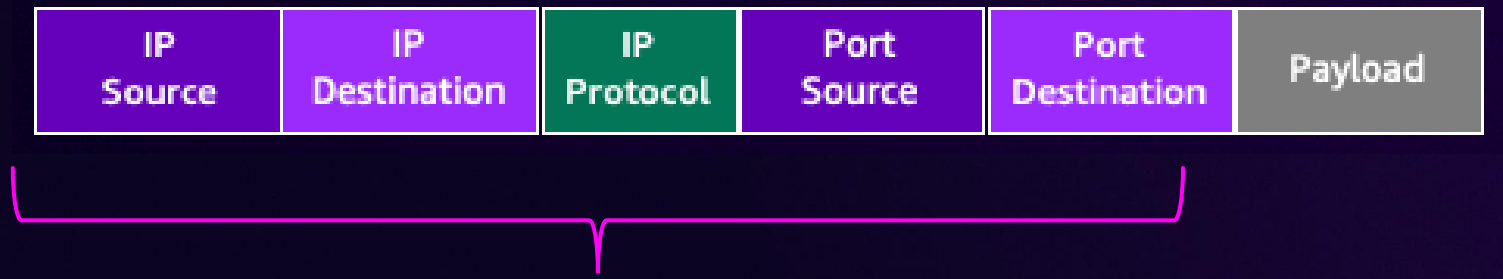
ENI

TCP Socket Opened => 5-Tuple

[Source IP, Destination IP, Protocol, Source Port, Destination Port]

Kernel Routing to Interface





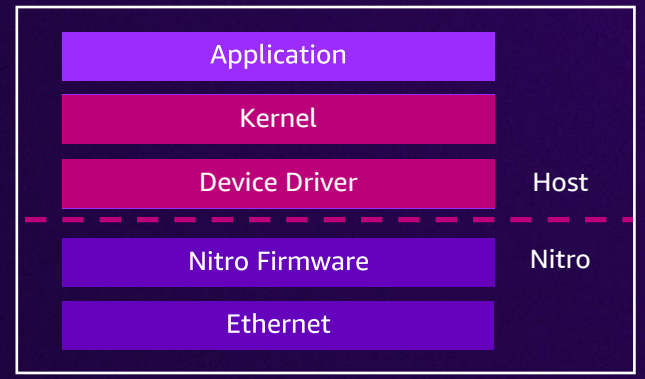
ENI

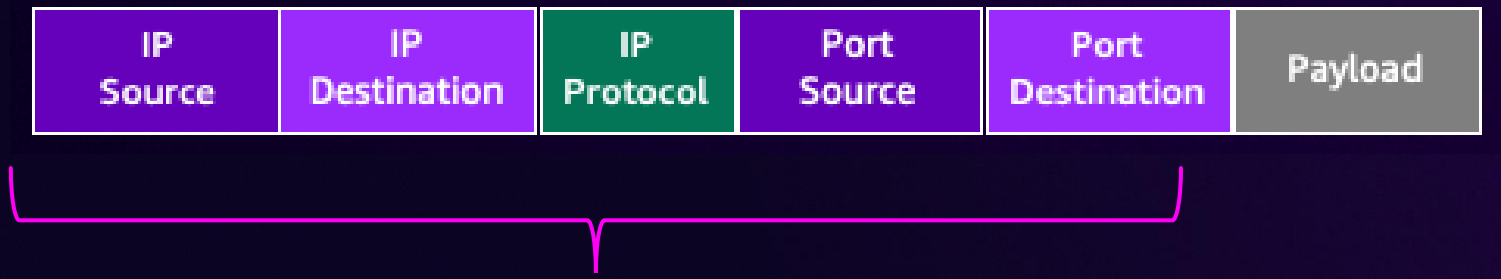
TCP Socket Opened => 5-Tuple

[Source IP, Destination IP, Protocol, Source Port, Destination Port]

Routing Packet to OS Device Interface

Kernel Routing to Interface





ENI

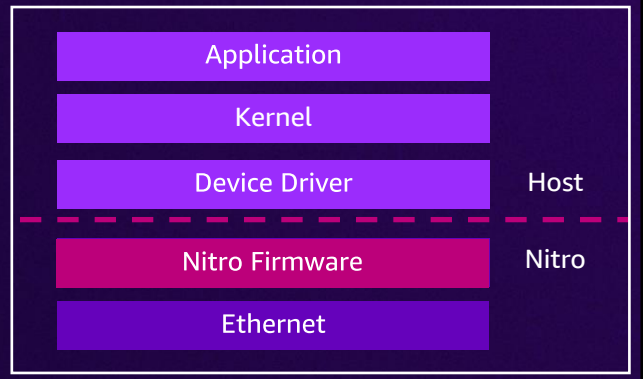
TCP Socket Opened => 5-Tuple

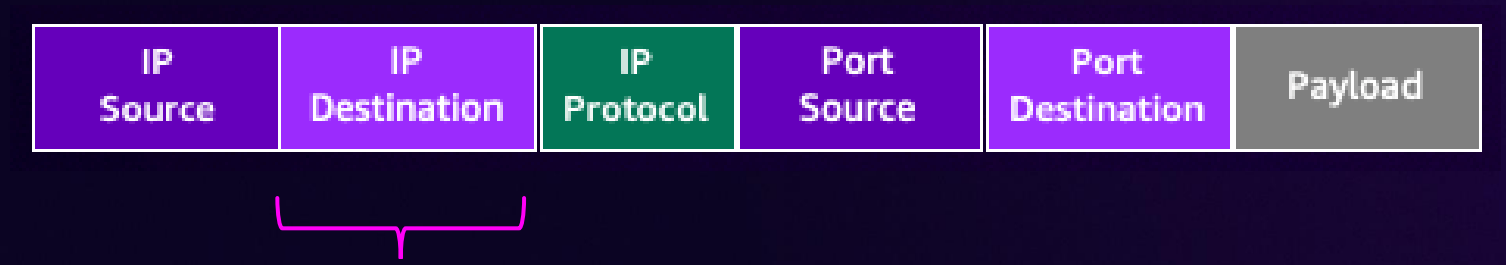
[Source IP, Destination IP, Protocol, Source Port, Destination Port]

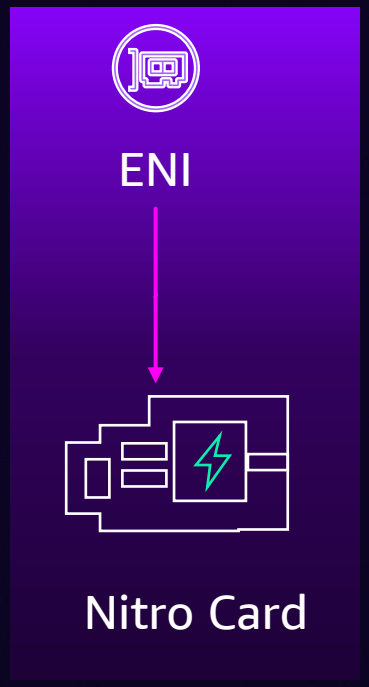
Routing Packet to OS Device Interface

/dev/ens5 mapped to Elastic Network Interface (ENI)

Elastic Network Adapter



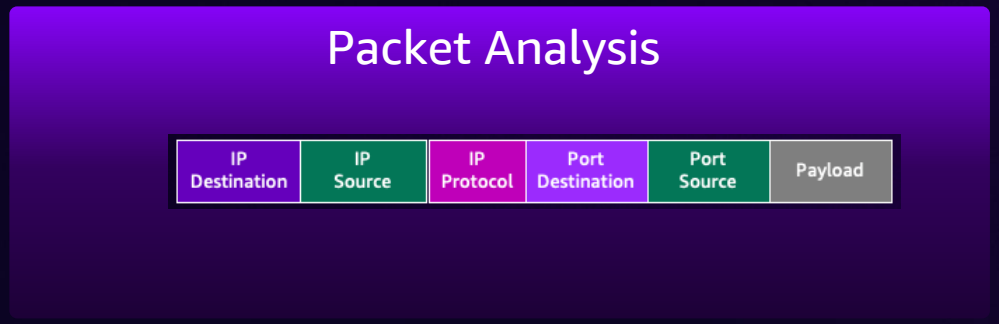




ENA

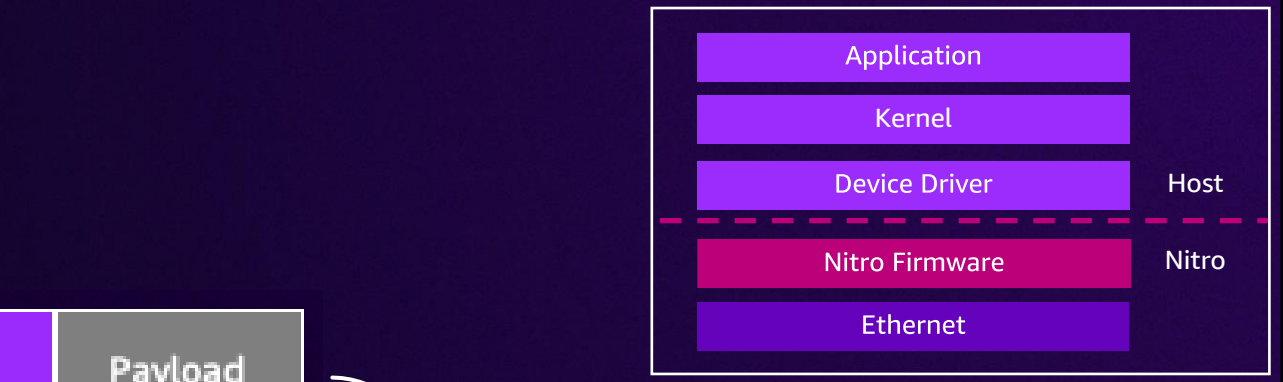
Elastic Network Interface managed by Elastic Network Adapter (ENA)

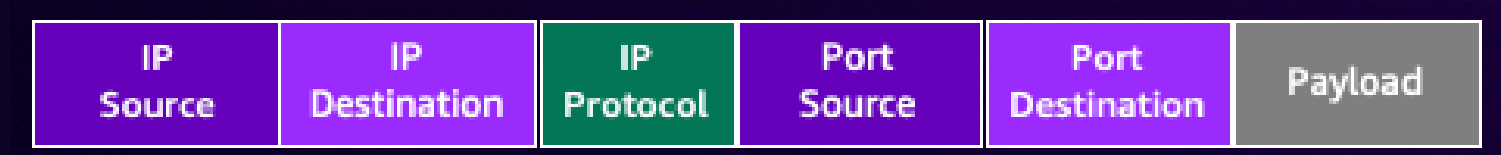
Starting with Packet Analysis



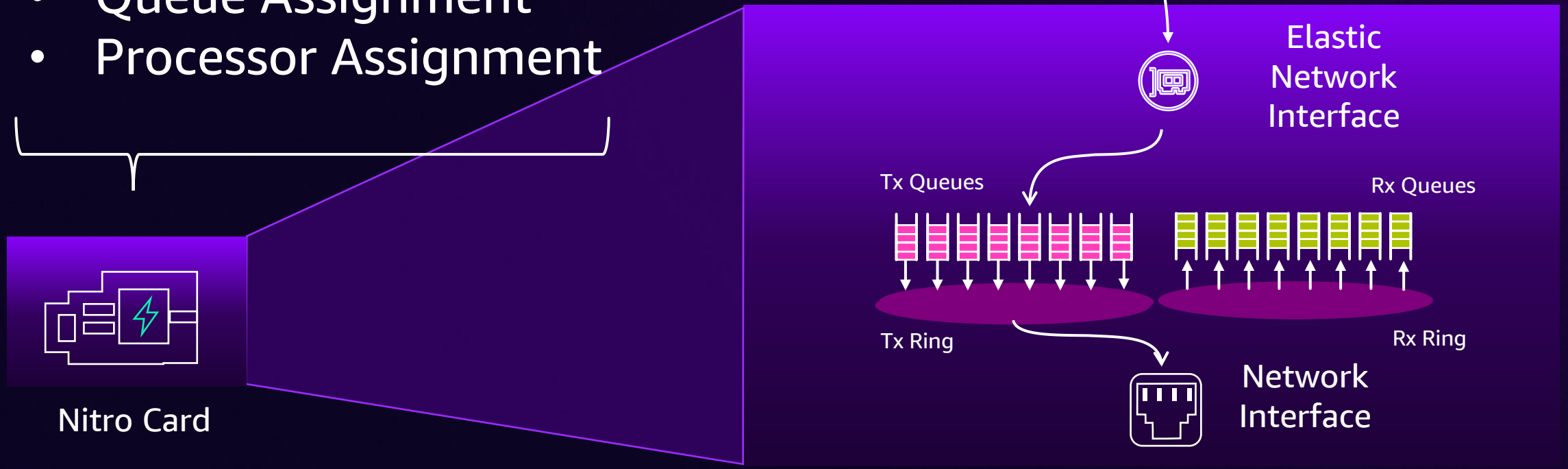
- Packet Analysis
- VPC Overview Flows
- Packet Processing
- Nitro State Machine

Inside Nitro

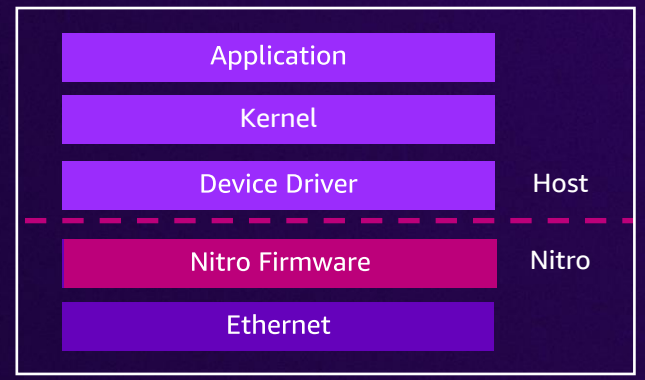


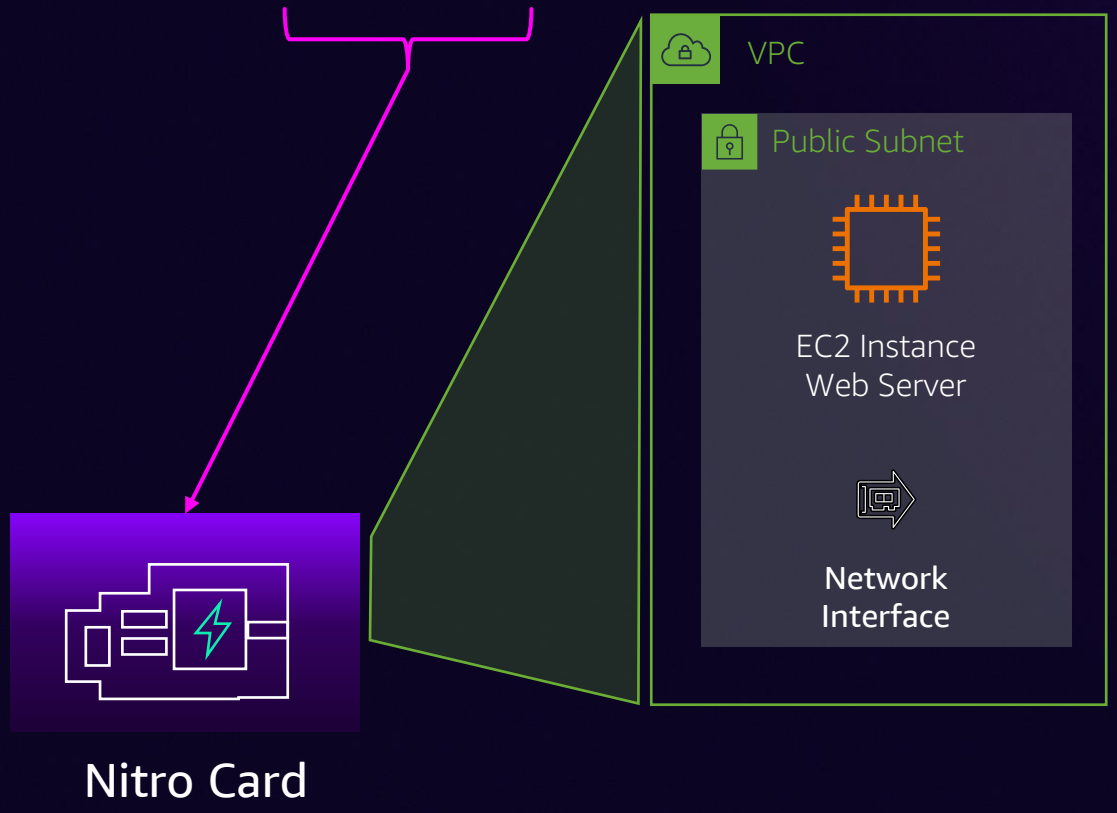


- Hash on 5-Tuple
- Queue Assignment
- Processor Assignment



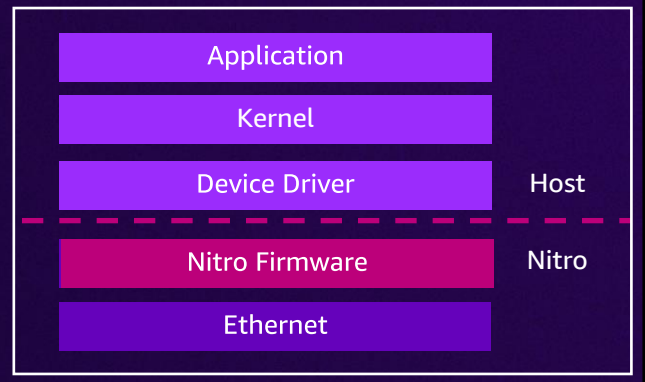
Virtual Private Cloud in Nitro



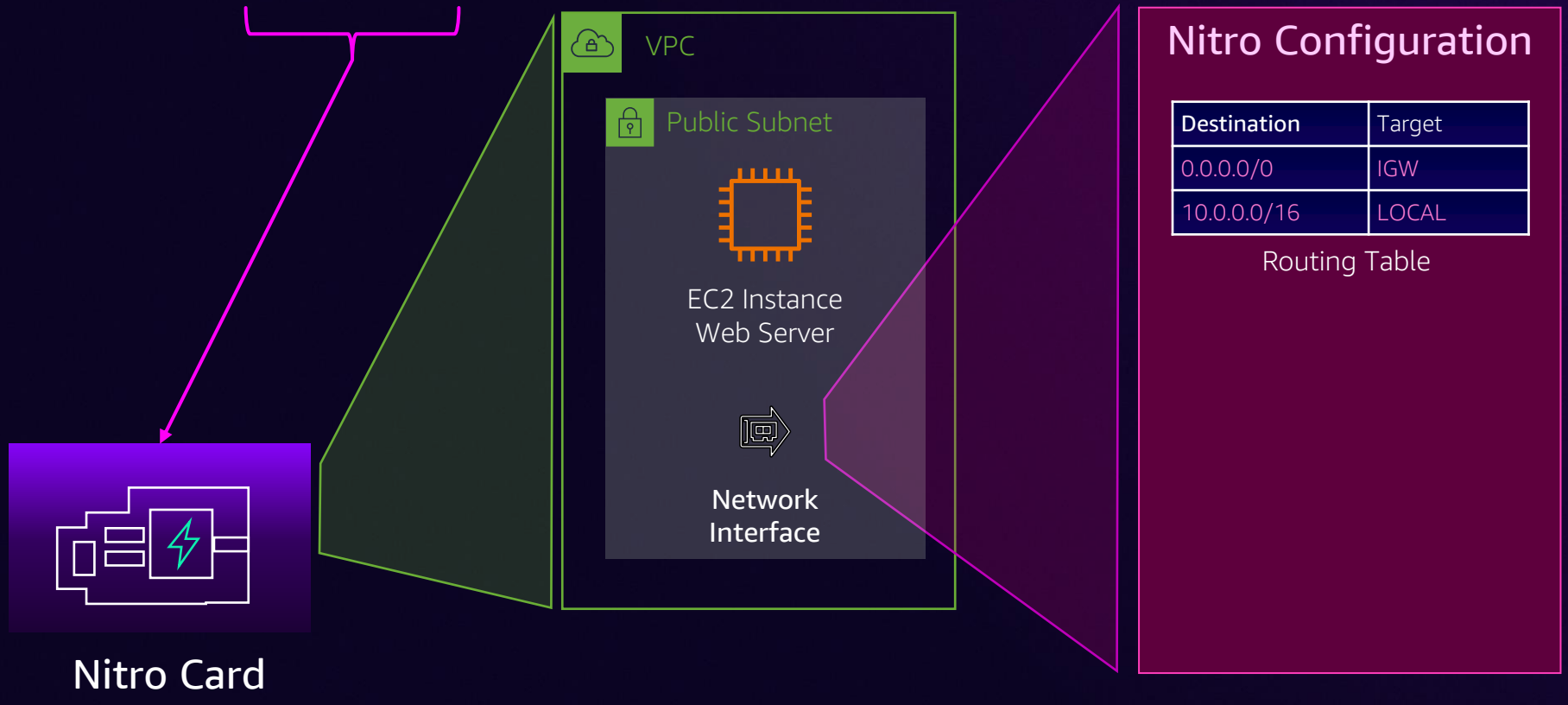


Virtual Private Cloud and Subnets Represented in Nitro Cards

Flow Setup Processes



IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload

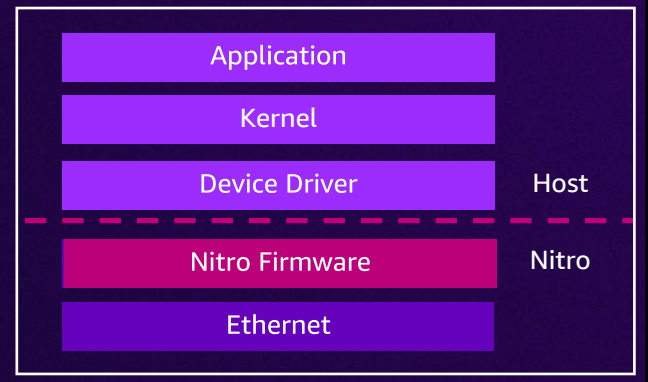


Nitro Configuration

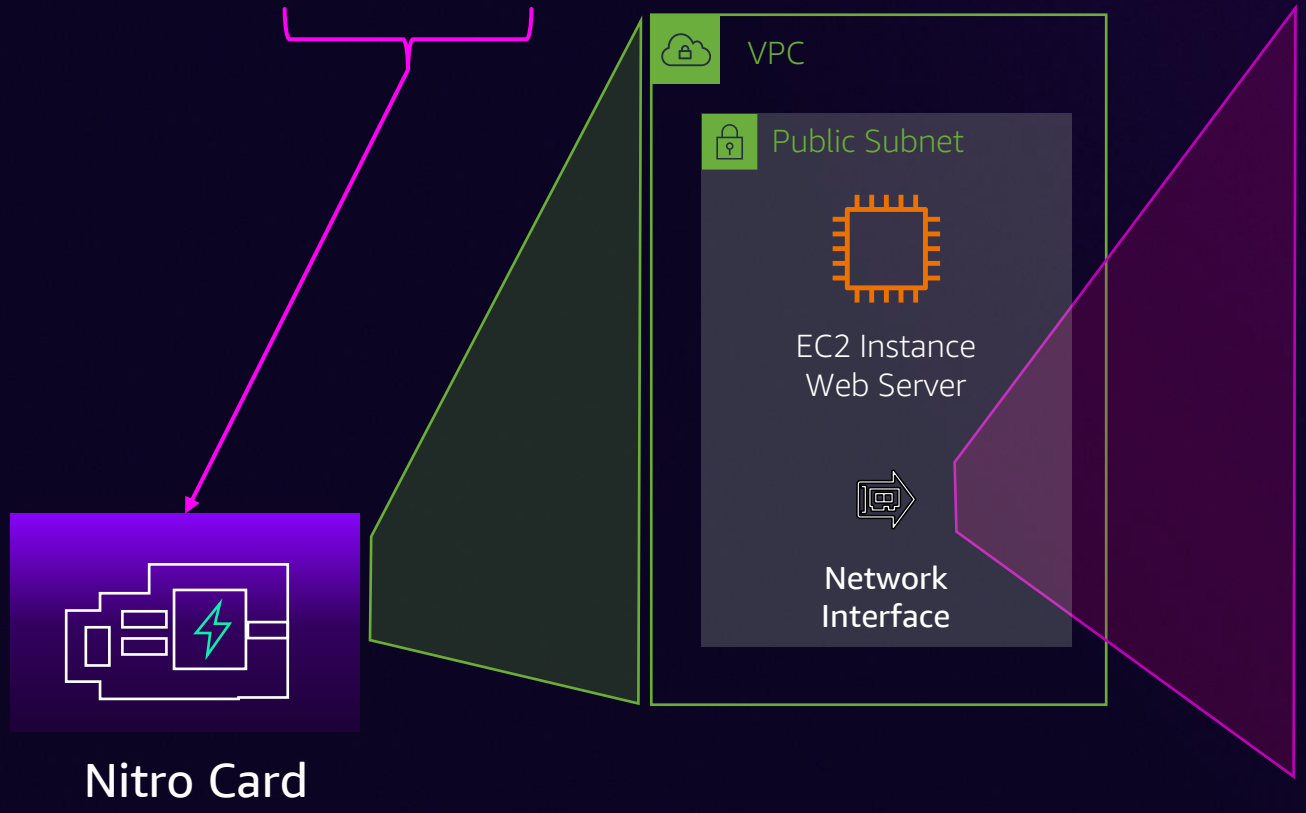
Destination	Target
0.0.0.0/0	IGW
10.0.0.0/16	LOCAL

Routing Table

Flow Setup Processes



IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload



Nitro Configuration

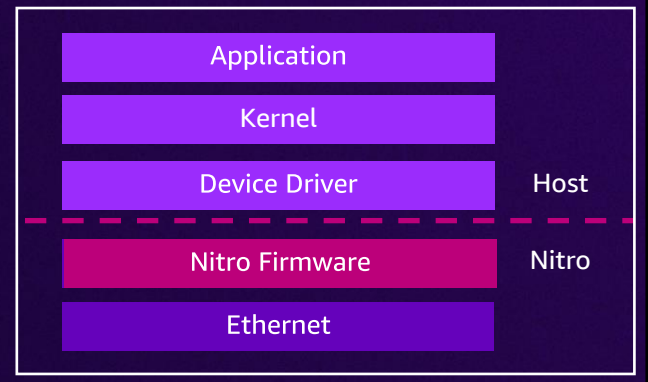
Destination	Target
0.0.0.0/0	IGW
10.0.0.0/16	LOCAL

Routing Table

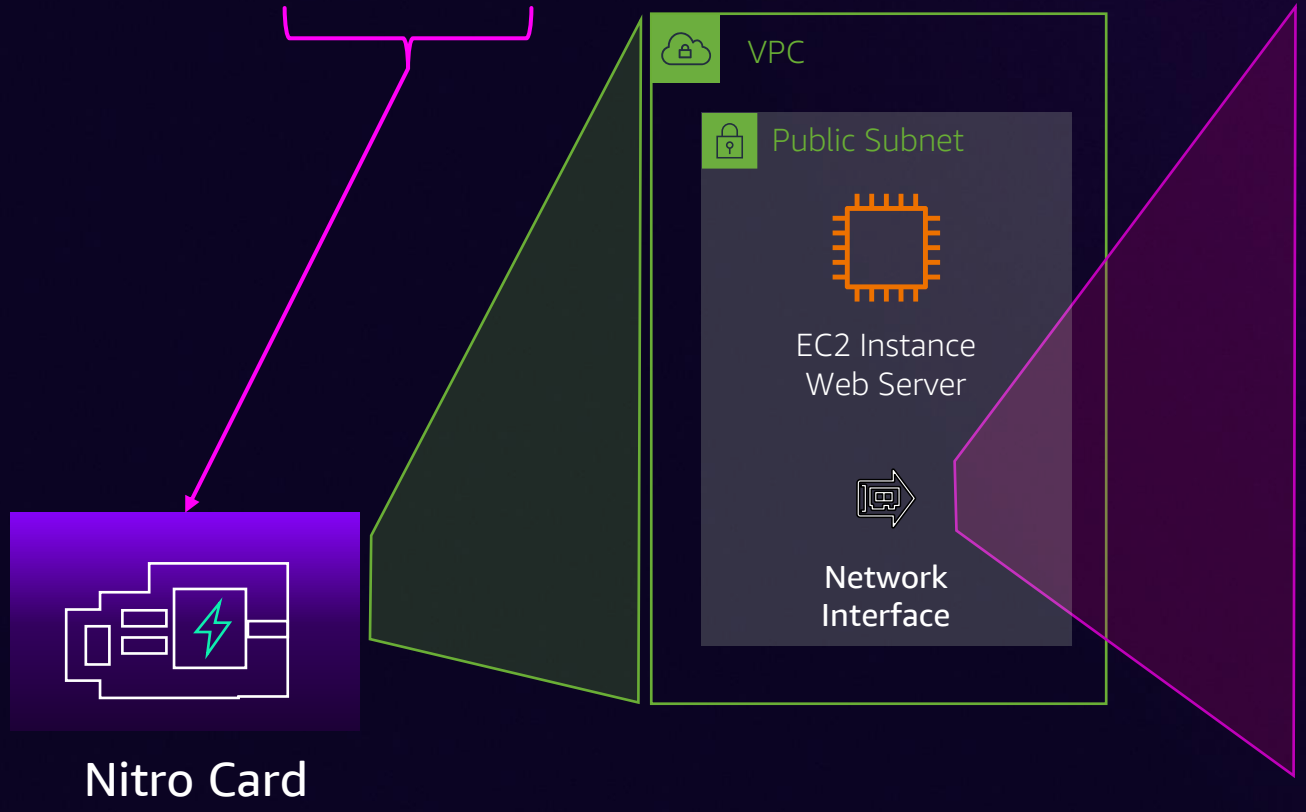
Rule	Type	Proto	Port	Src	State
100	IPv4	All	All	0/0	Allow
*	IPv4	All	All	0/0	Deny

Network Access Control List

Flow Setup Processes



IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload



Nitro Configuration

Destination	Target
0.0.0.0/0	IGW
10.0.0.0/16	LOCAL

Routing Table

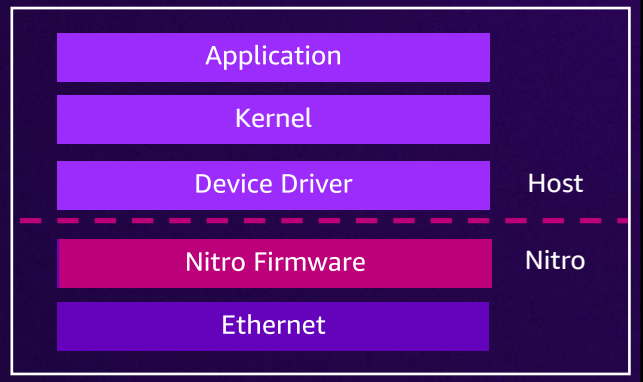
Rule	Type	Proto	Port	Src	State
100	IPv4	All	All	0/0	Allow
*	IPv4	All	All	0/0	Deny

Network Access Control List

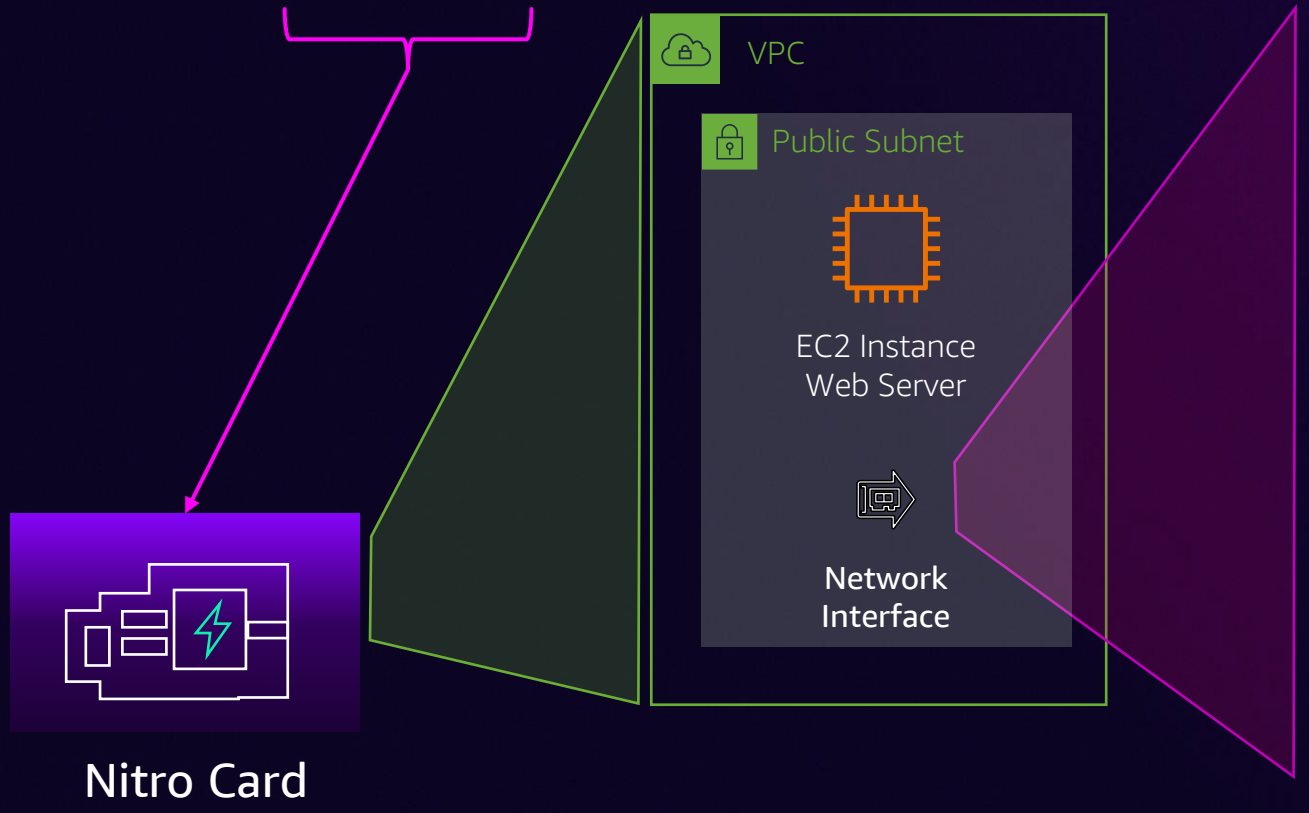
Source	Protocol	Port range
Any	TCP	HTTPS

Security Group

Flow Setup Processes



IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload



Nitro Configuration

Destination	Target
0.0.0.0/0	IGW
10.0.0.0/16	LOCAL

Routing Table

Rule	Type	Proto	Port	Src	State
100	IPv4	All	All	0/0	Allow
*	IPv4	All	All	0/0	Deny

Network Access Control List

Source	Protocol	Port range
Any	TCP	HTTPS

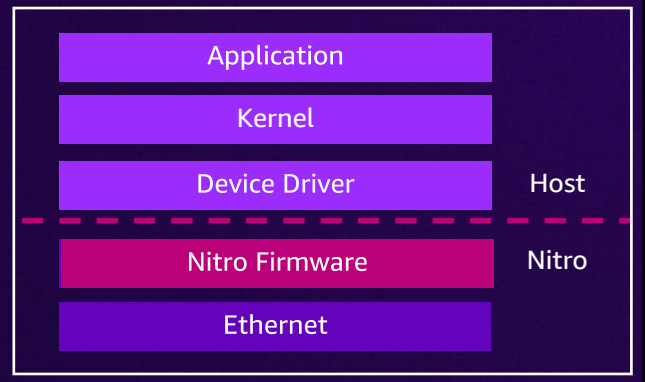
Security Group

Nitro State

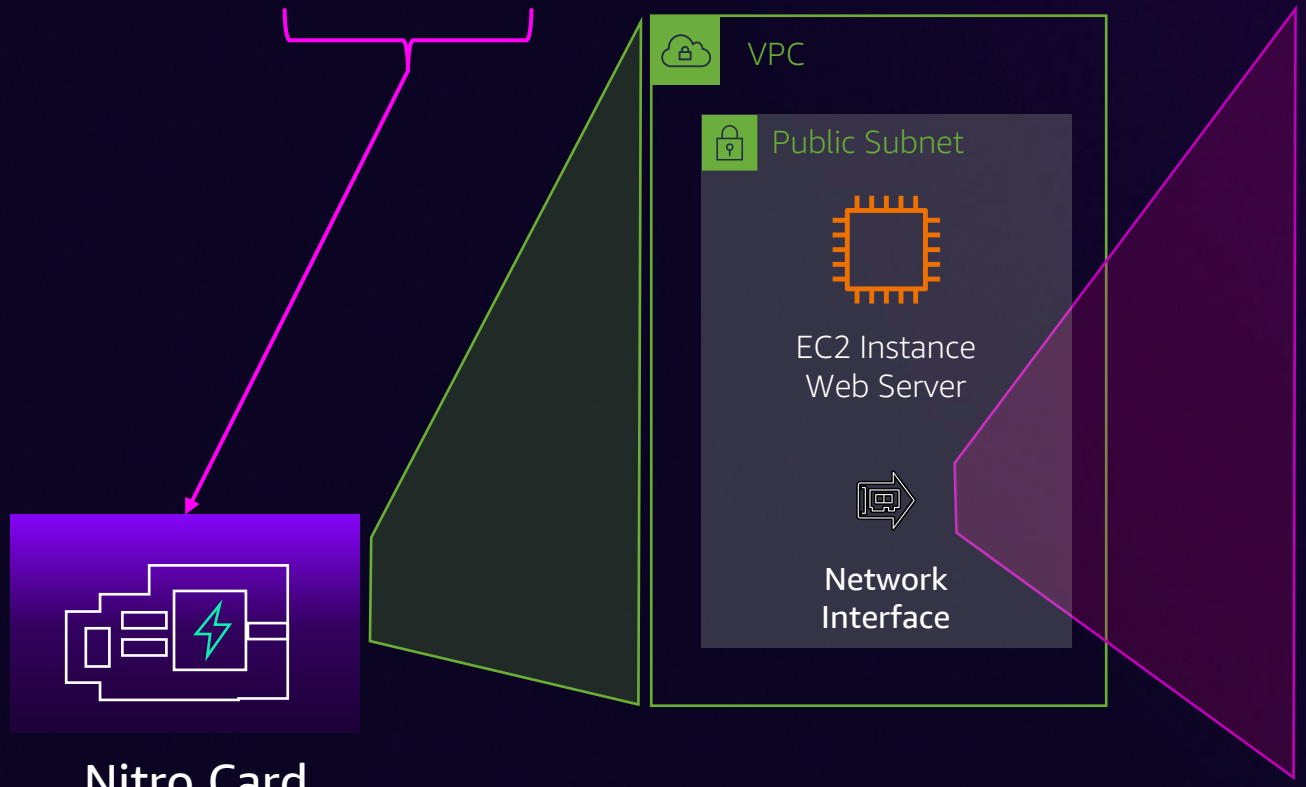
Adjacency	Target
IGW	Unknown
App Server ENI	Unknown

Mapping State

Flow Setup Processes



IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload



Nitro State

Adjacency	Target
IGW	Unknown
App Server ENI	IP-Address

Mapping State

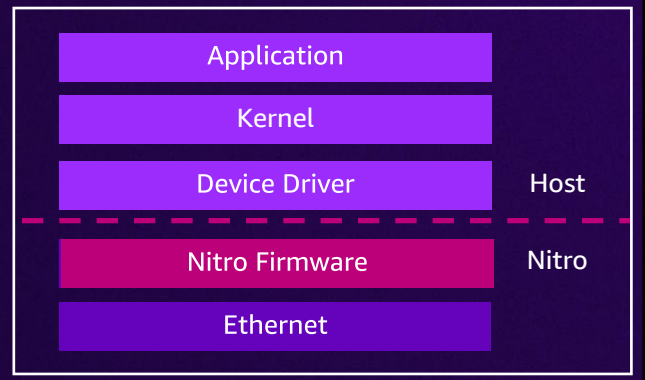
Remote Mapping Service

Lookup

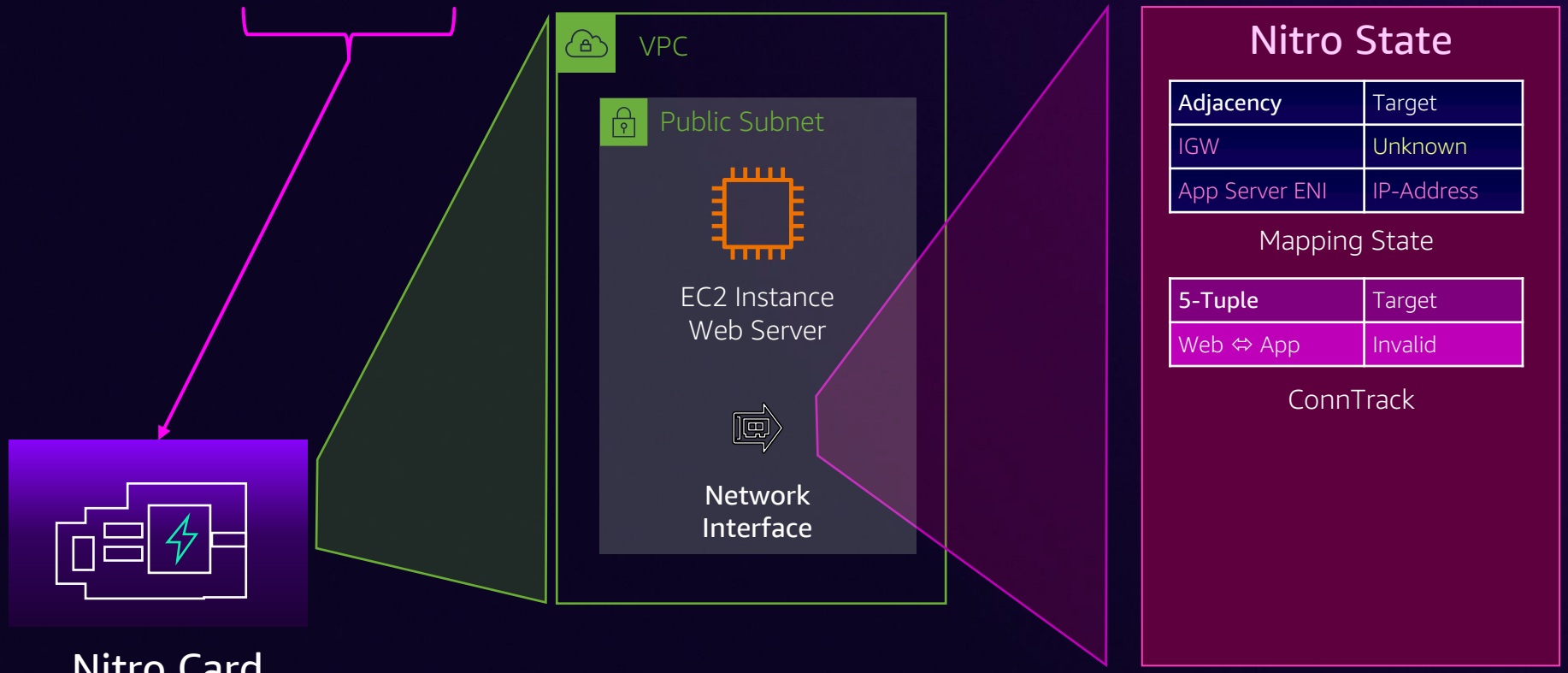
Target	IP
Web Server	IP-Address
App Server	IP-Address
Database	IP Address

Endpoint Mapping

Flow Setup Processes



IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload



Nitro State

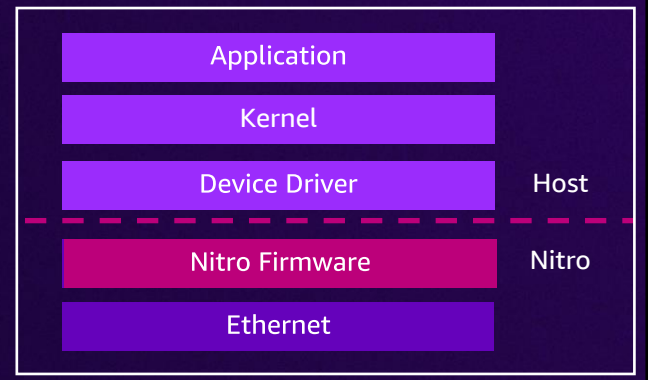
Adjacency	Target
IGW	Unknown
App Server ENI	IP-Address

Mapping State

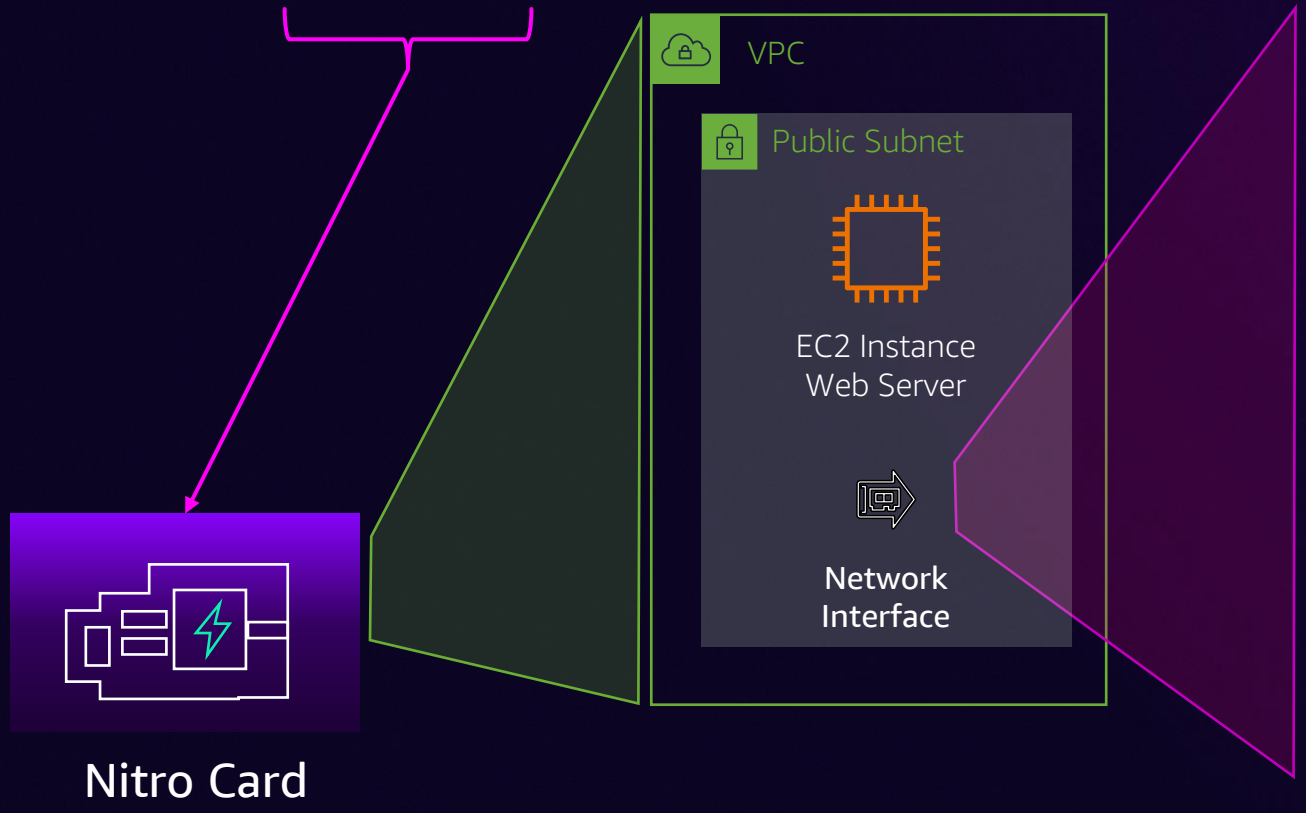
5-Tuple	Target
Web ⇔ App	Invalid

ConnTrack

Flow Setup Processes



IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload



Nitro State

Adjacency	Target
IGW	Unknown
App Server ENI	IP-Address

Mapping State

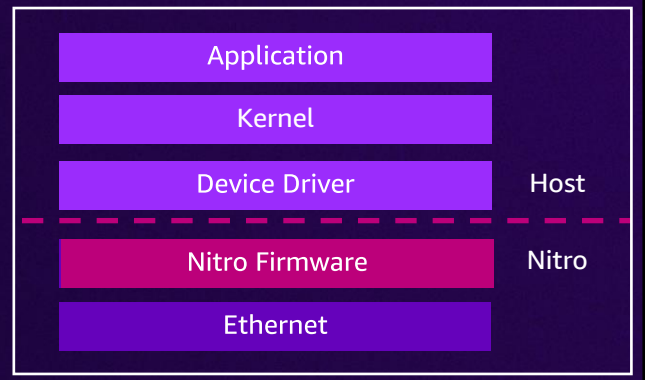
5-Tuple	Target
Web ⇔ App	Invalid

ConnTrack

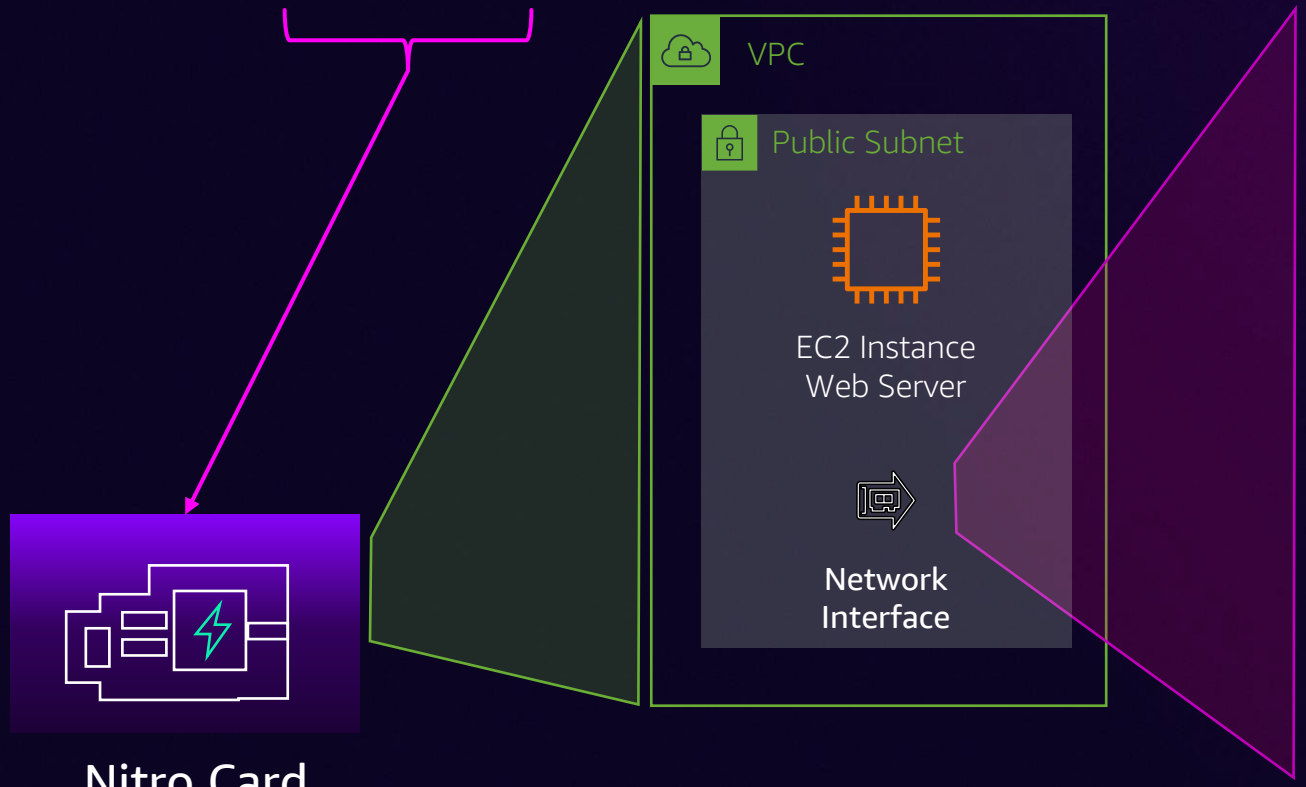
Flow Cache	State
Web to App	Pending

Flow Cache

Flow Setup Processes



IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload



Nitro State

Adjacency	Target
IGW	Unknown
App Server ENI	IP-Address

Mapping State

5-Tuple	Target
Web ⇔ App	Invalid

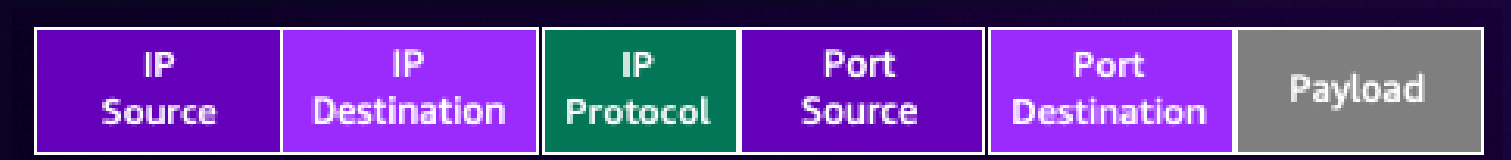
ConnTrack

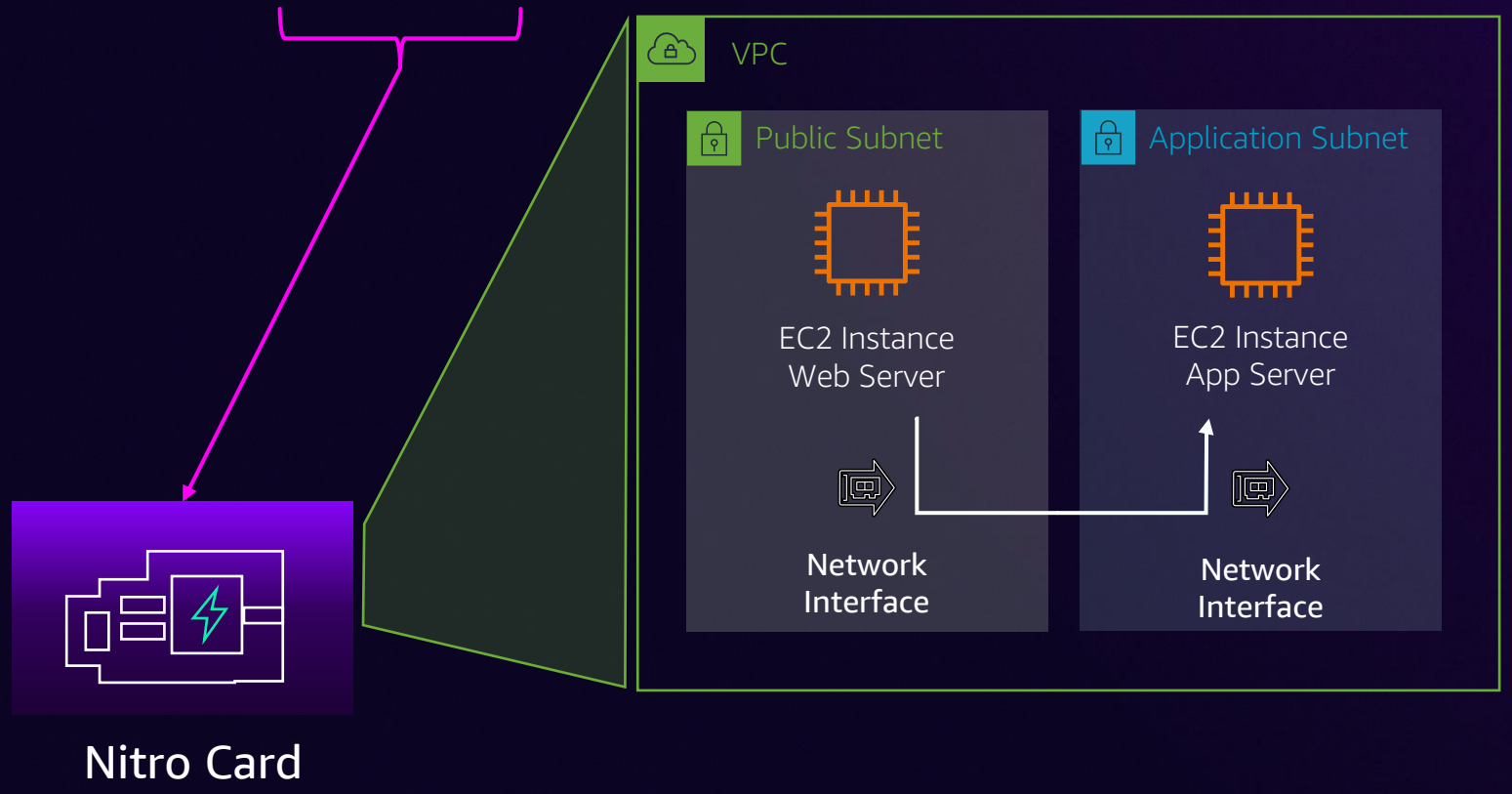
Flow Cache	State
Web to App	Pending

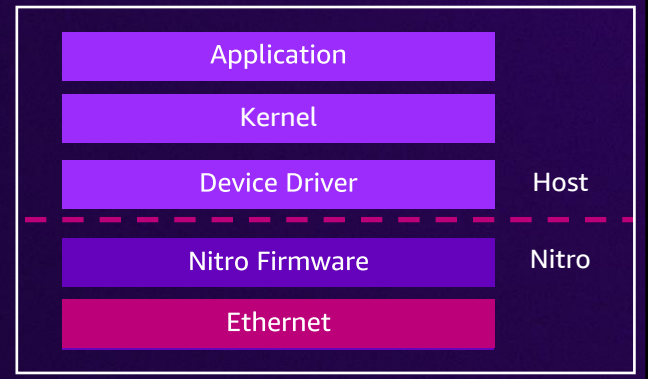
Flow Cache

Initial Security Group Eval: Processing-Path With lower Packets Per Second

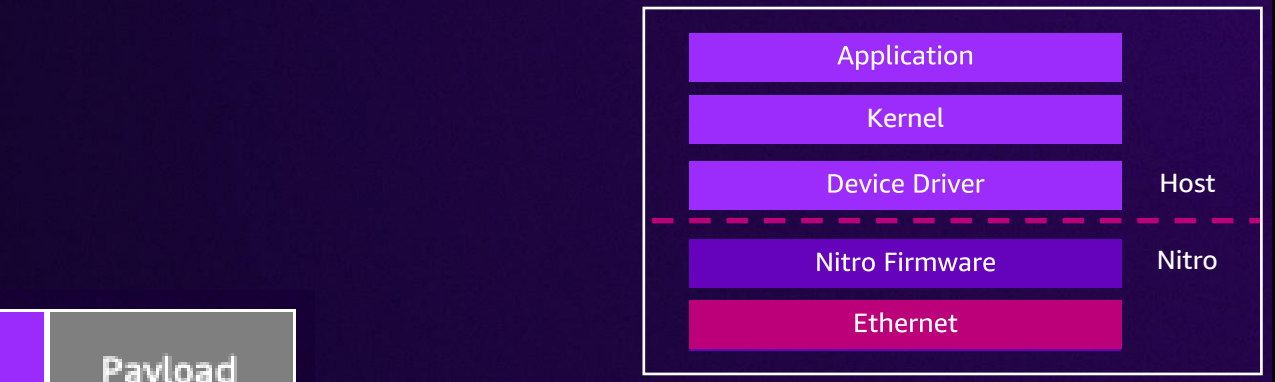
Flow Setup Processes

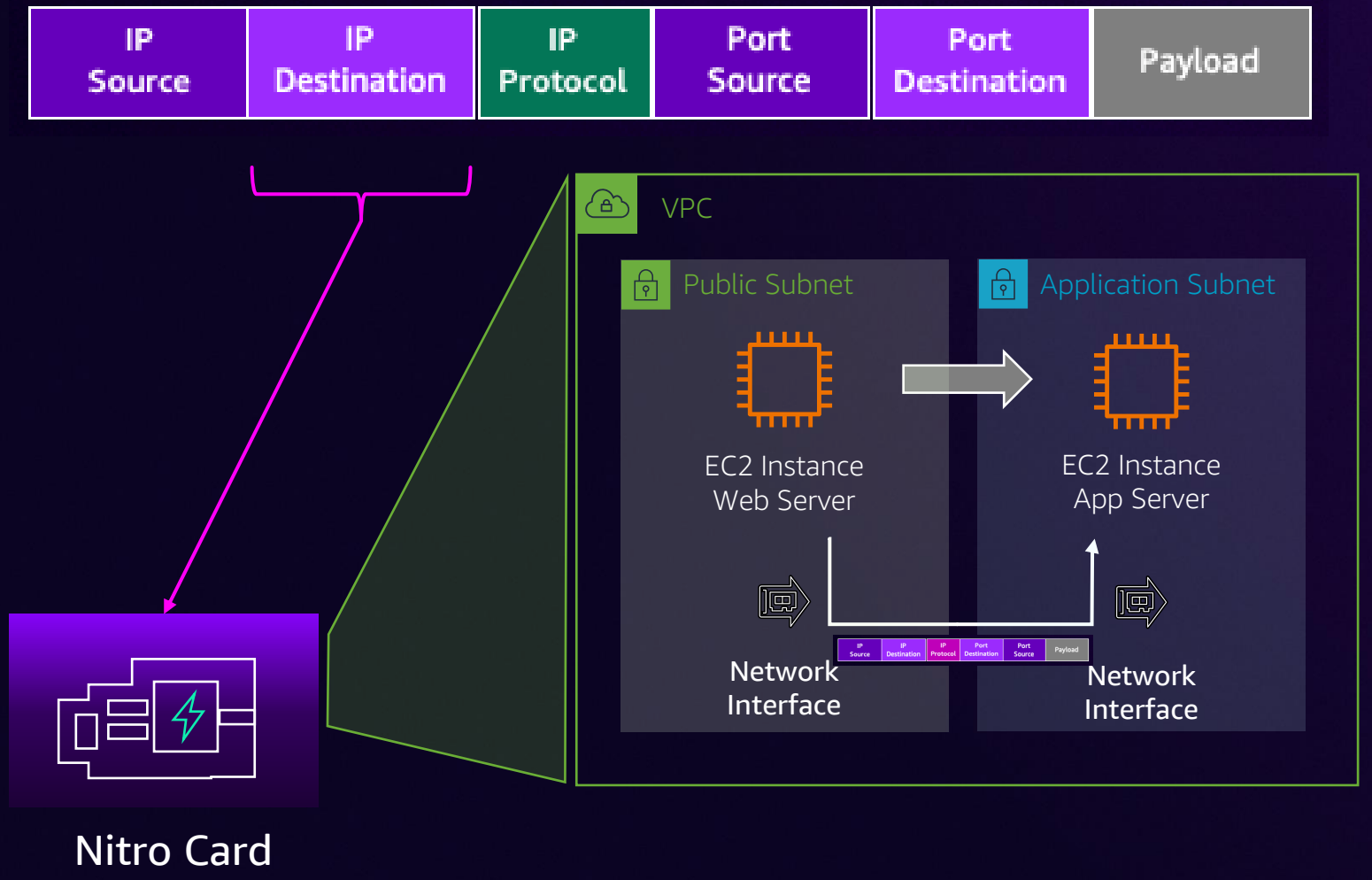




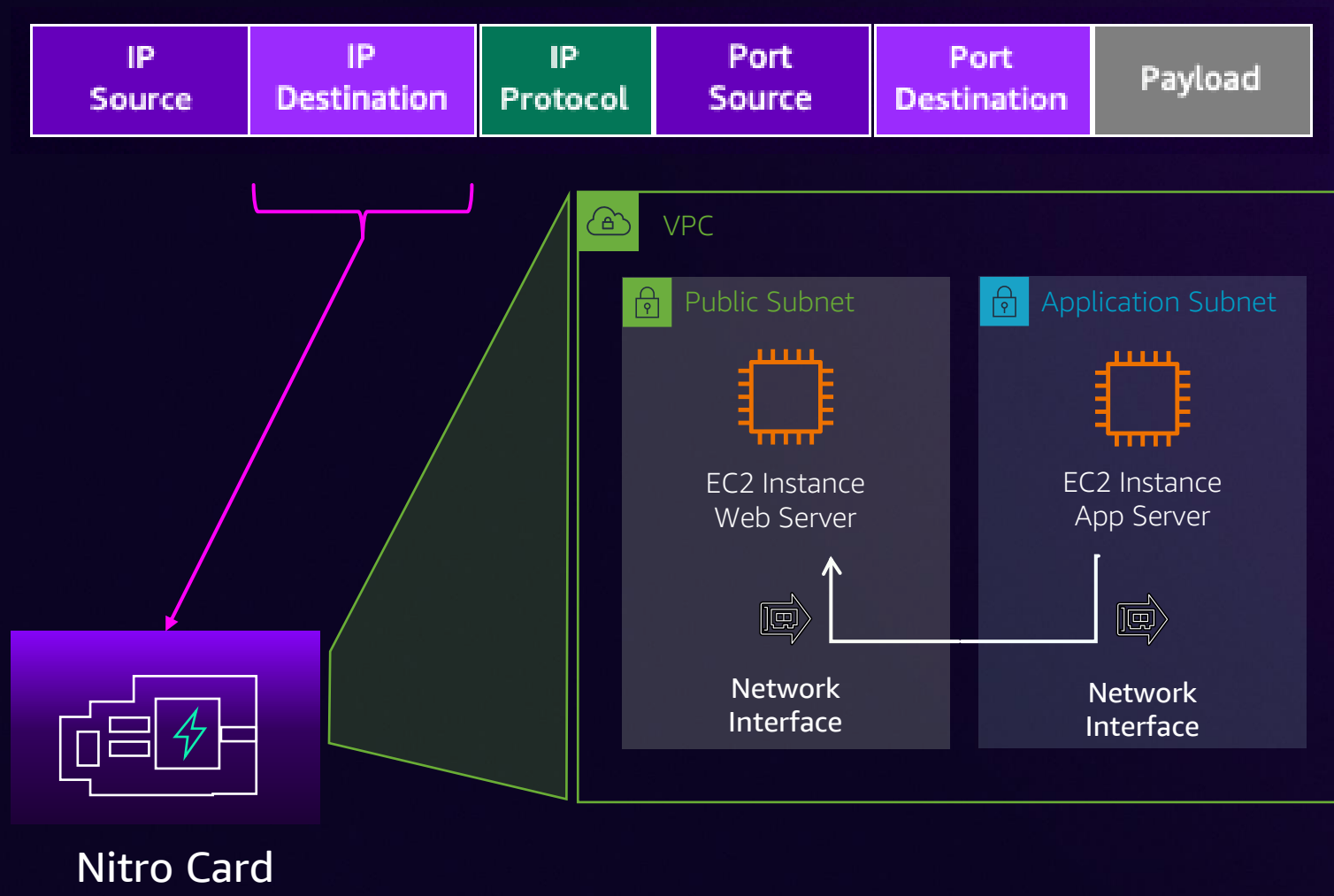


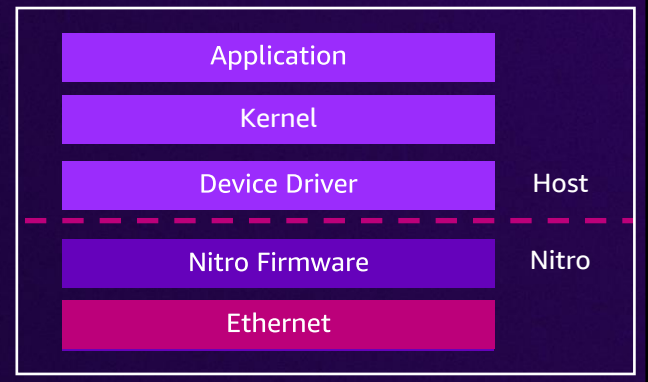
Flow Setup Processes



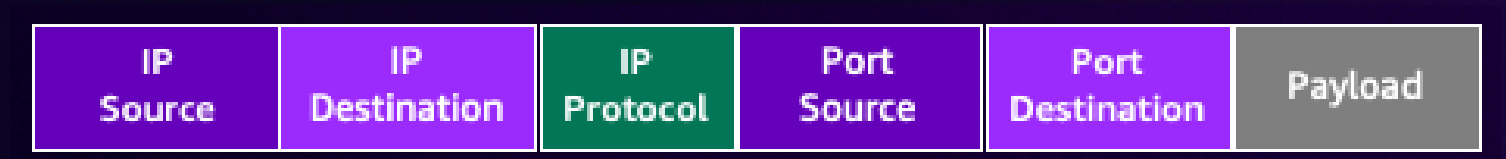


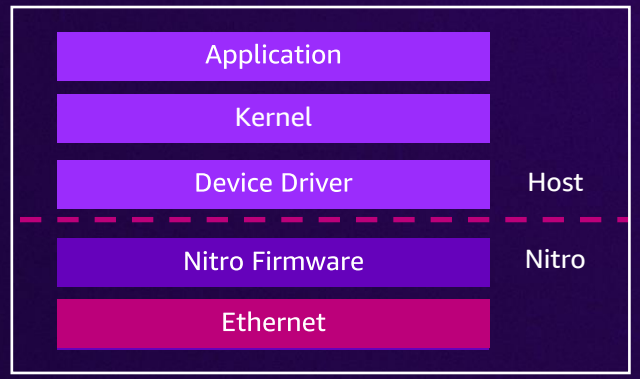
Flow Setup Processes

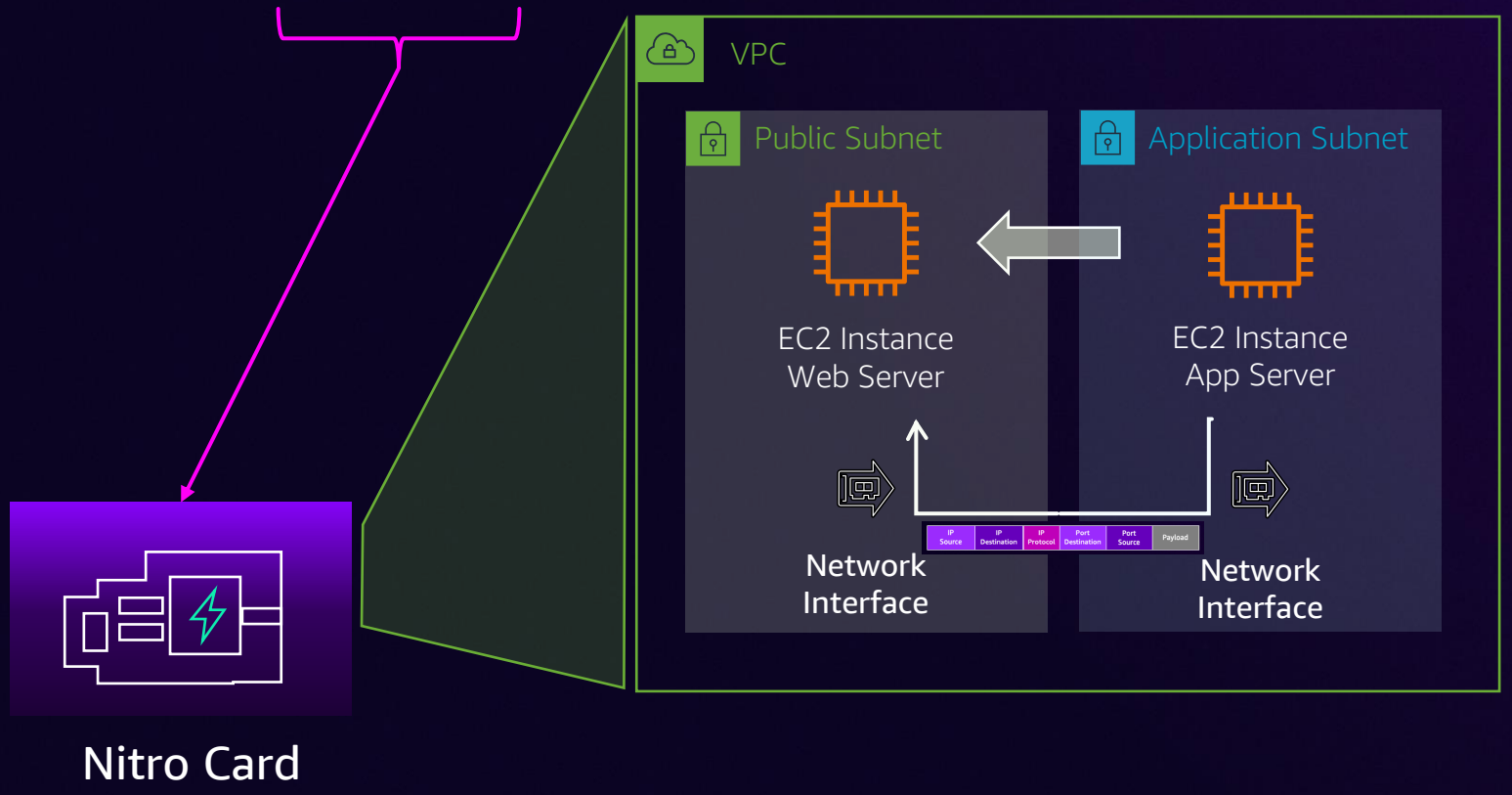




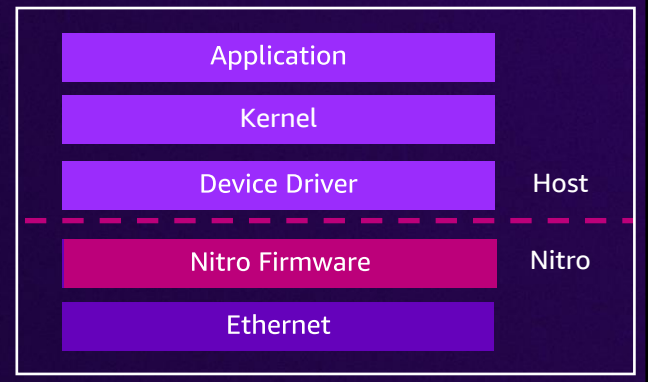
Flow Setup Processes



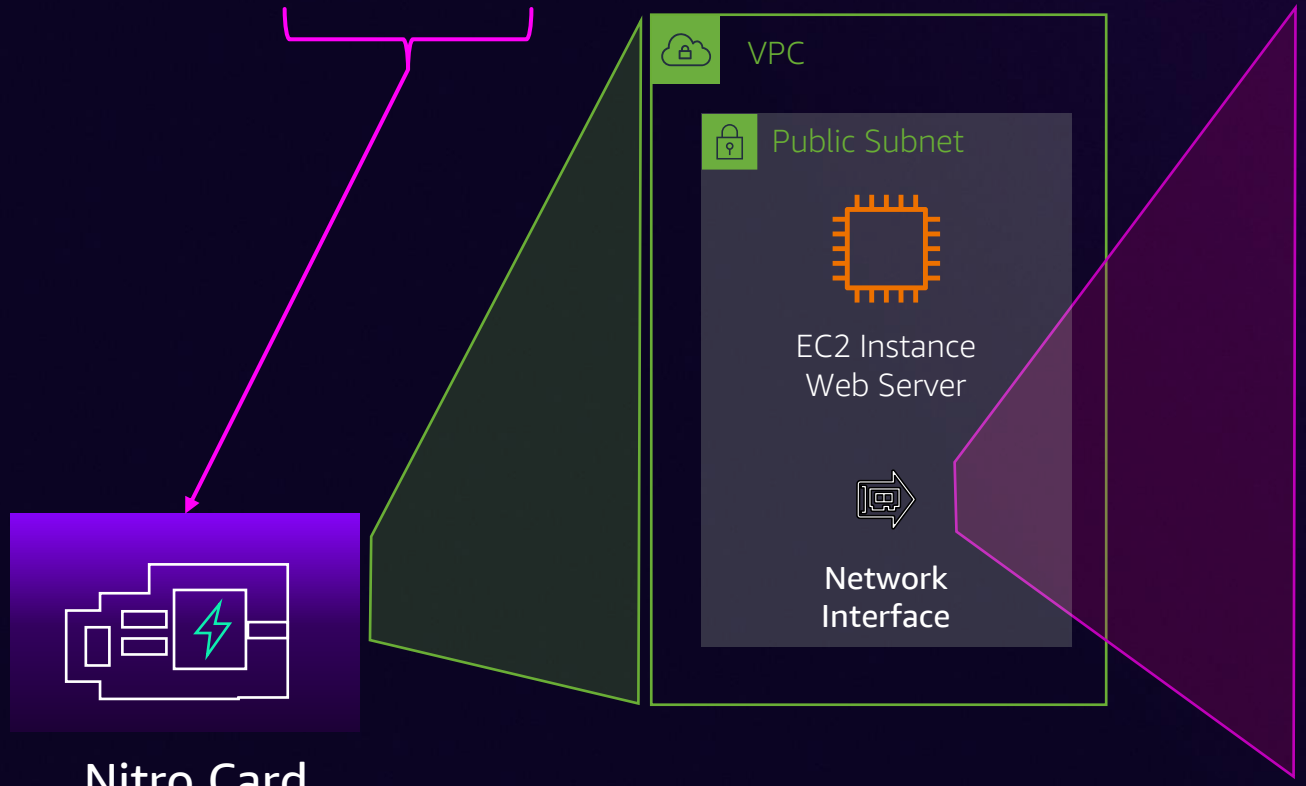




Flow Setup Complete



IP Source	IP Destination	IP Protocol	Port Source	Port Destination	Payload



Nitro State

Adjacency	Target
IGW	Unknown
App Server ENI	IP-Address

Mapping State

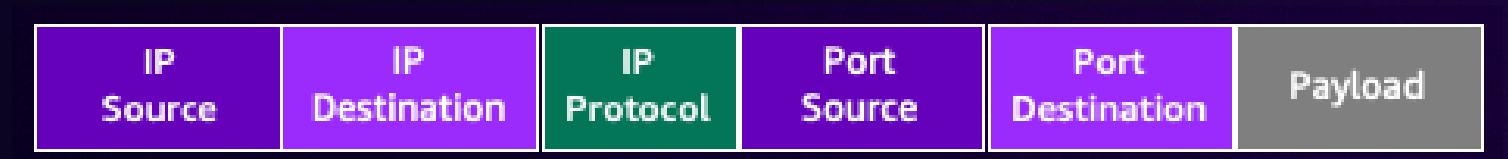
5-Tuple	Target
Web ⇔ App	Established

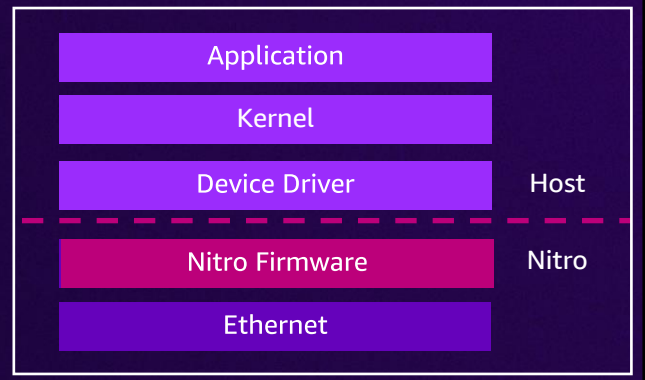
ConnTrack

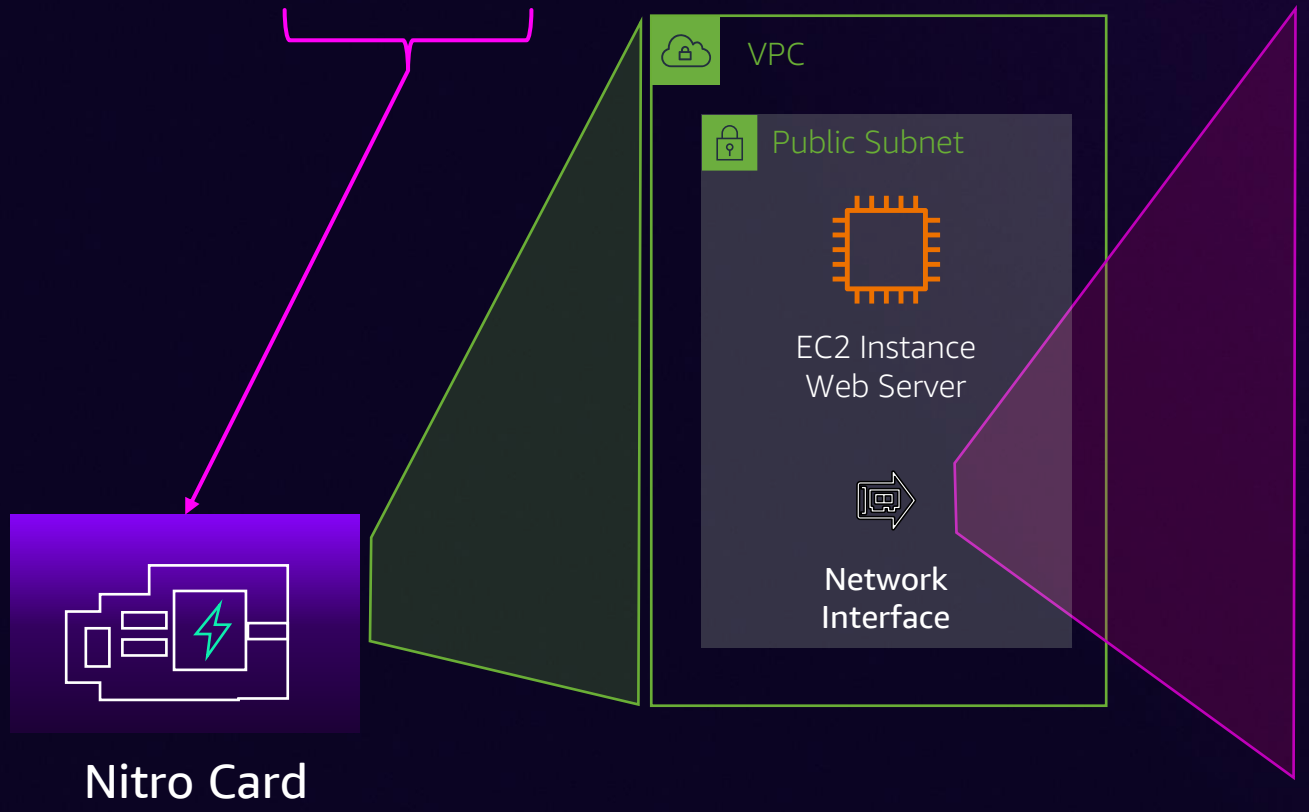
Flow Cache	State
Web to App	Complete
App to Web	Complete

Flow Cache

Flow Setup Complete







Nitro State

Adjacency	Target
IGW	Unknown
App Server ENI	IP-Address

Mapping State

5-Tuple	Target
Web ⇔ App	Established

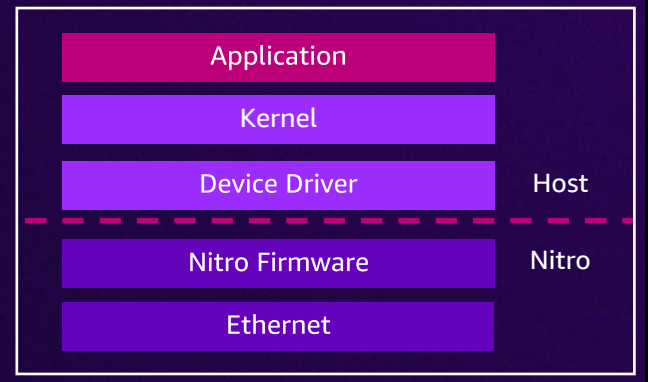
ConnTrack

Flow Cache	State
Web to App	Complete
App to Web	Complete

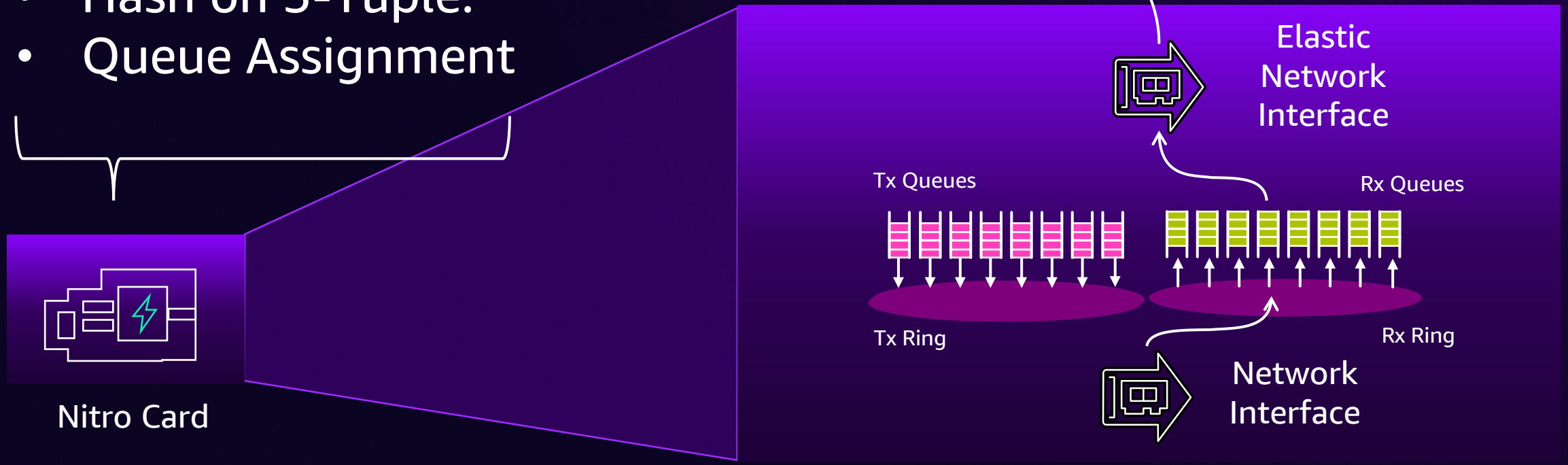
Flow Cache

Established Security Group Eval: Accelerated Flow With Higher Packets Per Second

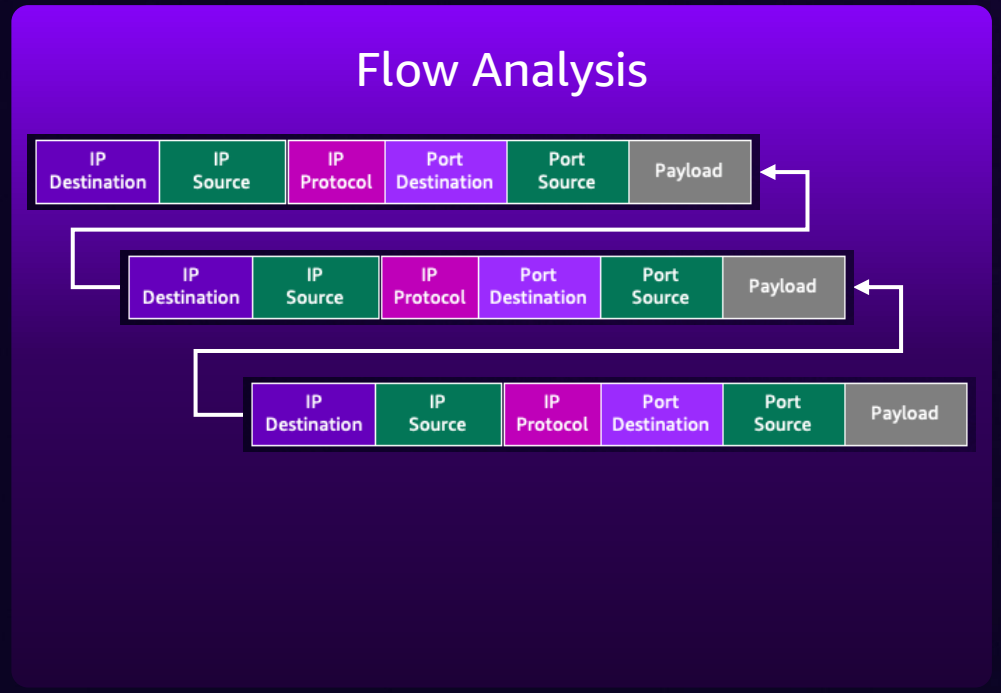
Flow Complete Inside Nitro



- Hash on 5-Tuple:
- Queue Assignment



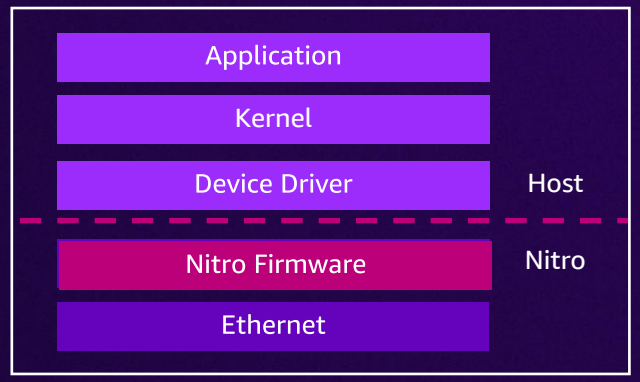
Flow Analysis

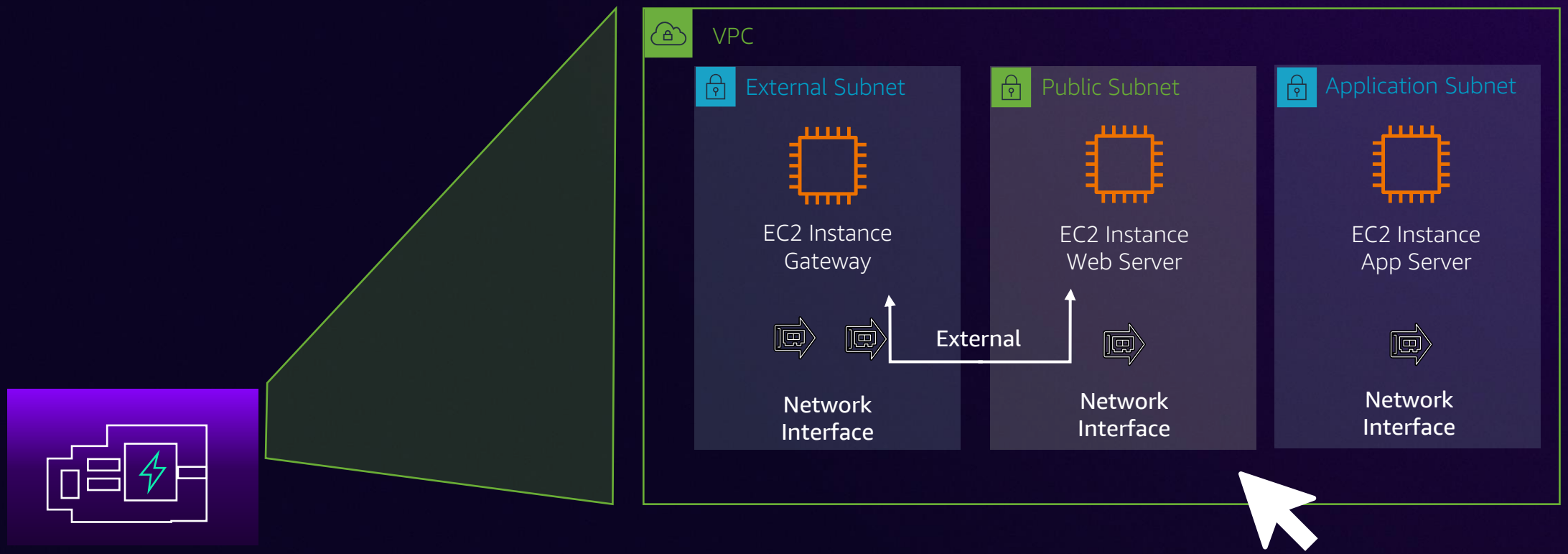


- Flow Specifications
- Micro bursting
- Flow State Anomalies
- “Whale” Flows

Single flow Specifications

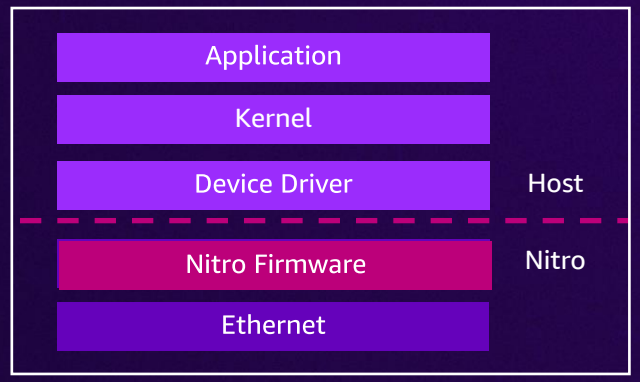
- External Flow – 5Gbps

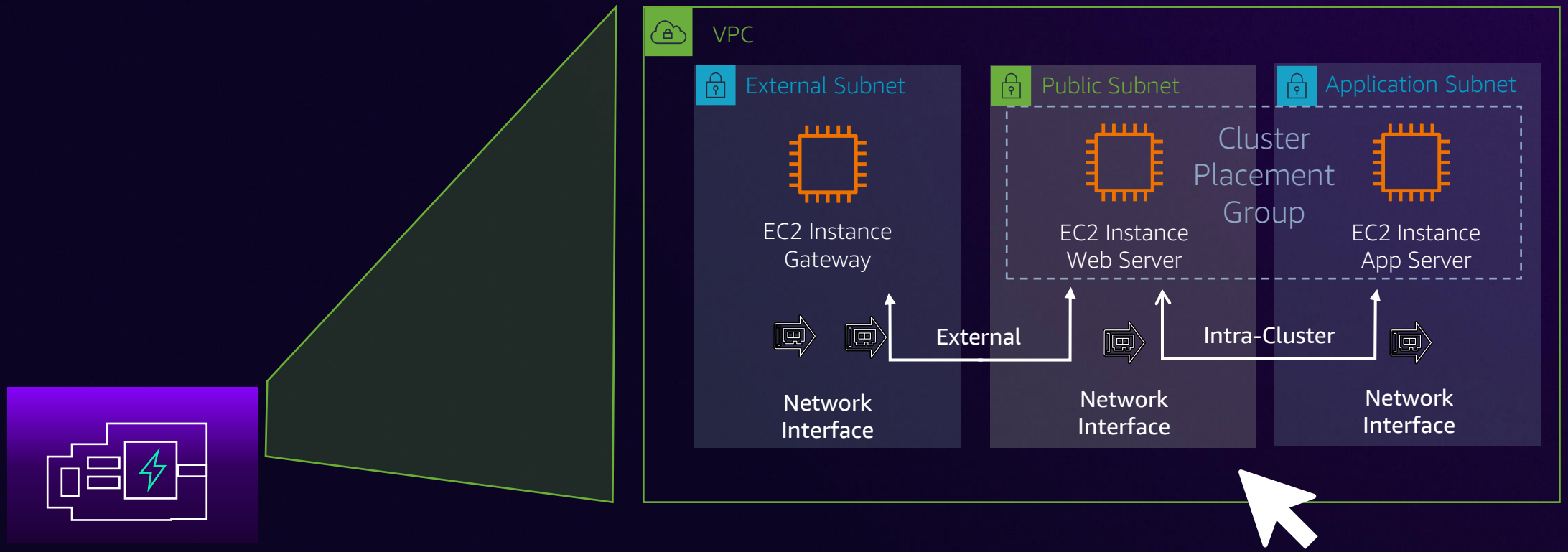




Single flow Specifications

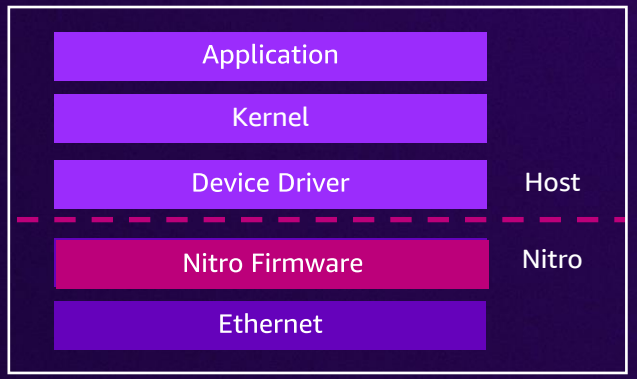
- External Flow – 5Gbps
- Cluster Placement Group – 10Gbps

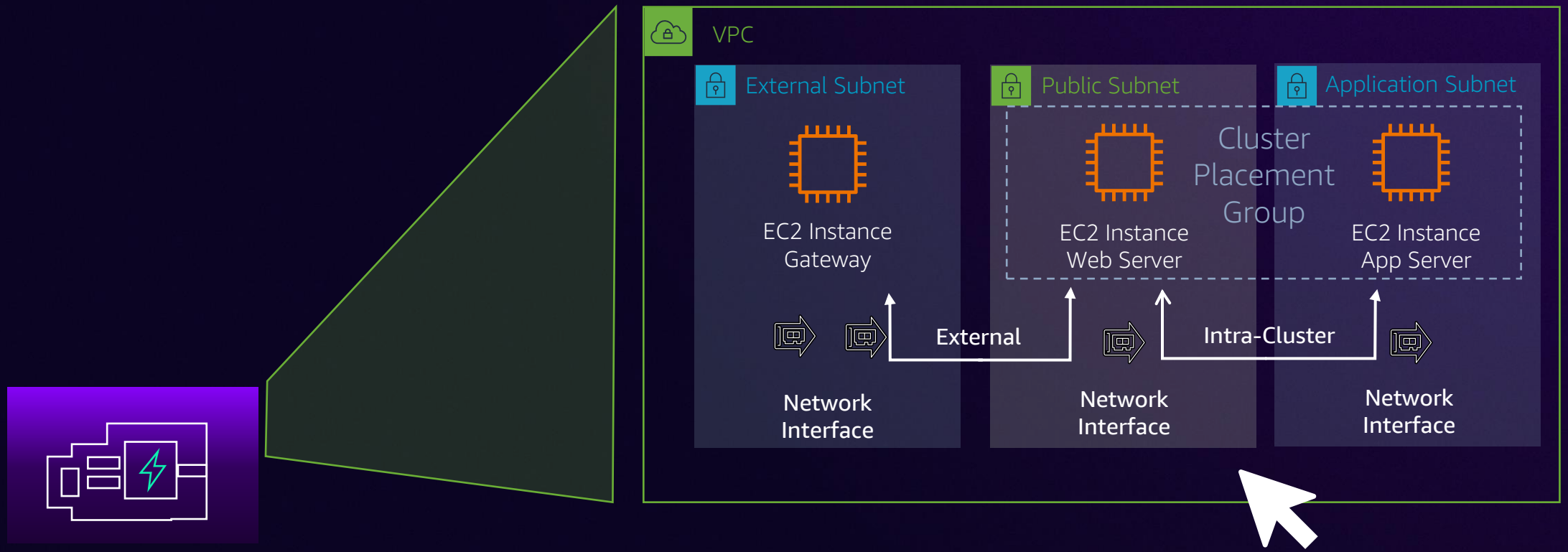




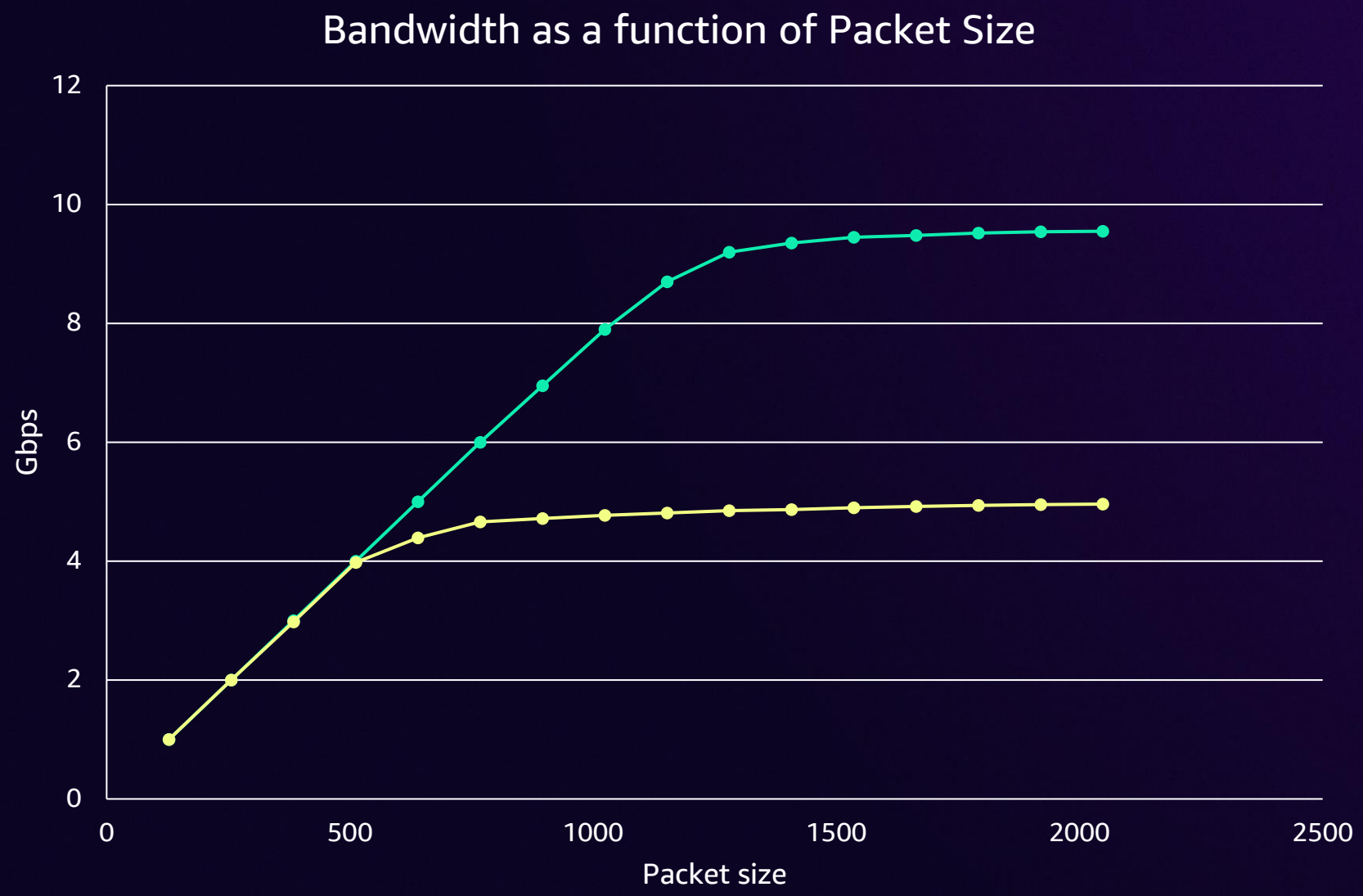
Single flow Specifications

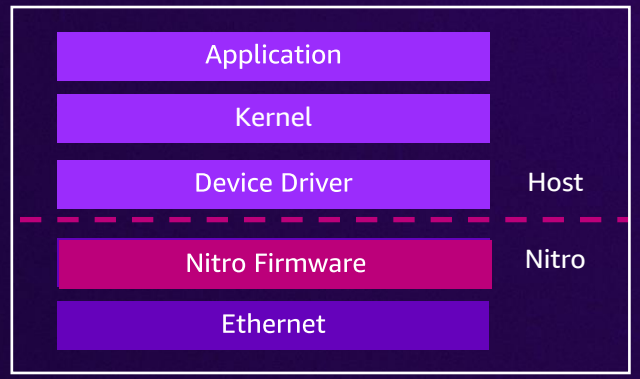
- External Flow – 5Gbps
- Cluster Placement Group – 10Gbps
- ENA Express Flow – 25Gbps



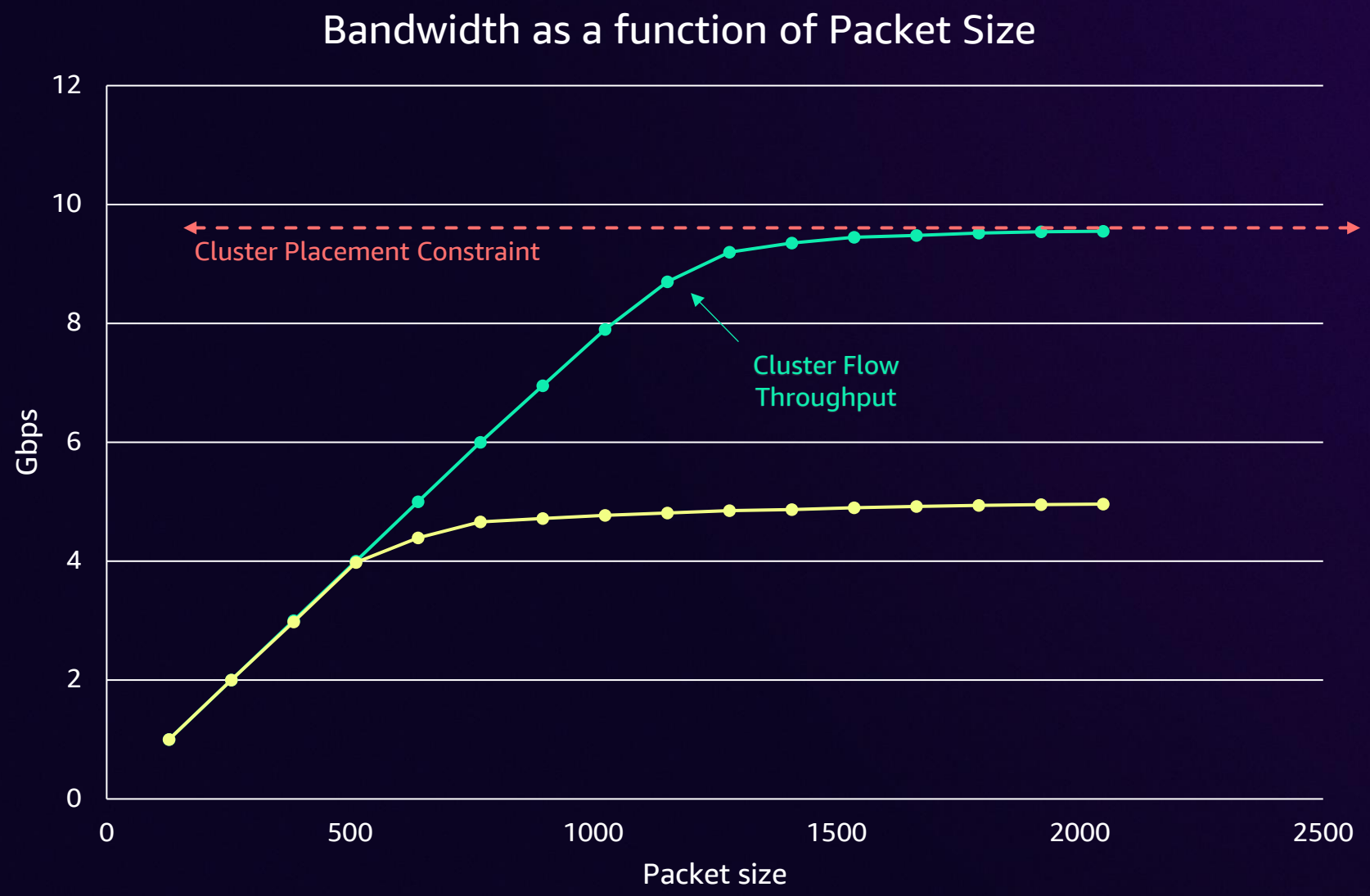


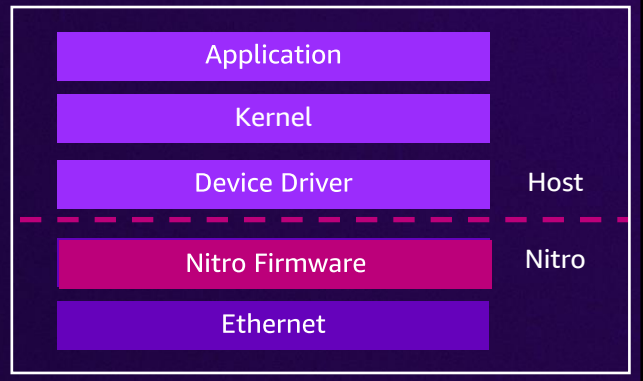
Flow Limits – PPS versus Bandwidth



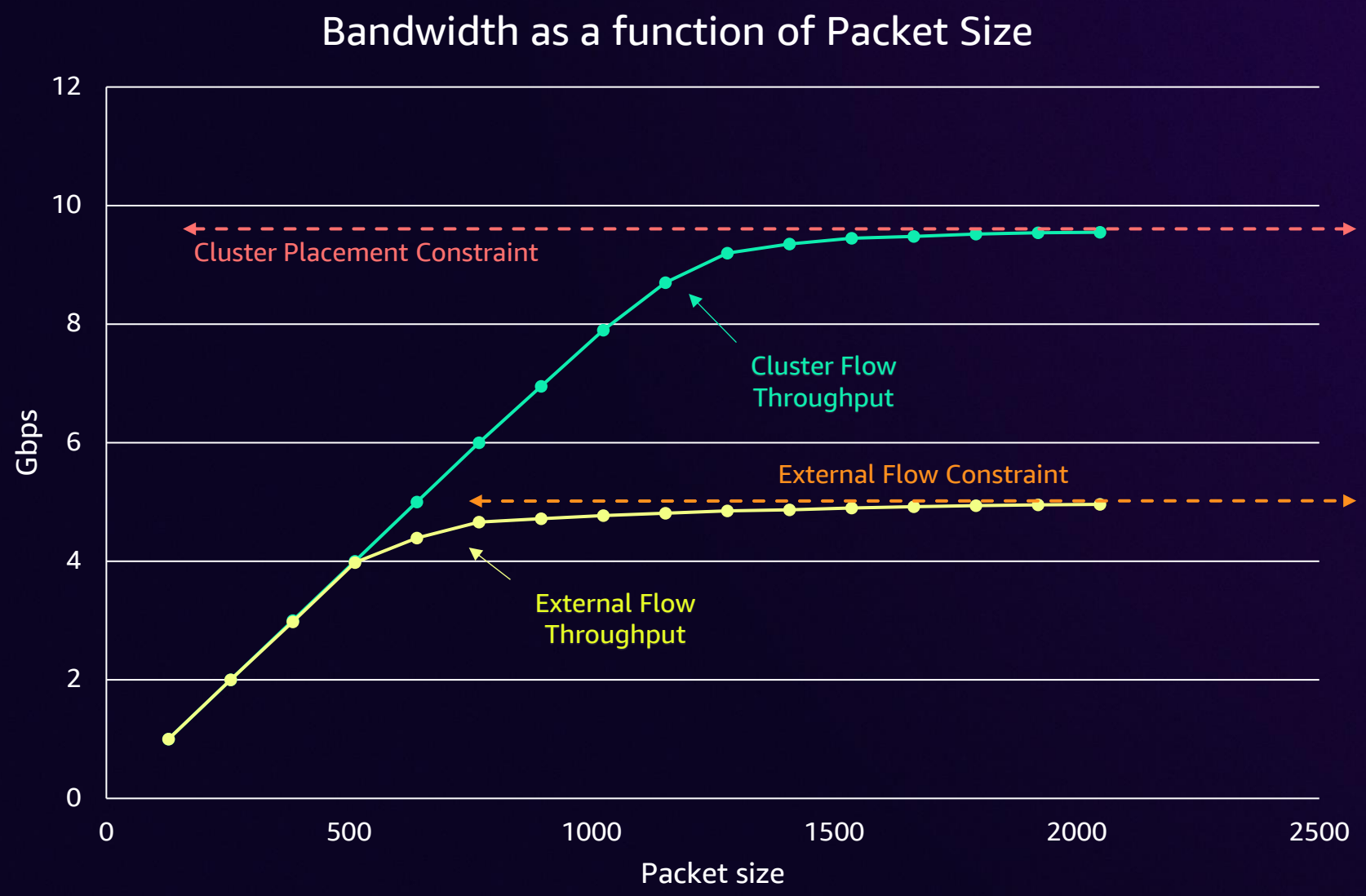


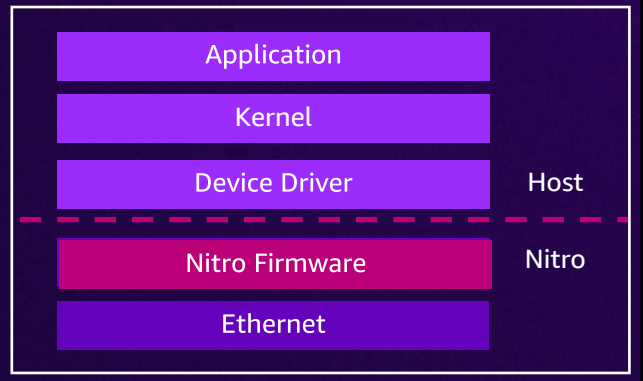
Flow Limits – PPS versus Bandwidth



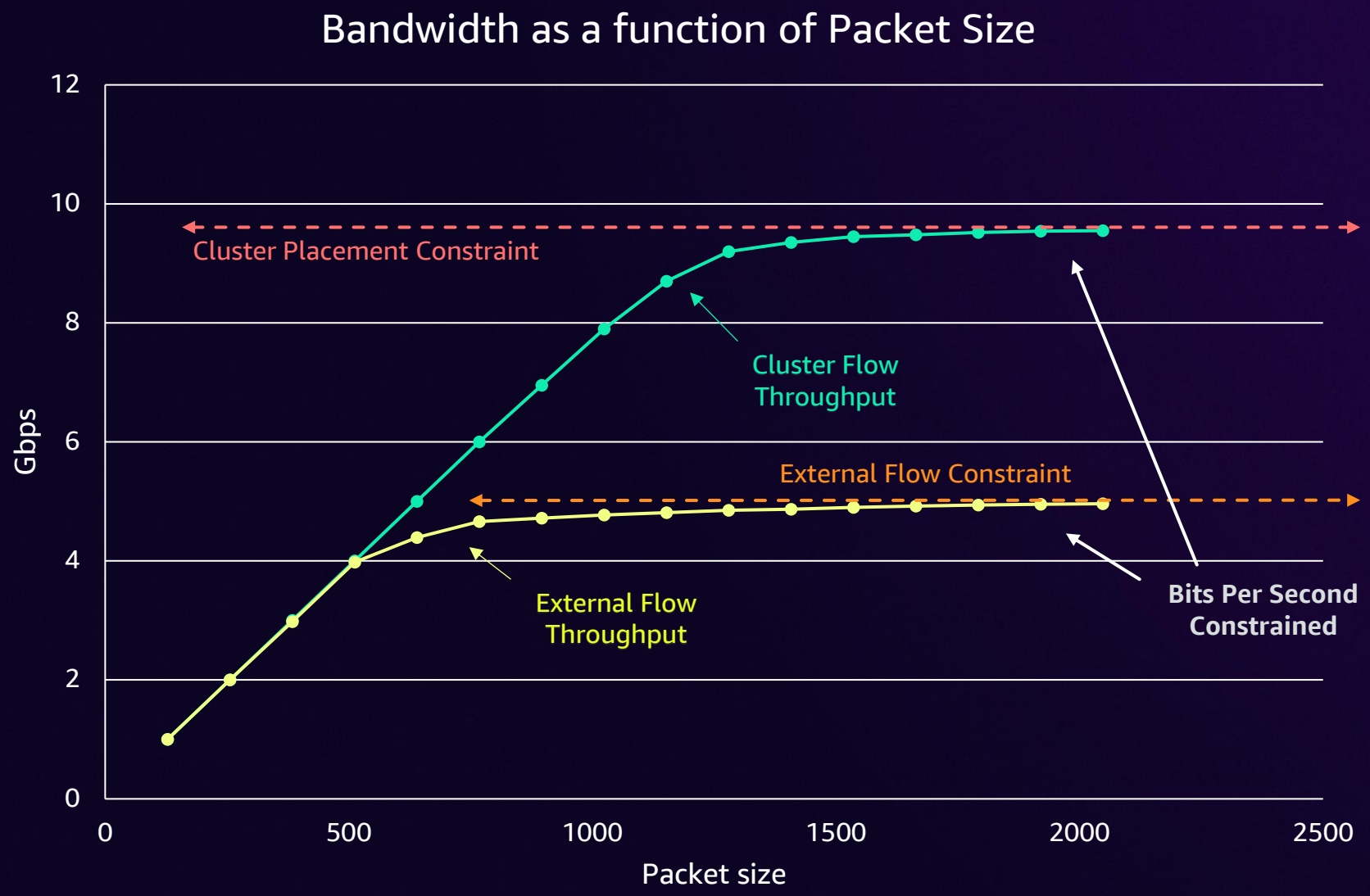


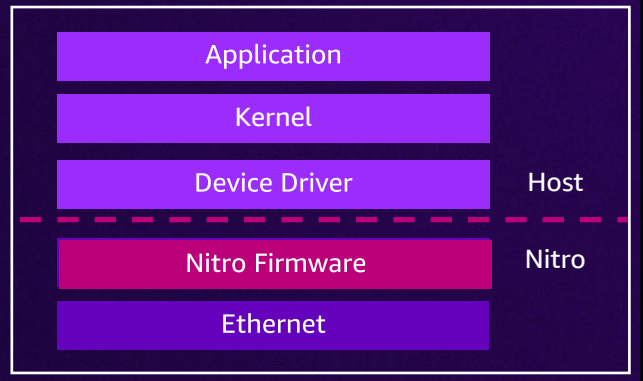
Flow Limits – PPS versus Bandwidth



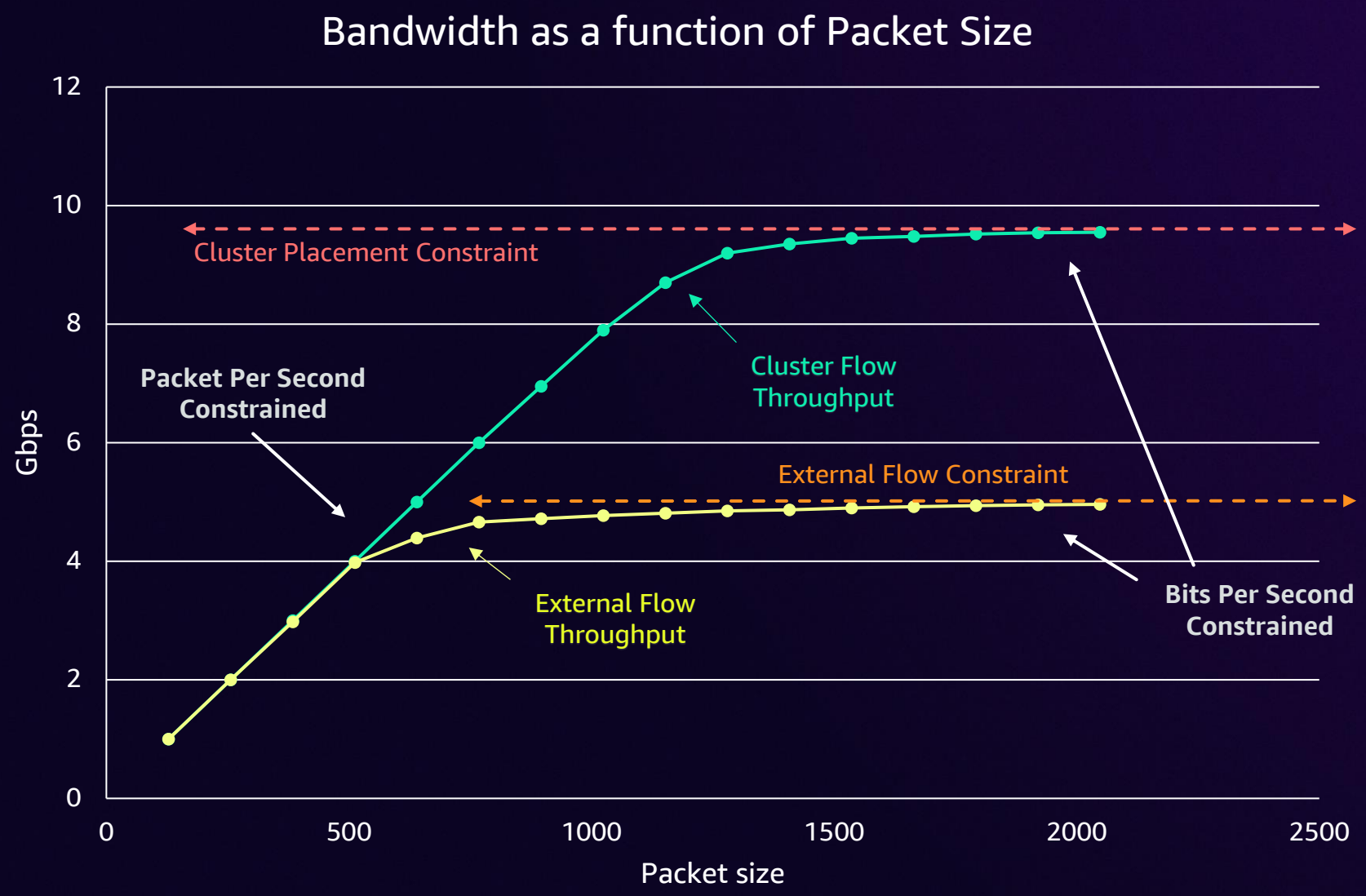


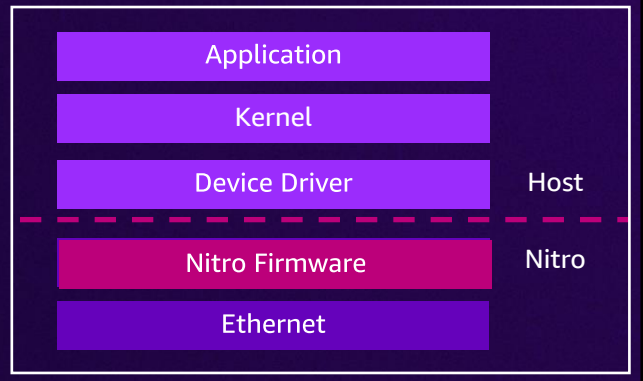
Flow Limits – PPS versus Bandwidth



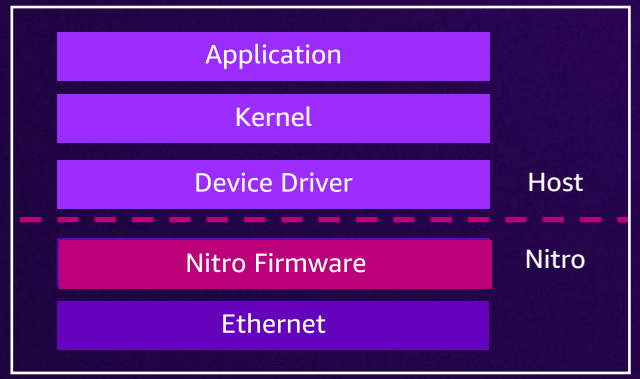


Flow Limits – PPS versus Bandwidth



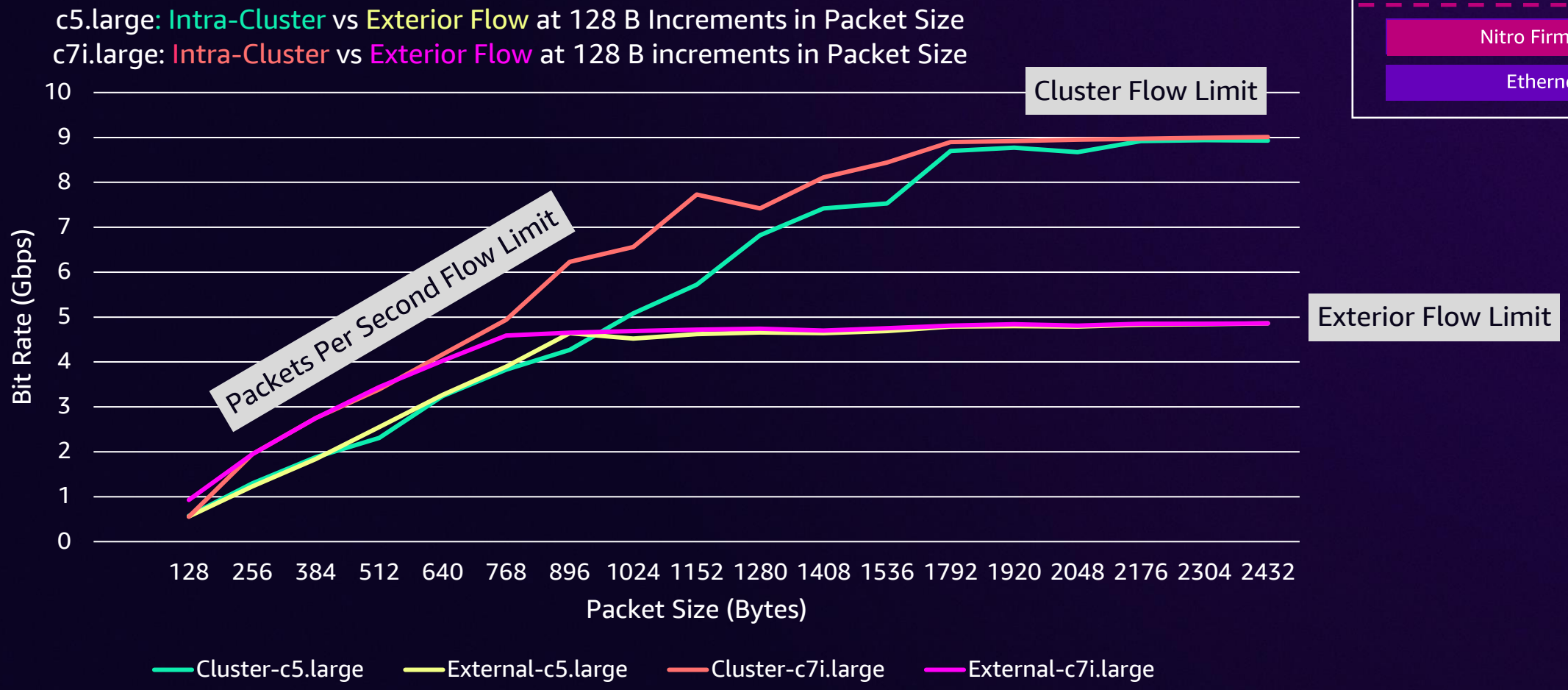


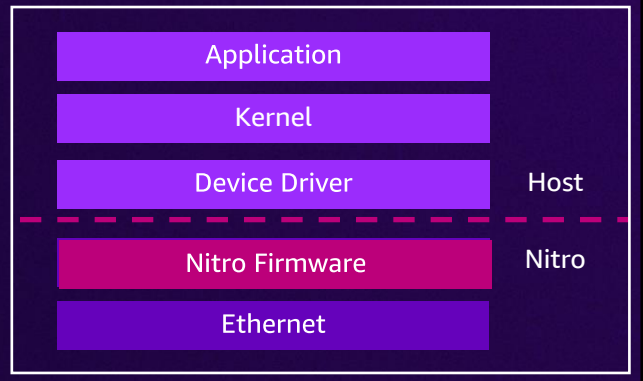
Test Method: c5.large vs c7i.large



CLIENT: iperf3 -c < app_server | gateway > -P 1 -l < buffer >
RECEIVER: iperf3 -s

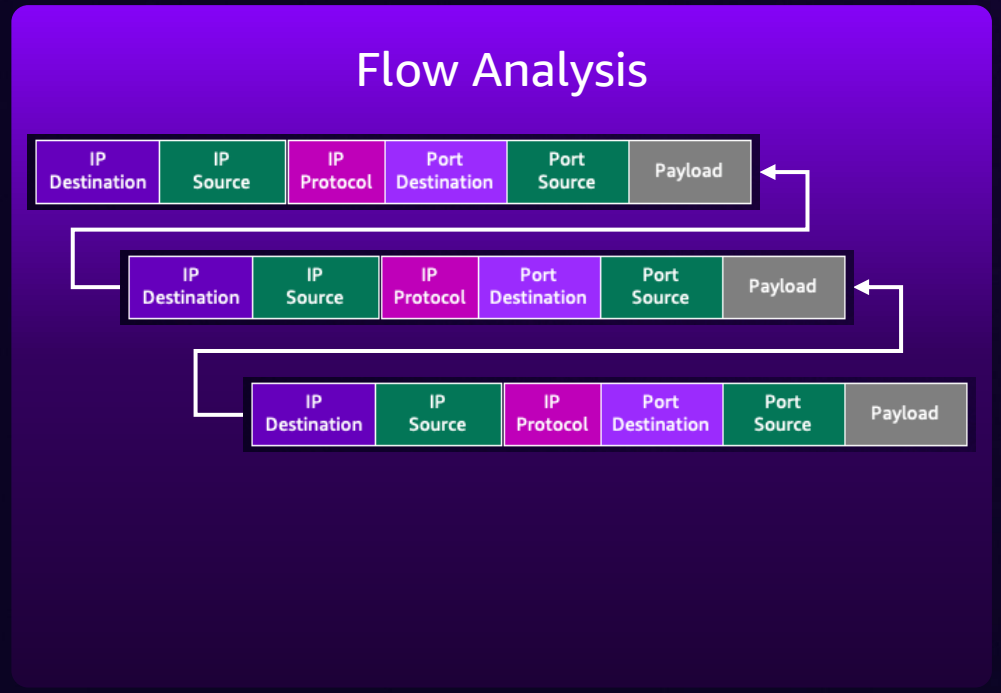
Test Method: c5.large vs c7i.large





CLIENT: iperf3 -c < app_server | gateway > -P 1 -l < buffer >
RECEIVER: iperf3 -s

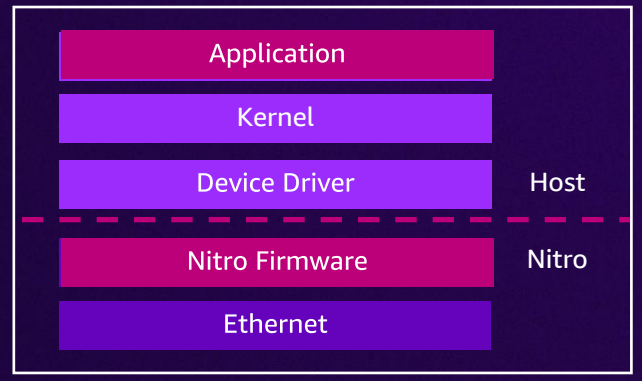
Flow Analysis



- Flow Specifications
- Micro bursting
- Flow State Anomalies
- “Whale” Flows

Microbursts

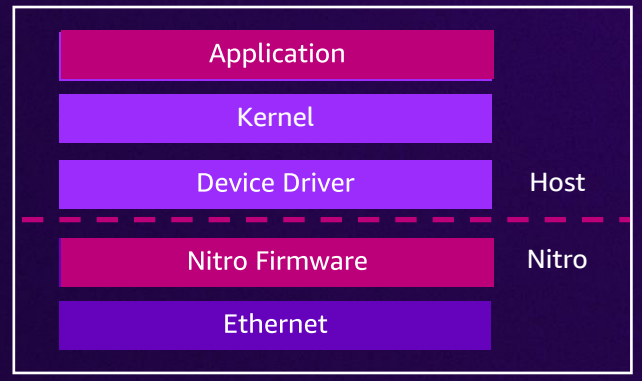
What is it?



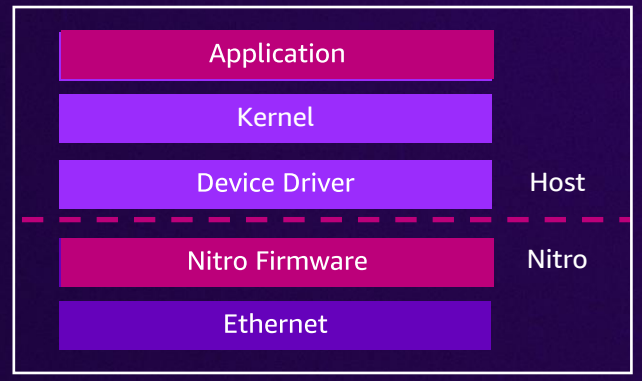
Microbursts

What is it?

How do you diagnose it?



Microbursts

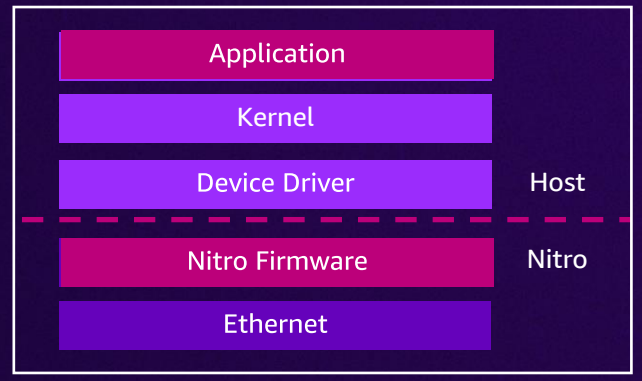


What is it?

How do you diagnose it?

What causes it?

Microbursts



What is it?

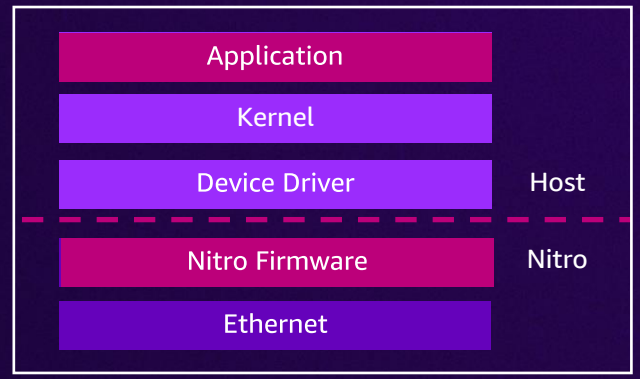
How do you diagnose it?

What causes it?

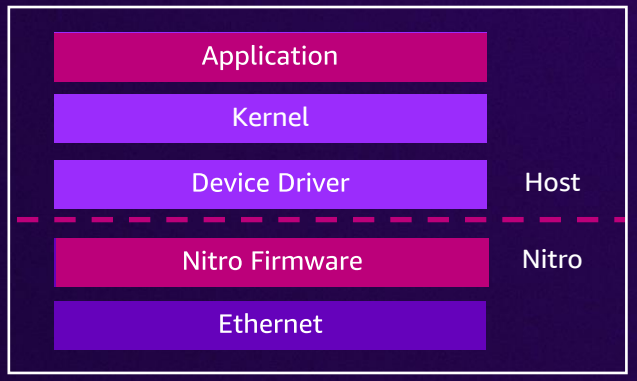
How do you prevent it?

Measuring Network Bursts

What actually happens

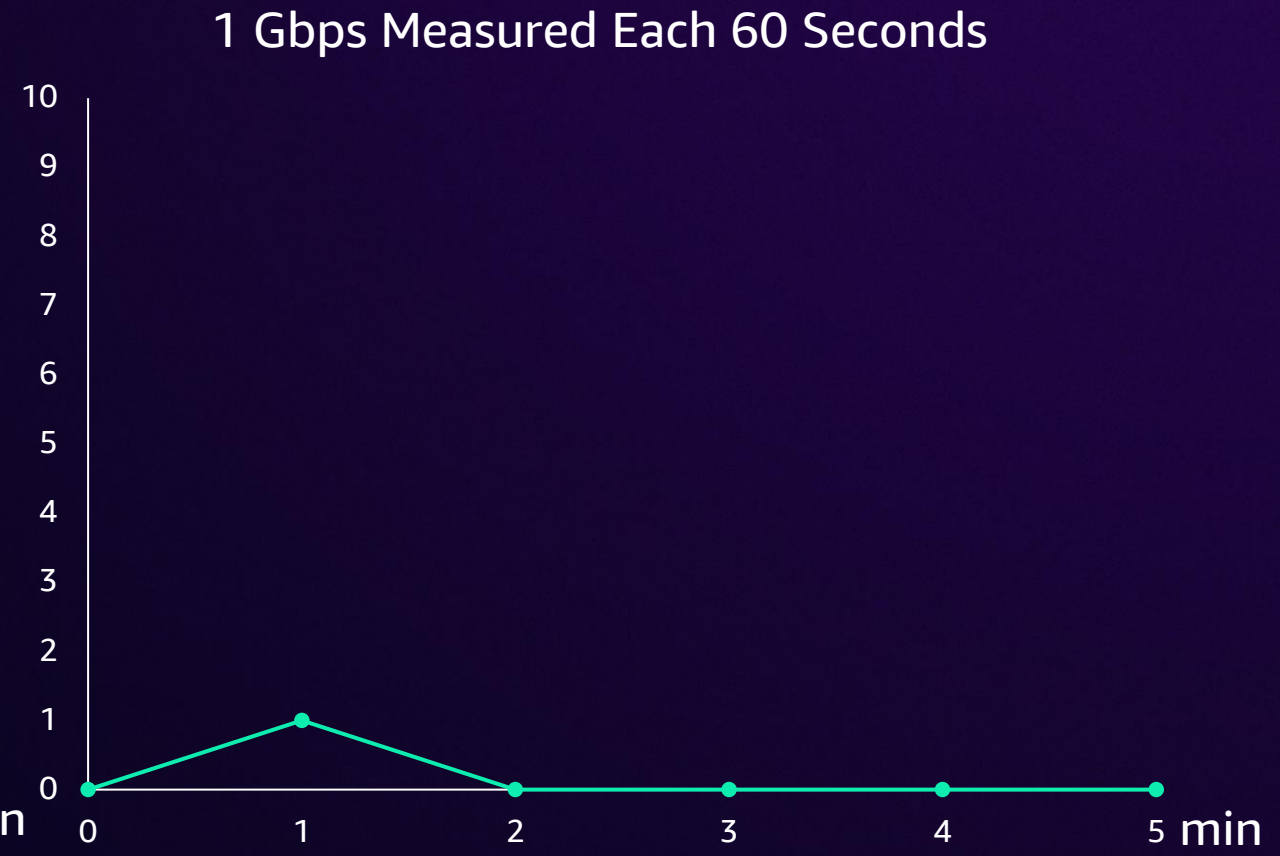


Measuring Network Bursts



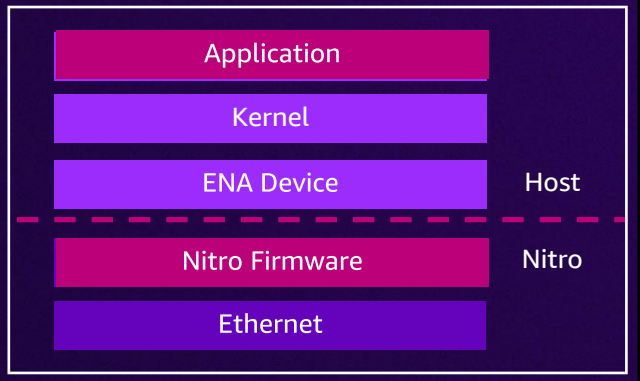
What actually happens

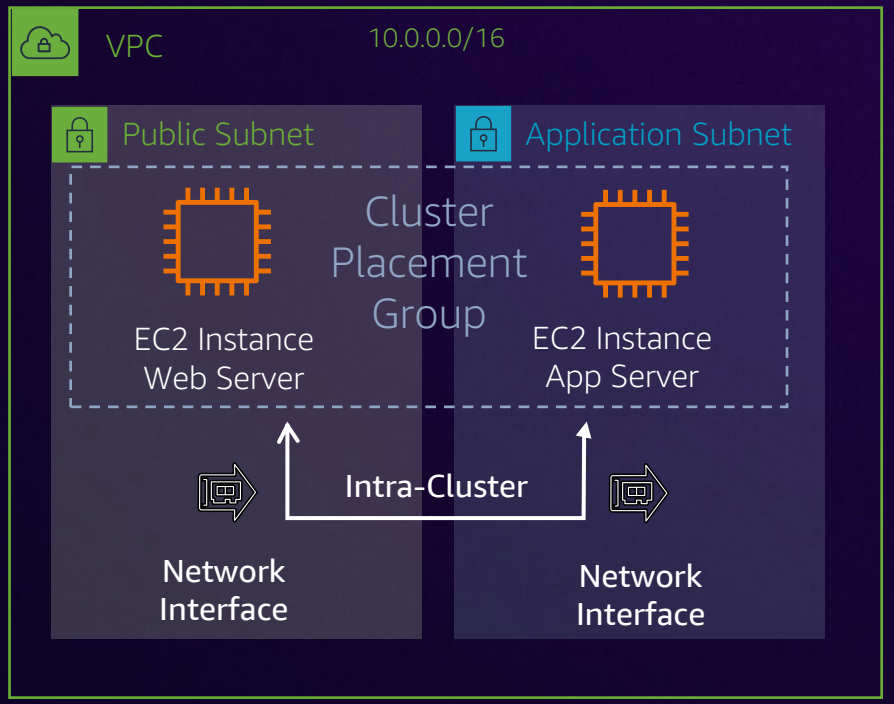
What your metrics show you



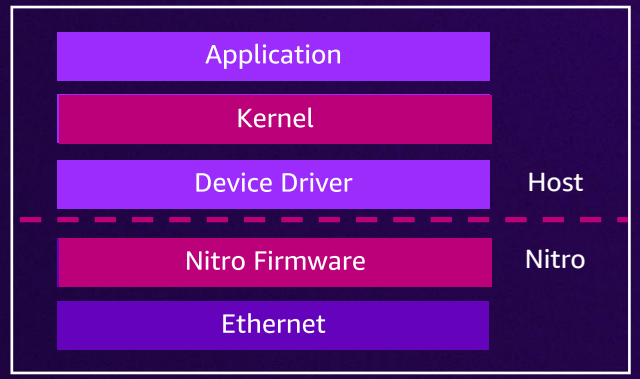
Test Topology

- Tool: iperf3
- Instance c7i.large
- Transmit to Receiver
- Large MTU to hit Burst Limit
- Short Test Interval
- Setup up CloudWatch Metrics
- Metric Expressions (Mbps)





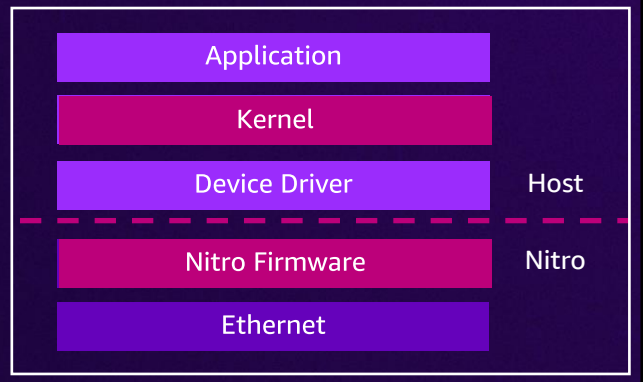
Test Method (iperf and CloudWatch)



IPerf3 Metrics: 6 sec @ 9530 Mbps

```
[ec2-user@web-serveriperf3 –c app-server -t 6 -M 9000 -P 1

Connecting to host app-server, port 5200

[  5] local web-server port 45742 connected to app-server port 5200

[ ID]          Interval          Transfer          Bitrate          Retr          Cwnd

[  5]          0.00-1.00   sec  1.11 GBytes     9.53 Gbits/sec    0          1.60 MBytes

[  5]          1.00-2.00   sec  1.11 GBytes     9.53 Gbits/sec    0          1.69 MBytes

[  5]          2.00-3.00   sec  1.11 GBytes     9.54 Gbits/sec    0          1.69 Mbytes

[  5]          3.00-4.00   sec  1.11 GBytes     9.53 Gbits/sec    0          1.69 MBytes

[  5]          4.00-5.00   sec  1.11 GBytes     9.52 Gbits/sec    0          1.83 MBytes

[  5]          5.00-6.00   sec  1.11 GBytes     9.53 Gbits/sec    0          1.83 MBytes

- - - - - - - - - - - - - - - - - - - - - - - - -

[ ID]          Interval          Transfer          Bitrate          Retr

[  5]          0.00-6.00   sec  6.66 GBytes     9.53 Gbits/sec    0

iperf Done.
```

Test Method (iperf and CloudWatch)

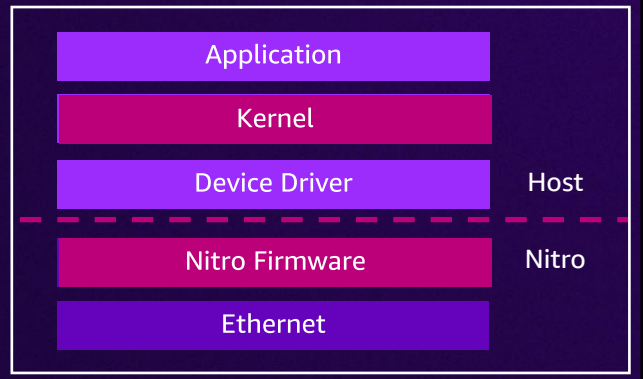


-t = duration sec
-M max segment size
-P streams

IPerf3 Metrics: 6 sec @ 9530 Mbps

```
[ec2-user@web-serveriperf3 –c app-server -t 6 -M 9000 -P 1
Connecting to host app-server, port 5200
[  5] local web-server port 45742 connected to app-server port 5200
[ ID]          Interval          Transfer          Bitrate          Retr          Cwnd
[  5]          0.00-1.00   sec  1.11 GBytes       9.53 Gbits/sec   0             1.60 MBytes
[  5]          1.00-2.00   sec  1.11 GBytes       9.53 Gbits/sec   0             1.69 MBytes
[  5]          2.00-3.00   sec  1.11 GBytes       9.54 Gbits/sec   0             1.69 Mbytes
[  5]          3.00-4.00   sec  1.11 GBytes       9.53 Gbits/sec   0             1.69 MBytes
[  5]          4.00-5.00   sec  1.11 GBytes       9.52 Gbits/sec   0             1.83 MBytes
[  5]          5.00-6.00   sec  1.11 GBytes       9.53 Gbits/sec   0             1.83 MBytes
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID]          Interval          Transfer          Bitrate          Retr
[  5]          0.00-6.00   sec  6.66 GBytes       9.53 Gbits/sec   0
iperf Done.
```

Test Method (iperf and CloudWatch)

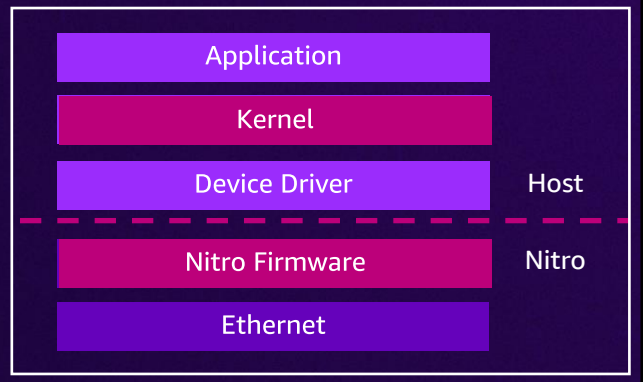


-t = duration sec
-M max segment size
-P streams

IPerf3 Metrics: 6 sec @ 9530 Mbps

```
[ec2-user@web-serveriperf3 –c app-server -t 6 -M 9000 -P 1
Connecting to host app-server, port 5200
[  5] local web-server port 45742 connected to app-server port 5200
[ ID]          Interval          Transfer          Bitrate          Retr          Cwnd
[  5]          0.00-1.00   sec  1.11 GBytes       9.53 Gbits/sec   0             1.60 MBytes
[  5]          1.00-2.00   sec  1.11 GBytes       9.53 Gbits/sec   0             1.69 MBytes
[  5]          2.00-3.00   sec  1.11 GBytes       9.54 Gbits/sec   0             1.69 Mbytes
[  5]          3.00-4.00   sec  1.11 GBytes       9.53 Gbits/sec   0             1.69 MBytes
[  5]          4.00-5.00   sec  1.11 GBytes       9.52 Gbits/sec   0             1.83 MBytes
[  5]          5.00-6.00   sec  1.11 GBytes       9.53 Gbits/sec   0             1.83 MBytes
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID]          Interval          Transfer          Bitrate          Retr
[  5]          0.00-6.00   sec  6.66 GBytes       9.53 Gbits/sec   0
iperf Done.
```

Test Method (iperf and CloudWatch)

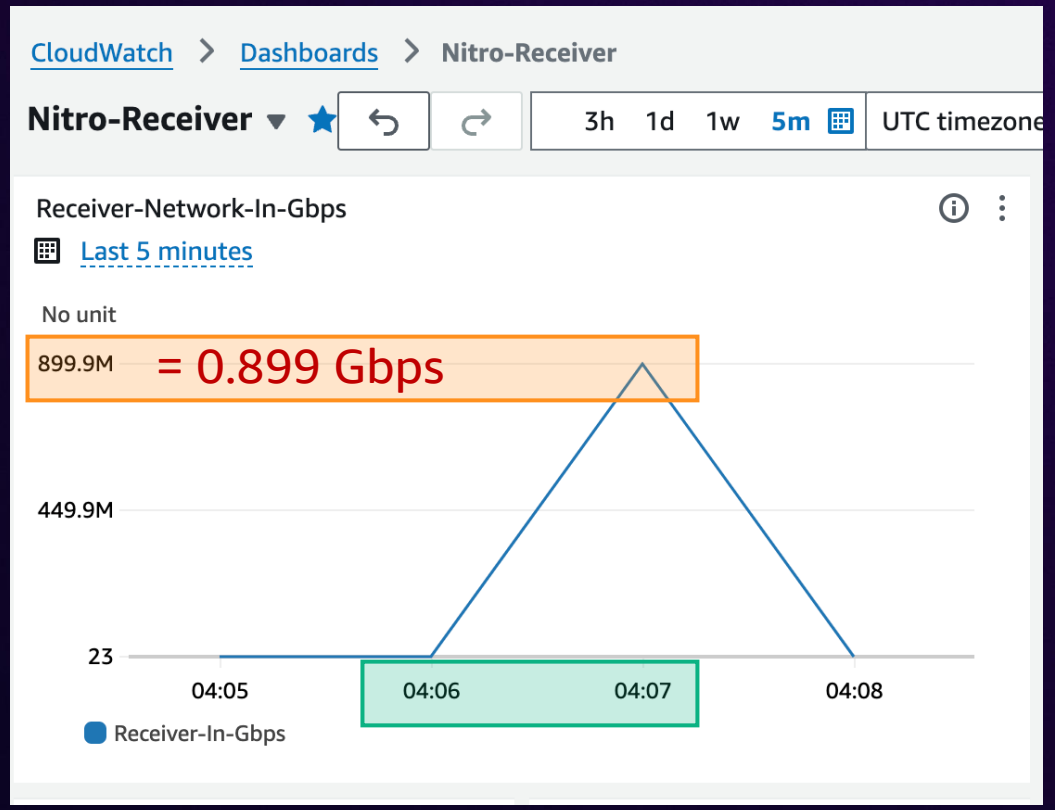


-t = duration sec
-M max segment size
-P streams

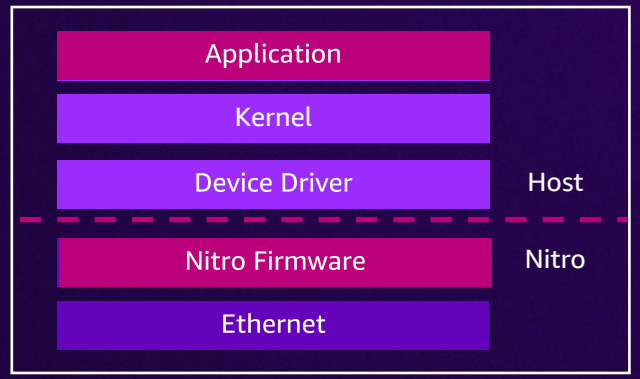
IPerf3 Metrics: 6 sec @ 9530 Mbps

```
[ec2-user@web-serveriperf3 –c app-server -t 6 -M 9000 -P 1
Connecting to host app-server, port 5200
[  5] local web-server port 45742 connected to app-server port 5200
[ ID]        Interval          Transfer      Bitrate         Retr     Cwnd
[  5]        0.00-1.00   sec  1.11 GBytes   9.53 Gbits/sec  0        1.60 MBytes
[  5]        1.00-2.00   sec  1.11 GBytes   9.53 Gbits/sec  0        1.69 MBytes
[  5]        2.00-3.00   sec  1.11 GBytes   9.54 Gbits/sec  0        1.69 Mbytes
[  5]        3.00-4.00   sec  1.11 GBytes   9.53 Gbits/sec  0        1.69 MBytes
[  5]        4.00-5.00   sec  1.11 GBytes   9.52 Gbits/sec  0        1.83 MBytes
[  5]        5.00-6.00   sec  1.11 GBytes   9.53 Gbits/sec  0        1.83 MBytes
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID]        Interval          Transfer      Bitrate         Retr
[  5]        0.00-6.00   sec  6.66 GBytes   9.53 Gbits/sec  0
iperf Done.
```

CloudWatch Metrics: 60sec @ 899 Mbps



Test Method : iperf microburst

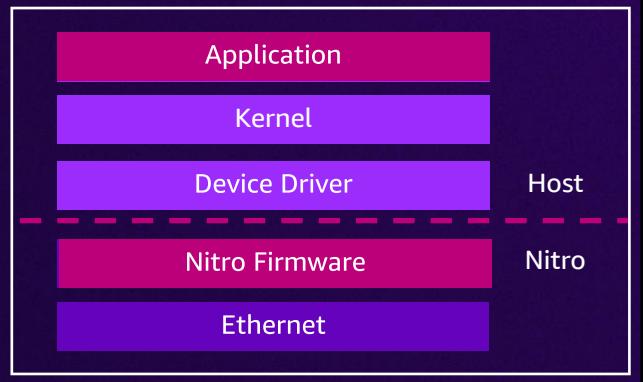


IPerf3 Metrics: 1 sec with 8 flows @ 10 Gbps target

```
[ec2-user@web-server]$ iperf3 -c app-server -t 1 -p 5201 -M 9000 -P 8 -u -b 10G
[ ID]   Interval           Transfer     Bitrate         Total Datagrams
[  5]   0.00-1.00   sec   397 MBytes     3.33 Gbits/sec  46261
[  7]   0.00-1.00   sec   391 MBytes     3.28 Gbits/sec  45610
[  9]   0.00-1.00   sec   390 MBytes     3.27 Gbits/sec  45444
[ 11]   0.00-1.00   sec   390 MBytes     3.27 Gbits/sec  45459
[ 13]   0.00-1.00   sec   389 MBytes     3.26 Gbits/sec  45279
[ 15]   0.00-1.00   sec   391 MBytes     3.28 Gbits/sec  45577
[ 17]   0.00-1.00   sec   389 MBytes     3.26 Gbits/sec  45271
[ 19]   0.00-1.00   sec   389 MBytes     3.27 Gbits/sec  45378
[SUM]   0.00-1.00   sec  3.05 GBytes     26.2 Gbits/sec  364279
```

- iperf Done.

Test Method : iperf microburst



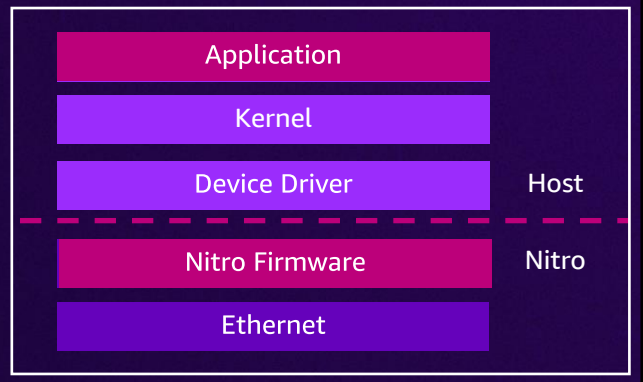
IPerf3 Metrics: 1 sec with 8 flows @ 10 Gbps target

```
[ec2-user@web-server]$ iperf3 -c app-server -t 1 -p 5201 -M 9000 -P 8 -u -b 10G
[ ID]   Interval           Transfer     Bitrate         Total Datagrams
[  5]   0.00-1.00   sec   397 MBytes     3.33 Gbits/sec  46261
[  7]   0.00-1.00   sec   391 MBytes     3.28 Gbits/sec  45610
[  9]   0.00-1.00   sec   390 MBytes     3.27 Gbits/sec  45444
[ 11]   0.00-1.00   sec   390 MBytes     3.27 Gbits/sec  45459
[ 13]   0.00-1.00   sec   389 MBytes     3.26 Gbits/sec  45279
[ 15]   0.00-1.00   sec   391 MBytes     3.28 Gbits/sec  45577
[ 17]   0.00-1.00   sec   389 MBytes     3.26 Gbits/sec  45271
[ 19]   0.00-1.00   sec   389 MBytes     3.27 Gbits/sec  45378
[SUM]   0.00-1.00   sec  3.05 GBytes     26.2 Gbits/sec  364279
```

- iperf Done.

-t = duration in sec
-M max segment size
-P streams
-u UDP
-b bit rate

Test Method : iperf microburst



IPerf3 Metrics: 1 sec with 8 flows @ 10 Gbps target

```
[ec2-user@web-server]$ iperf3 -c app-server -t 1 -p 5201 -M 9000 -P 8 -u -b 10G
[ ID]   Interval           Transfer     Bitrate         Total Datagrams
[  5]   0.00-1.00   sec   397 MBytes     3.33 Gbits/sec  46261
[  7]   0.00-1.00   sec   391 MBytes     3.28 Gbits/sec  45610
[  9]   0.00-1.00   sec   390 MBytes     3.27 Gbits/sec  45444
[ 11]   0.00-1.00   sec   390 MBytes     3.27 Gbits/sec  45459
[ 13]   0.00-1.00   sec   389 MBytes     3.26 Gbits/sec  45279
[ 15]   0.00-1.00   sec   391 MBytes     3.28 Gbits/sec  45577
[ 17]   0.00-1.00   sec   389 MBytes     3.26 Gbits/sec  45271
[ 19]   0.00-1.00   sec   389 MBytes     3.27 Gbits/sec  45378
[SUM]   0.00-1.00   sec  3.05 GBytes     26.2 Gbits/sec  364279
```

- iperf Done.

-t = duration in sec
-M max segment size
-P streams
-u UDP
-b bit rate

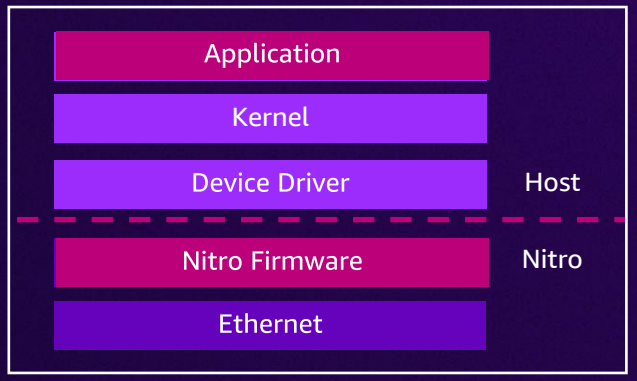
What causes it?

Issue:

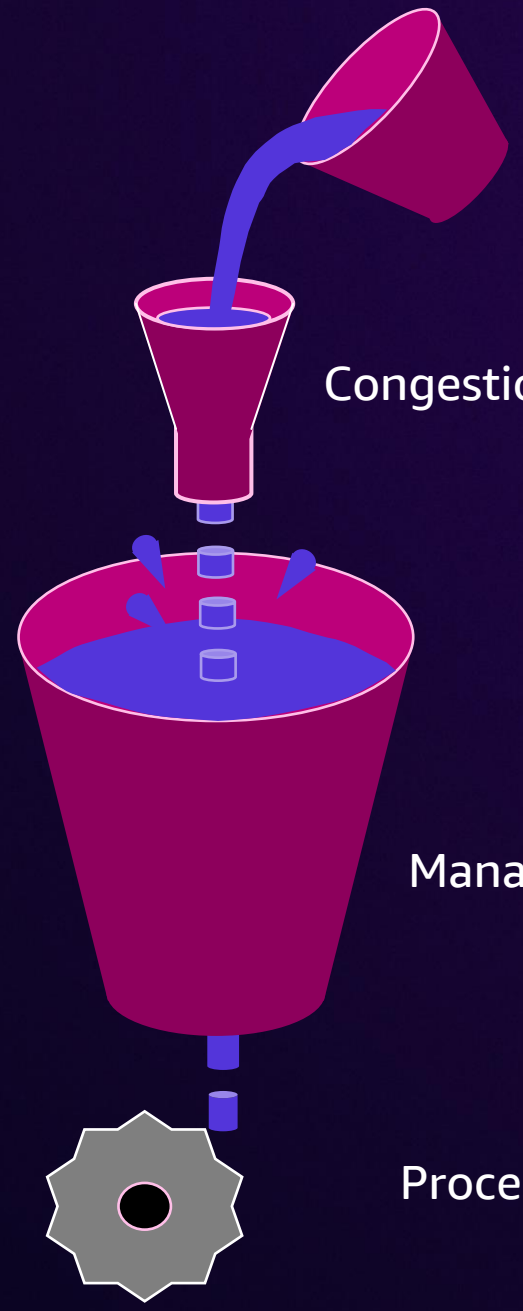
Packet Arrival Rate Exceeds Packet Processing Rate

Impact:

- Partially Full Queue → Delay
- Completely Full Queue → Drop



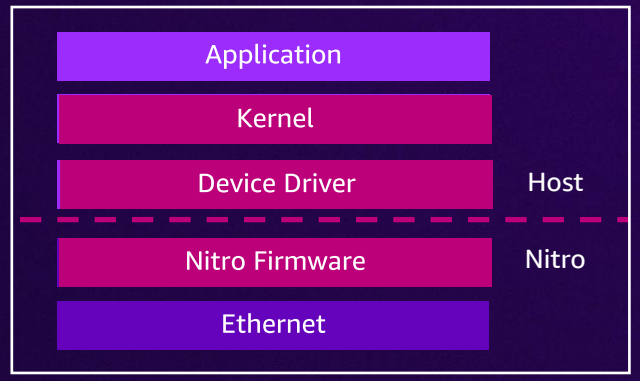
Preventing Microbursts



Congestion Controls

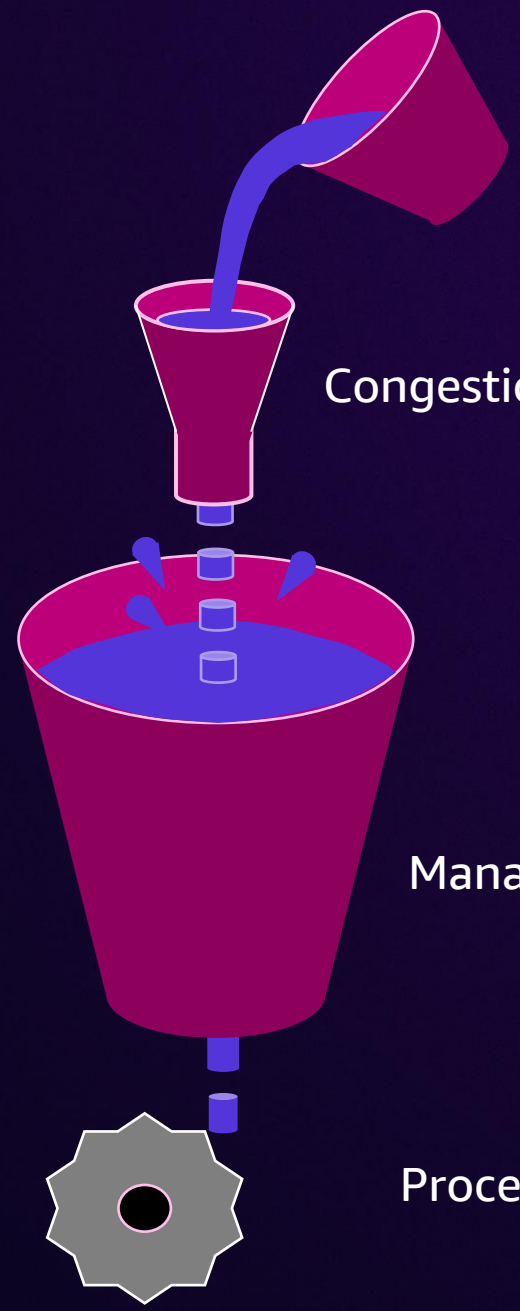
Managed Depth

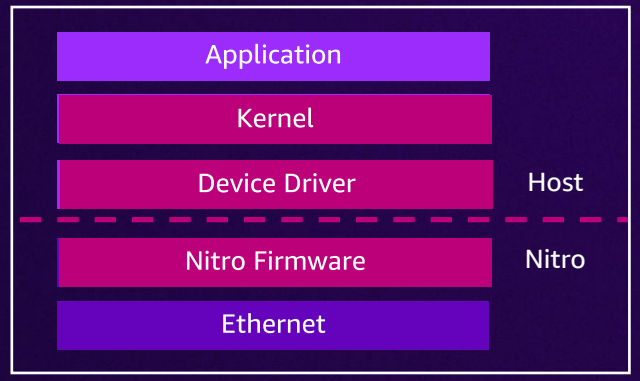
Processing Rate



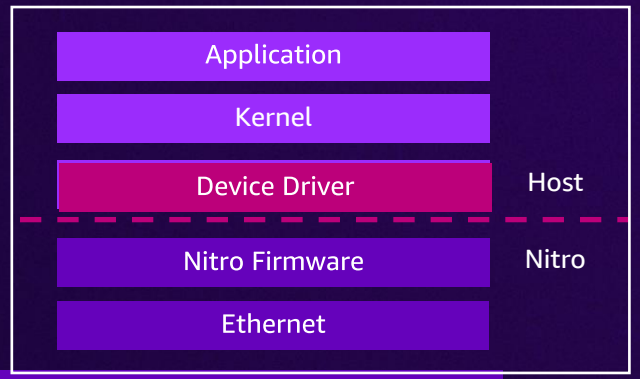
Preventing Microbursts

- General RX
  - Increase Queue Depth
- General TX
  - Traffic Control
    - Queuing discs
    - Fair queueing
- Nitro v4+ RX
  - ENA Express

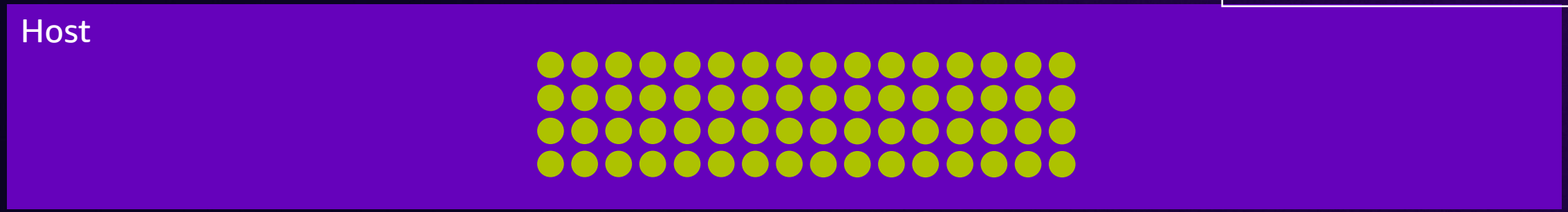




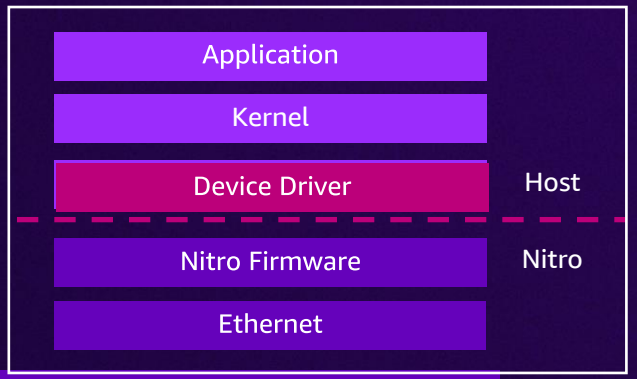
Nitro ENA Rx Rings and Tx Rings



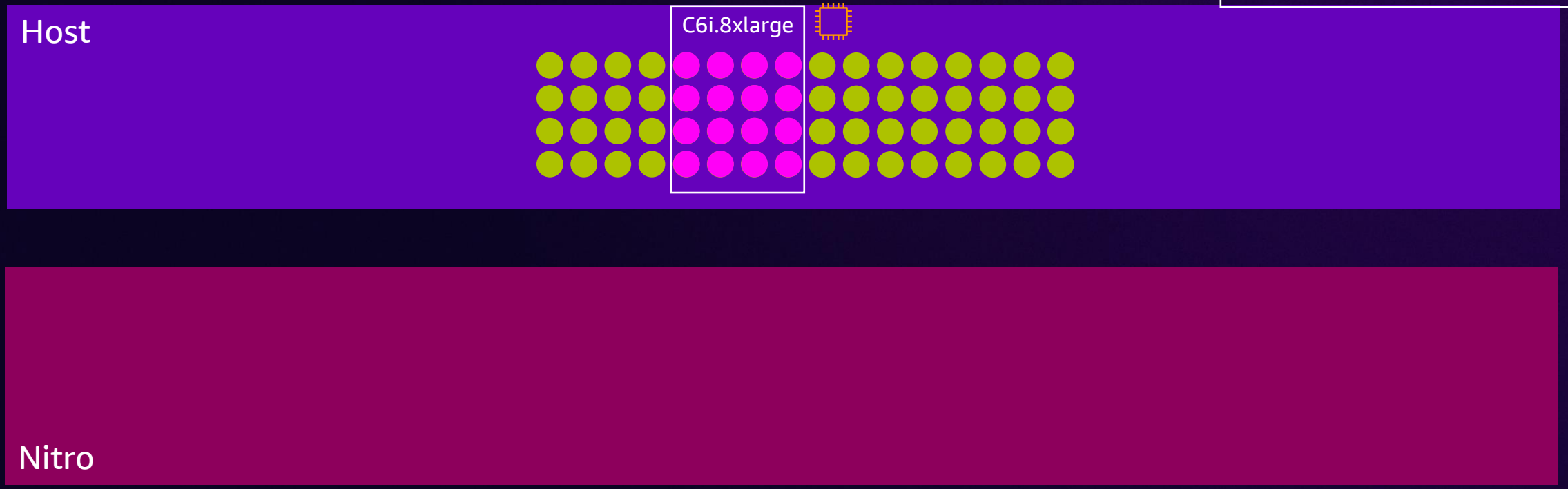
Example: c6i Host (64 core)



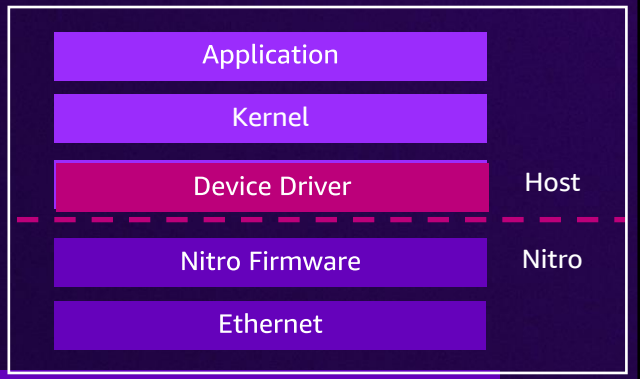
Nitro ENA Rx Rings and Tx Rings



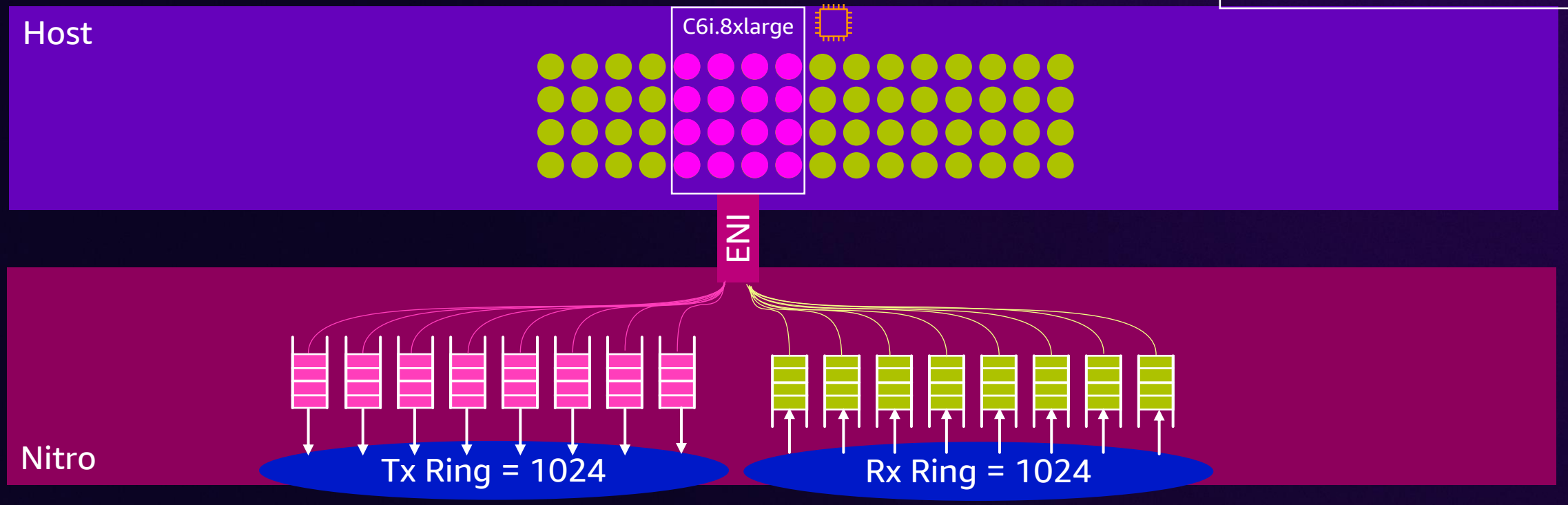
Example: c6i Host (64 core) with c6i.8xlarge (16 cores)



Nitro ENA Rx Rings and Tx Rings



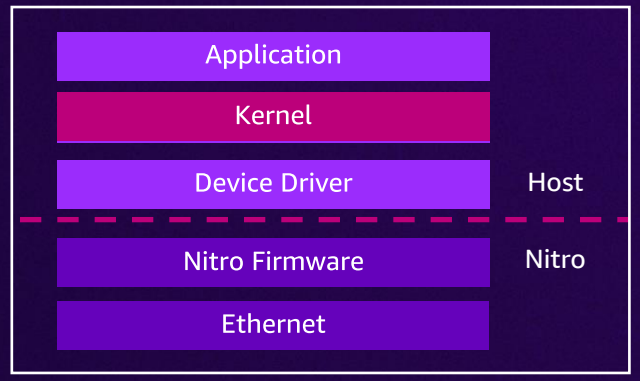
Example: c6i Host (64 core) with c6i.8xlarge (16 cores)



Tx Ring Default = 512 [256 – 1024]

Rx Ring Default = 1024 [256 – 16,384]

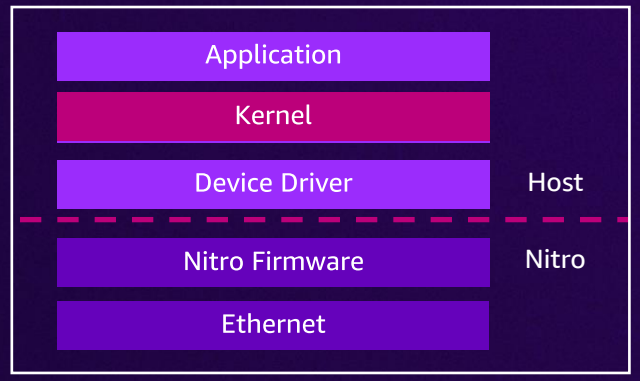
Traffic Control – TX, c7i.large



Without Traffic Control

```
[ec2-user@ip-10-200-1-235 ~]iperf3 -c App_Server -t 30 -P 10
```

Traffic Control – TX, c7i.large

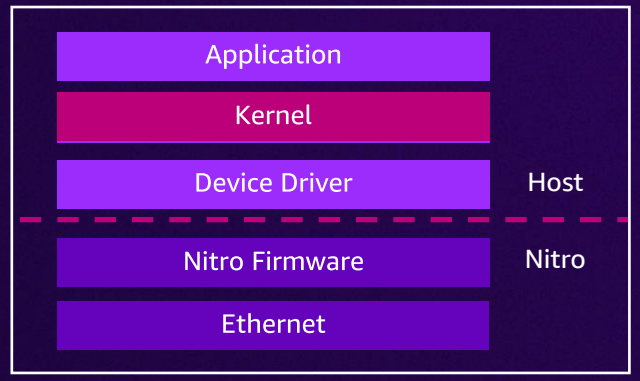


Without Traffic Control

```
[ec2-user@ip-10-200-1-235 ~]iperf3 -c App_Server -t 30 -P 10
```

-t – duration
-P = streams

Traffic Control – TX, c7i.large



Without Traffic Control

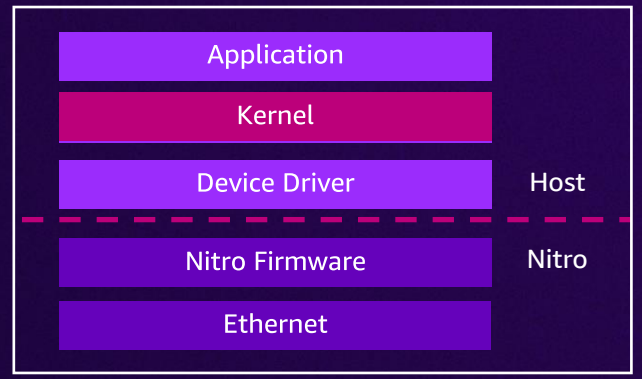
```
[ec2-user@ip-10-200-1-235 ~]iperf3 -c App_Server -t 30 -P 10
```

```
[ ID] Interval Transfer Bitrate
[ 4] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[ 6] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[ 8] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[ 10] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[ 12] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[ 14] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[ 16] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[ 18] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[ 20] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[ 22] 0.00-30.00 sec 32.5 GBytes 9.31 Gbits/sec
[SUM] 0.00-30.00 sec 325 GBytes 93.1 Gbits/sec
```

-t – duration
-P = streams

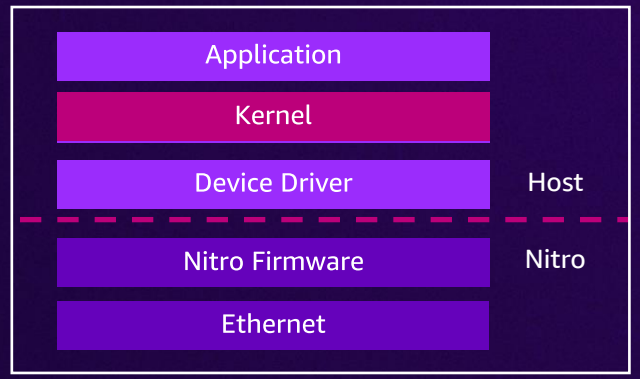
Flow Flooding

Traffic Control – TX



Enable Traffic Control, class of traffic, and enable fair queuing

Traffic Control – TX

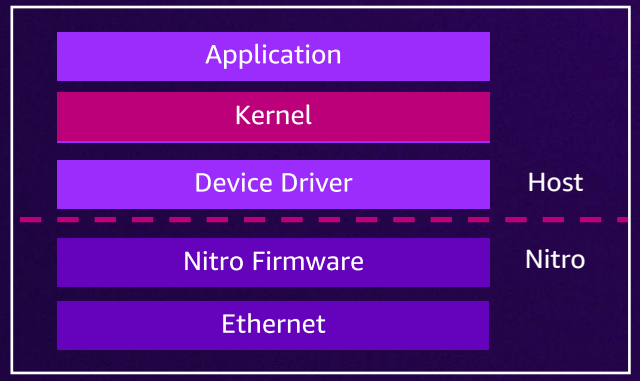


Enable Traffic Control, class of traffic, and enable fair queuing

Token bucket, 10 classes of traffic

```
tc qdisc add dev eth0 root handle 1: htb default 10

tc class add dev eth0 parent 1: classid 1:10 htb rate 10gbit ceil 10gbit

tc qdisc add dev eth0 parent 1:10 handle 20: fq_codel flows 1024 limit 10240 quantum 8000 maxrate 1gbit
```

Traffic Control – TX



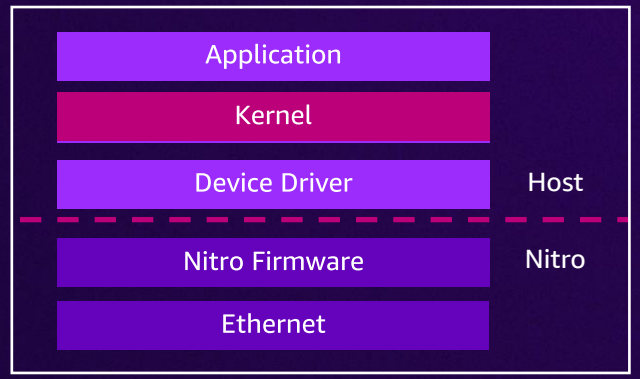
Enable Traffic Control, class of traffic, and enable fair queuing

Token bucket, 10 classes of traffic

```
tc qdisc add dev eth0 root handle 1: htb default 10
tc class add dev eth0 parent 1: classid 1:10 htb rate 10gbit ceil 10gbit
tc qdisc add dev eth0 parent 1:10 handle 20: fq_codel flows 1024 limit 10240 quantum 8000 maxrate 1gbit
```

max throughput 10Gps

Traffic Control – TX



Enable Traffic Control, class of traffic, and enable fair queuing

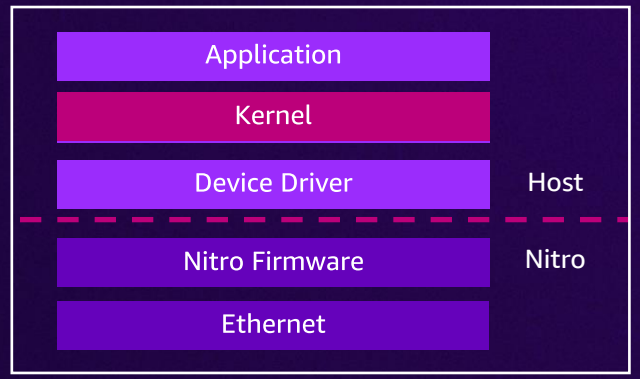
Token bucket, 10 classes of traffic

```
tc qdisc add dev eth0 root handle 1: htb default 10
tc class add dev eth0 parent 1: classid 1:10 htb rate 10gbit ceil 10gbit
tc qdisc add dev eth0 parent 1:10 handle 20: fq_codel flows 1024 limit 10240 quantum 8000 maxrate 1gbit
```

max throughput 10Gps

Max 1Gbps per flow

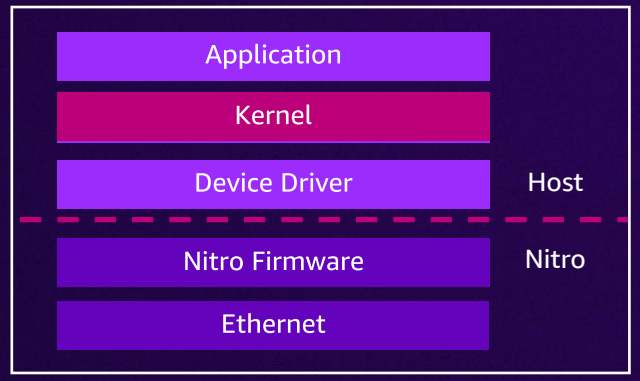
Traffic Control – TX



With Traffic Control

```
[ec2-user@ip-10-200-1-235 ~] iperf3 -c App_Server -t 30 -P 10
```

Traffic Control – TX

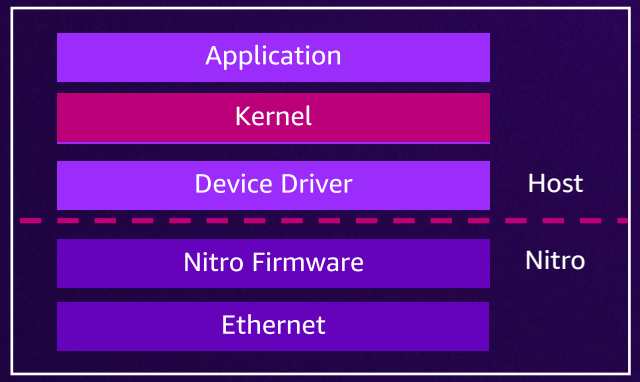


With Traffic Control

```
[ec2-user@ip-10-200-1-235 ~] iperf3 -c App_Server -t 30 -P 10
```

-t – duration
-P = streams

Traffic Control – TX



With Traffic Control

```
[ec2-user@ip-10-200-1-235 ~] iperf3 -c App_Server -t 30 -P 10
```

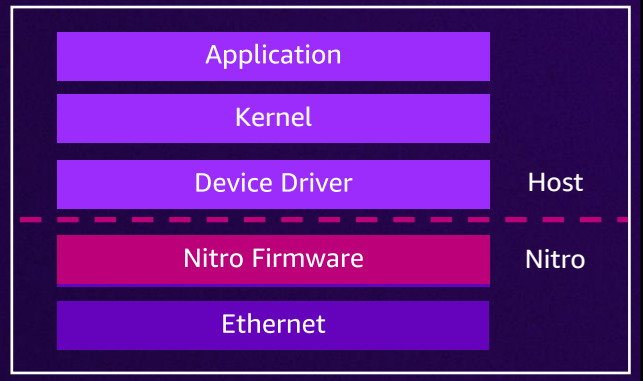
```
[ ID] Interval Transfer Bitrate
[ 4] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[ 6] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[ 8] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[ 10] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[ 12] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[ 14] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[ 16] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[ 18] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[ 20] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[ 22] 0.00-30.00 sec 3.37 GBytes 964 Mbits/sec
[SUM] 0.00-30.00 sec 33.7 GBytes 9.64 Gbits/sec
```

-t – duration
-P = streams

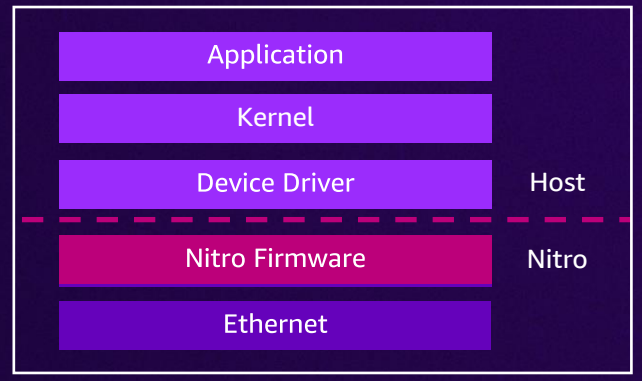
Flow Traffic Control

ENA Express 101

- Enables up to 25Gbps Flow Specification

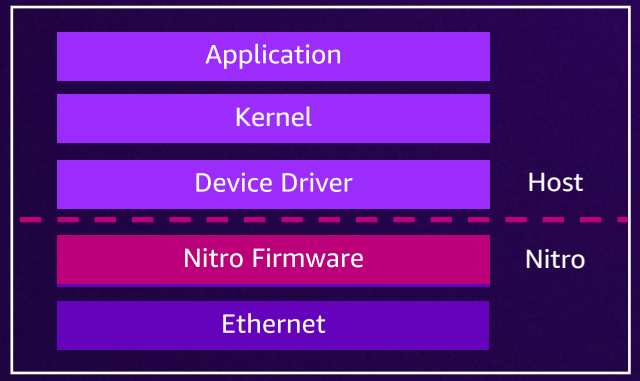


ENA Express 101



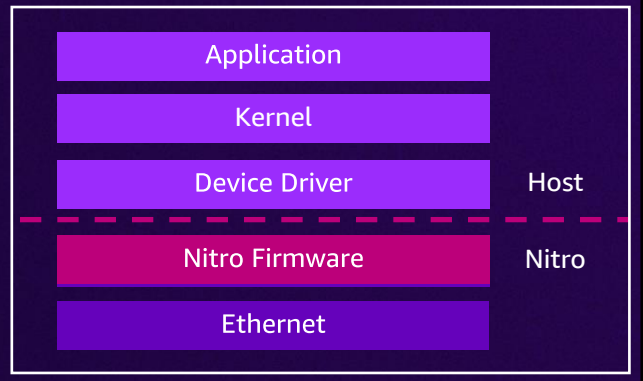
- Enables up to 25Gbps Flow Specification
- Built in Congestion Control
  - Reacts well to “incasts” or ingress microbursts
  - Improves P99+ tail latency

ENA Express 101



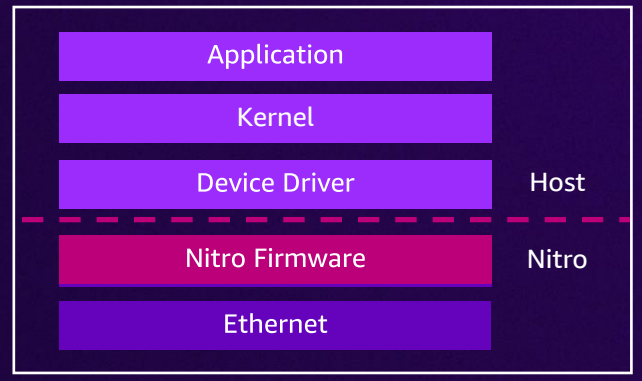
- Enables up to 25Gbps Flow Specification
- Built in Congestion Control
  - Reacts well to “incasts” or ingress microbursts
  - Improves P99+ tail latency
- Supported on NitroV4+ Instances

ENA Express 101



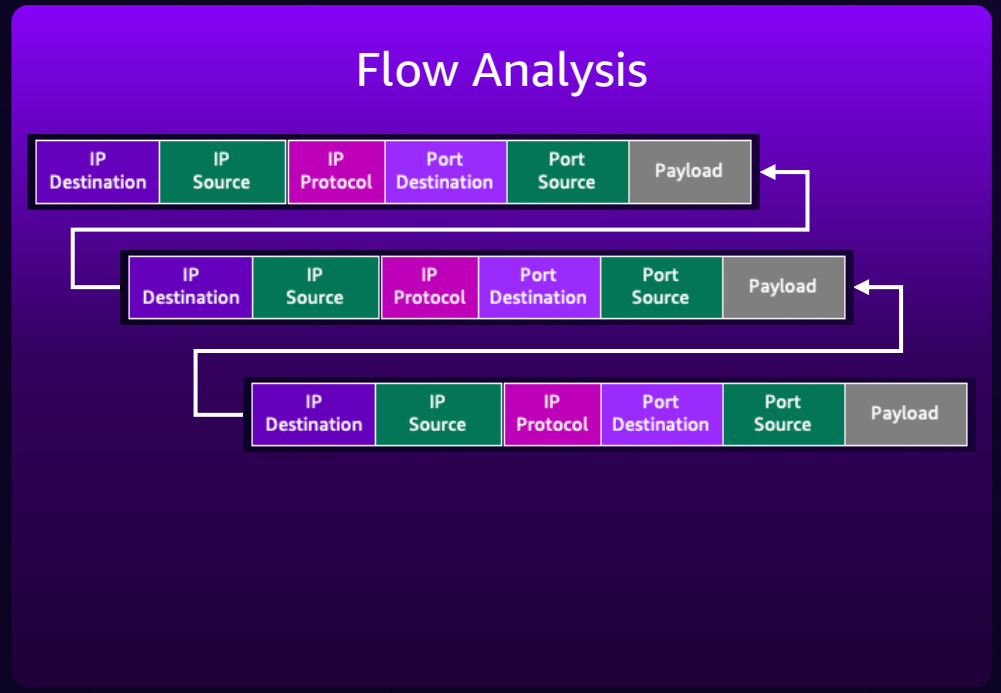
- Enables up to 25Gbps Flow Specification
- Built in Congestion Control
  - Reacts well to “incasts” or ingress microbursts
  - Improves P99+ tail latency
- Supported on NitroV4+ Instances
- Customer Opt-in

ENA Express 101



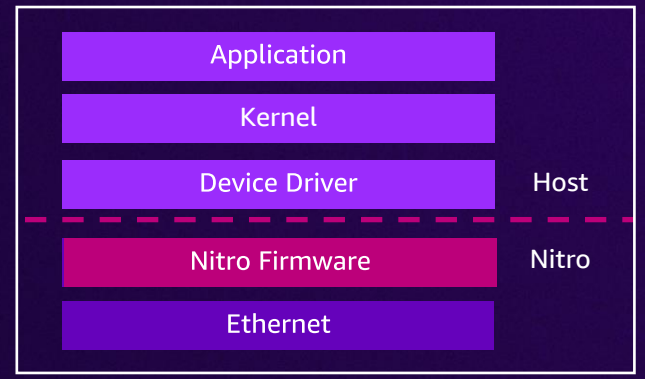
- Enables up to 25Gbps Flow Specification
- Built in Congestion Control
  - Reacts well to “incasts” or ingress microbursts
  - Improves P99+ tail latency
- Supported on NitroV4+ Instances
- Customer Opt-in
- AZ local

Flow Analysis

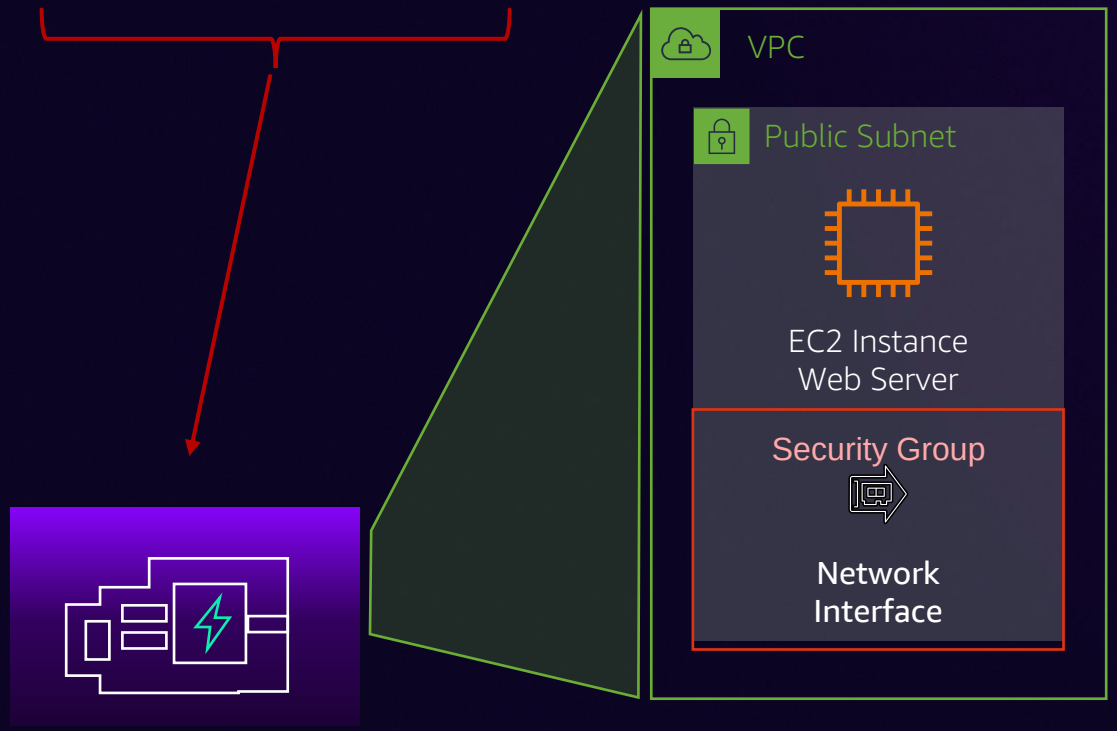


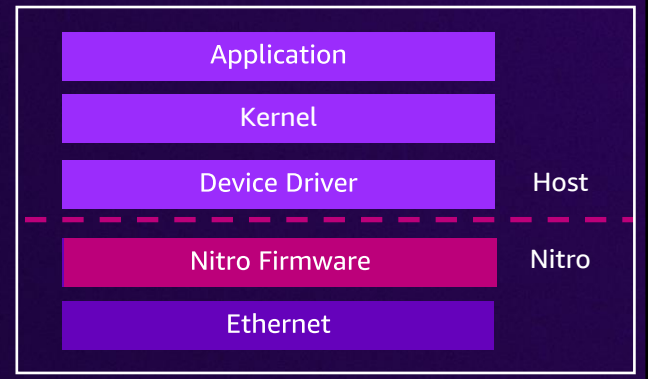
- Flow Limits
- Micro bursting
- Flow State Anomalies
- “Whale” Flows

VPC Mutations

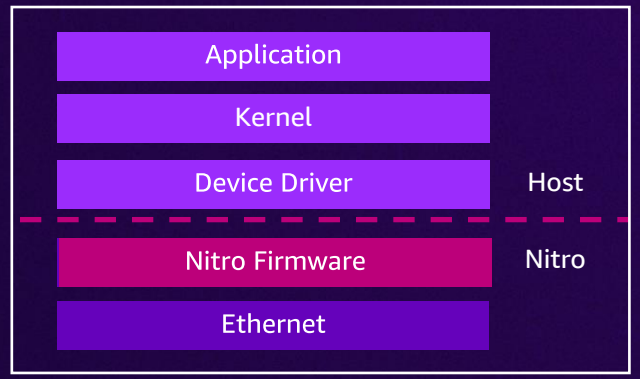


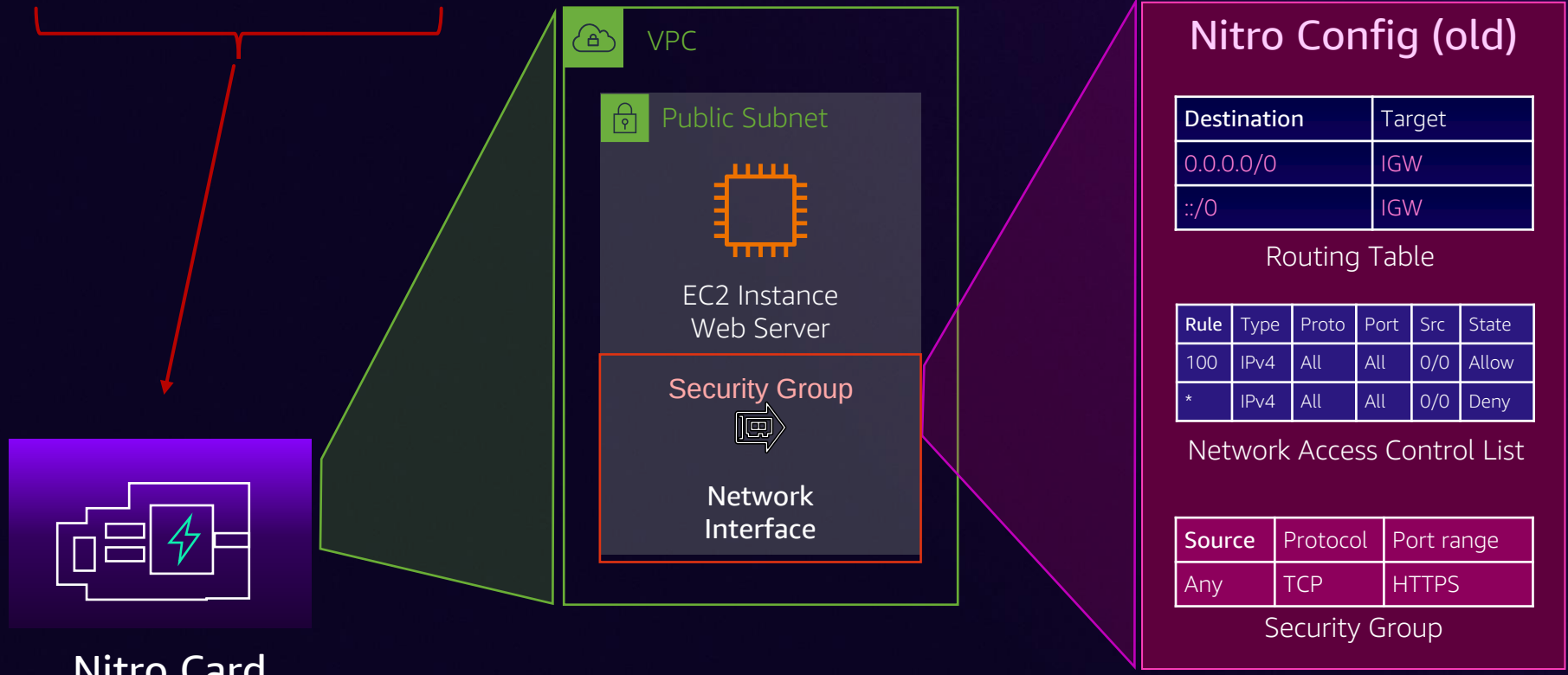
VPC Mutations





VPC Mutations





Nitro Config (old)

Destination	Target
0.0.0.0/0	IGW
::/0	IGW

Routing Table

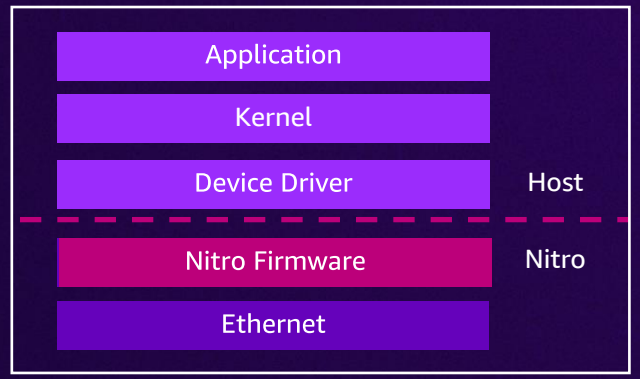
Rule	Type	Proto	Port	Src	State
100	IPv4	All	All	0/0	Allow
*	IPv4	All	All	0/0	Deny

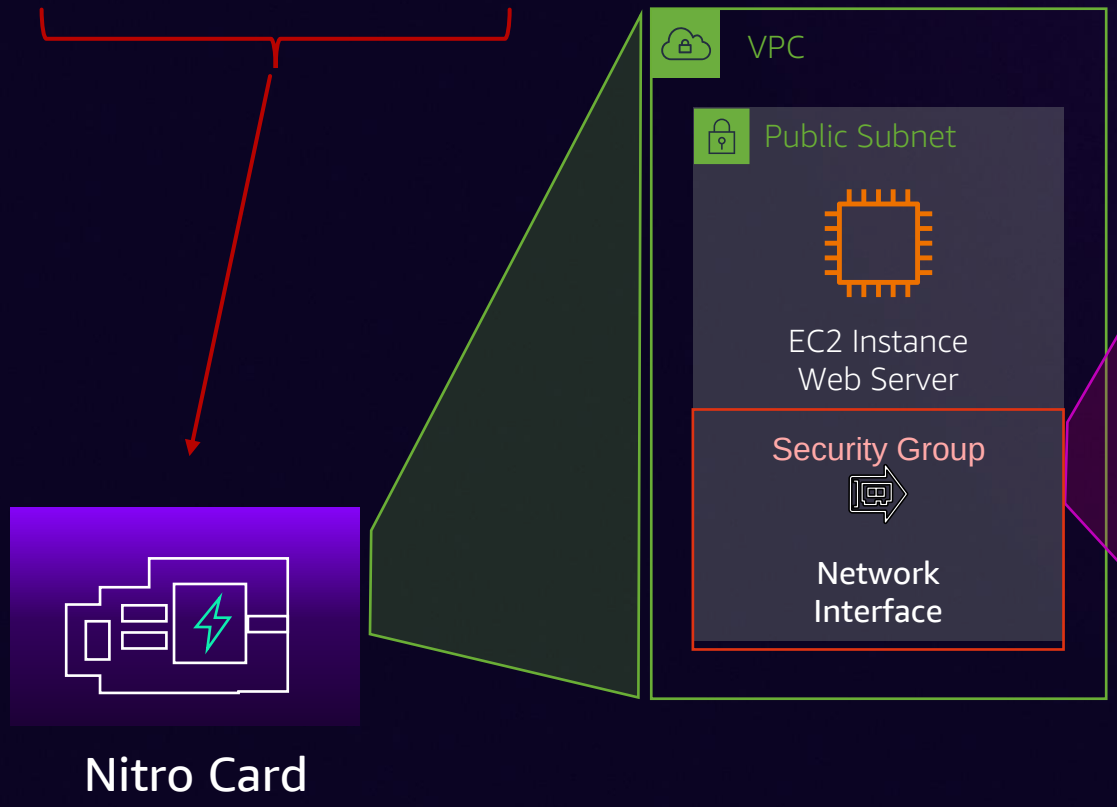
Network Access Control List

Source	Protocol	Port range
Any	TCP	HTTPS

Security Group

VPC Mutations





Nitro Config (old)

Destination	Target
0.0.0.0/0	IGW
::/0	IGW

Routing Table

Rule	Type	Proto	Port	Src	State
100	IPv4	All	All	0/0	Allow
*	IPv4	All	All	0/0	Deny

Network Access Control List

Source	Protocol	Port range
Any	TCP	HTTPS

Security Group

Nitro Config (new)

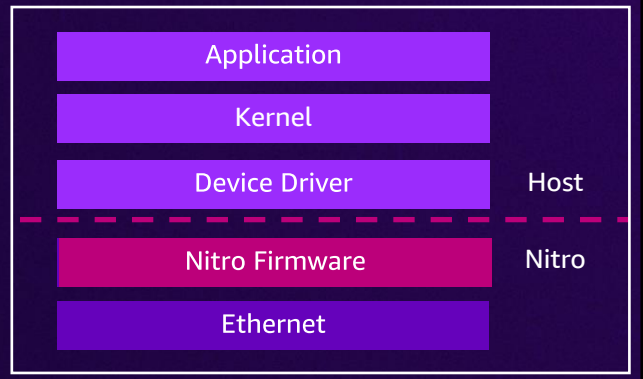
Destination	Target
0.0.0.0/0	IGW
::/0	IGW

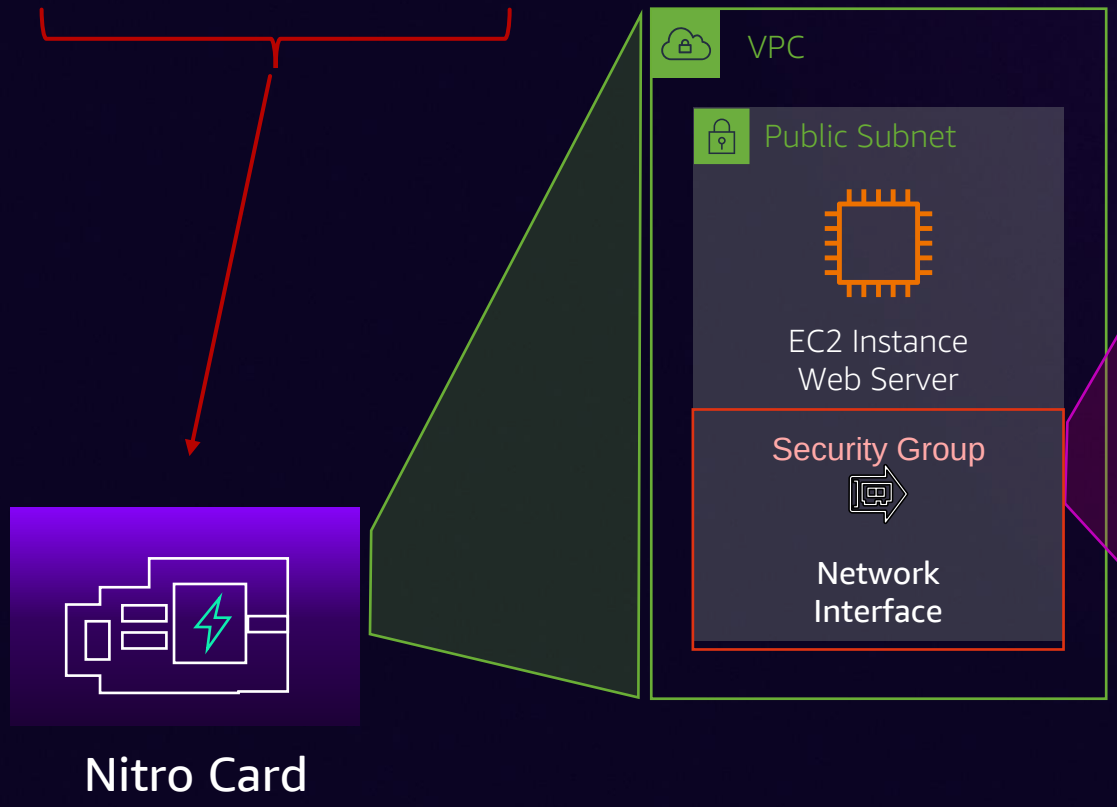
Routing Table

Rule	Type	Proto	Port	Src	State
100	IPv4	TCP	443	0/0	Allow
*	IPv4	All	All	0/0	Deny

Network Access Control List

VPC Mutations





Nitro Config (old)

Destination	Target
0.0.0.0/0	IGW
::/0	IGW

Routing Table

Rule	Type	Proto	Port	Src	State
100	IPv4	All	All	0/0	Allow
*	IPv4	All	All	0/0	Deny

Network Access Control List

Source	Protocol	Port range
Any	TCP	HTTPS

Security Group

Nitro Config (new)

Destination	Target
0.0.0.0/0	IGW
::/0	IGW

Routing Table

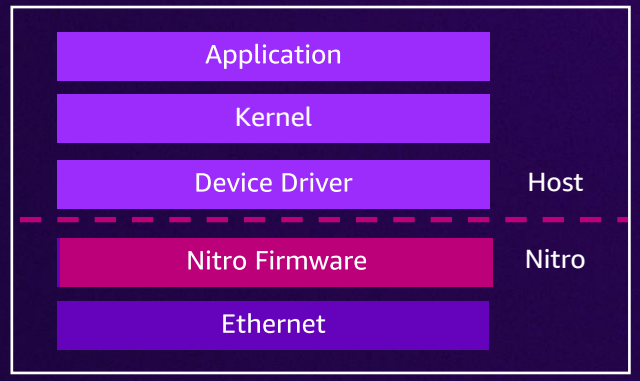
Rule	Type	Proto	Port	Src	State
100	IPv4	TCP	443	0/0	Allow
*	IPv4	All	All	0/0	Deny

Network Access Control List

Source	Protocol	Port range
10.0.0.0/16	TCP	HTTPS

Security Group

Untracked Connections

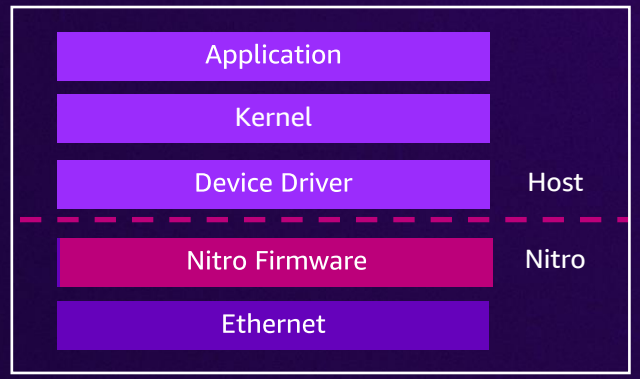


Security Group Configuration

Inbound

Protocol type	Port number	Source
TCP	22 (SSH)	203.0.113.1/32
TCP	80 (HTTP)	0.0.0.0/0
TCP	80 (HTTP)	::/0
ICMP	All	0.0.0.0/0

Untracked Connections



Security Group Configuration

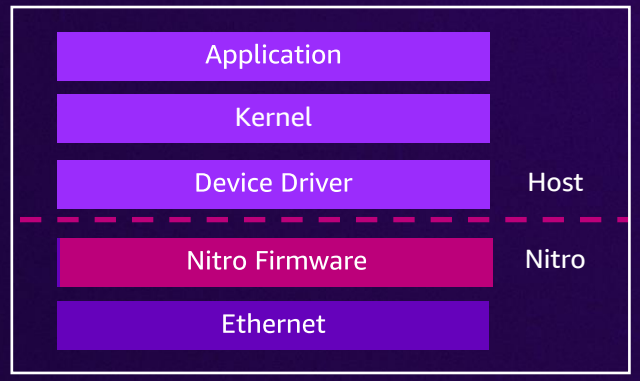
Inbound

Protocol type	Port number	Source
TCP	22 (SSH)	203.0.113.1/32
TCP	80 (HTTP)	0.0.0.0/0
TCP	80 (HTTP)	::/0
ICMP	All	0.0.0.0/0

Outbound

Protocol type	Port number	Destination
All	All	0.0.0.0/0
All	All	::/0

Untracked Connections



Security Group Configuration

Inbound

Protocol type	Port number	Source
TCP	22 (SSH)	203.0.113.1/32
TCP	80 (HTTP)	0.0.0.0/0
TCP	80 (HTTP)	::/0
ICMP	All	0.0.0.0/0

TCP Port 22 – Tracked

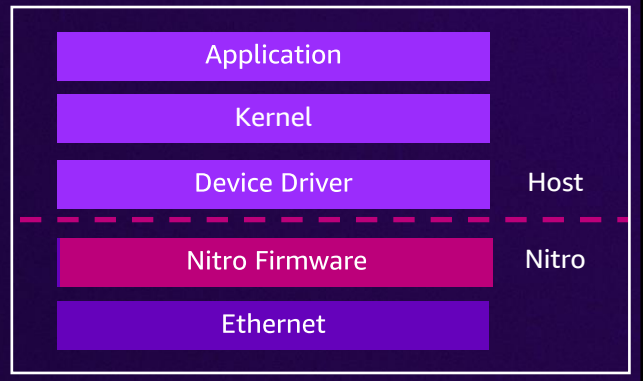
TCP Port 80 – Untracked

ICMP – Always tracked

Outbound

Protocol type	Port number	Destination
All	All	0.0.0.0/0
All	All	::/0

Untracked Connections



Security Group Configuration

Inbound

Protocol type	Port number	Source
TCP	22 (SSH)	203.0.113.1/32
TCP	80 (HTTP)	0.0.0.0/0
TCP	80 (HTTP)	::/0
ICMP	All	0.0.0.0/0

TCP Port 22 – Tracked

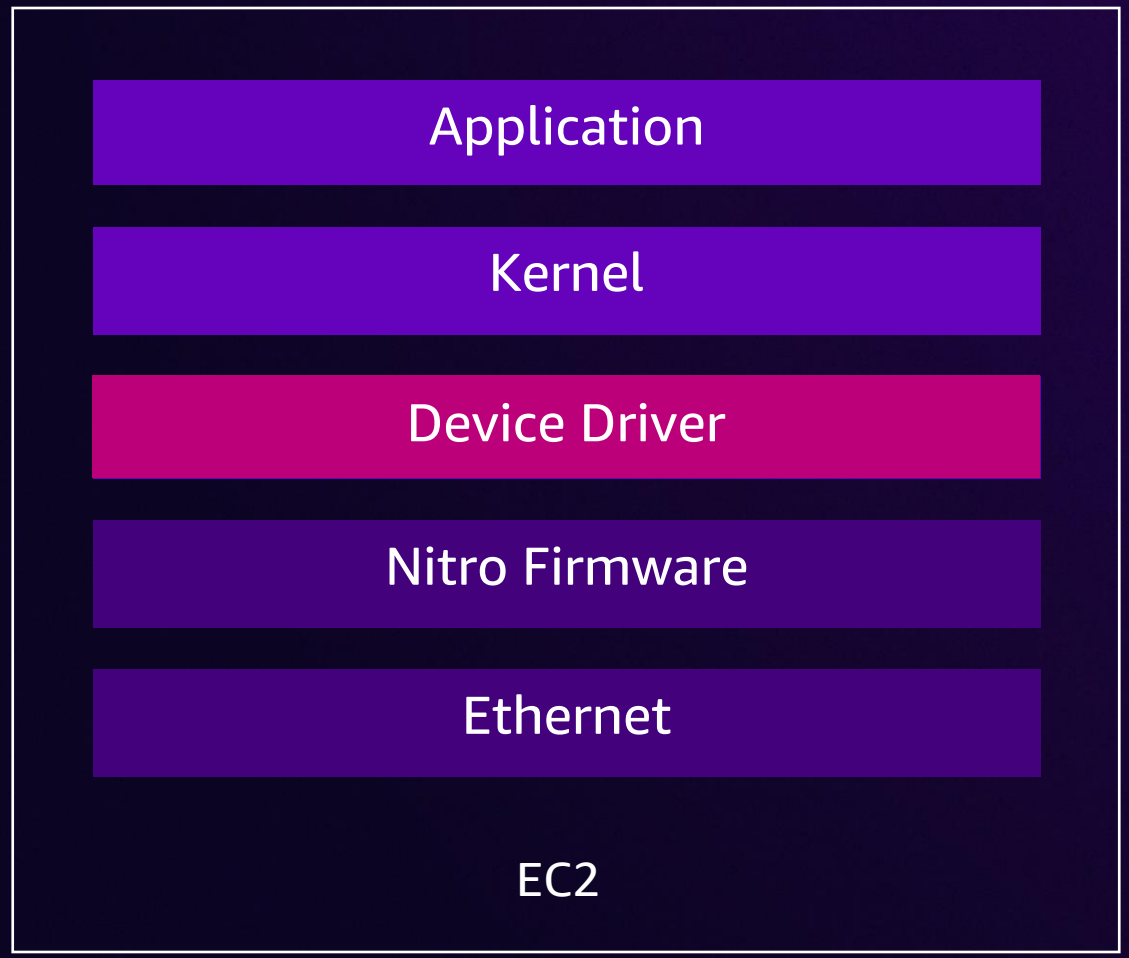
TCP Port 80 – Untracked

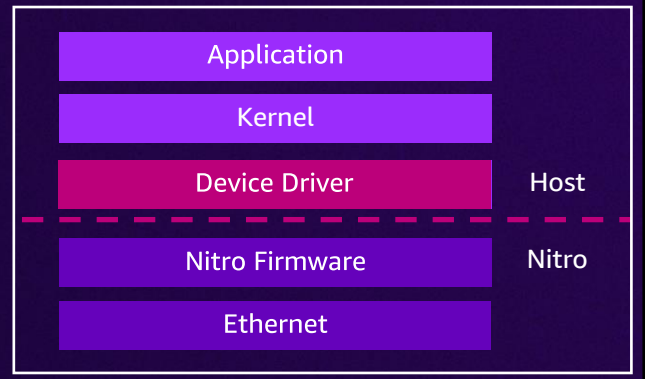
ICMP – Always tracked

Outbound

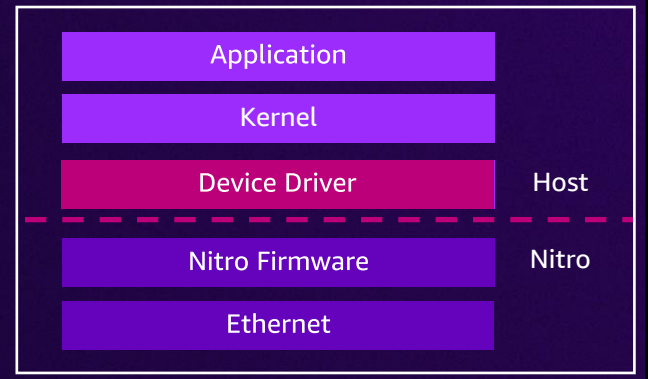
Protocol type	Port number	Destination
All	All	0.0.0.0/0
All	All	::/0

Minimum Driver Requirements



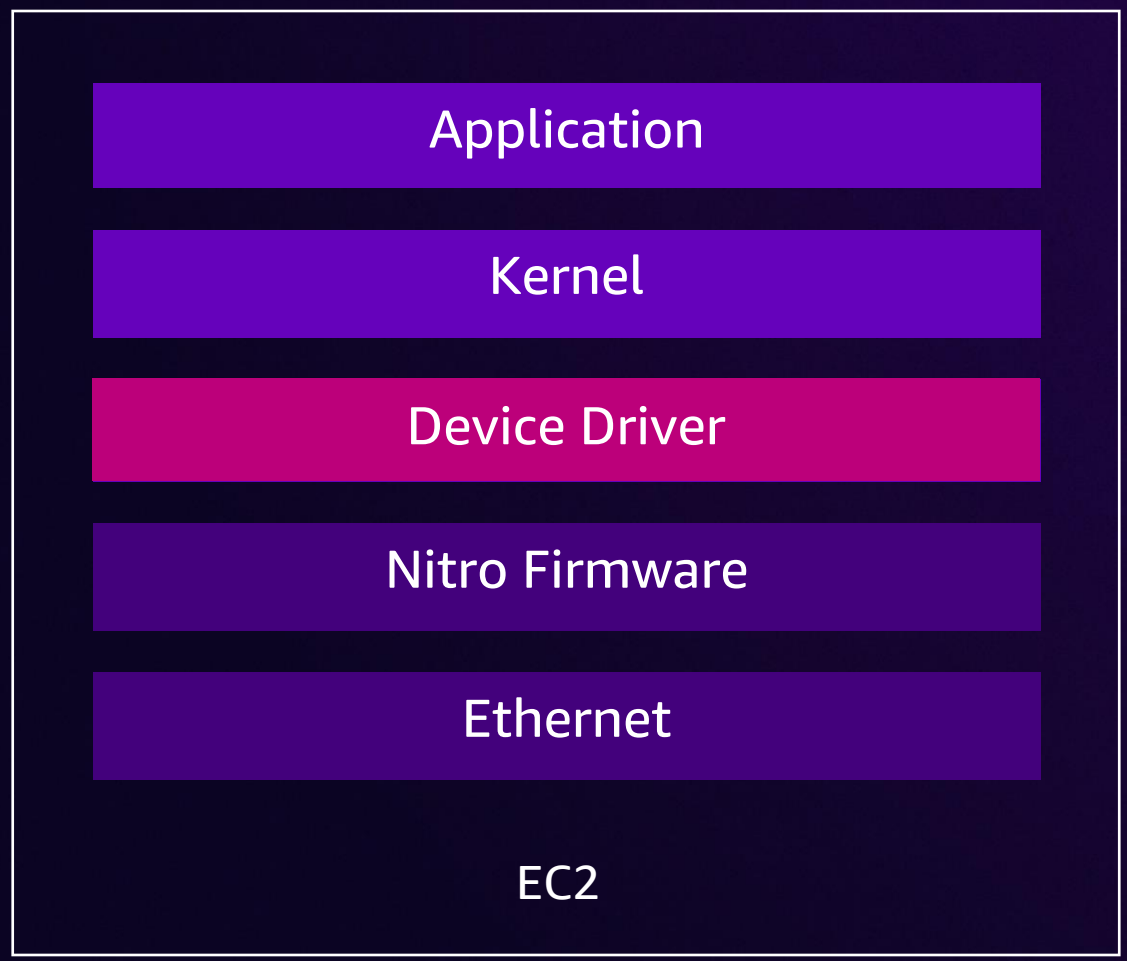


Minimum Driver Requirements

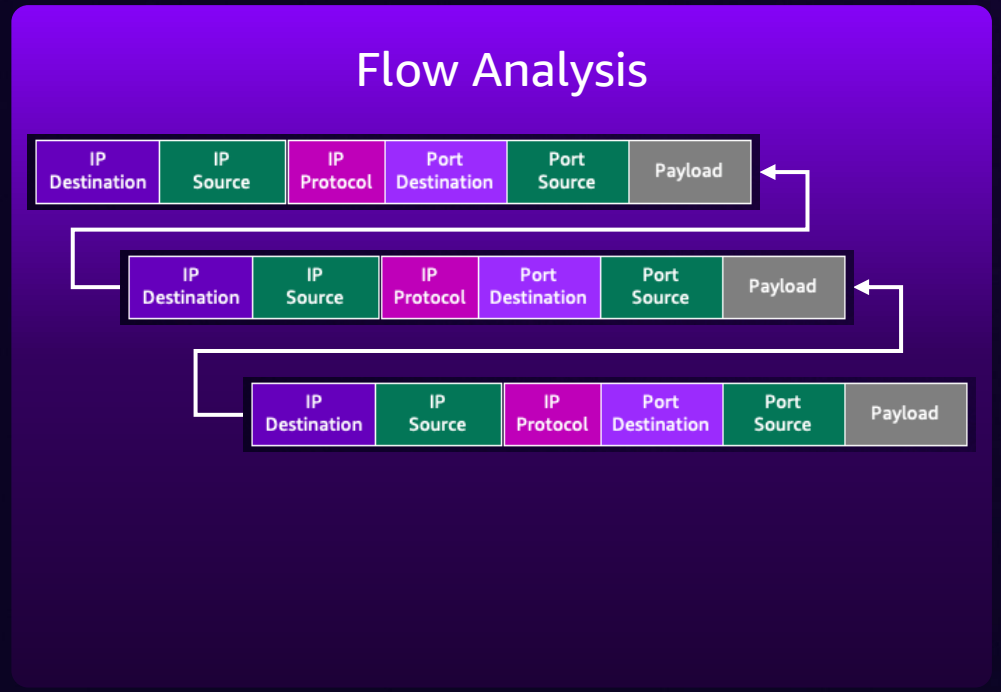


ENA Driver Version 2.2.9

Nitro V4 – Recommend
Nitro V5 - Required

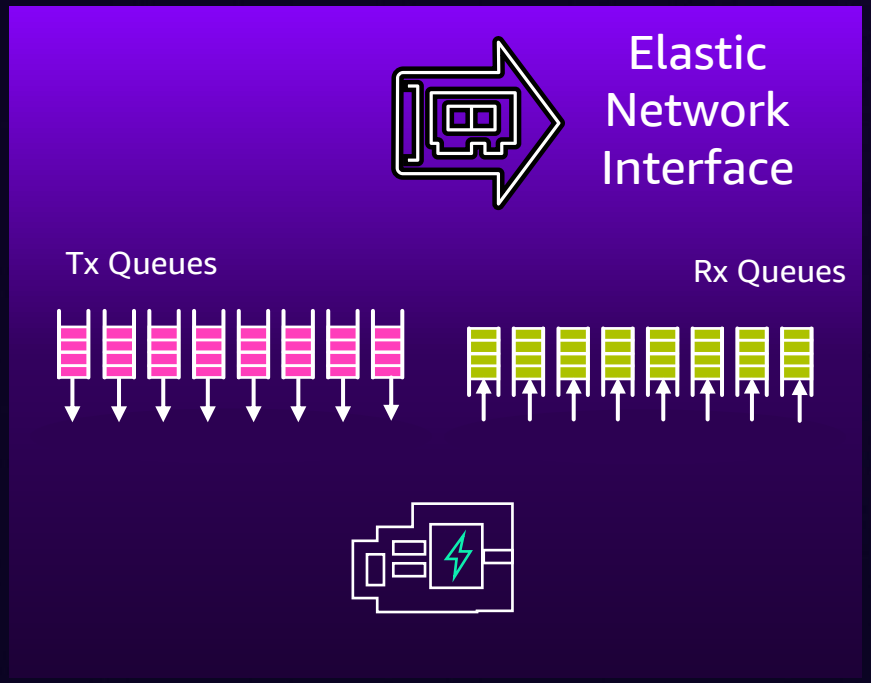


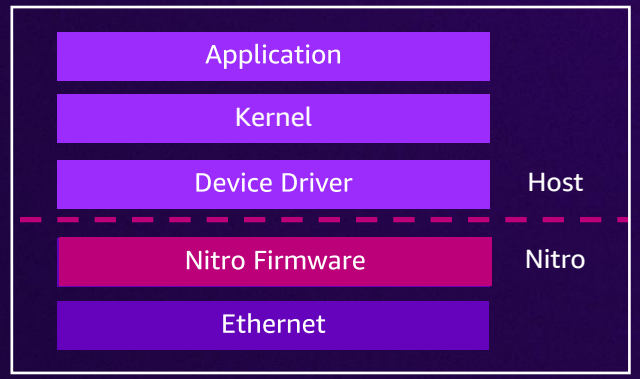
Flow Analysis



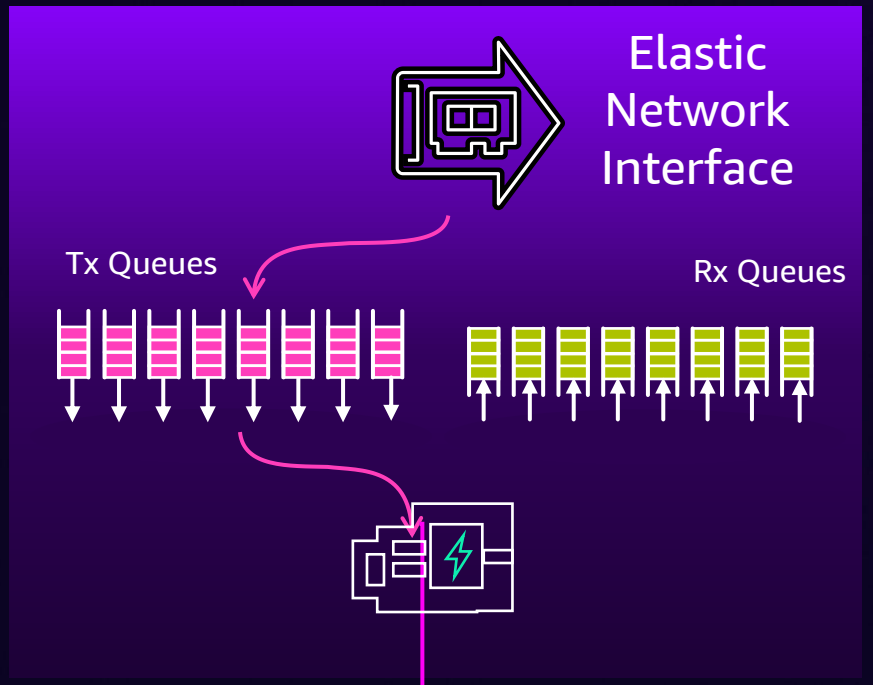
- Flow Limits
- Micro bursting
- Flow State Anomalies
- “Whale” Flows

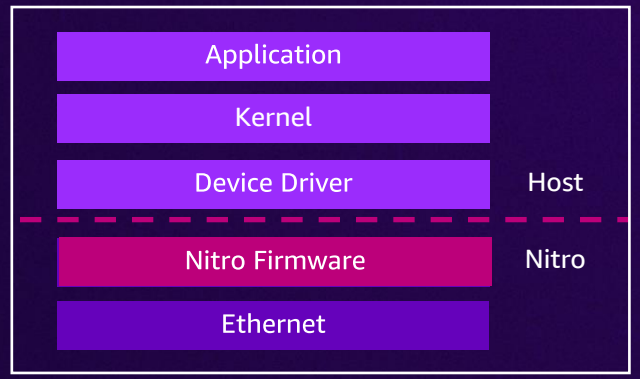
Nitro handing of Tunnel Hashing



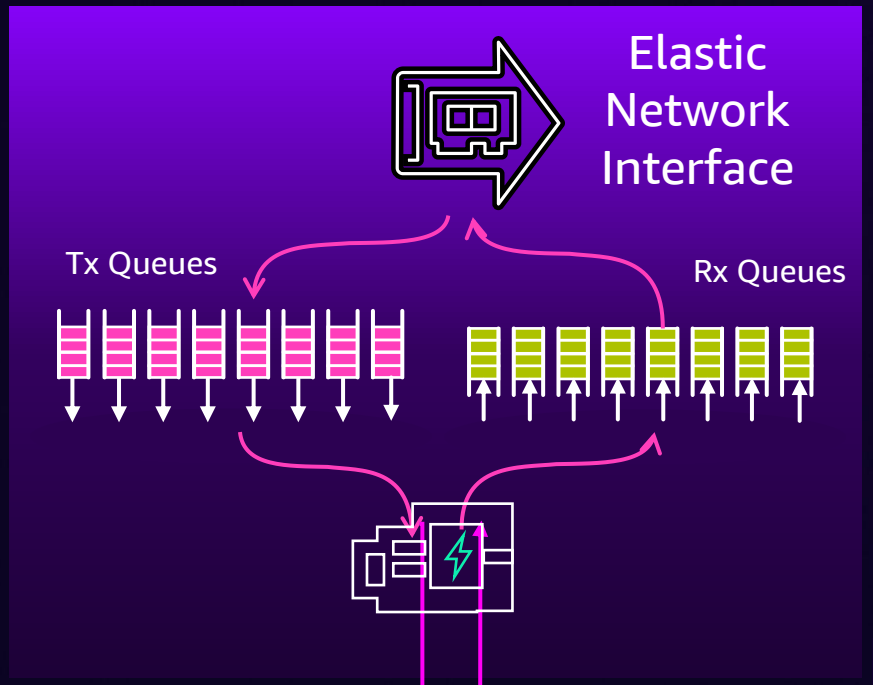


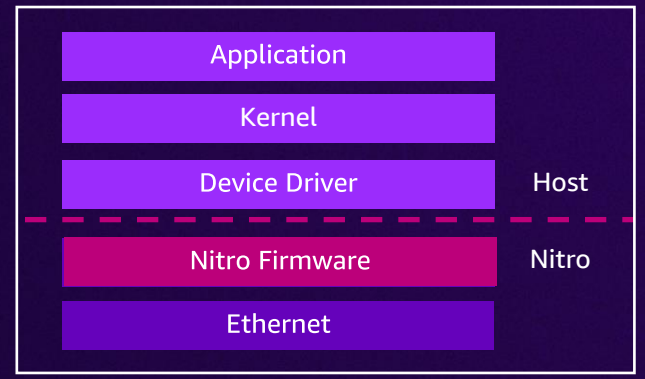
Nitro handing of Tunnel Hashing





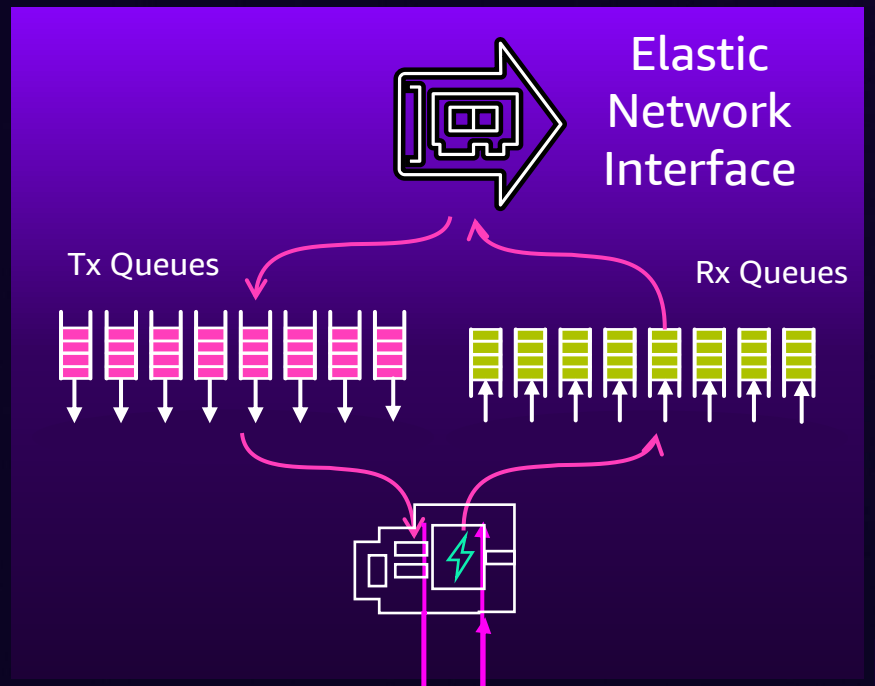
Nitro handing of Tunnel Hashing

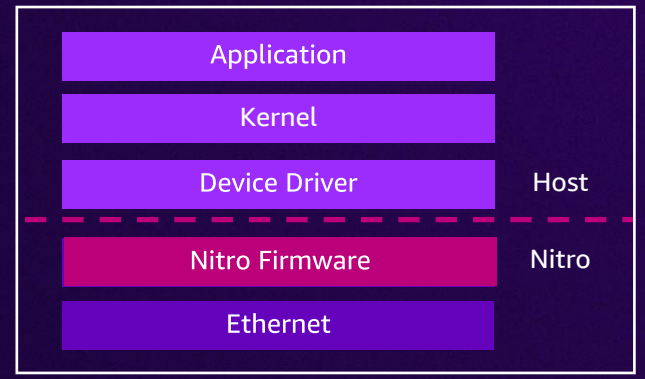




IP Source	IP Destination	IP Tunnel	IP Packet	Payload
IP Source	IP Destination	IP Tunnel	IP Packet	Payload

Nitro handing of Tunnel Hashing

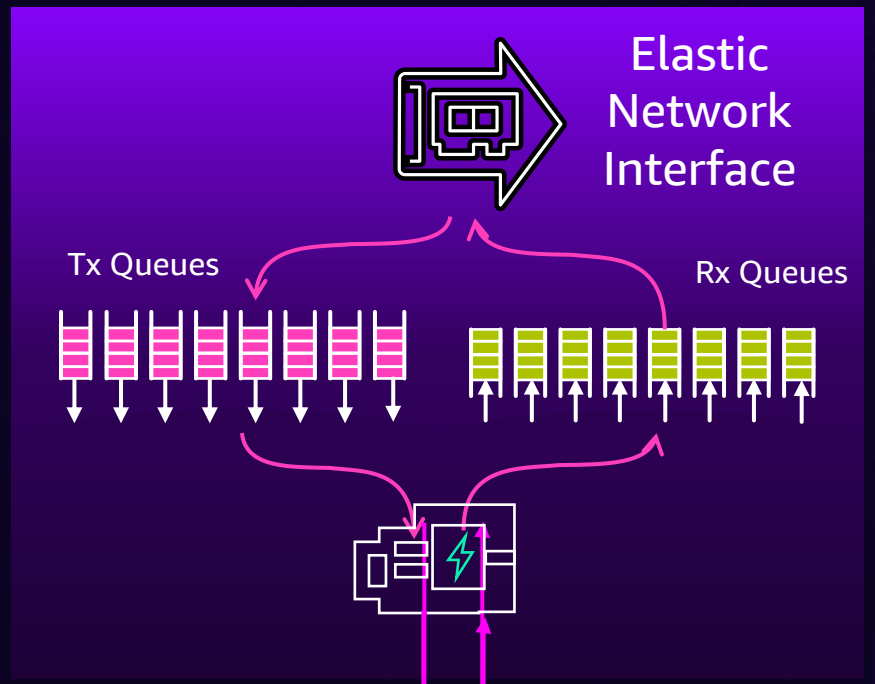


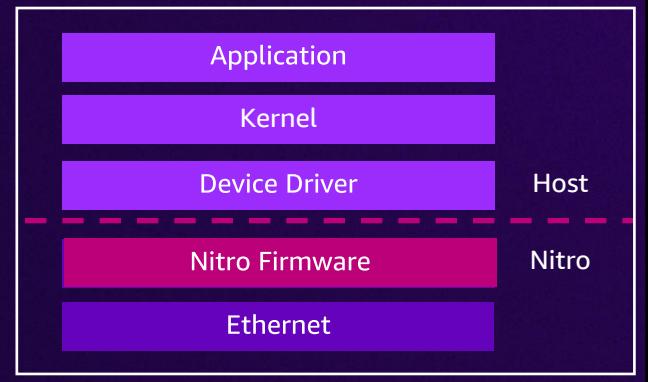


IP Source	IP Destination	IP Tunnel	IP Packet	Payload
IP Source	IP Destination	IP Tunnel	IP Packet	Payload

IP Source	IP Destination	IP Tunnel	IP Packet	Payload
IP Source	IP Destination	IP Tunnel	IP Packet	Payload

Nitro handing of Tunnel Hashing



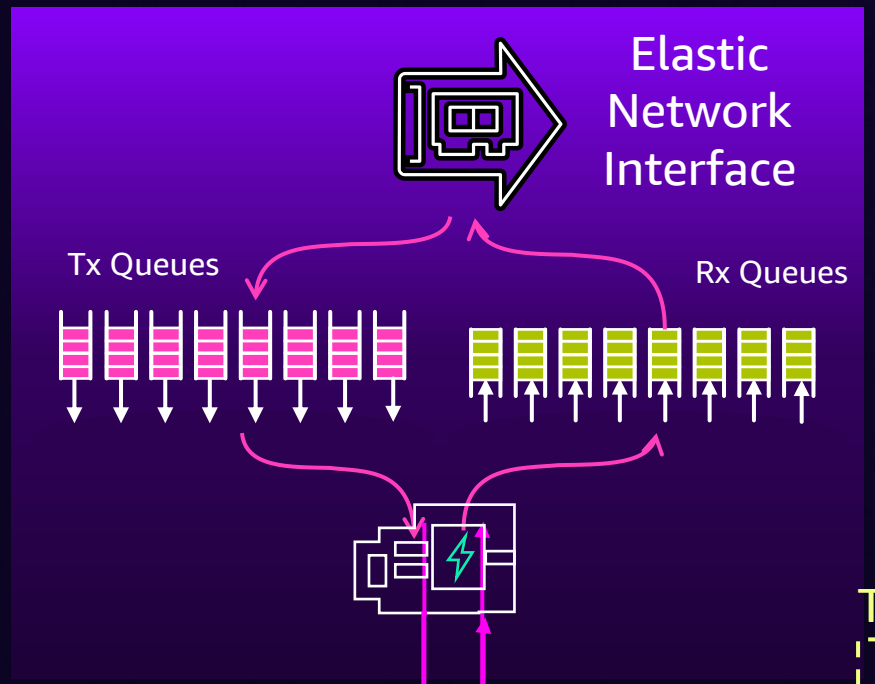


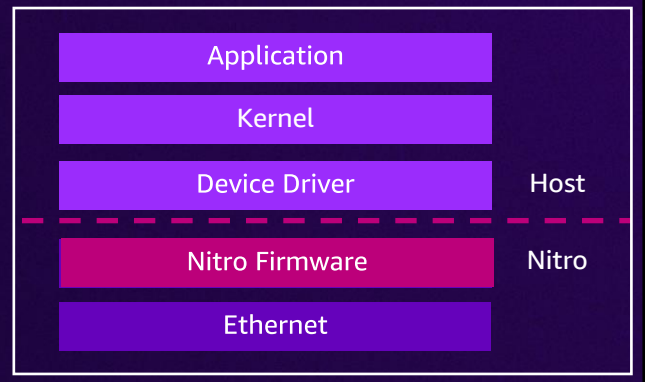
Unique Packet 5-Tuple

IP Source	IP Destination	IP Tunnel	IP Packet	Payload
IP Source	IP Destination	IP Tunnel	IP Packet	Payload

IP Source	IP Destination	IP Tunnel	IP Packet	Payload
IP Source	IP Destination	IP Tunnel	IP Packet	Payload

Nitro handing of Tunnel Hashing





Common Packet 3-Tuple

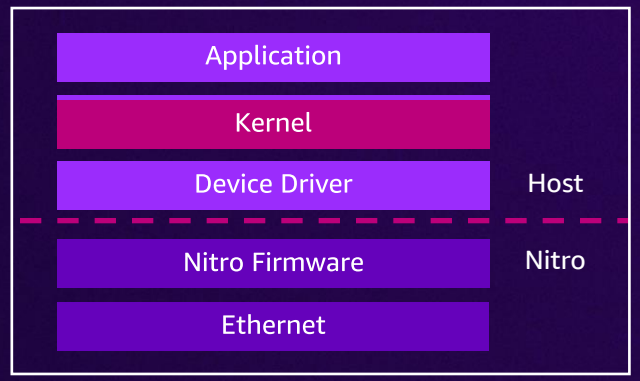
Unique Packet 5-Tuple

Tunnel Flow Evaluation

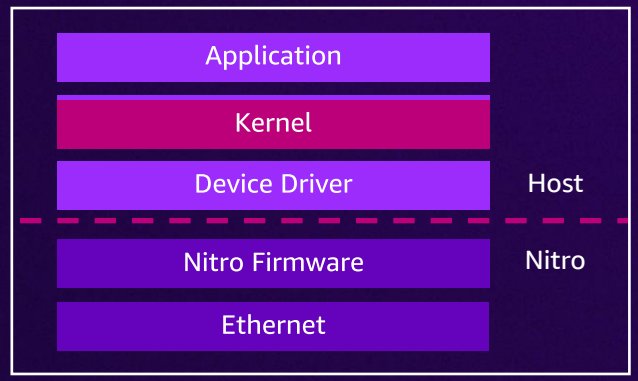
IP Source	IP Destination	IP Tunnel	IP Packet	Payload
IP Source	IP Destination	IP Tunnel	IP Packet	Payload

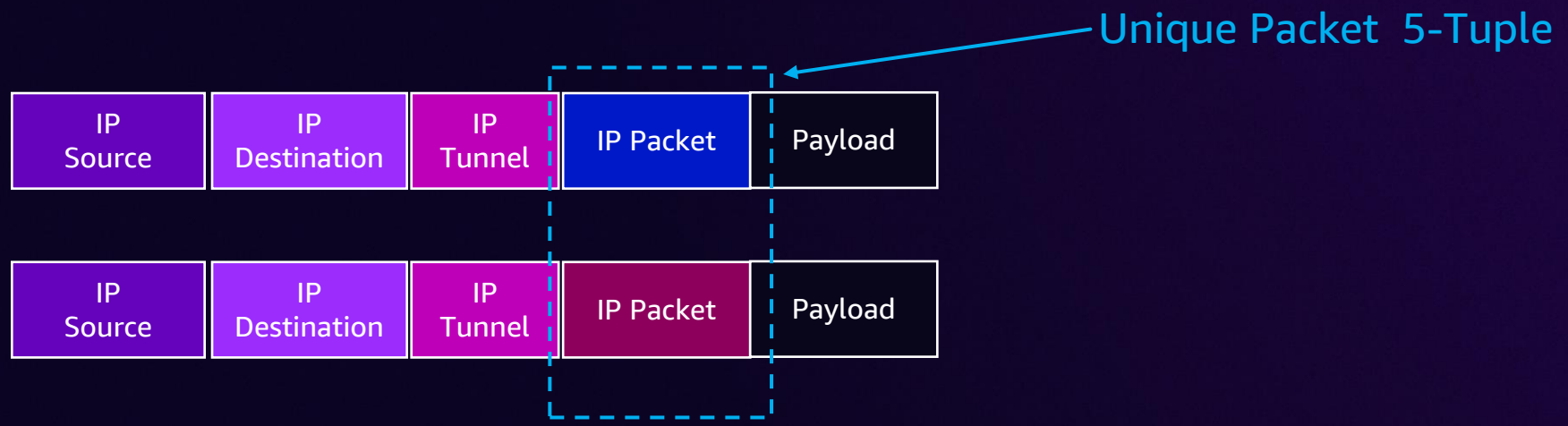
IP Source	IP Destination	IP Tunnel	IP Packet	Payload
IP Source	IP Destination	IP Tunnel	IP Packet	Payload

Optimizing for Large Single Flows

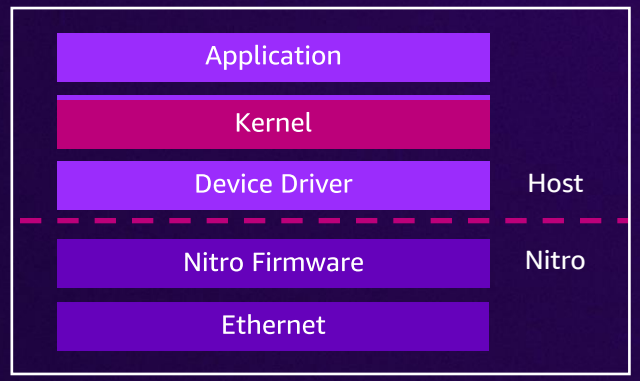


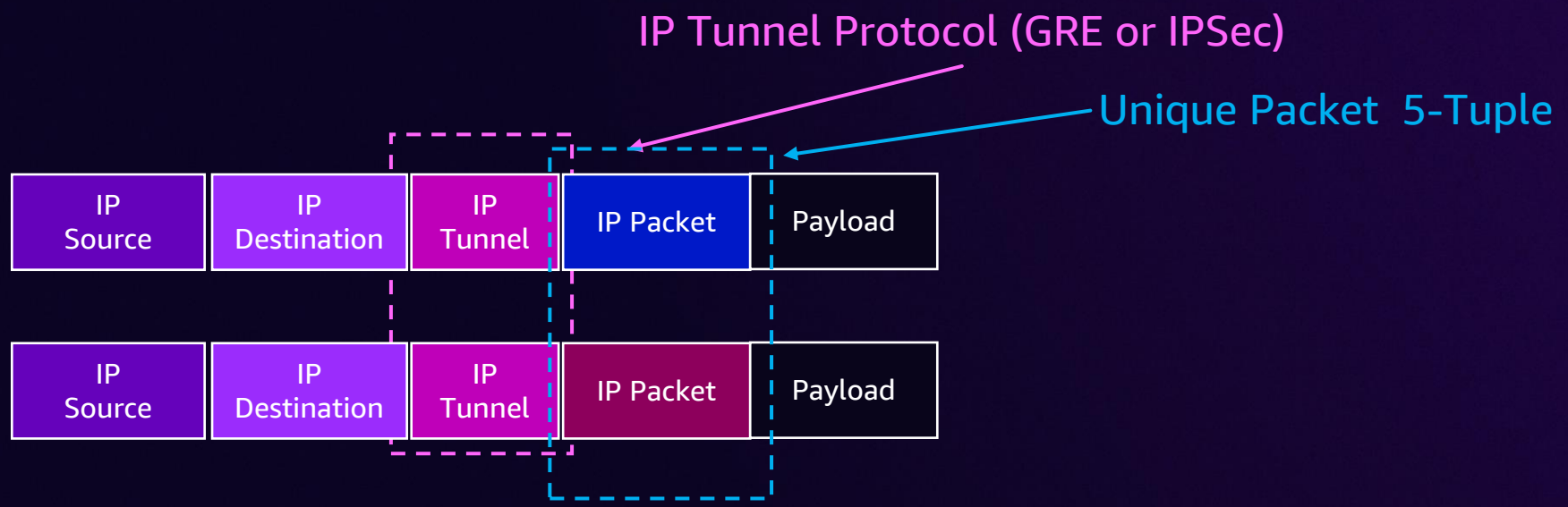
Optimizing for Large Single Flows



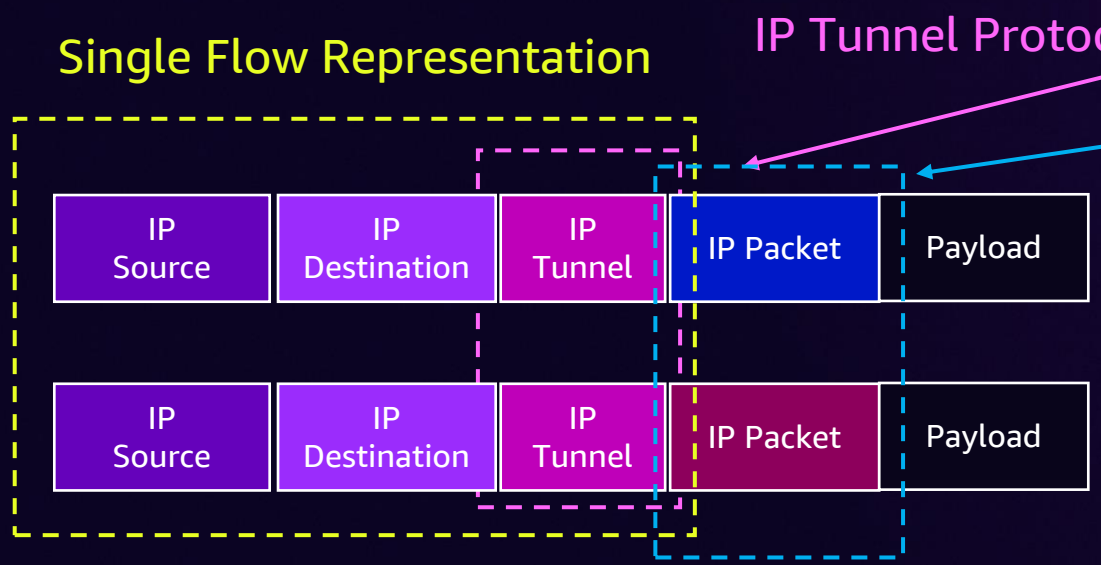


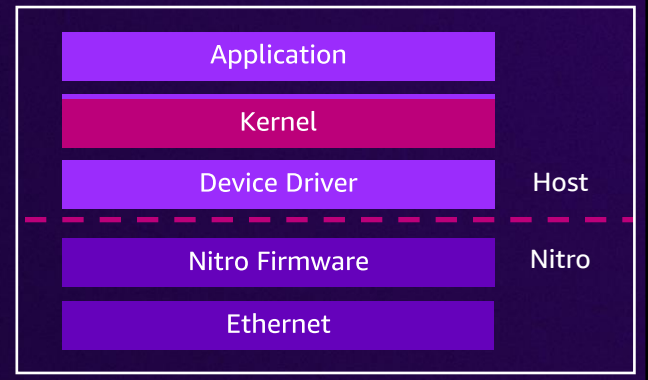
Optimizing for Large Single Flows



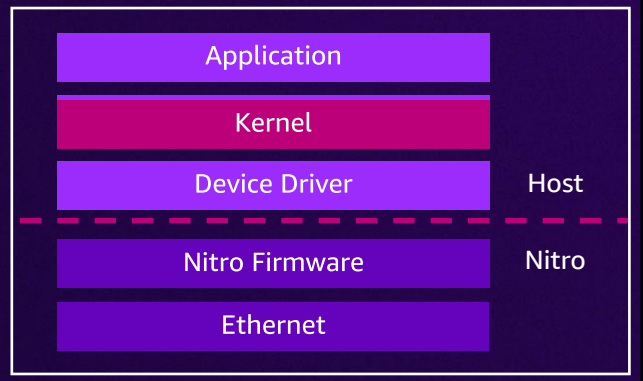


Optimizing for Large Single Flows





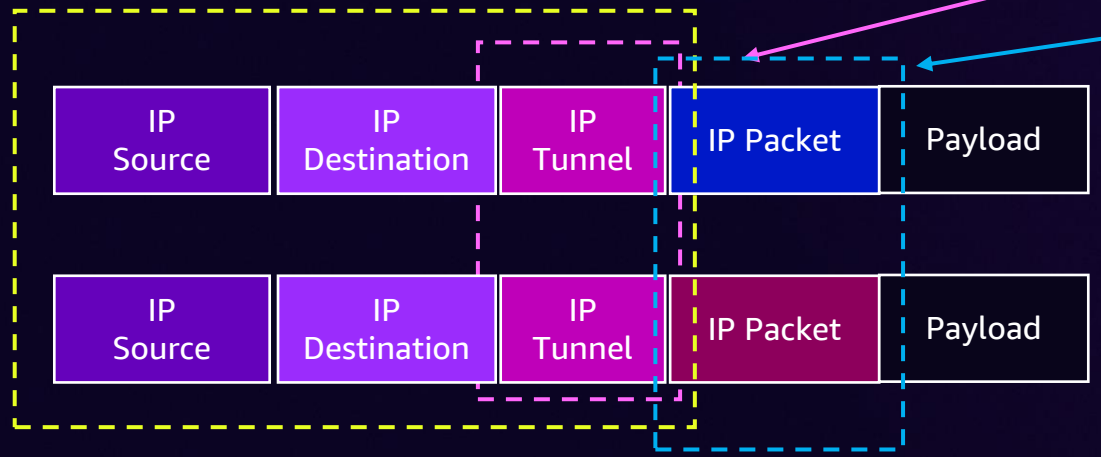
Optimizing for Large Single Flows

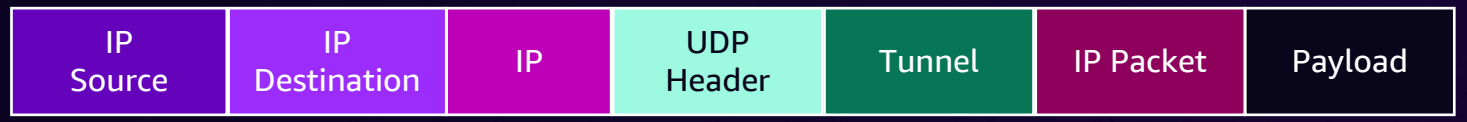


Single Flow Representation

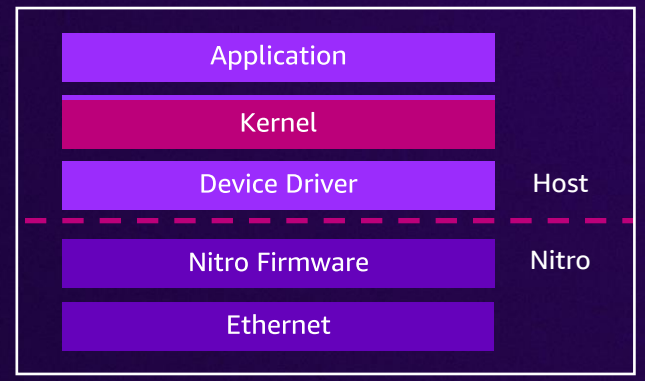
IP Tunnel Protocol (GRE or IPSec)

Unique Packet 5-Tuple





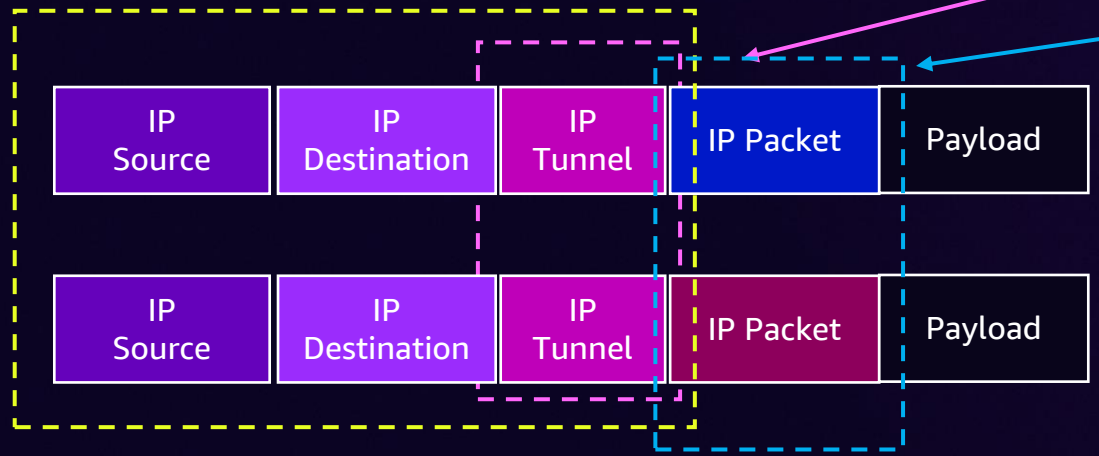
Optimizing for Large Single Flows



Single Flow Representation

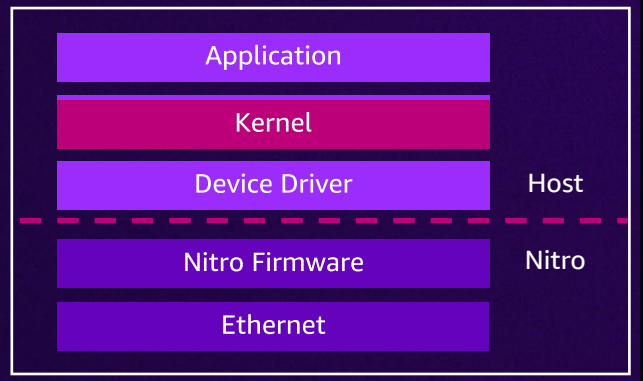
IP Tunnel Protocol (GRE or IPSec)

Unique Packet 5-Tuple



Unique Packet 5-Tuple

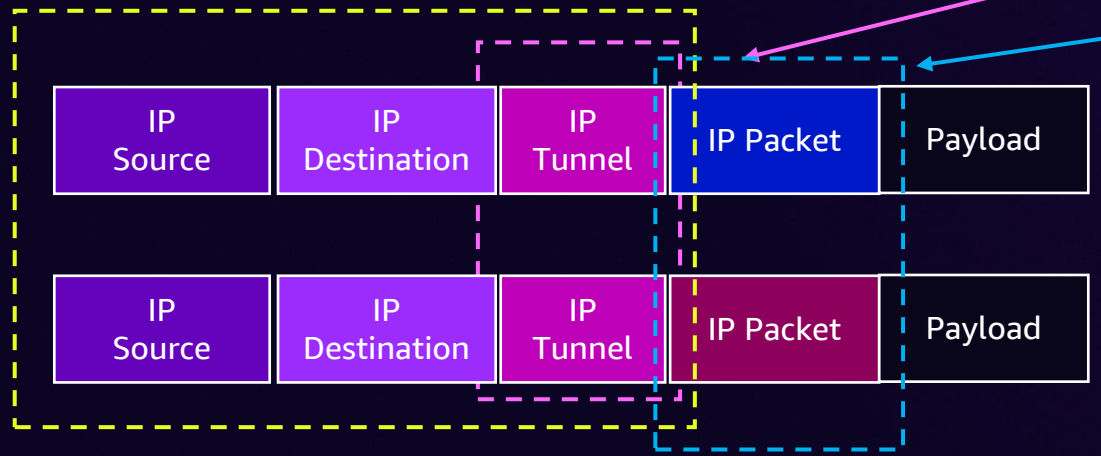
Optimizing for Large Single Flows



Single Flow Representation

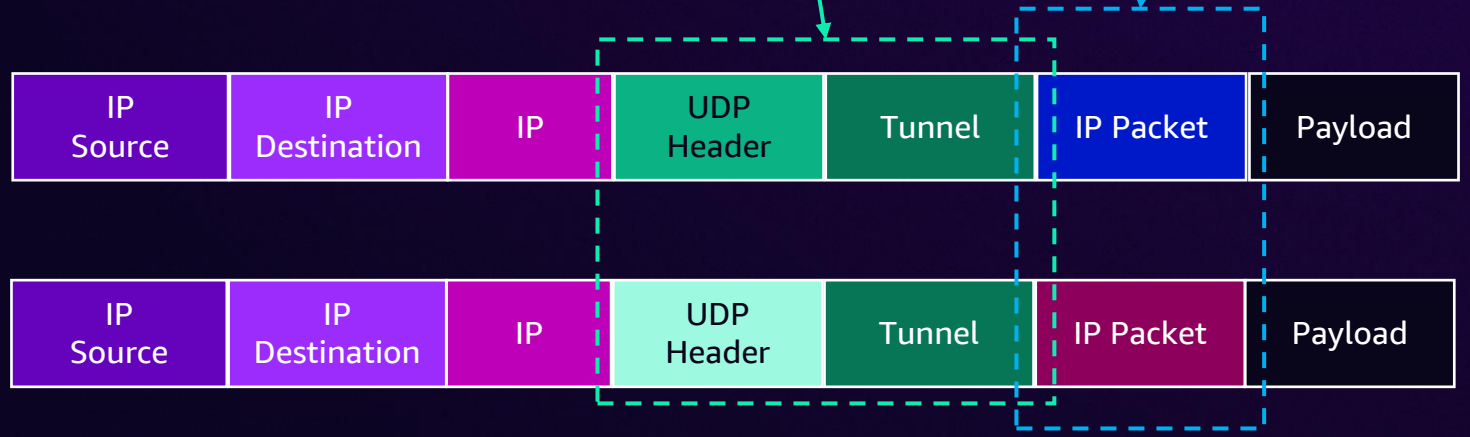
IP Tunnel Protocol (GRE or IPSec)

Unique Packet 5-Tuple

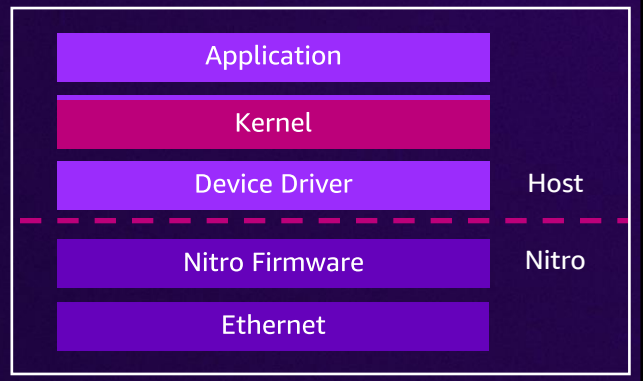


Unique Packet 5-Tuple

IP Tunnel Protocol
(VxLAN or GENEVE or GTP)
with Unique Source Port



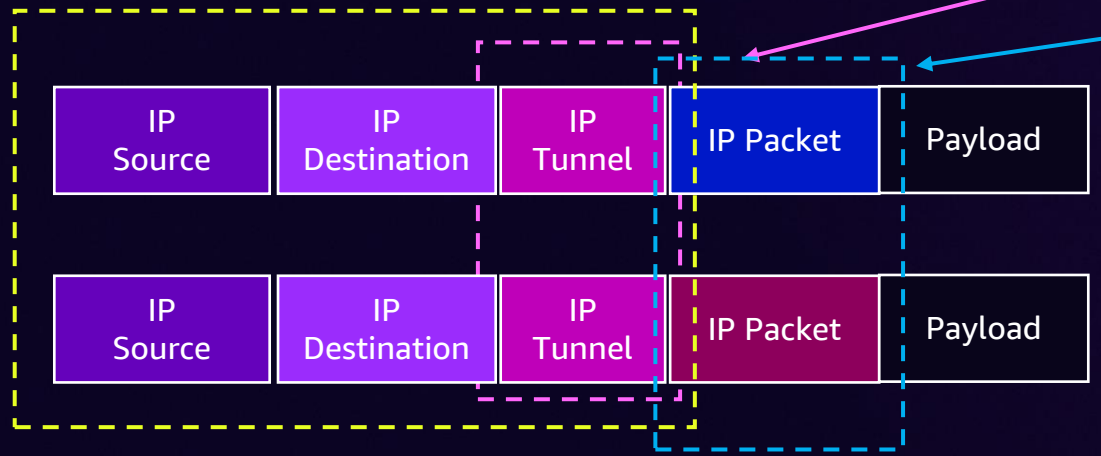
Optimizing for Large Single Flows



Single Flow Representation

IP Tunnel Protocol (GRE or IPSec)

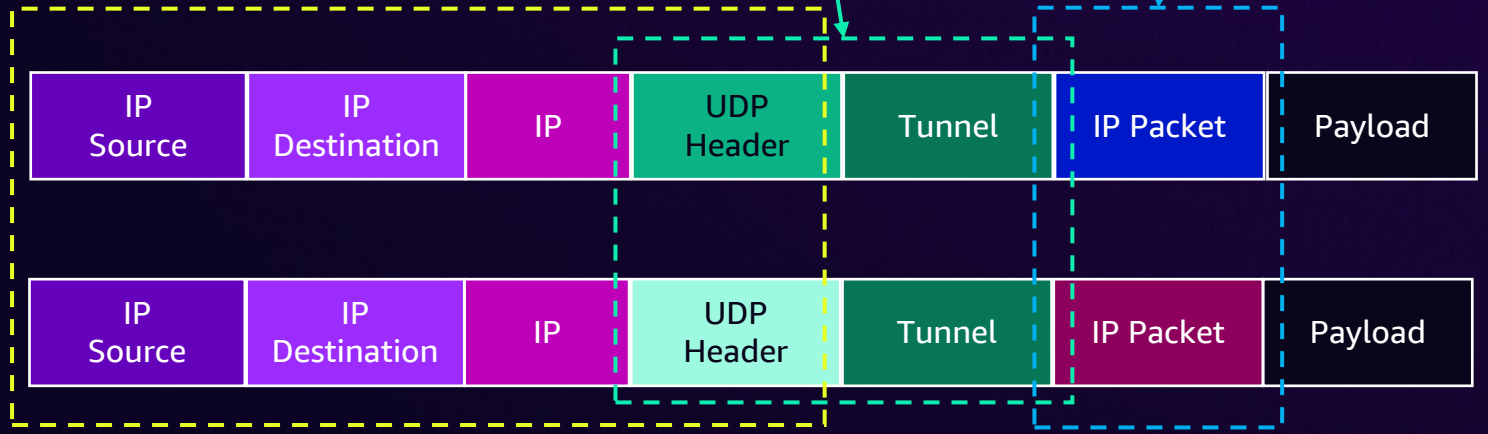
Unique Packet 5-Tuple



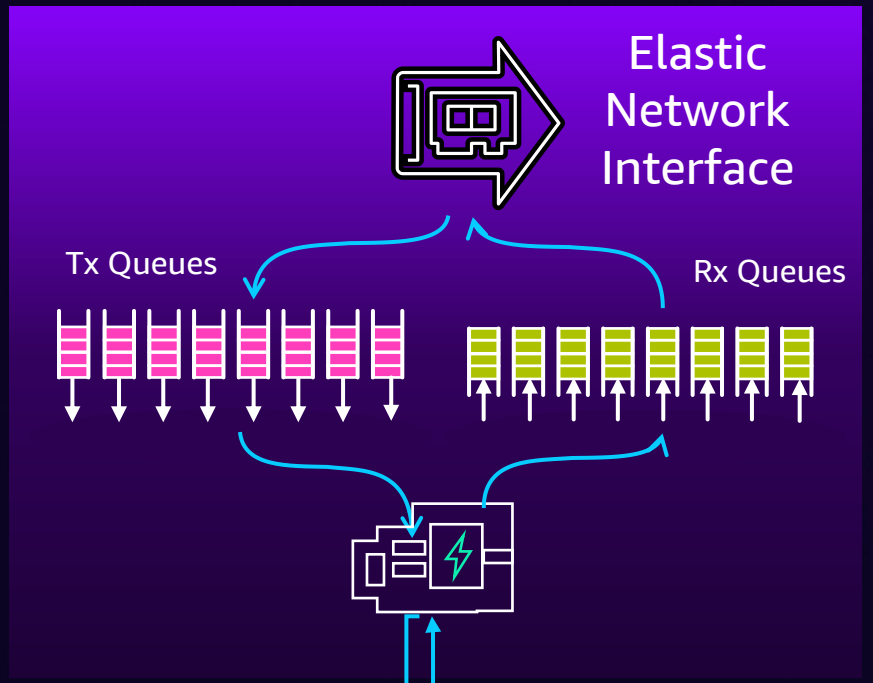
Unique Packet 5-Tuple

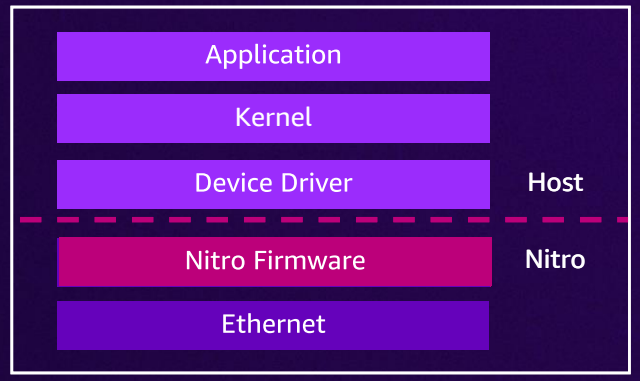
IP Tunnel Protocol
(VxLAN or GENEVE or GTP)
with Unique Source Port

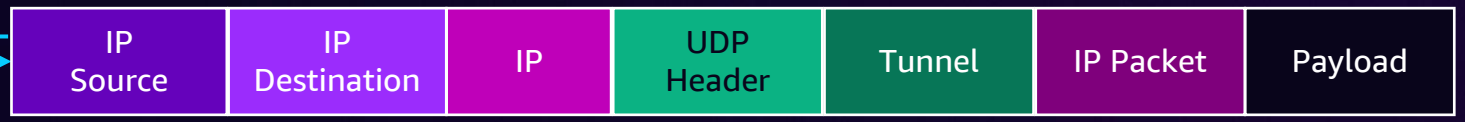
Multiple 5-Tuple Representation



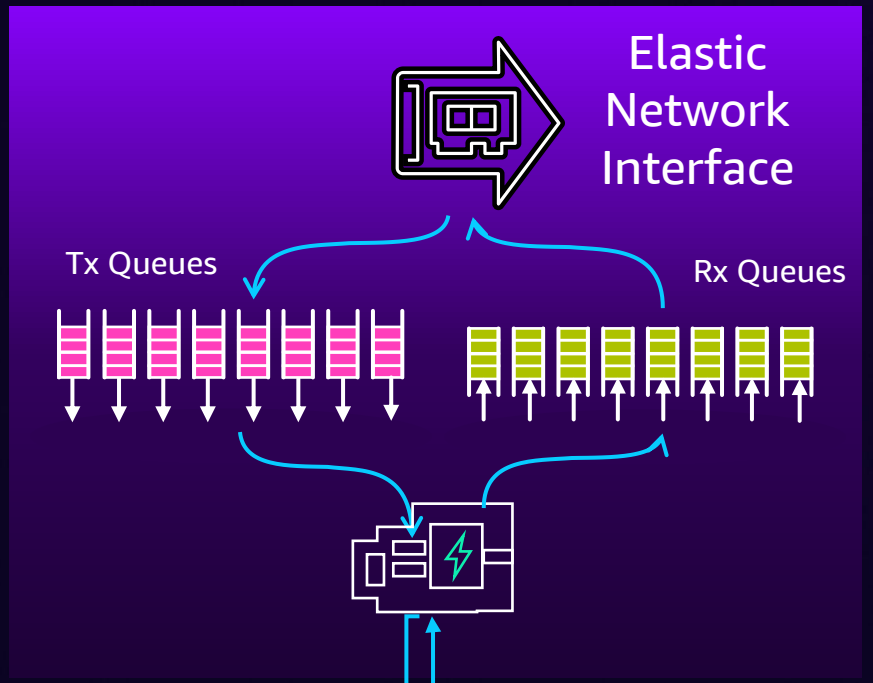
Nitro handing of Tunnel Hashing

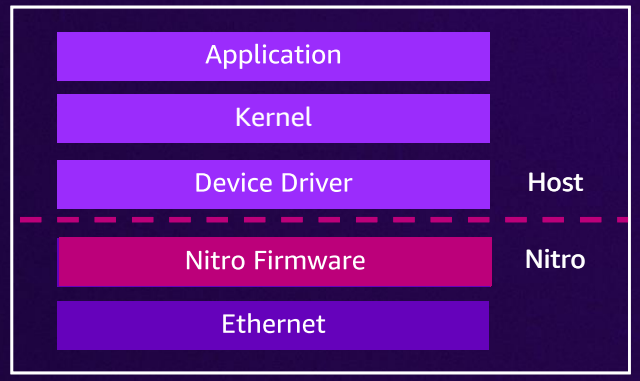






Nitro handing of Tunnel Hashing

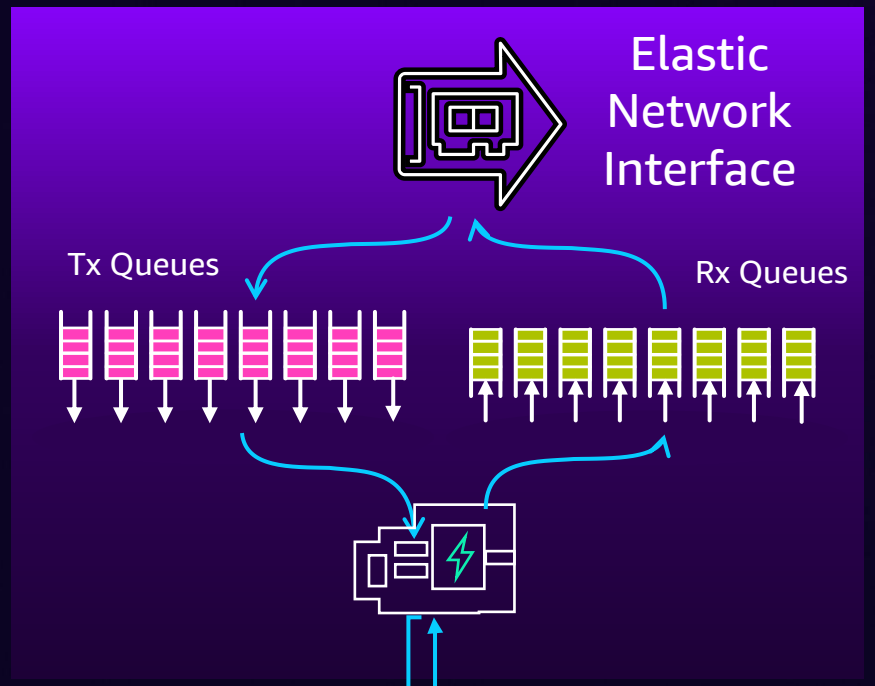


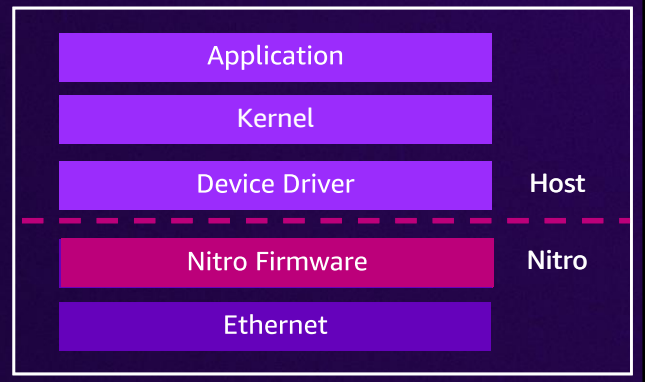


IP Source	IP Destination	IP	UDP Header	Tunnel	IP Packet	Payload

IP Source	IP Destination	IP	UDP Header	Tunnel	IP Packet	Payload

Nitro handing of Tunnel Hashing



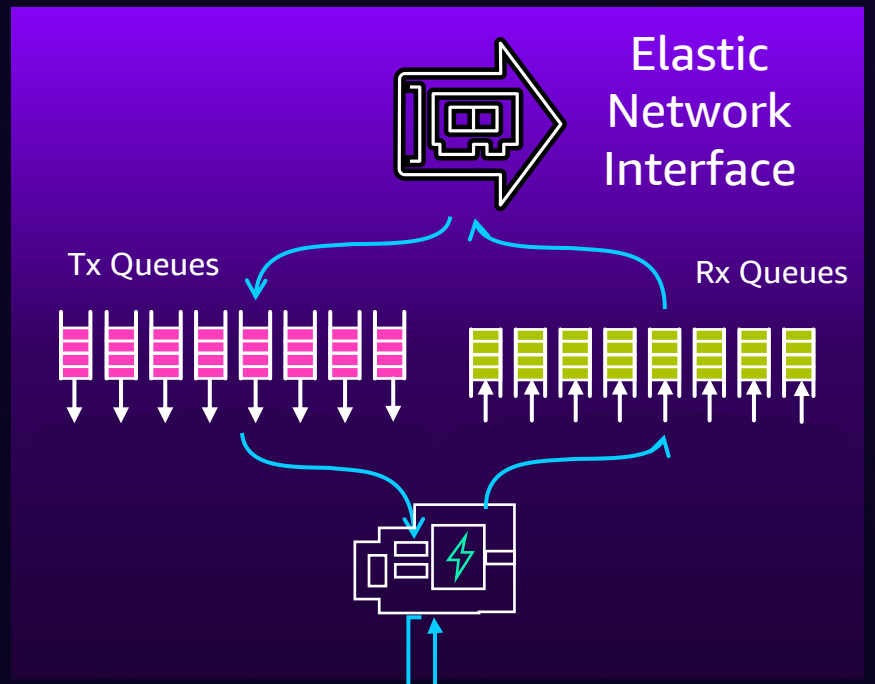


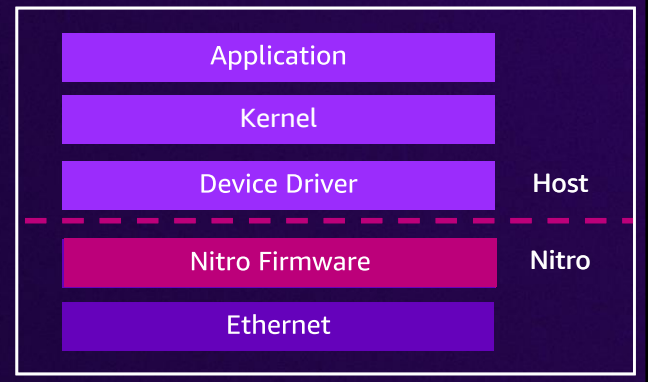
Unique Packet 5-Tuple

IP Source	IP Destination	IP	UDP Header	Tunnel	IP Packet	Payload

IP Source	IP Destination	IP	UDP Header	Tunnel	IP Packet	Payload

Nitro handing of Tunnel Hashing



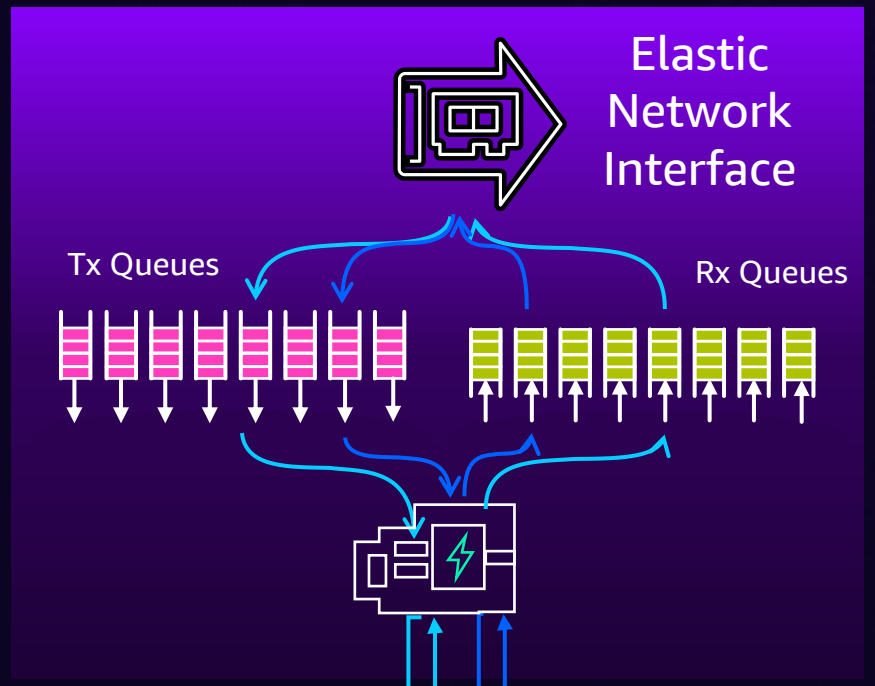


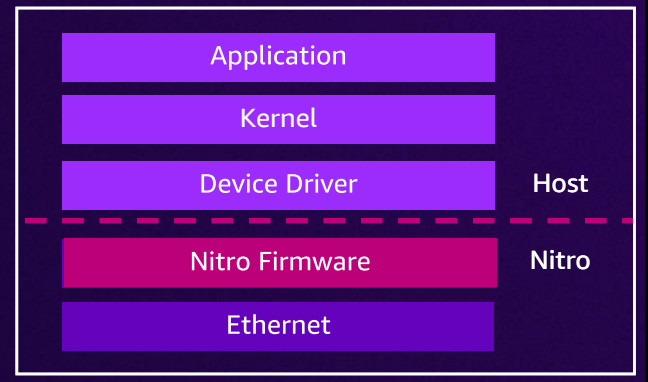
Unique Packet 5-Tuple

Multiple 5-Tuple Representation

IP Source	IP Destination	IP	UDP Header	Tunnel	IP Packet	Payload
IP Source	IP Destination	IP	UDP Header	Tunnel	IP Packet	Payload

Nitro handing of Tunnel Hashing





Unique Packet 5-Tuple

Multiple 5-Tuple Representation

IP Source	IP Destination	IP	UDP Header	Tunnel	IP Packet	Payload
IP Source	IP Destination	IP	UDP Header	Tunnel	IP Packet	Payload

Multi-flow Analysis

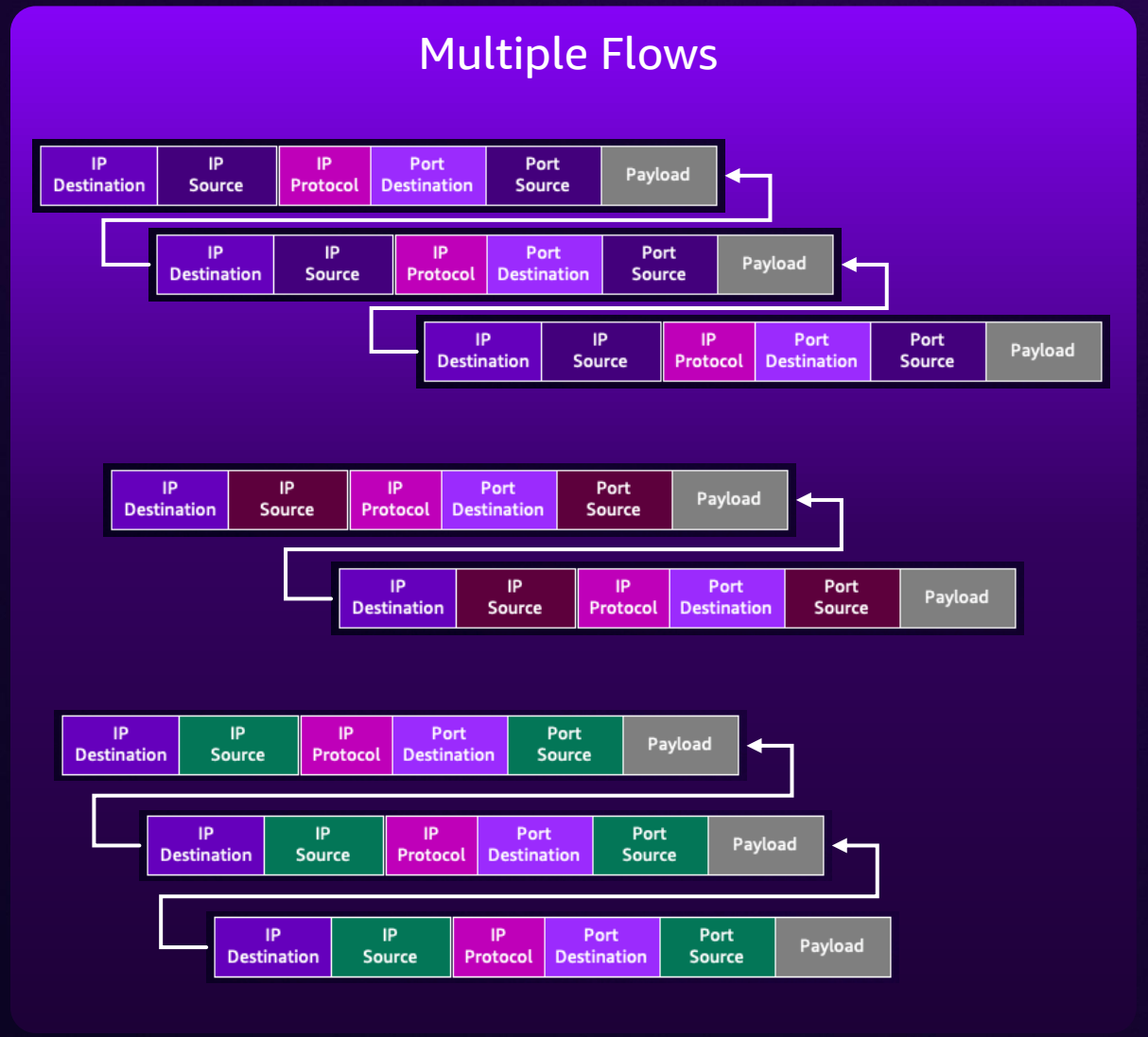
- Number of Flows
- Burst Bandwidth

Spec

- Baseline Bandwidth

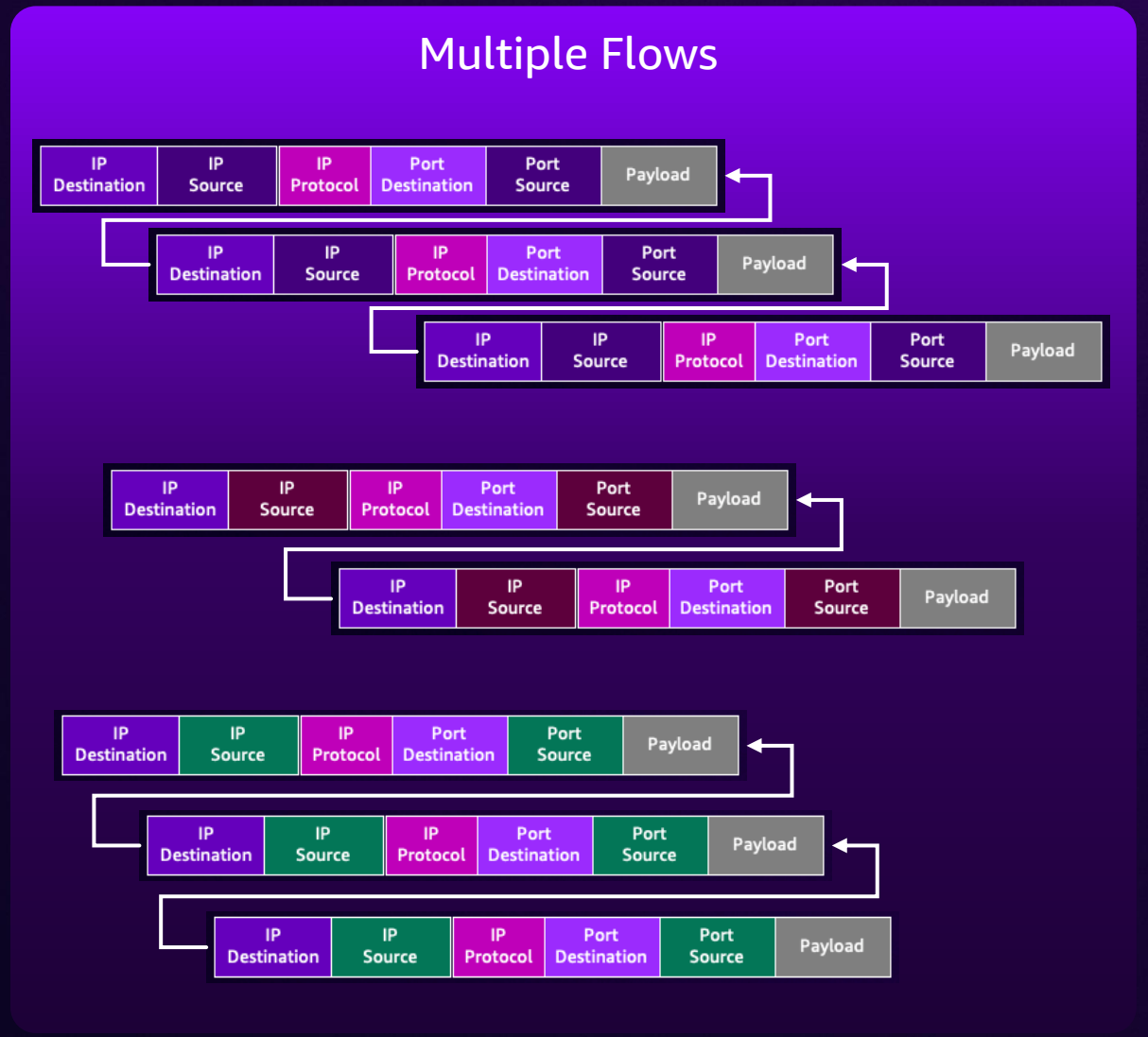
- Packets Per Second
- Shared Nitro Resources

Test

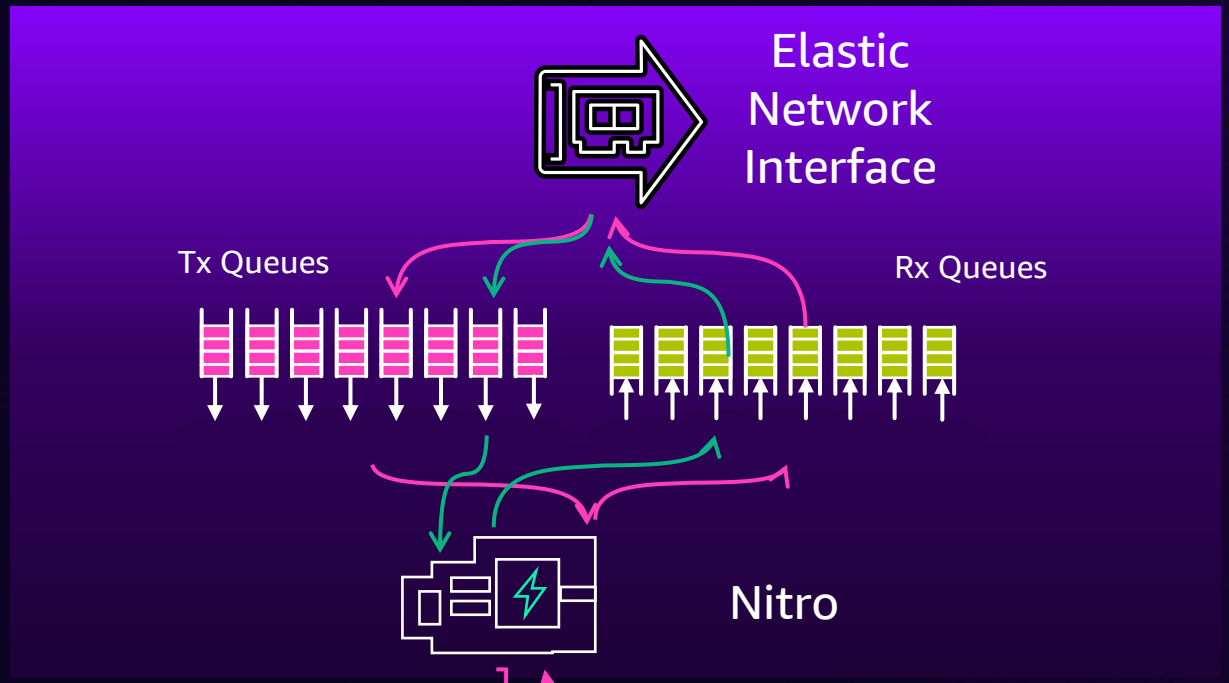


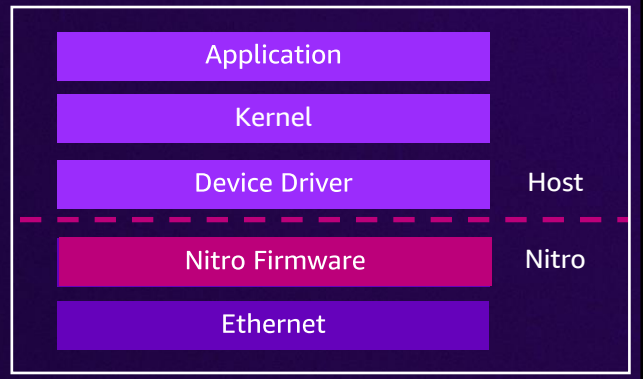
Multi-flow Analysis

- Number of Flows
- Burst Bandwidth
- Baseline Bandwidth
- Packets Per Second
- Shared Nitro Resources



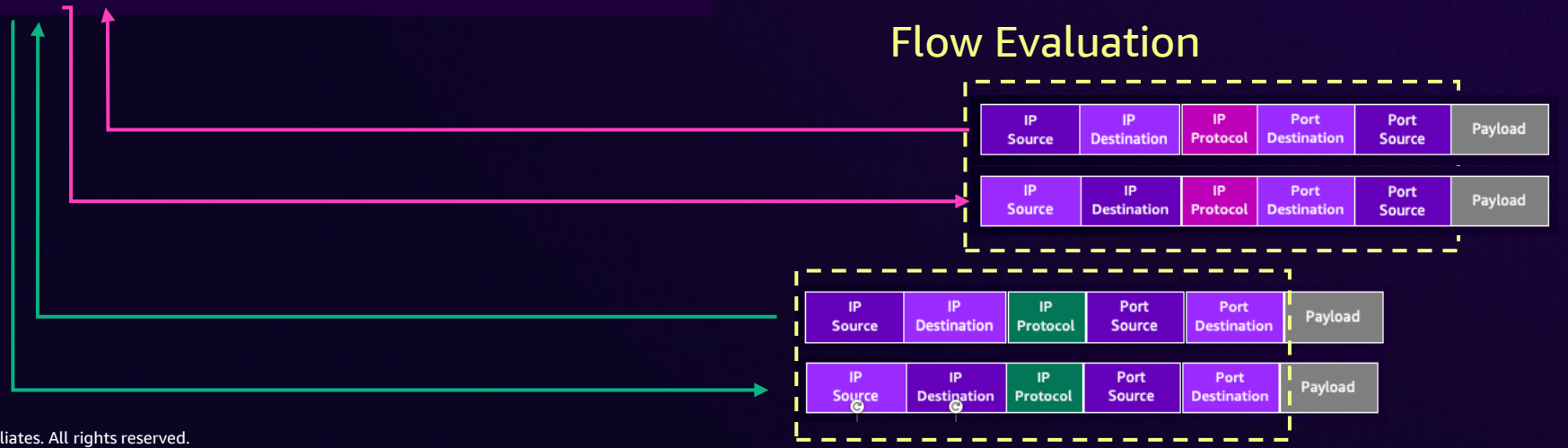
ENA Handling of Discrete Flows





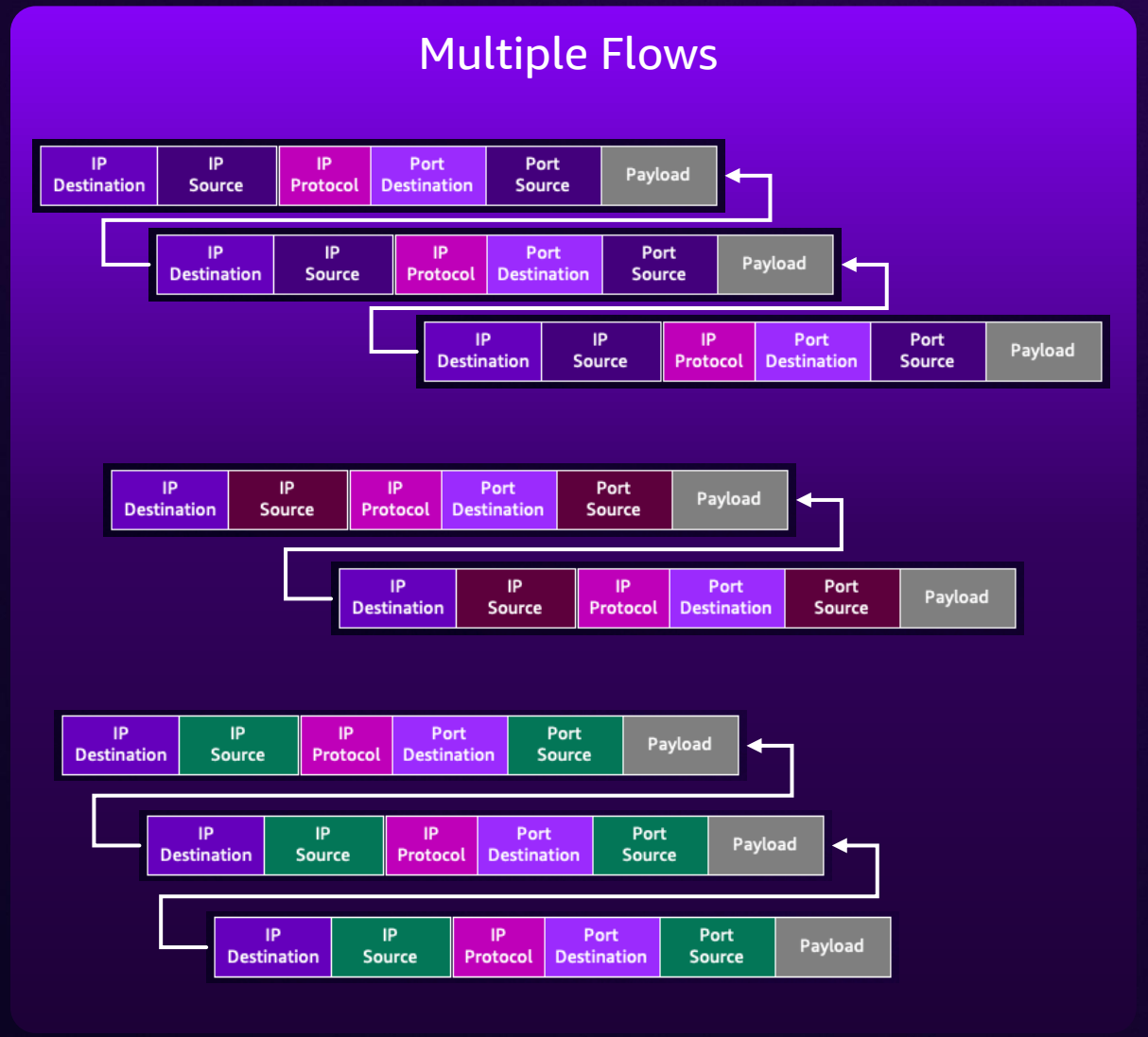
Nitro Processing

- Unique Flow Hashes
- Independent Assignment
  - Queue
  - Nitro Processor
- Flow Symmetry

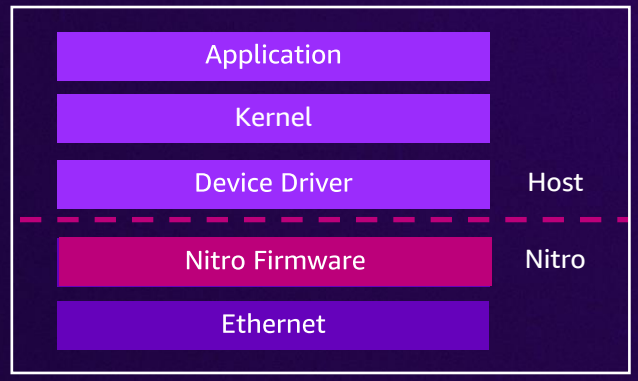


Multi-flow Analysis

- Number of Flows
- Burst Bandwidth
- Baseline Bandwidth
- Packets Per Second
- Shared Nitro Resources

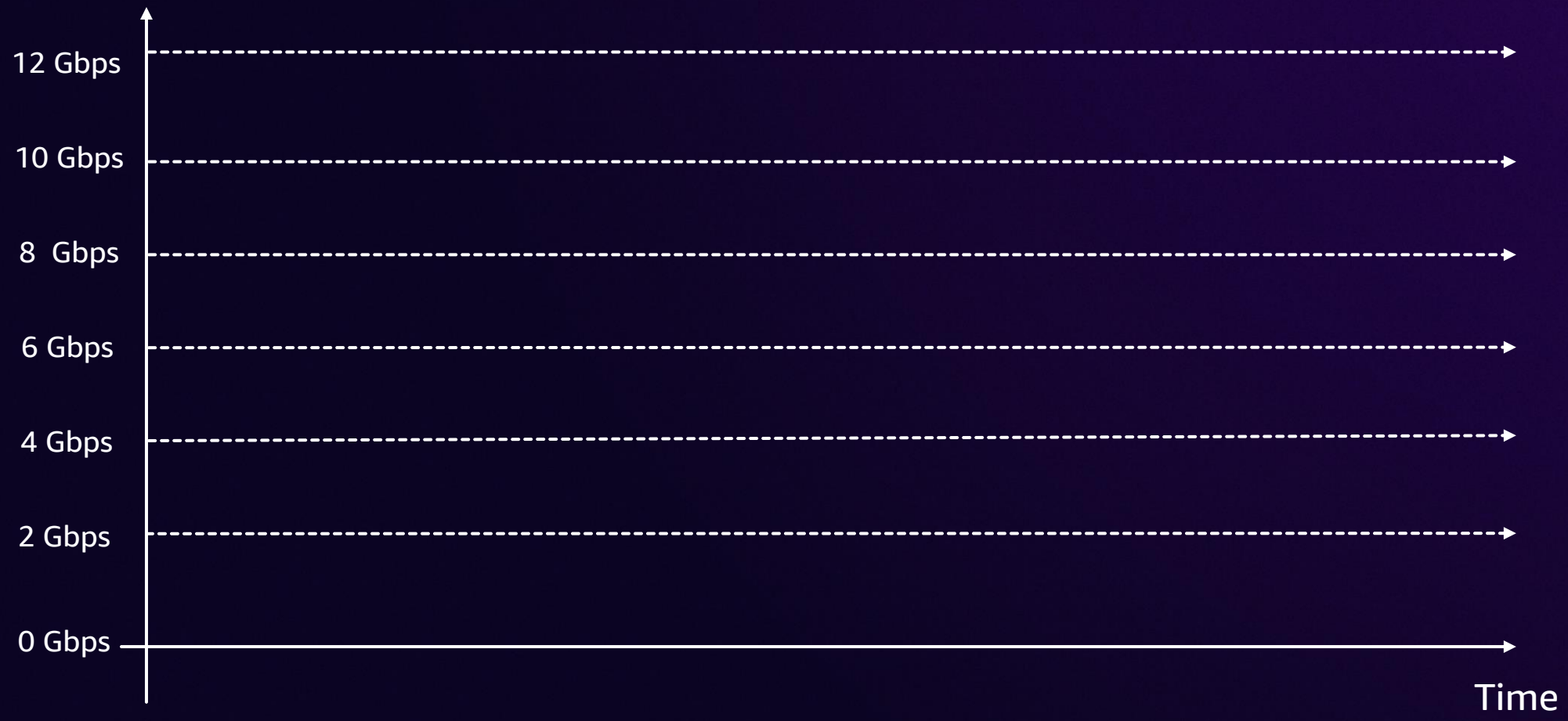


Instance Bandwidth Specifications

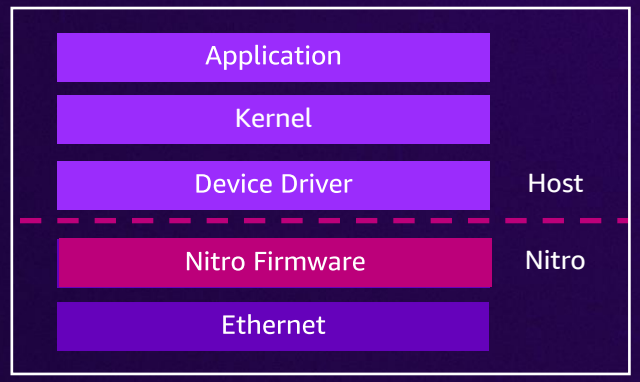


Example: c7i.large Instance Bandwidth

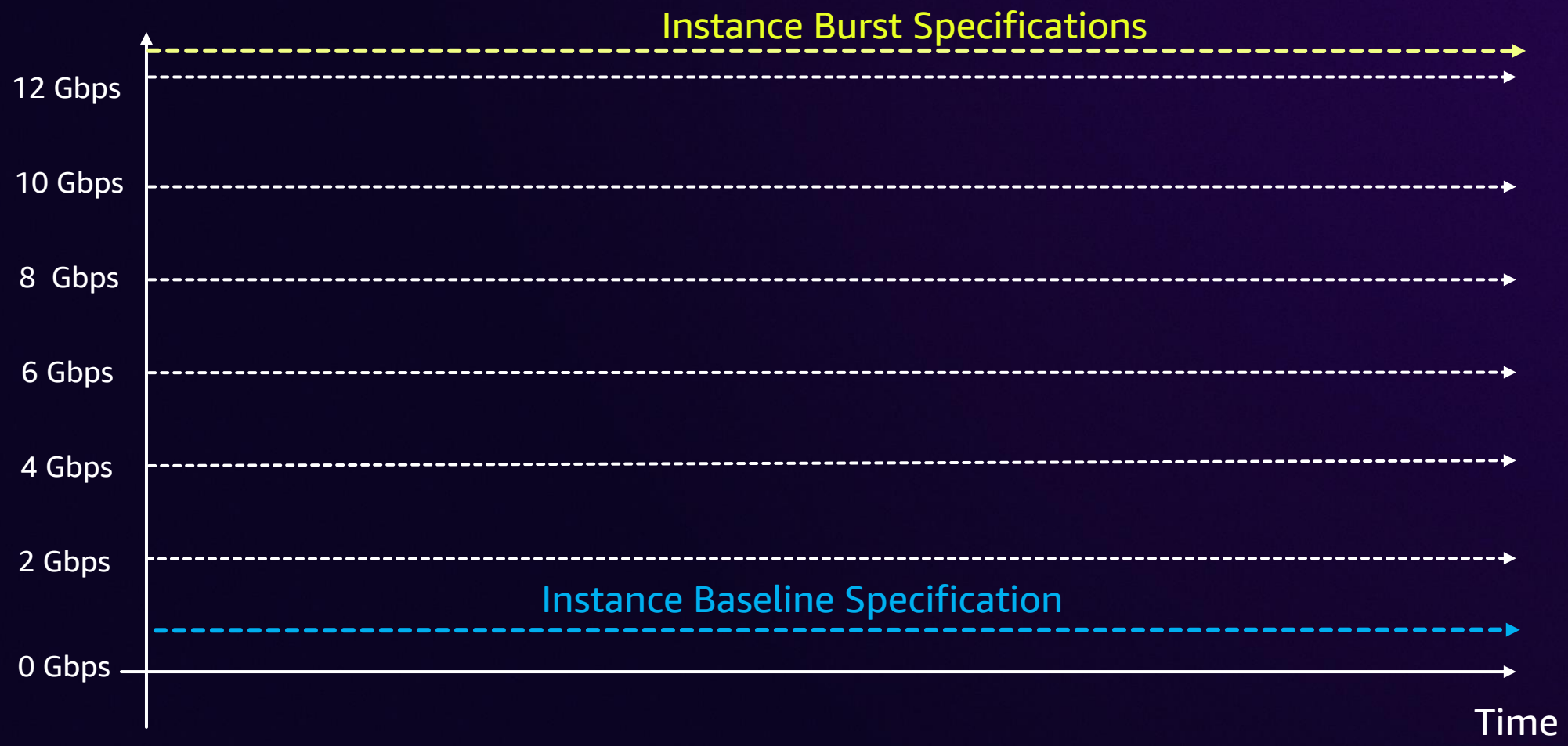
Specification: Burst 12.5 Gbps / Baseline 0.780 Gbps



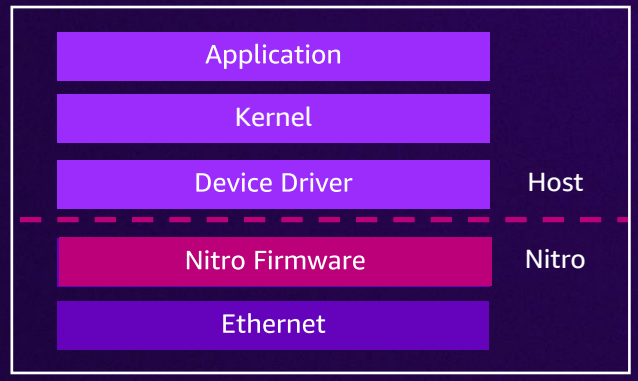
Instance Bandwidth Specifications



Example: c7i.large Instance Bandwidth
Specification: Burst 12.5 Gbps / Baseline 0.780 Gbps

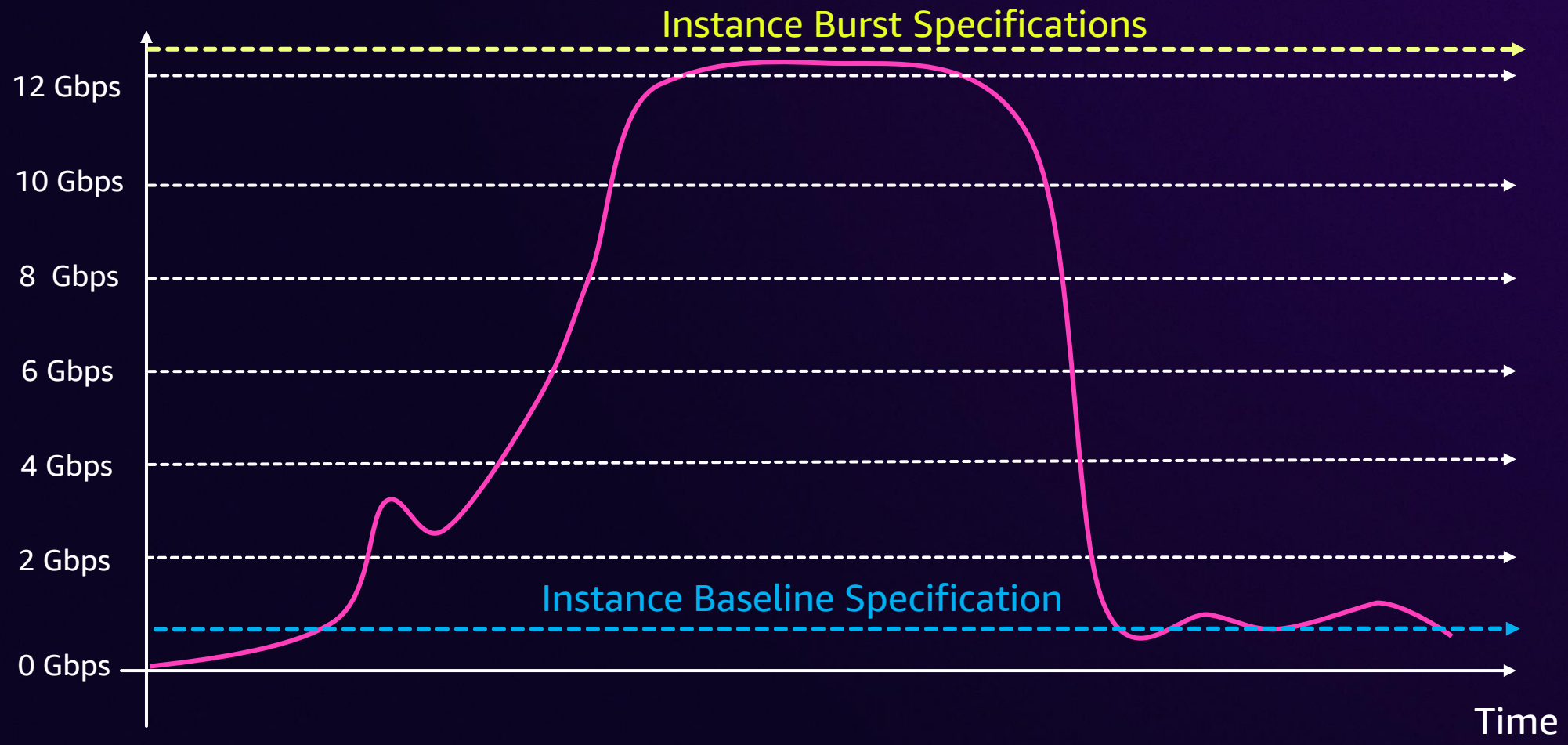


Instance Bandwidth Specifications

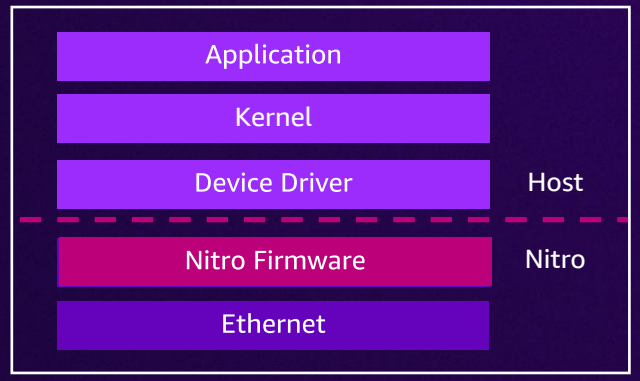


Example: c7i.large Instance Bandwidth

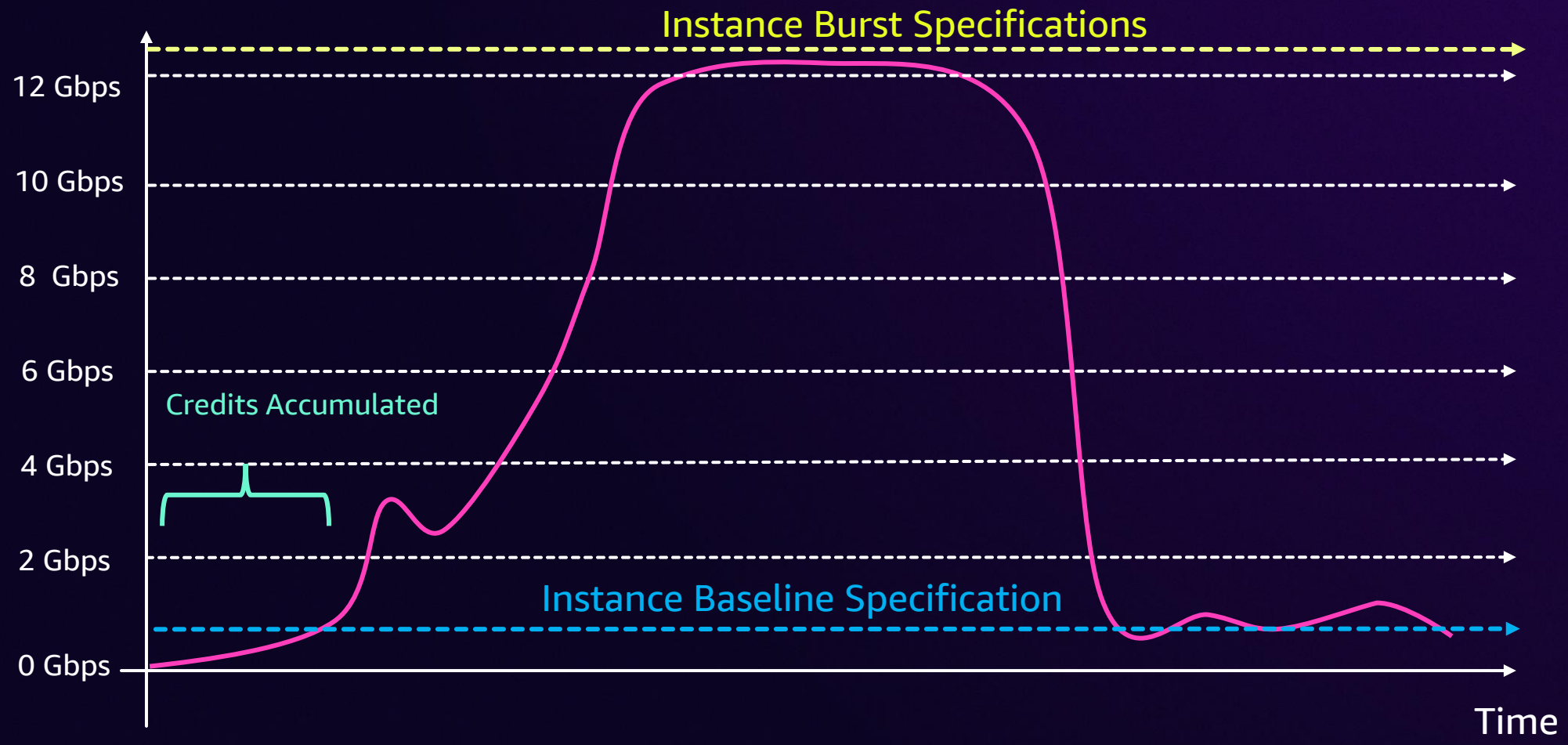
Specification: Burst 12.5 Gbps / Baseline 0.780 Gbps



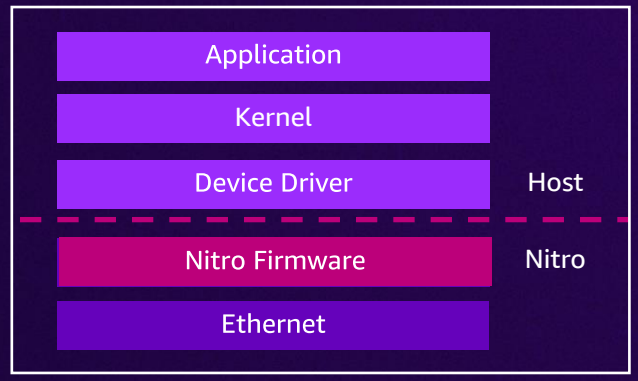
Instance Bandwidth Specifications



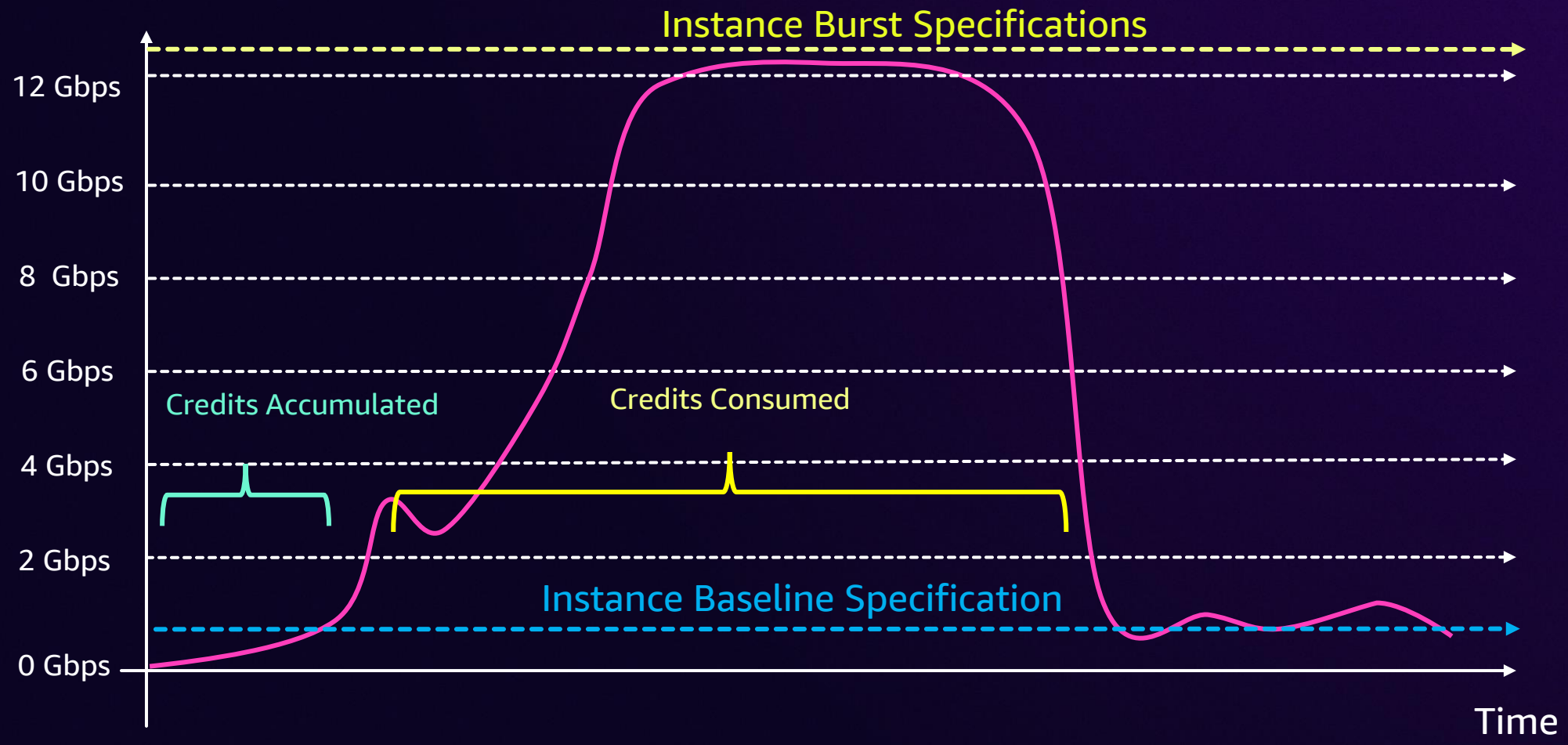
Example: c7i.large Instance Bandwidth
Specification: Burst 12.5 Gbps / Baseline 0.780 Gbps



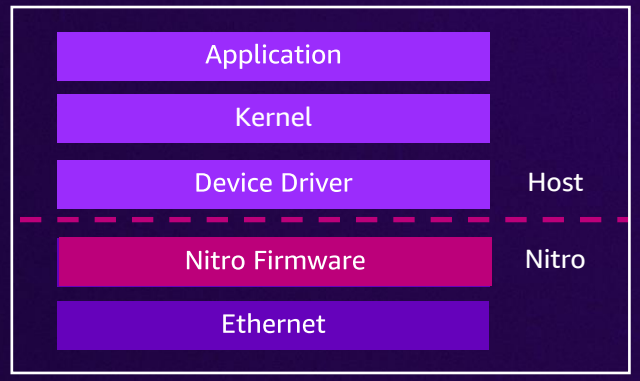
Instance Bandwidth Specifications



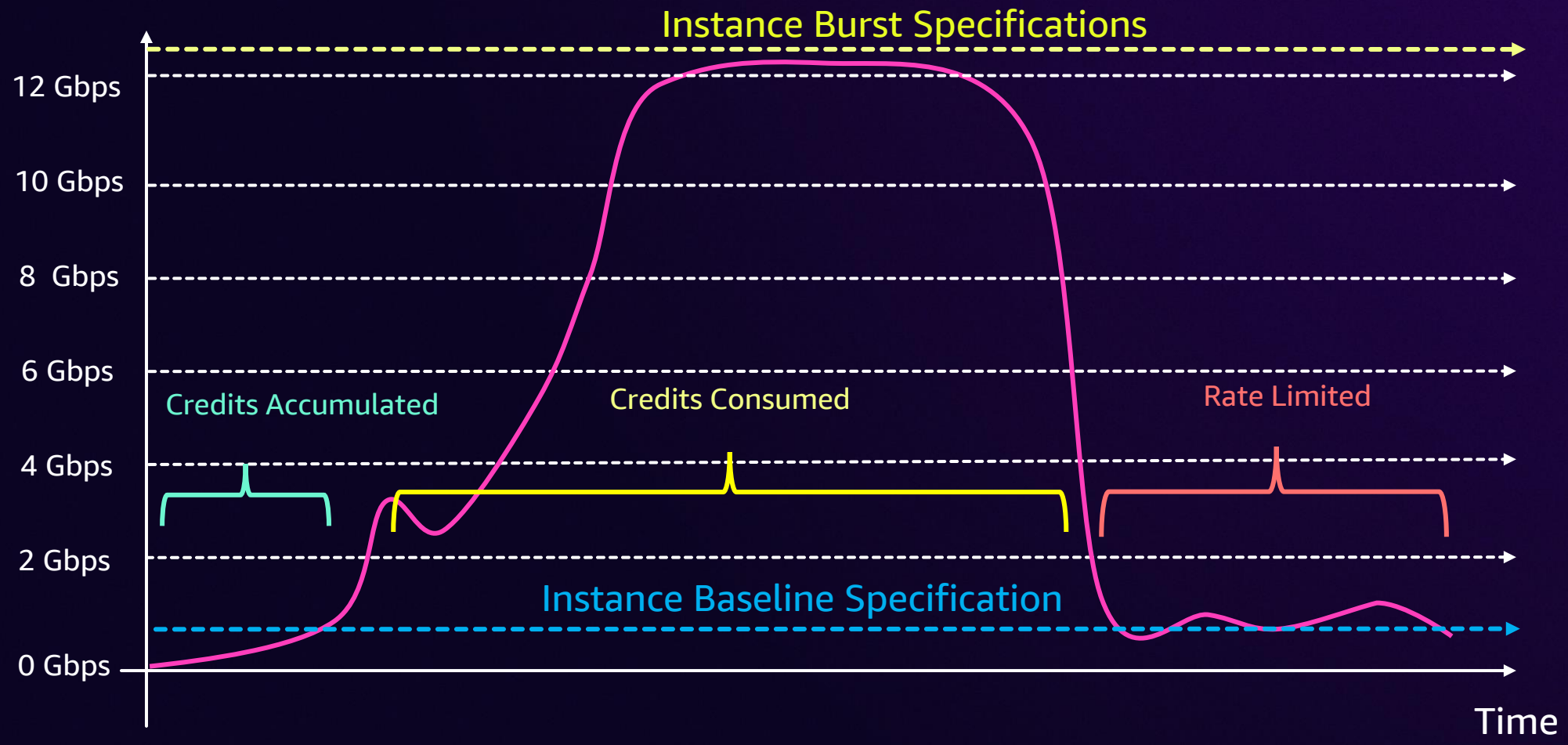
Example: c7i.large Instance Bandwidth
Specification: Burst 12.5 Gbps / Baseline 0.780 Gbps



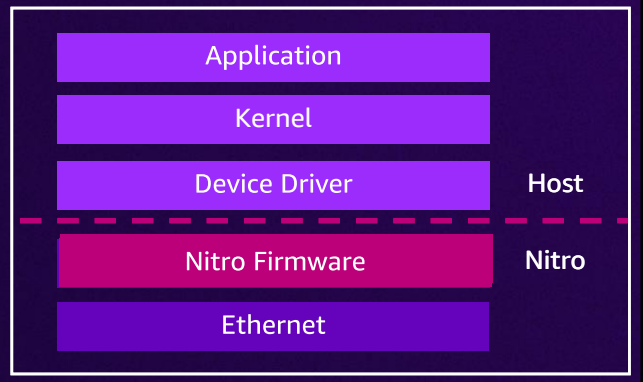
Instance Bandwidth Specifications



Example: c7i.large Instance Bandwidth
Specification: Burst 12.5 Gbps / Baseline 0.780 Gbps

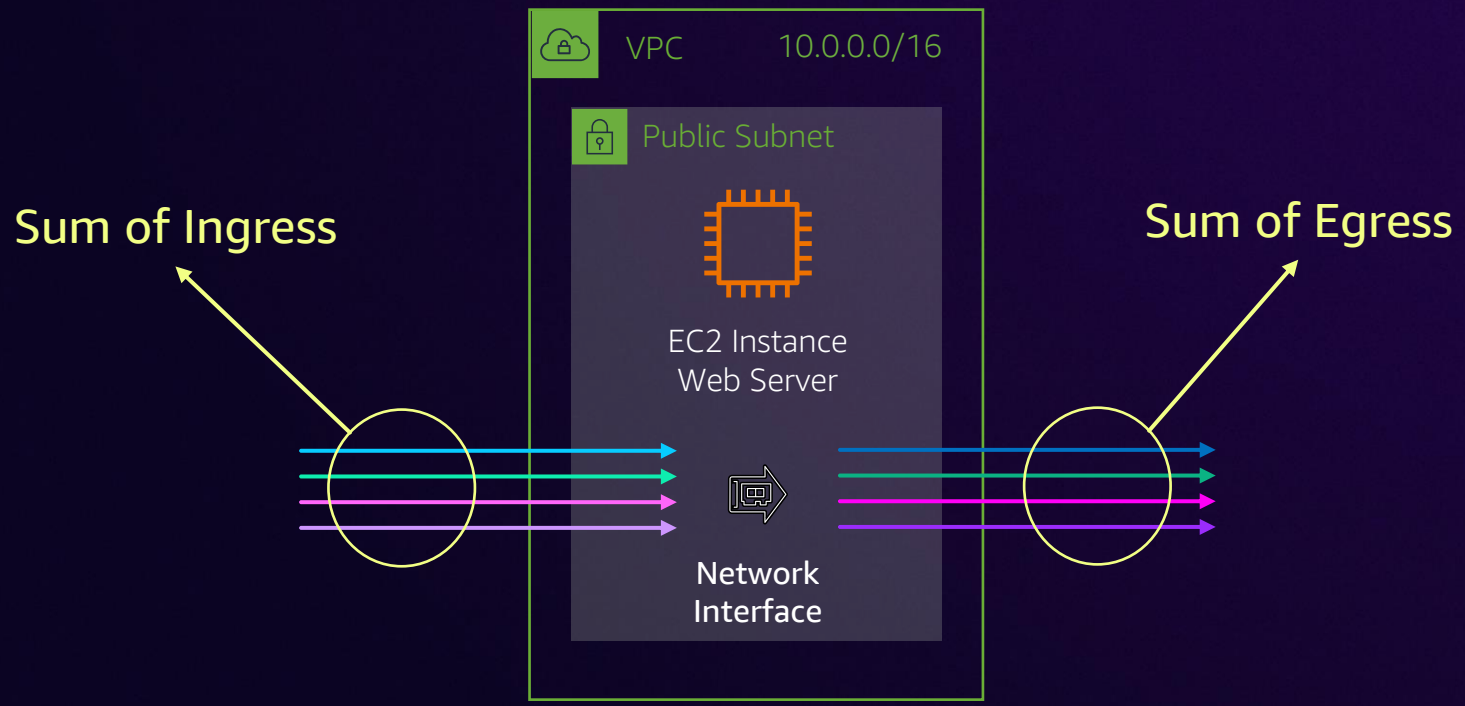


Instance Burst Bandwidth Capabilities



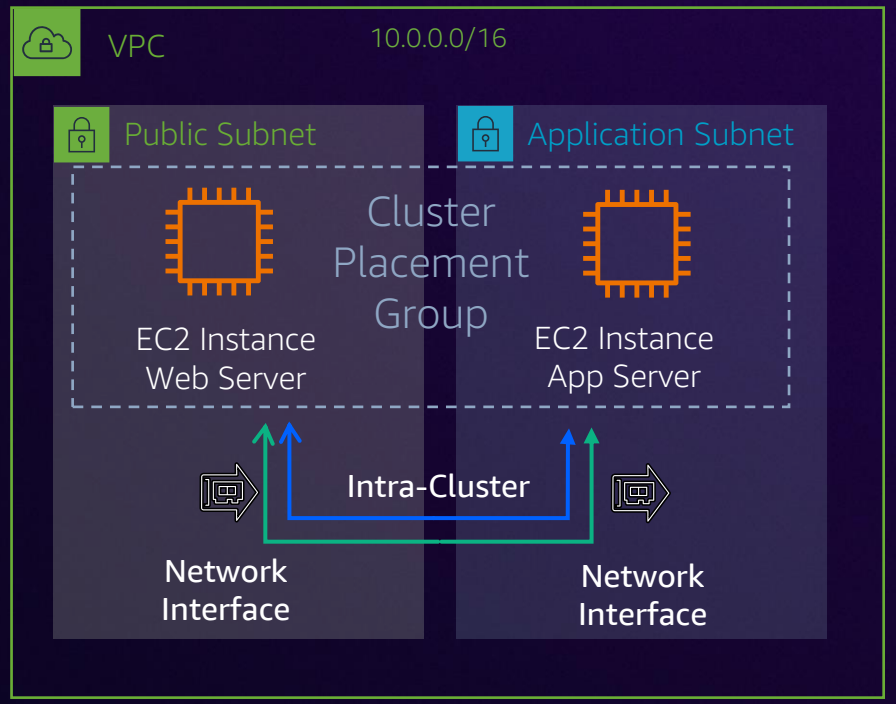
Burst bandwidth

Burst credits availability

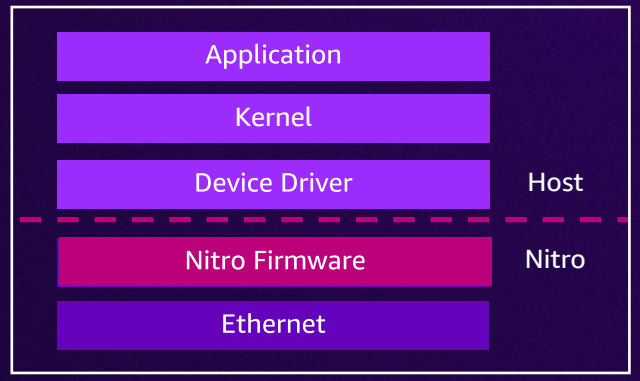


Multi-Flow Test Topology

- Tool: iperf3
- Instance c7i.large
  - Burst 12.5 Gbps, Baseline 780 Mbps
- Transmit to Receiver
- Large MTU to hit Burst Limit
- Rate Limited to Instance Spec
- Aggregate Bandwidth Split



Test : iperf3 on c7i.large (burst)



Single Flow, Single Queue

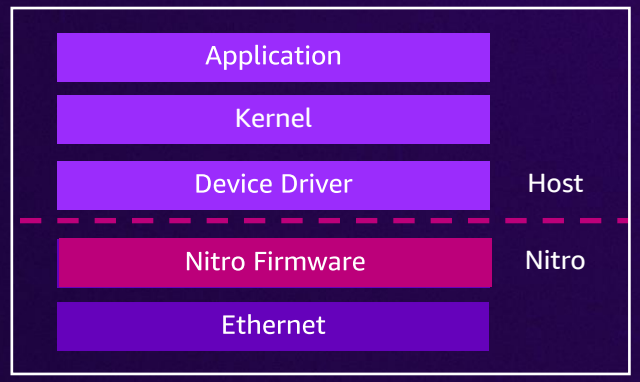
```
[ec2-user@web-server ~]$ iperf3 -c app_server -t 1 -p 5201 -l 9000 -P 1
[ ID]         Interval                Transfer                Bitrate
[  5]         0.00-1.00   sec         1.11 GBytes             9.54 Gbits/sec
```

Test : iperf3 on c7i.large (burst)

Single Flow, Single Queue

```
[ec2-user@web-server ~]$ iperf3 -c app_server -t 1 -p 5201 -l 9000 -P 1
[ ID]         Interval                Transfer              Bitrate
[  5]         0.00-1.00   sec         1.11 GBytes           9.54 Gbits/sec
```

-t – duration
-p = port
-l = buffer length
-P = streams



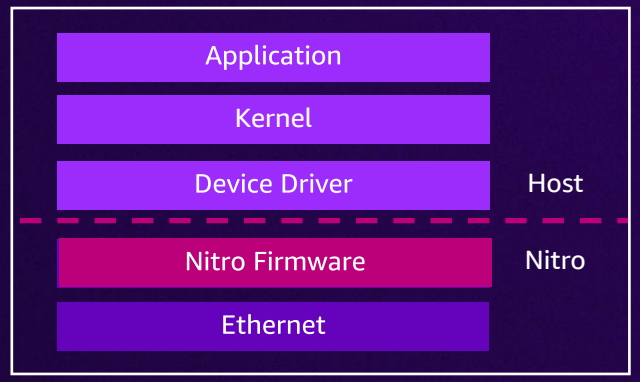
Test : iperf3 on c7i.large (burst)

Single Flow, Single Queue

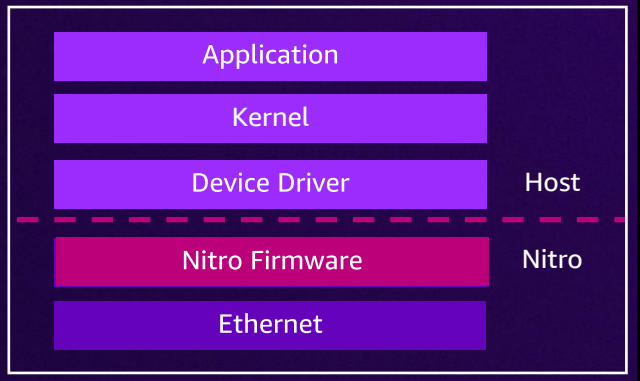
```
[ec2-user@web-server ~]$ iperf3 -c app_server -t 1 -p 5201 -l 9000 -P 1
[ ID]         Interval                Transfer              Bitrate
[  5]         0.00-1.00   sec         1.11 GBytes           9.54 Gbits/sec
```

-t – duration
-p = port
-l = buffer length
-P = streams

Intra-Cluster
Flow Limit



Test : iperf3 on c7i.large (burst)



Single Flow, Single Queue

```
[ec2-user@web-server ~]$ iperf3 -c app_server -t 1 -p 5201 -l 9000 -P 1
[ ID]        Interval                Transfer              Bitrate
[  5]        0.00-1.00   sec         1.11 GBytes           9.54 Gbits/sec
```

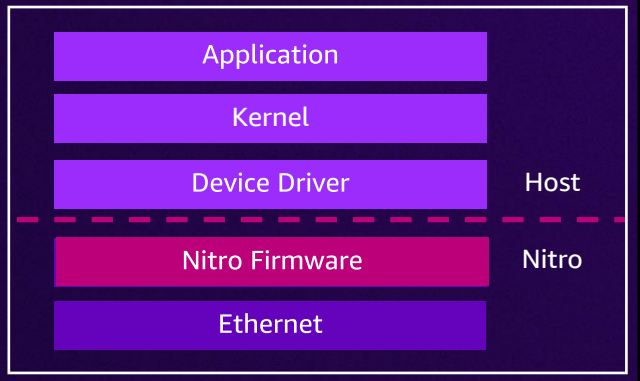
-t – duration
-p = port
-l = buffer length
-P = streams

Intra-Cluster Flow Limit

Two Flows, Two Queues

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 1 -p 5201 -l 9000 -P 2
[  5] local web-server port 35722 connected to app-server port 5201
[  7] local web-server port 35726 connected to app-server port 5201
[ ID]        Interval                Transfer              Bitrate
[  5]        0.00-1.00   sec         745 MBytes            6.24 Gbits/sec
[  7]        0.00-1.00   sec         743 MBytes            6.22 Gbits/sec
[SUM]        0.00-1.00   sec         1.45 GBytes           12.5 Gbits/sec
```

Test : iperf3 on c7i.large (burst)



Single Flow, Single Queue

```
[ec2-user@web-server ~]$ iperf3 -c app_server -t 1 -p 5201 -l 9000 -P 1
[ ID]        Interval                Transfer              Bitrate
[  5]        0.00-1.00   sec         1.11 GBytes           9.54 Gbits/sec
```

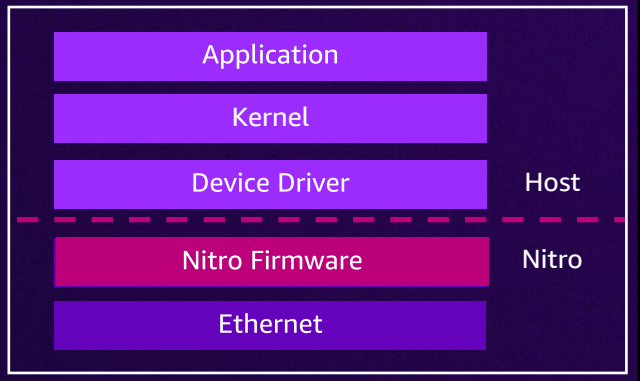
-t – duration
-p = port
-l = buffer length
-P = streams

Intra-Cluster Flow Limit

Two Flows, Two Queues

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 1 -p 5201 -l 9000 -P 2
[  5] local web-server port 35722 connected to app-server port 5201
[  7] local web-server port 35726 connected to app-server port 5201
[ ID]        Interval                Transfer              Bitrate
[  5]        0.00-1.00   sec         745 MBytes            6.24 Gbits/sec
[  7]        0.00-1.00   sec         743 MBytes            6.22 Gbits/sec
[SUM]        0.00-1.00   sec         1.45 GBytes           12.5 Gbits/sec
```

Test : iperf3 on c7i.large (burst)



Single Flow, Single Queue

```
[ec2-user@web-server ~]$ iperf3 -c app_server -t 1 -p 5201 -l 9000 -P 1
[ ID]        Interval                Transfer              Bitrate
[  5]        0.00-1.00   sec         1.11 GBytes           9.54 Gbits/sec
```

-t – duration
-p = port
-l = buffer length
-P = streams

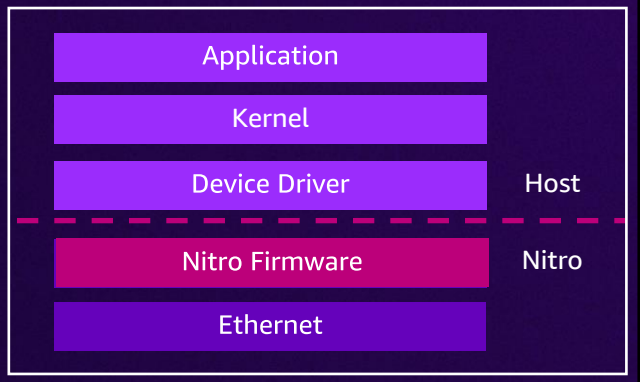
Intra-Cluster Flow Limit

Two Flows, Two Queues

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 1 -p 5201 -l 9000 -P 2
[  5] local web-server port 35722 connected to app-server port 5201
[  7] local web-server port 35726 connected to app-server port 5201
[ ID]        Interval                Transfer              Bitrate
[  5]        0.00-1.00   sec         745 MBytes            6.24 Gbits/sec
[  7]        0.00-1.00   sec         743 MBytes            6.22 Gbits/sec
[SUM]        0.00-1.00   sec         1.45 GBytes           12.5 Gbits/sec
```

Instance Aggregate Burst Flow Limit

Test : iperf3 on c7i.large (burst)



Single Flow, Single Queue

```
[ec2-user@web-server ~]$ iperf3 -c app_server -t 1 -p 5201 -l 9000 -P 1
[ ID]        Interval                Transfer              Bitrate
[  5]        0.00-1.00   sec         1.11 GBytes           9.54 Gbits/sec
```

-t – duration
-p = port
-l = buffer length
-P = streams

Enhanced Network Metrics
- bandwidth_allowance_in
- bandwidth_allowance_out

Two Flows, Two Queues

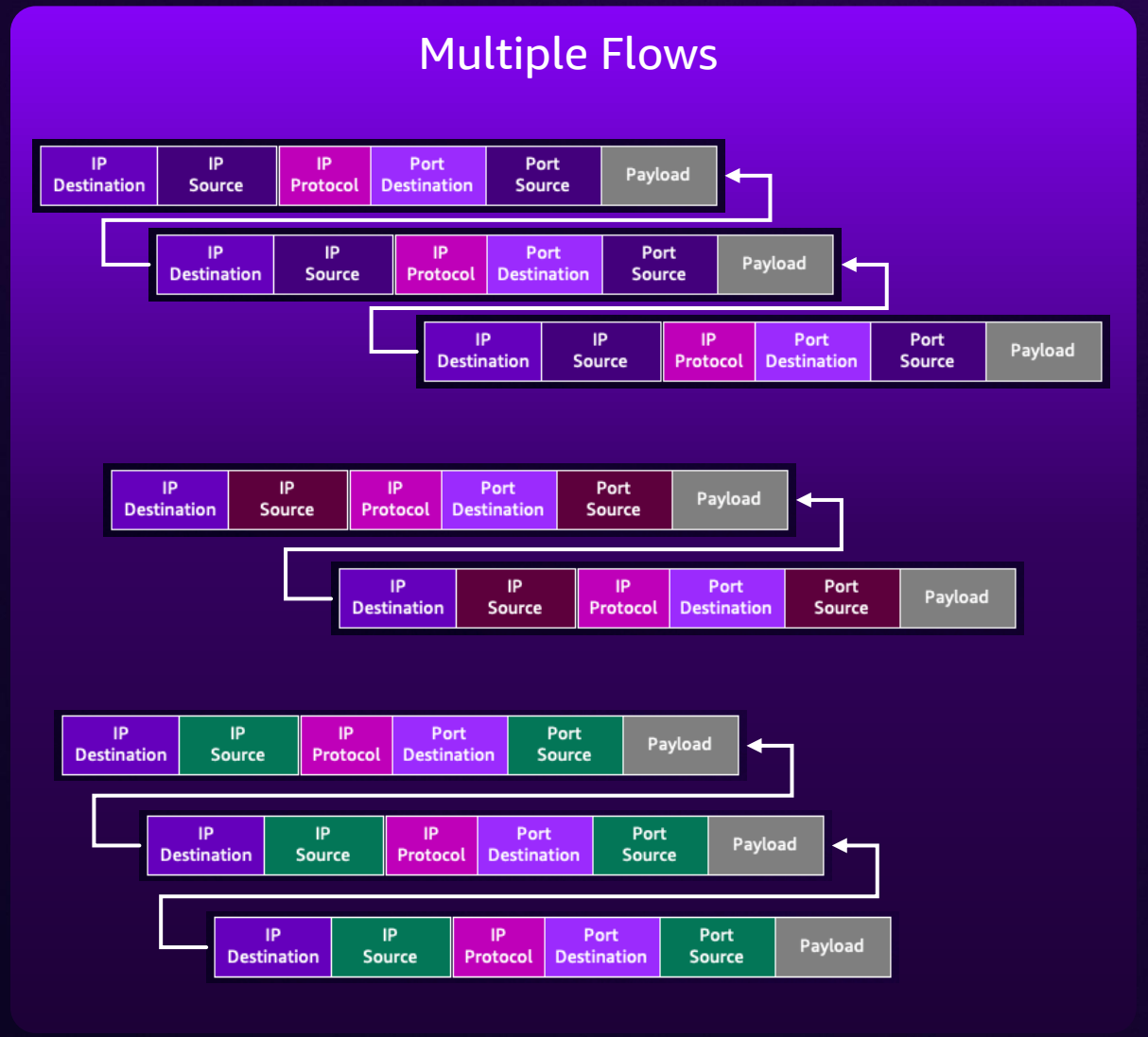
```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 1 -p 5201 -l 9000 -P 2
[  5] local web-server port 35722 connected to app-server port 5201
[  7] local web-server port 35726 connected to app-server port 5201
[ ID]        Interval                Transfer              Bitrate
[  5]        0.00-1.00   sec         745 MBytes            6.24 Gbits/sec
[  7]        0.00-1.00   sec         743 MBytes            6.22 Gbits/sec
[SUM]        0.00-1.00   sec         1.45 GBytes           12.5 Gbits/sec
```

Intra-Cluster Flow Limit

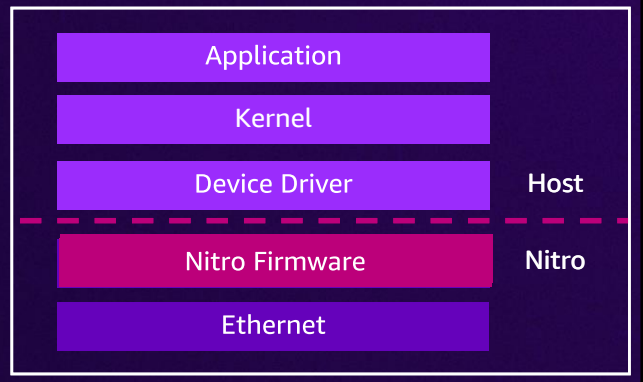
Instance Aggregate Burst Flow Limit

Multi-flow Analysis

- Number of Flows
- Burst Bandwidth
- Baseline Bandwidth
- Packets Per Second
- Shared Nitro Resources

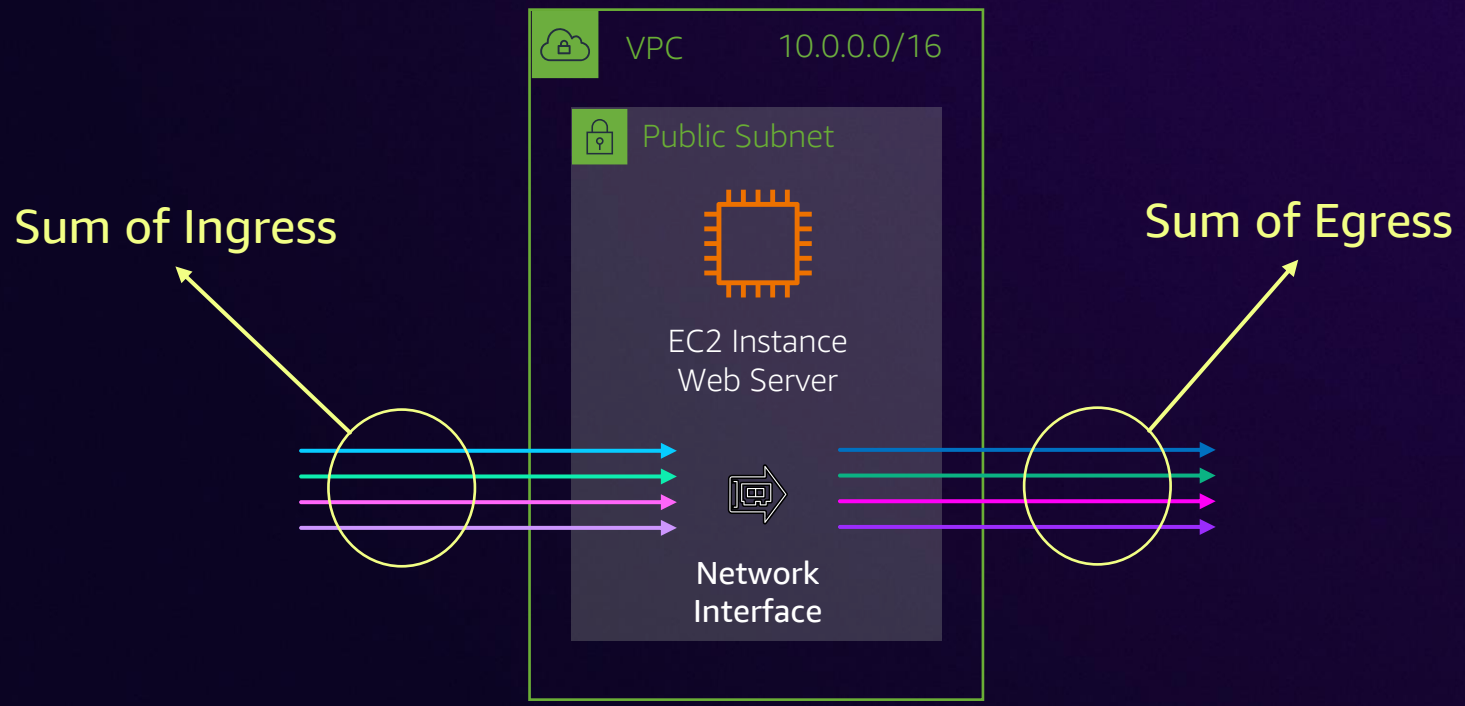


Instance Bandwidth Baseline

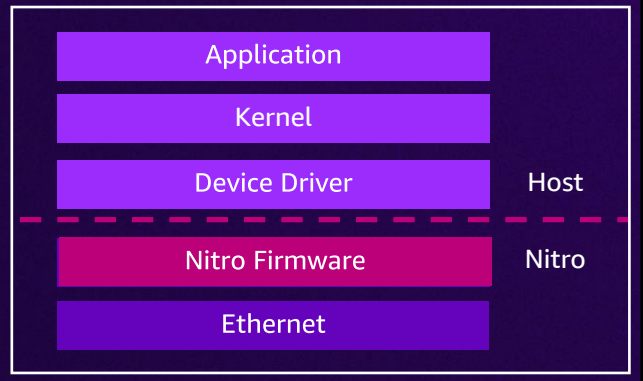


Baseline bandwidth

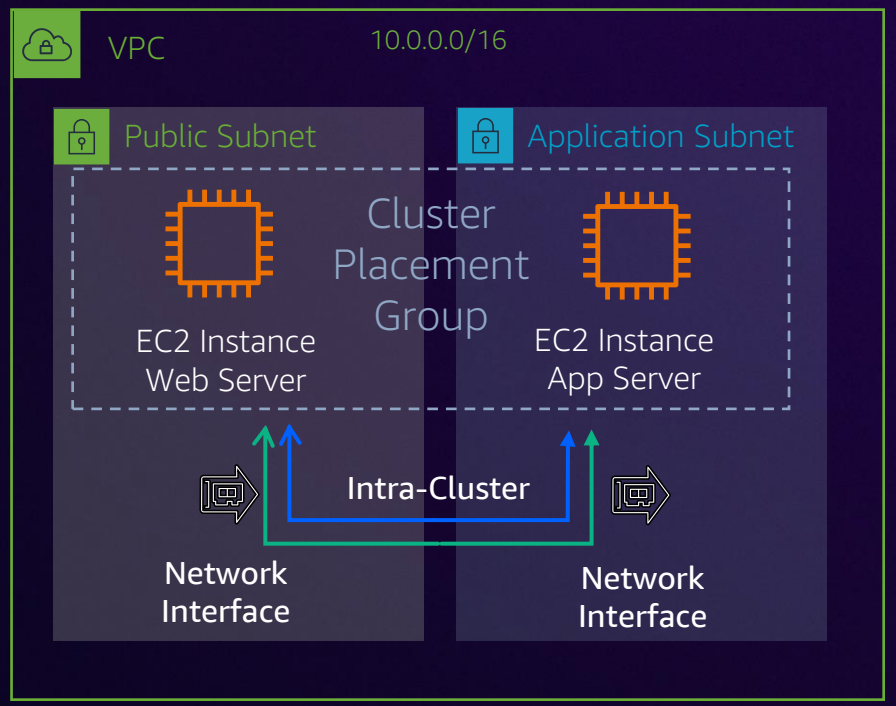
Burst credits consumed



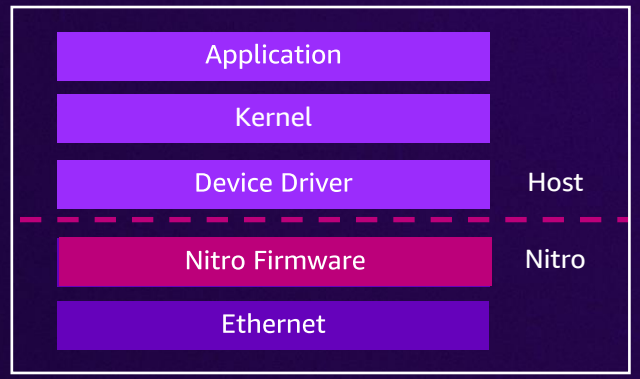
Multi-Flow Test Topology



- Tool: iperf3
- Instance c7i.large
  - Burst 12.5 Gbps, Baseline 780 Mbps
- Transmit to Receiver
- Large MTU to hit Burst Limit
- Time Period to Consume Credits
- Burst until Credit Exhaustion
- Rate Limited to Baseline
- Aggregate Bandwidth Split



Test : iperf3 on c7i.large (baseline)

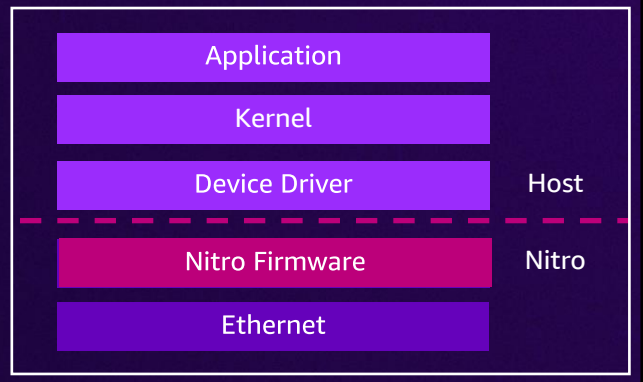


Two Flows, Two Queues [1]

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 600 -p 5201 -l 9000 -P 2
Connecting to host app-server7, port 5201
[  5] local web-server port 47410 connected to app-server port 5201
[  7] local web-server port 47424 connected to app-server port 5201
[ ID]        Interval                Transfer              Bitrate
[ 5]         304.00-305.00  sec      43.9 MBytes           368 Mbits/sec
[ 7]         304.00-305.00 sec       48.1 MBytes           403 Mbits/sec
[SUM]        304.00-305.00 sec       92.0 MBytes           772 Mbits/sec
```

[1] Test executed immediately after credits consumed

Test : iperf3 on c7i.large (baseline)



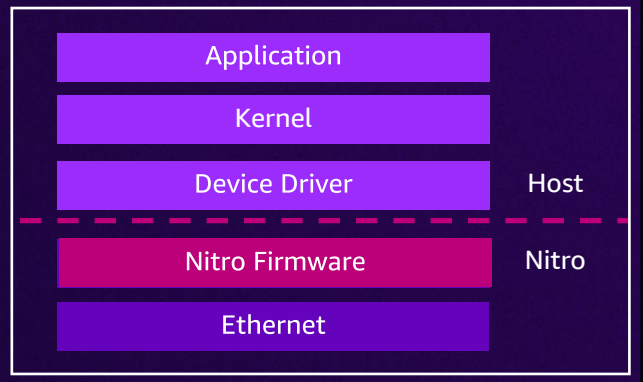
Two Flows, Two Queues [1]

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 600 -p 5201 -l 9000 -P 2
Connecting to host app-server7, port 5201
[  5] local web-server port 47410 connected to app-server port 5201
[  7] local web-server port 47424 connected to app-server port 5201
[ ID]        Interval                Transfer              Bitrate
[ 5]         304.00-305.00  sec      43.9 MBytes           368 Mbits/sec
[ 7]         304.00-305.00 sec       48.1 MBytes           403 Mbits/sec
[SUM]        304.00-305.00 sec       92.0 MBytes           772 Mbits/sec
```

-t – duration
-p = port
-l = buffer length
-P = streams

[1] Test executed immediately after credits consumed

Test : iperf3 on c7i.large (baseline)



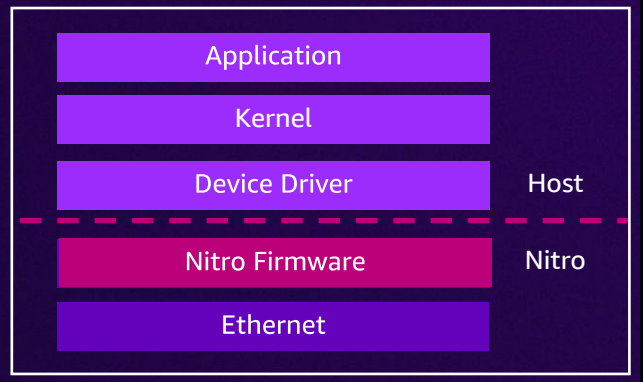
Two Flows, Two Queues [1]

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 600 -p 5201 -l 9000 -P 2
Connecting to host app-server7, port 5201
[  5] local web-server port 47410 connected to app-server port 5201
[  7] local web-server port 47424 connected to app-server port 5201
[ ID]        Interval                Transfer              Bitrate
[ 5]         304.00-305.00  sec      43.9 MBytes           368 Mbits/sec
[ 7]         304.00-305.00 sec       48.1 MBytes           403 Mbits/sec
[SUM]        304.00-305.00 sec       92.0 MBytes           772 Mbits/sec
```

-t – duration
-p = port
-l = buffer length
-P = streams

[1] Test executed immediately after credits consumed

Test : iperf3 on c7i.large (baseline)



Two Flows, Two Queues [1]

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 600 -p 5201 -l 9000 -P 2
Connecting to host app-server7, port 5201
[  5] local web-server port 47410 connected to app-server port 5201
[  7] local web-server port 47424 connected to app-server port 5201
[ ID]        Interval                Transfer              Bitrate
[ 5]         304.00-305.00  sec      43.9 MBytes           368 Mbits/sec
[ 7]         304.00-305.00 sec       48.1 MBytes           403 Mbits/sec
[SUM]        304.00-305.00 sec       92.0 MBytes           772 Mbits/sec
```

-t – duration
-p = port
-l = buffer length
-P = streams

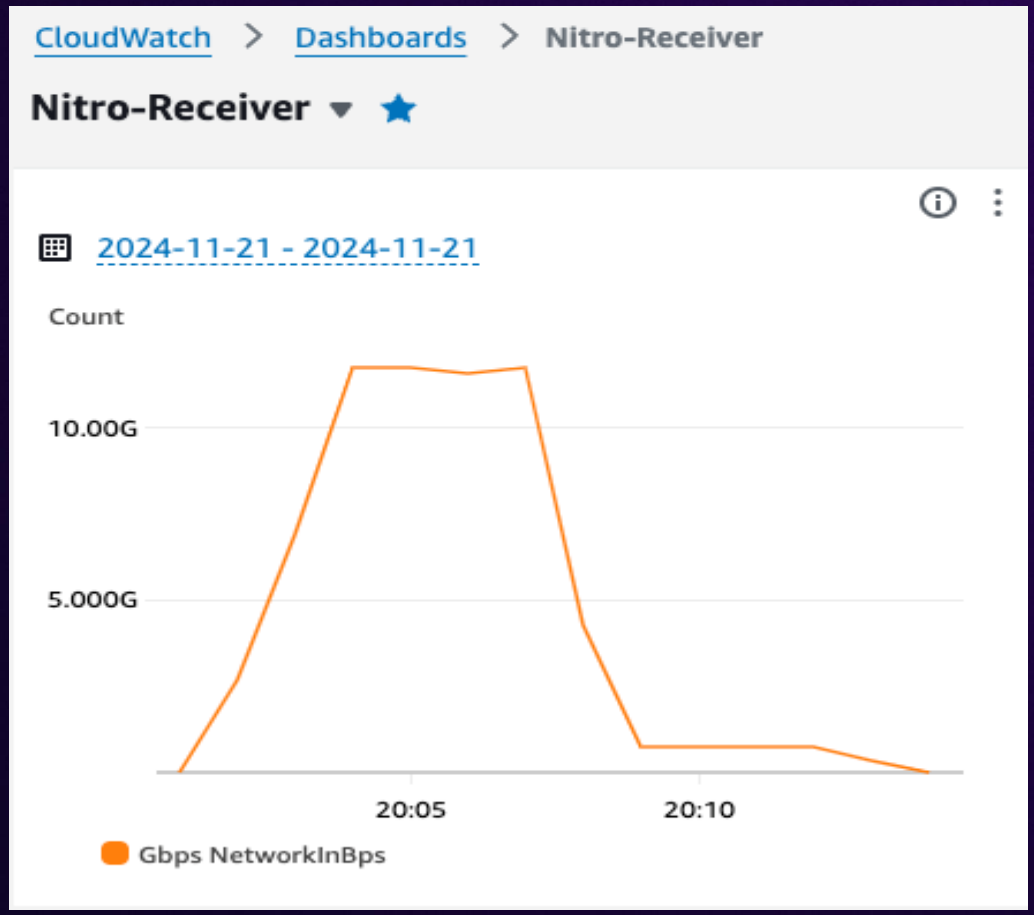
Enhanced Network Metrics
- bandwidth_allowance_in
- bandwidth_allowance_out

Instance Aggregate Baseline Flow Limit

[1] Test executed immediately after credits consumed

Test Method: iperf3 on c7i.large

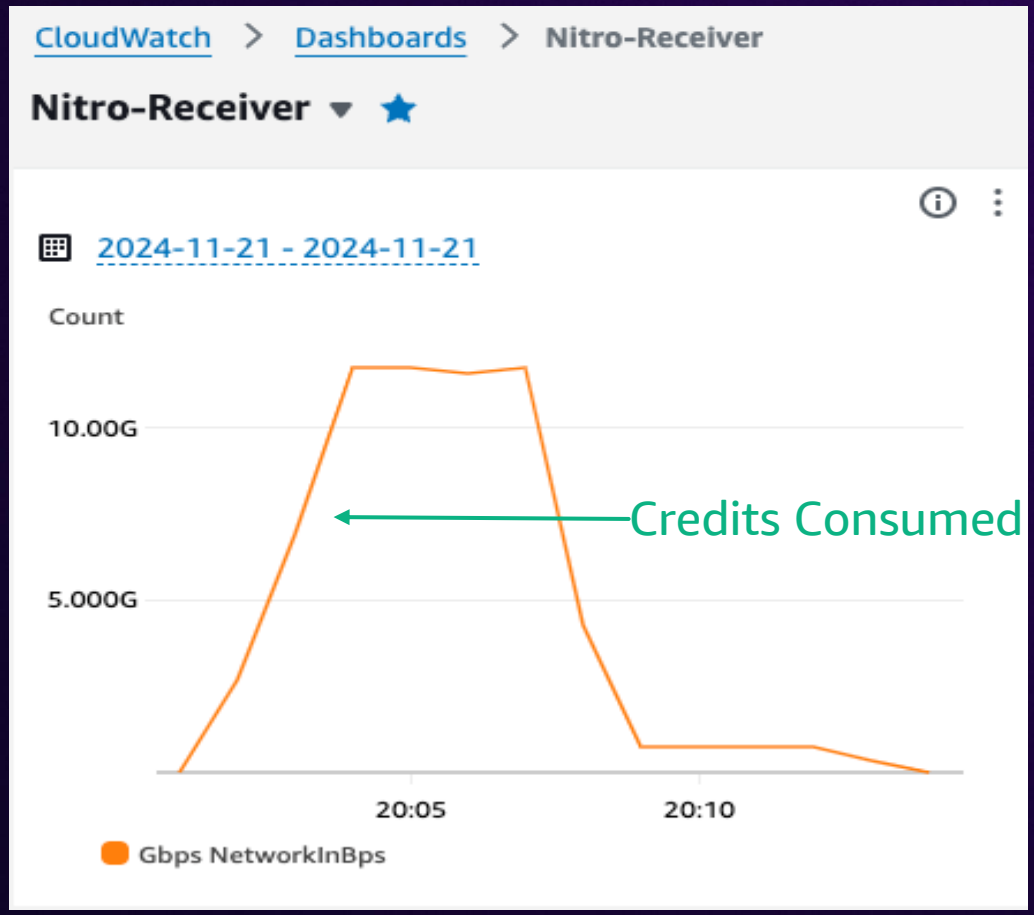
- Tool: iperf3
- Instance c7i.large
  - Burst 12.5 Gbps, Baseline 781 Mbps
- Transmit to Receiver
- Large MTU to hit Burst Limit
- Time Period to Consume Credits
- Burst until Credit Exhaustion
- Rate Limited to Baseline



CLIENT: iperf3 -c <receiver_ip> -t 600 -l 9000 –P 1
RECEIVER: iperf3 -s

Test Method: iperf3 on c7i.large

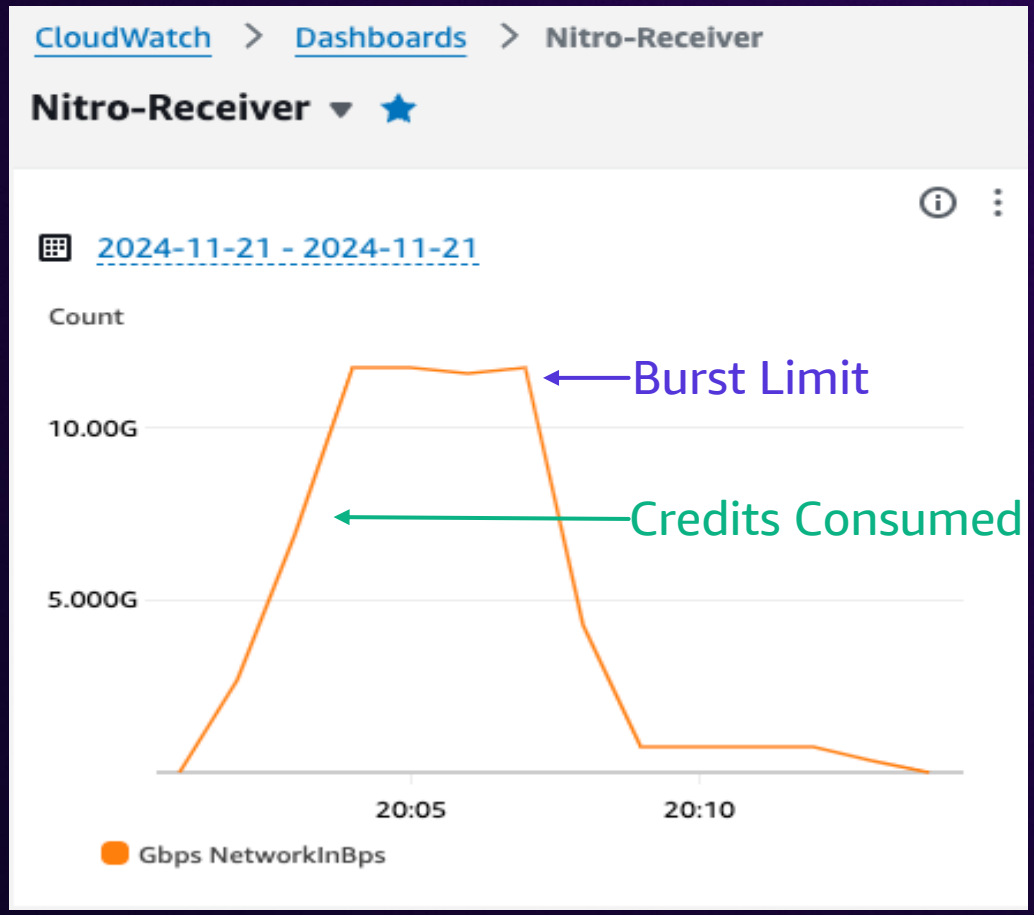
- Tool: iperf3
- Instance c7i.large
  - Burst 12.5 Gbps, Baseline 781 Mbps
- Transmit to Receiver
- Large MTU to hit Burst Limit
- Time Period to Consume Credits
- Burst until Credit Exhaustion
- Rate Limited to Baseline



CLIENT: iperf3 -c <receiver_ip> -t 600 -l 9000 –P 1
RECEIVER: iperf3 -s

Test Method: iperf3 on c7i.large

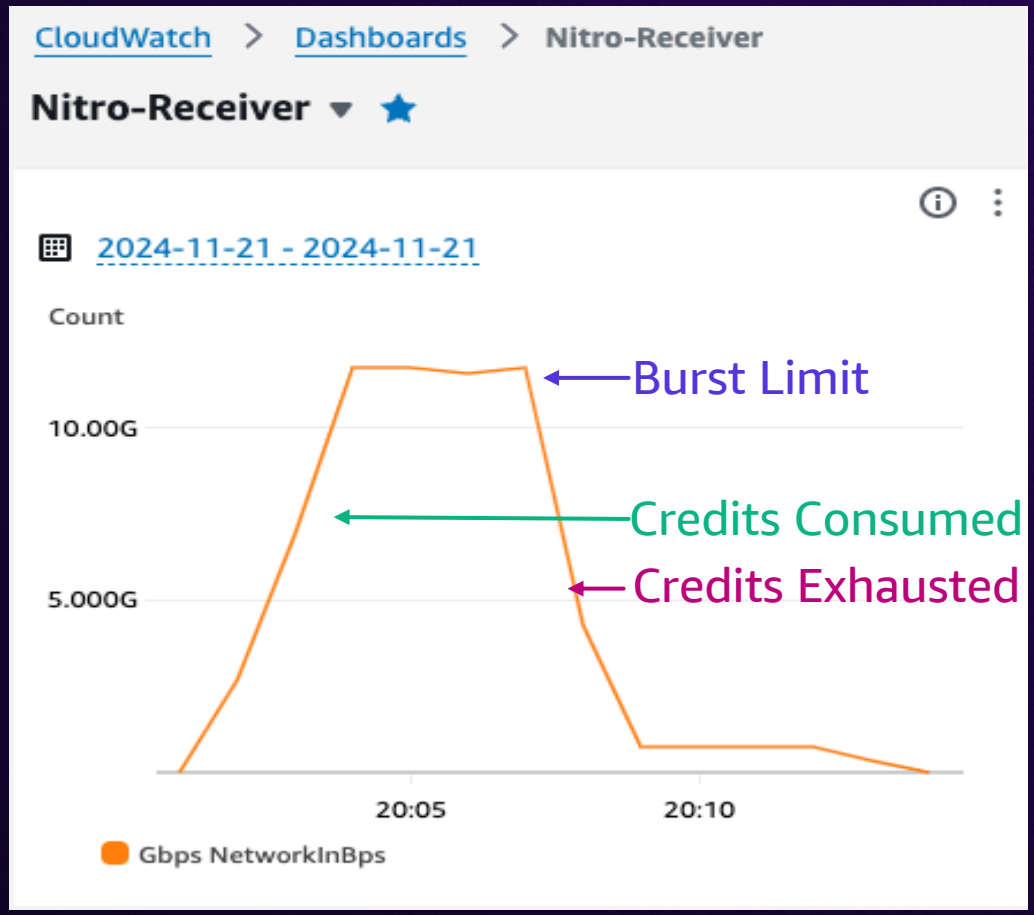
- Tool: iperf3
- Instance c7i.large
  - Burst 12.5 Gbps, Baseline 781 Mbps
- Transmit to Receiver
- Large MTU to hit Burst Limit
- Time Period to Consume Credits
- Burst until Credit Exhaustion
- Rate Limited to Baseline



CLIENT: iperf3 -c <receiver_ip> -t 600 -l 9000 –P 1
RECEIVER: iperf3 -s

Test Method: iperf3 on c7i.large

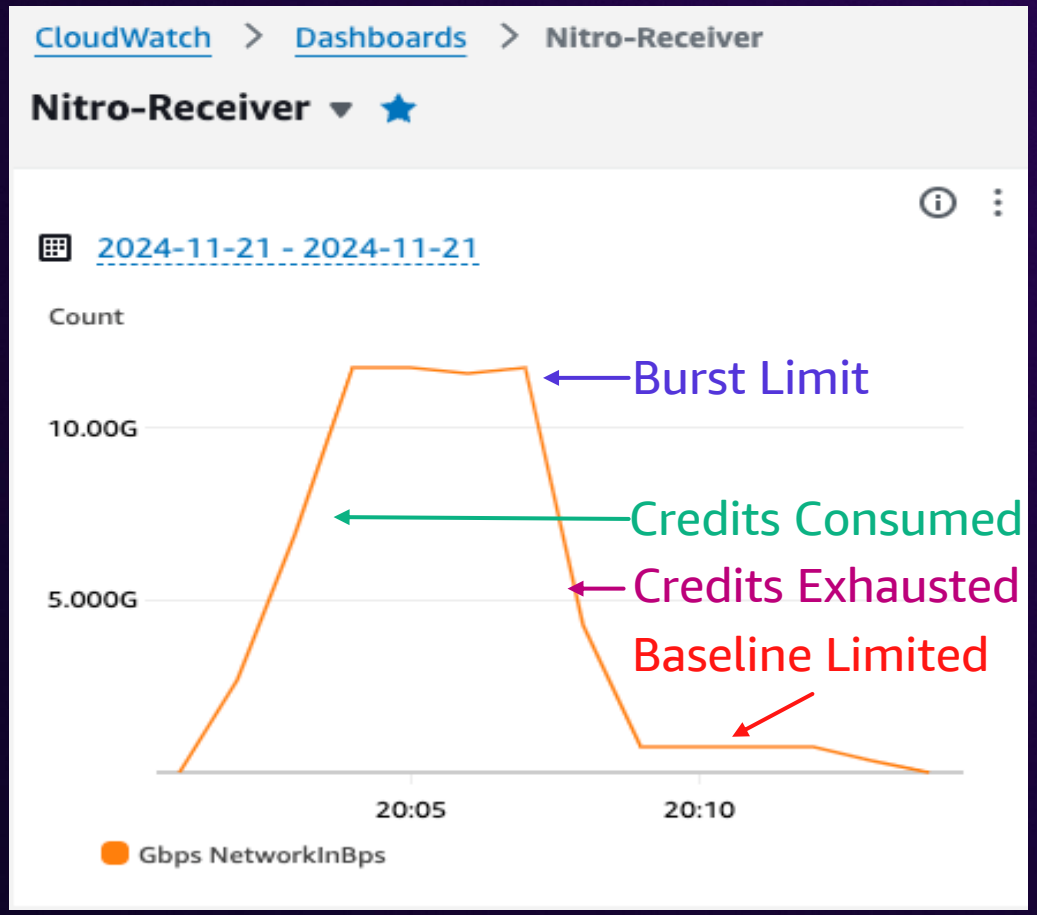
- Tool: iperf3
- Instance c7i.large
  - Burst 12.5 Gbps, Baseline 781 Mbps
- Transmit to Receiver
- Large MTU to hit Burst Limit
- Time Period to Consume Credits
- Burst until Credit Exhaustion
- Rate Limited to Baseline



CLIENT: iperf3 -c <receiver_ip> -t 600 -l 9000 –P 1
RECEIVER: iperf3 -s

Test Method: iperf3 on c7i.large

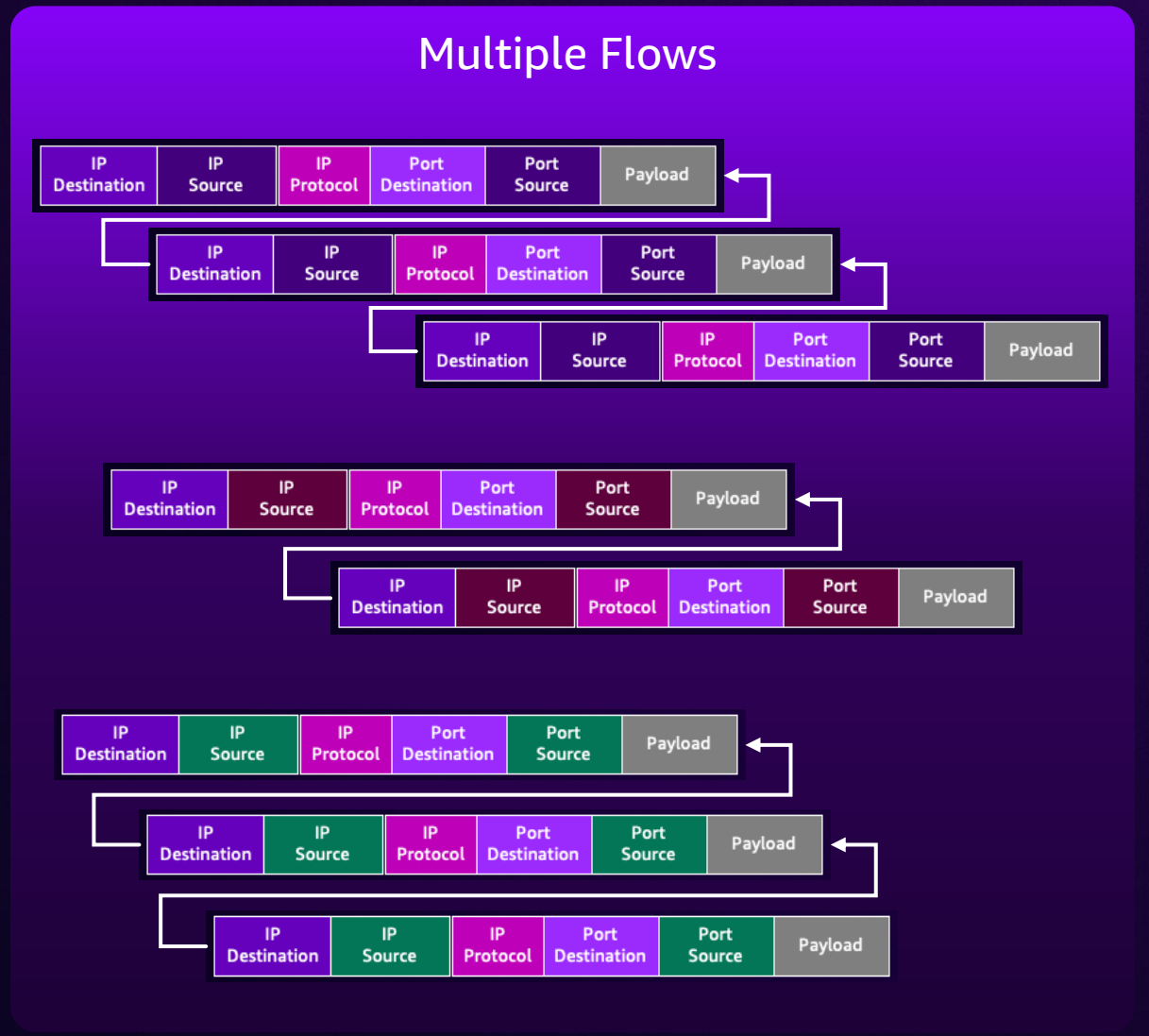
- Tool: iperf3
- Instance c7i.large
  - Burst 12.5 Gbps, Baseline 781 Mbps
- Transmit to Receiver
- Large MTU to hit Burst Limit
- Time Period to Consume Credits
- Burst until Credit Exhaustion
- Rate Limited to Baseline



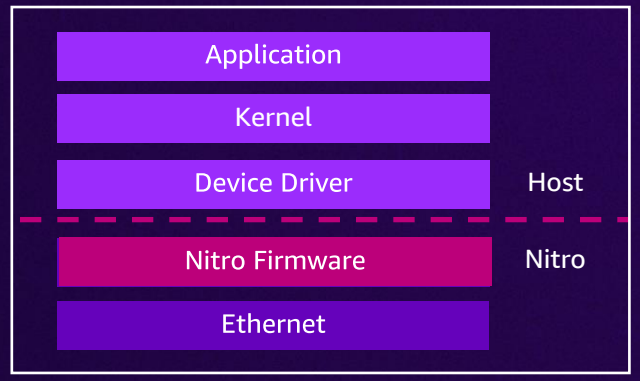
CLIENT: iperf3 -c <receiver_ip> -t 600 -l 9000 –P 1
RECEIVER: iperf3 -s

Multi-flow Analysis

- Number of Flows
- Baseline Bandwidth
- Burst Bandwidth
- Packets Per Second
- Shared Nitro Resources

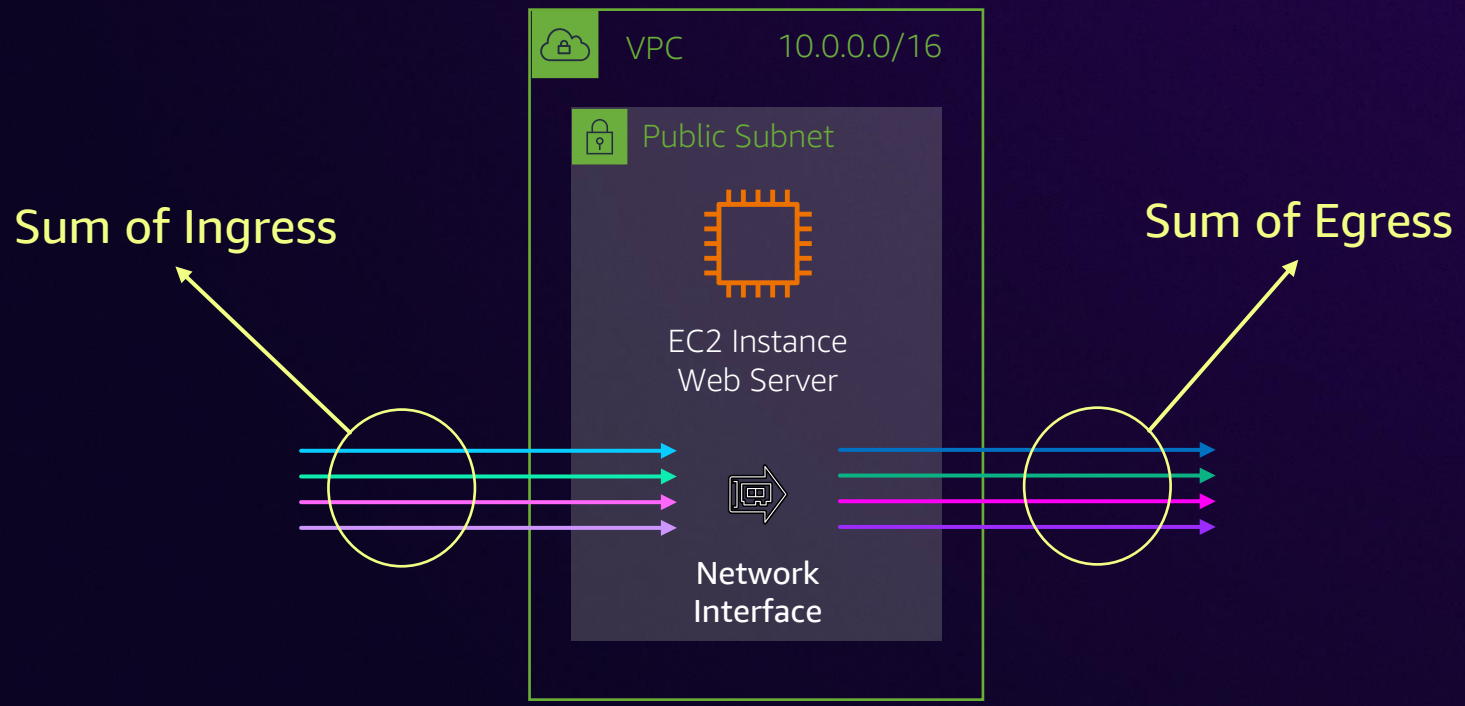


Instance Packets Per Second Targets



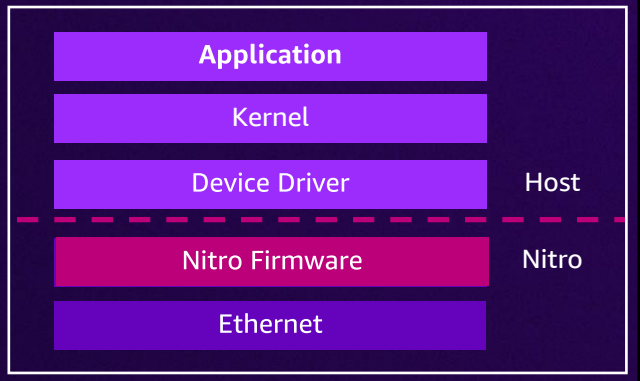
Target Packets Per Second

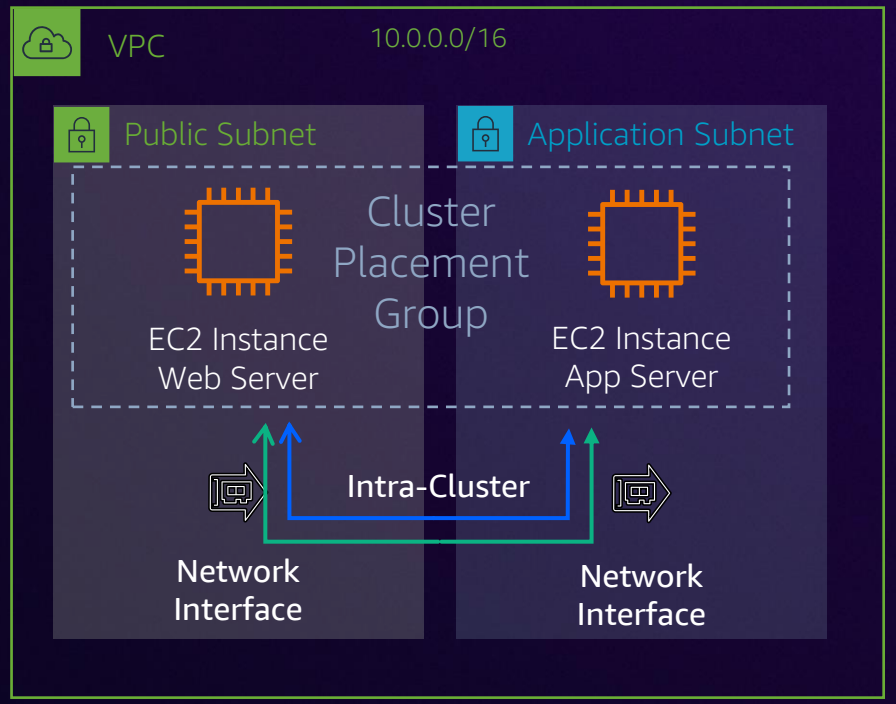
Flow Packets Per Second



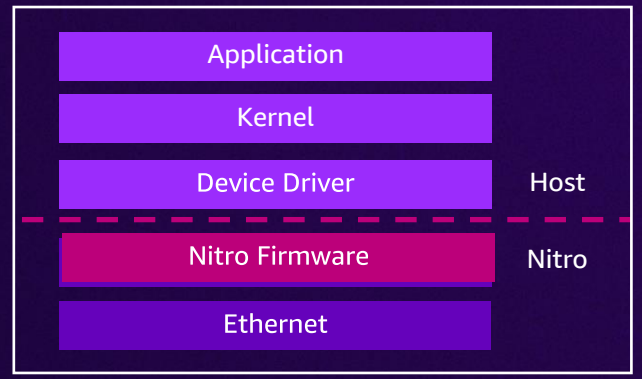
Multi-Flow Test Topology

- Tool: iperf3
- Instance c7i.large
- Transmit to Receiver
- Small MTU
  - Instance Limit
  - Flow Limits
- Objective
  - Aggregate PPS Across Queues
  - PPS Limit per Queue





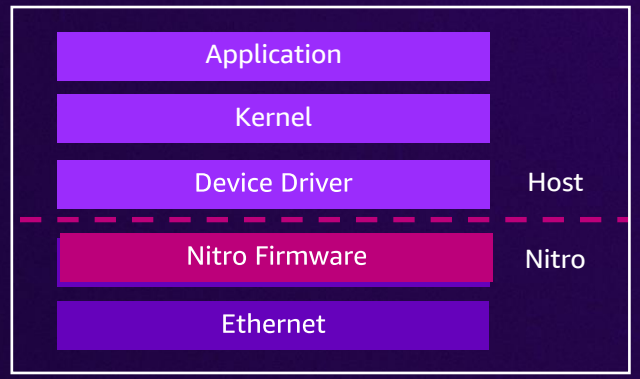
Test : iperf3 on c7i.large (pps)



One Flows, One Queue

```
[ec2-user@web-server ~]$iperf3 -c app-server -t 10 -p 5201 -M 128 -P 1
[  5] local web-server port 58284 connected to app-server port 5201
[ ID]        Interval                Transfer                Bitrate
[  5]        9.00-10.00  sec         111 MBytes              929 Mbits/sec
[SUM]        0.00-10.00   sec        1.08 Gbytes             928 Mbits/sec
```

Test : iperf3 on c7i.large (pps)

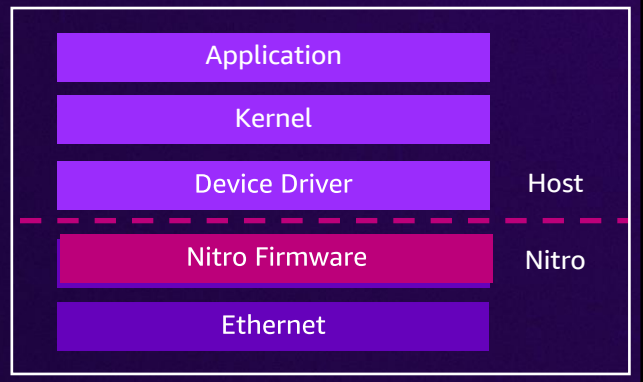


One Flows, One Queue

```
[ec2-user@web-server ~]$iperf3 -c app-server -t 10 -p 5201 -M 128 -P 1
[  5] local web-server port 58284 connected to app-server port 5201
[ ID]        Interval                Transfer                Bitrate
[  5]        9.00-10.00  sec        111 MBytes              929 Mbits/sec
[SUM]        0.00-10.00   sec       1.08 Gbytes             928 Mbits/sec
```

-t – duration
-p = port
-M = max segment
-P = streams

Test : iperf3 on c7i.large (pps)



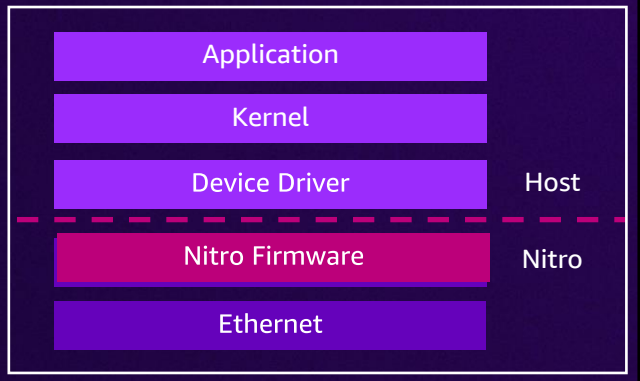
One Flows, One Queue

```
[ec2-user@web-server ~]$iperf3 -c app-server -t 10 -p 5201 -M 128 -P 1
[  5] local web-server port 58284 connected to app-server port 5201
[ ID]        Interval              Transfer          Bitrate
[  5]        9.00-10.00  sec       111 MBytes        929 Mbits/sec
[SUM]        0.00-10.00   sec      1.08 Gbytes       928 Mbits/sec
```

-t – duration
-p = port
-M = max segment
-P = streams

Queue
PPS Flow Limit

Test : iperf3 on c7i.large (pps)



One Flows, One Queue

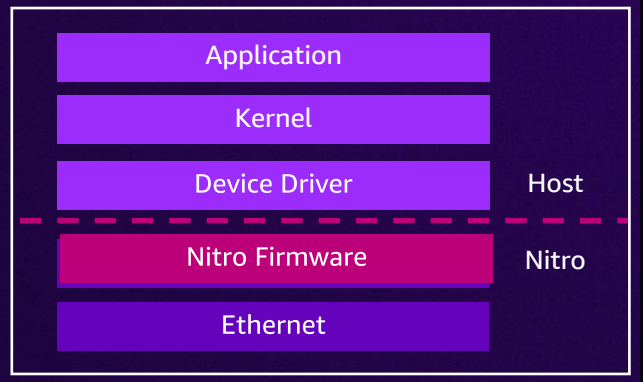
```
[ec2-user@web-server ~]$iperf3 -c app-server -t 10 -p 5201 -M 128 -P 1
[  5] local web-server port 58284 connected to app-server port 5201
[ ID]        Interval                Transfer                Bitrate
[  5]        9.00-10.00  sec         111 MBytes              929 Mbits/sec
[SUM]        0.00-10.00   sec        1.08 Gbytes             928 Mbits/sec
```

-t – duration
-p = port
-M = max segment
-P = streams

0.928 Gbps
/ 8 b/B / 128 B/pkt =
972 kpps

Queue
PPS Flow Limit

Test : iperf3 on c7i.large (pps)



One Flows, One Queue

```
[ec2-user@web-server ~]$iperf3 -c app-server -t 10 -p 5201 -M 128 -P 1
[  5] local web-server port 58284 connected to app-server port 5201
[ ID]        Interval                Transfer                Bitrate
[  5]        9.00-10.00  sec         111 MBytes              929 Mbits/sec
[SUM]        0.00-10.00   sec        1.08 Gbytes             928 Mbits/sec
```

-t – duration
-p = port
-M = max segment
-P = streams

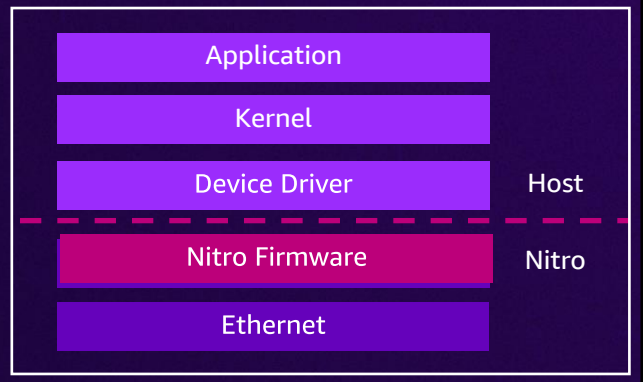
0.928 Gbps
/ 8 b/B / 128 B/pkt =
972 kpps

Queue
PPS Flow Limit

Two Flows, Two Queues

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 10 -p 5201 –M 128-P 2
[ ID]        Interval                Transfer                Bitrate
[  5]        9.00-10.00   sec        54.8 MBytes             471 Mbits/sec
[  7]        9.00-10.00   sec        56.1 MBytes             477 Mbits/sec
[SUM]        0.00-10.00   sec        1.08 GBytes             929 Mbits/sec

iperf Done.
```

Test : iperf3 on c7i.large (pps)



One Flows, One Queue

```
[ec2-user@web-server ~]$iperf3 -c app-server -t 10 -p 5201 -M 128 -P 1
[  5] local web-server port 58284 connected to app-server port 5201
[ ID]        Interval                Transfer              Bitrate
[  5]        9.00-10.00  sec        111 MBytes            929 Mbits/sec
[SUM]        0.00-10.00   sec       1.08 Gbytes           928 Mbits/sec
```

-t – duration
-p = port
-M = max segment
-P = streams

0.928 Gbps
/ 8 b/B / 128 B/pkt =
972 kpps

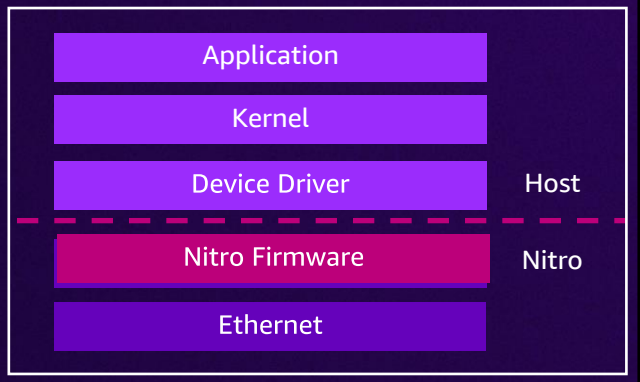
Queue
PPS Flow Limit

Two Flows, Two Queues

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 10 -p 5201 –M 128-P 2
[ ID]        Interval                Transfer              Bitrate
[  5]        9.00-10.00   sec        54.8 MBytes           471 Mbits/sec
[  7]        9.00-10.00   sec        56.1 MBytes           477 Mbits/sec
[SUM]        0.00-10.00   sec        1.08 GBytes           929 Mbits/sec

iperf Done.
```

Instance Aggregate
PPS Flow Limit

Test : iperf3 on c7i.large (pps)



One Flows, One Queue

```
[ec2-user@web-server ~]$iperf3 -c app-server -t 10 -p 5201 -M 128 -P 1
[  5] local web-server port 58284 connected to app-server port 5201
[ ID]        Interval                Transfer                Bitrate
[  5]        9.00-10.00  sec         111 MBytes              929 Mbits/sec
[SUM]        0.00-10.00   sec        1.08 Gbytes             928 Mbits/sec
```

-t – duration
-p = port
-M = max segment
-P = streams

0.928 Gbps / 8 b/B / 128 B/pkt = 972 kpps

Queue PPS Flow Limit

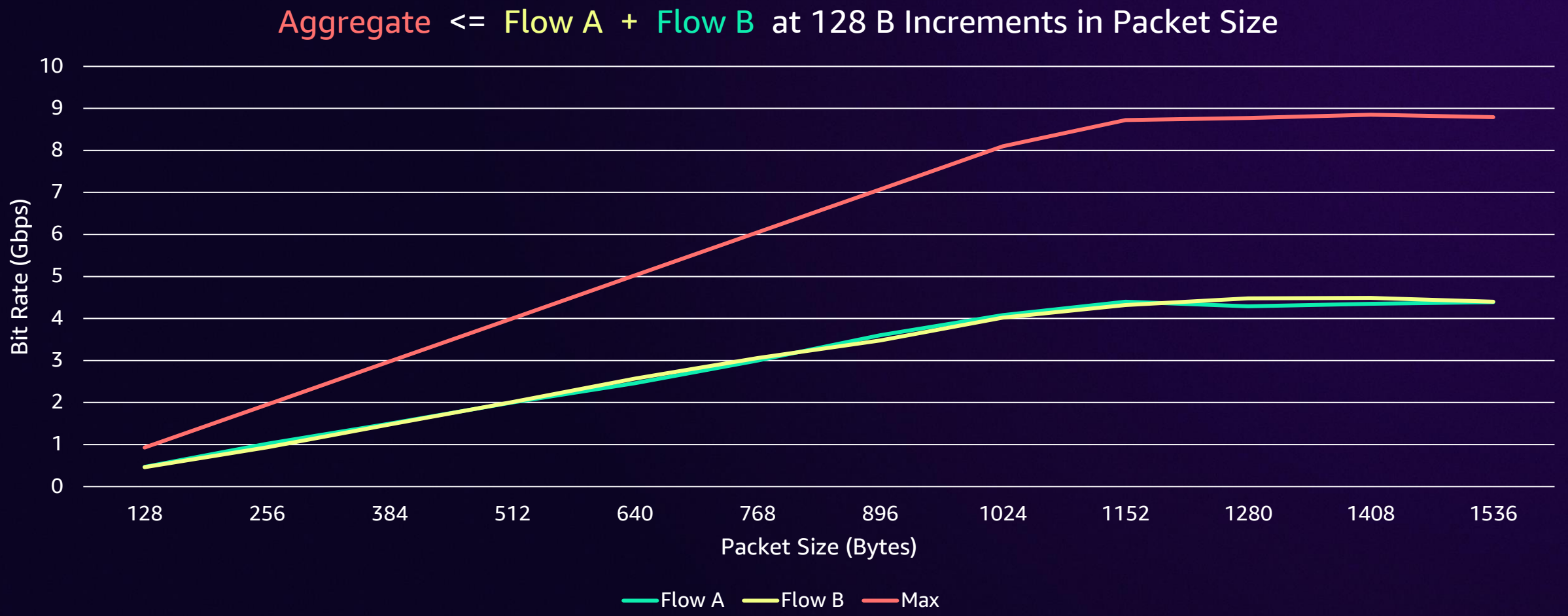
Two Flows, Two Queues

```
[ec2-user@web-server ~]$ iperf3 -c app-server -t 10 -p 5201 –M 128-P 2
[ ID]        Interval                Transfer                Bitrate
[  5]        9.00-10.00   sec        54.8 MBytes             471 Mbits/sec
[  7]        9.00-10.00   sec        56.1 MBytes             477 Mbits/sec
[SUM]        0.00-10.00   sec        1.08 GBytes             929 Mbits/sec

iperf Done.
```

Instance Aggregate PPS Flow Limit

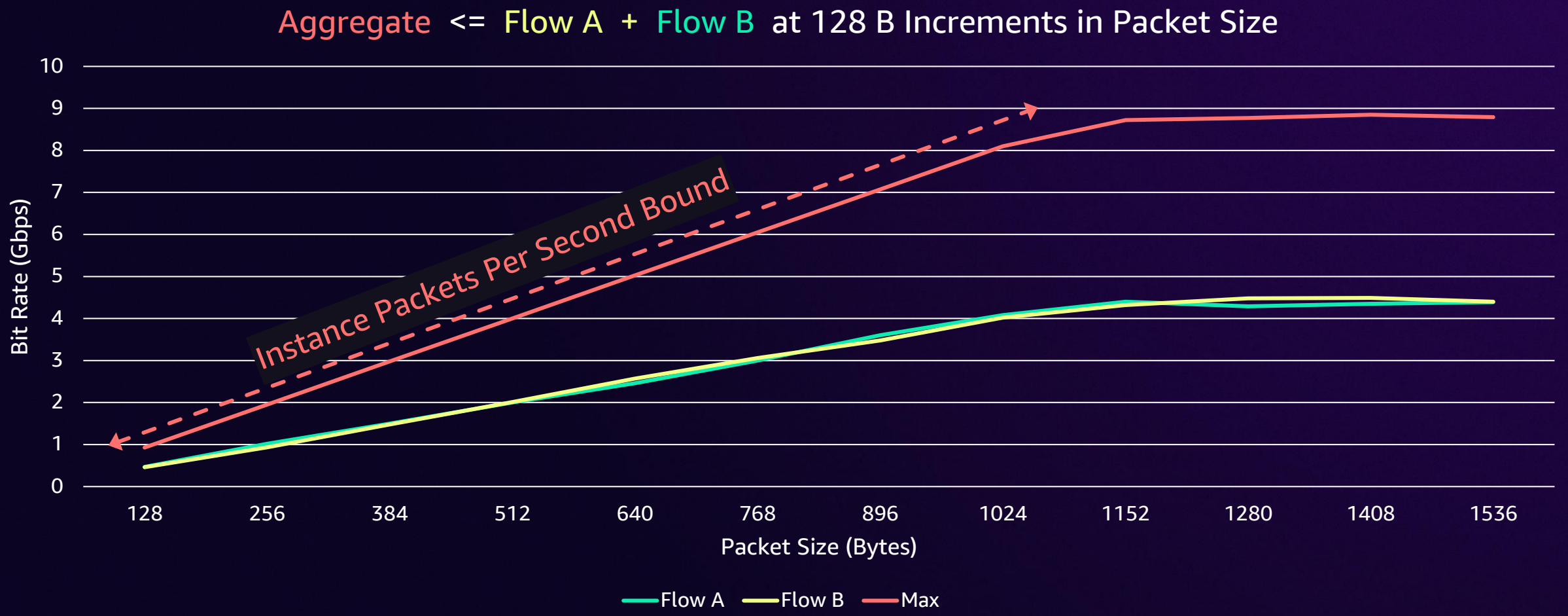
0.929 Gbps / 8 b/B / 128 B/pkt = 972 kpps

Test Method: iperf3 on c7i.large



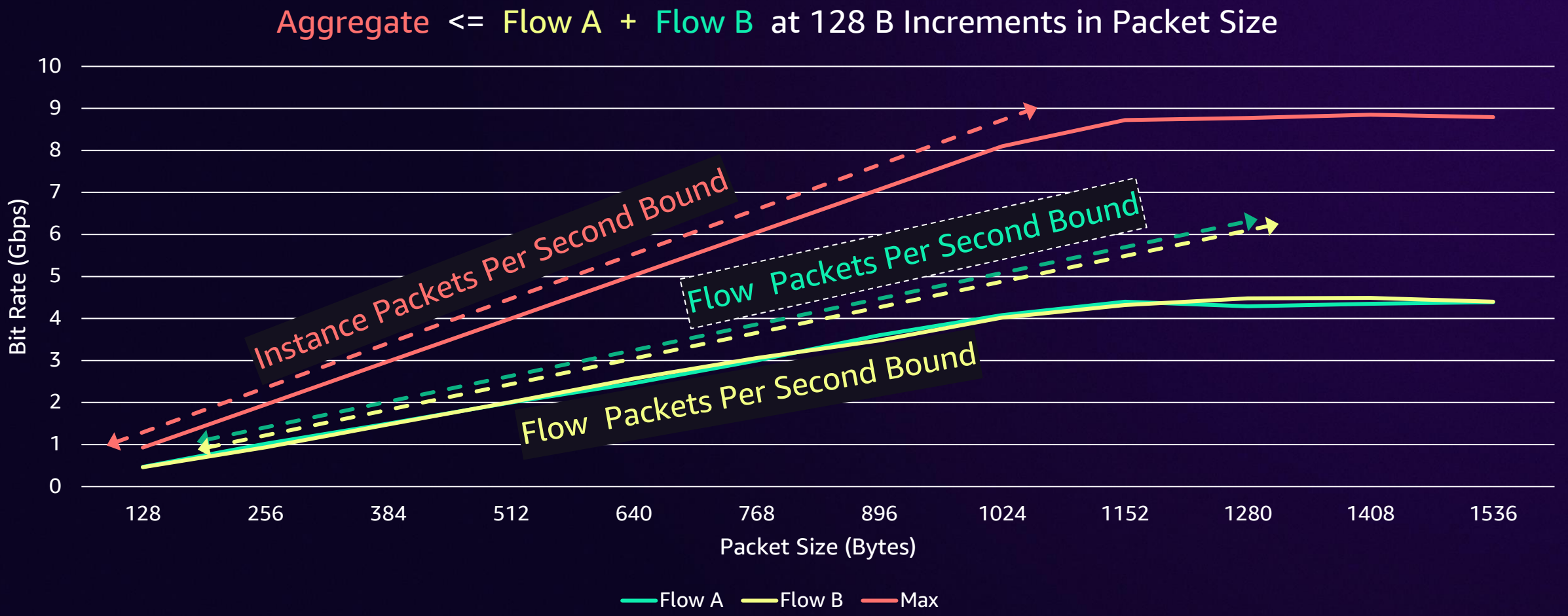
CLIENT: iperf3 -c <receiver> -M <segment_size> -P 2
RECEIVER: iperf3 -s

Test Method: iperf3 on c7i.large



CLIENT: iperf3 -c <receiver> -M <segment_size> -P 2
RECEIVER: iperf3 -s

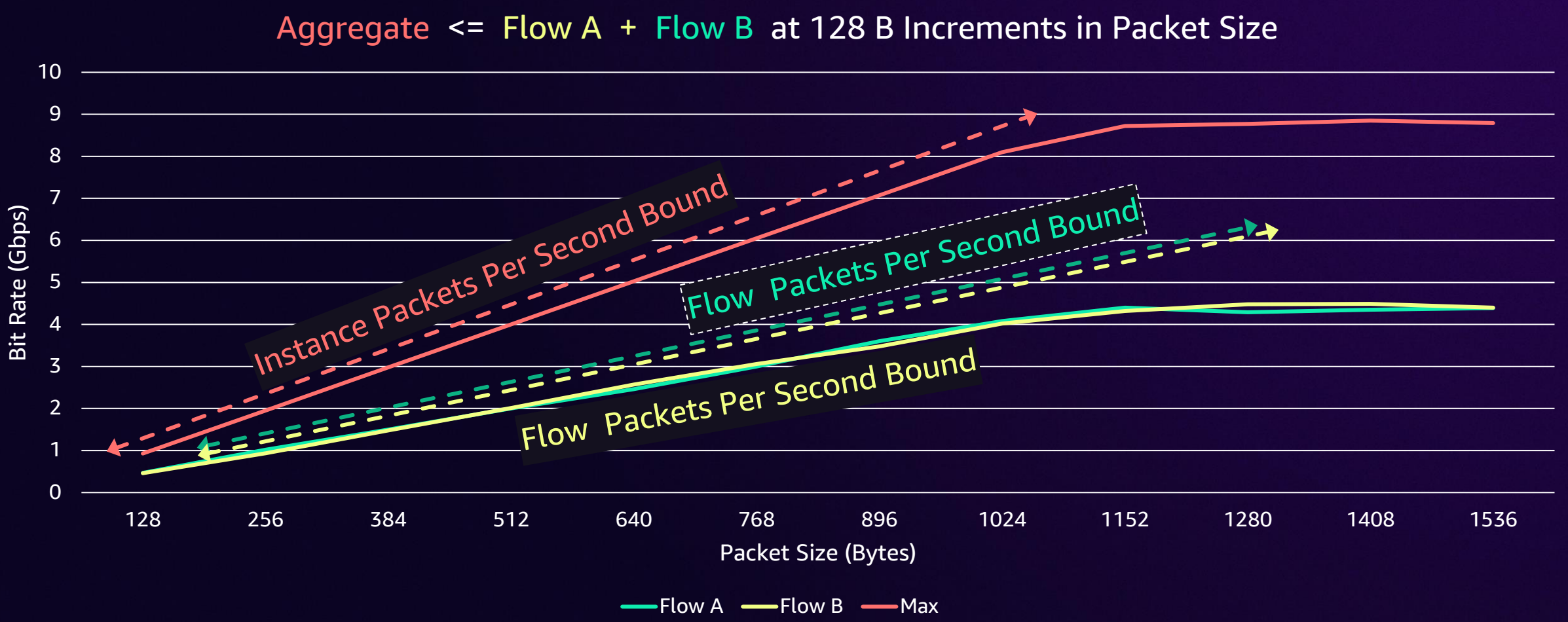
Test Method: iperf3 on c7i.large



CLIENT: iperf3 -c <receiver> -M <segment_size> -P 2
RECEIVER: iperf3 -s

Test Method: iperf3 on c7i.large

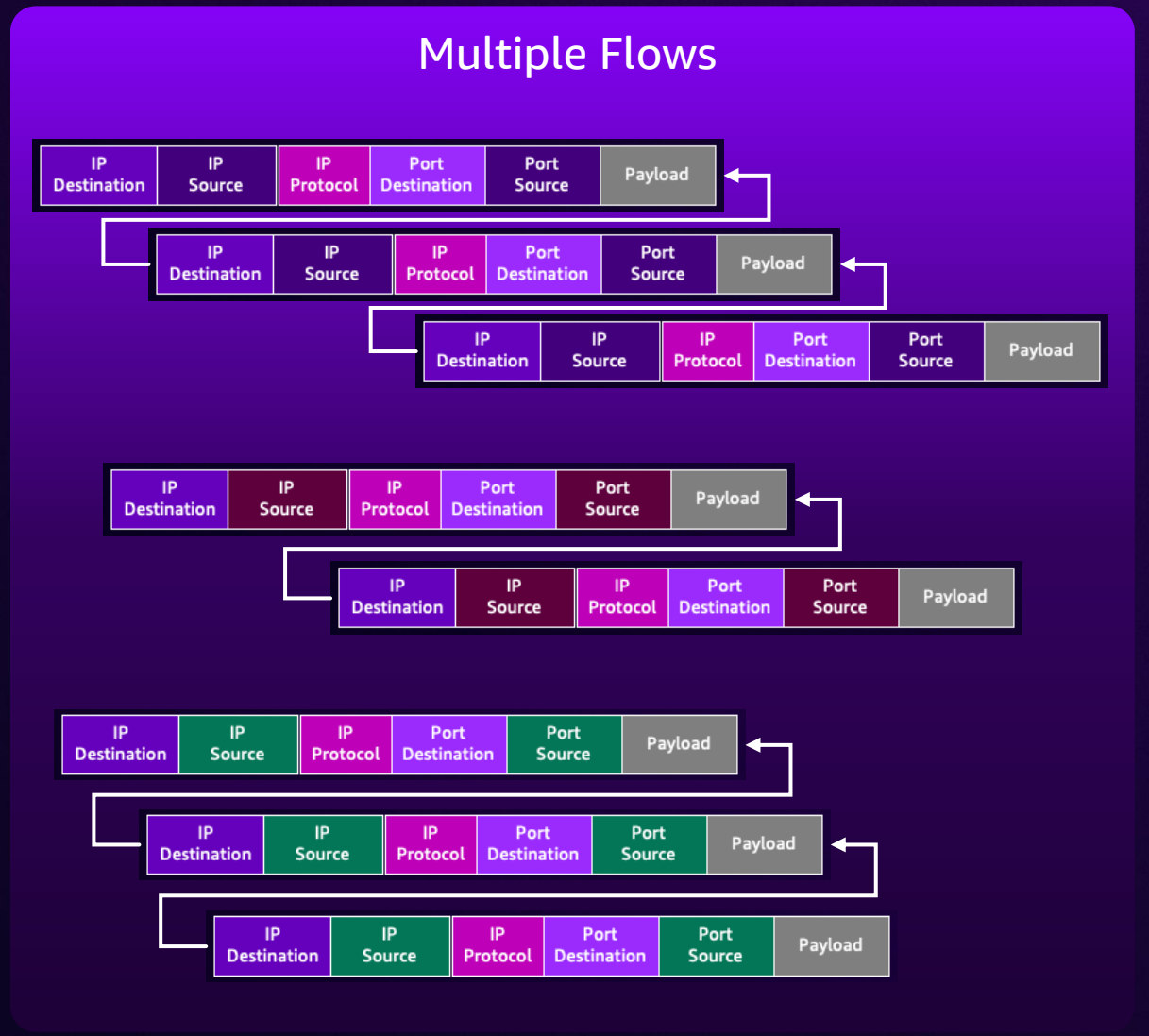
Enhanced ENA Metrics:
- pps_allowance_exceeded



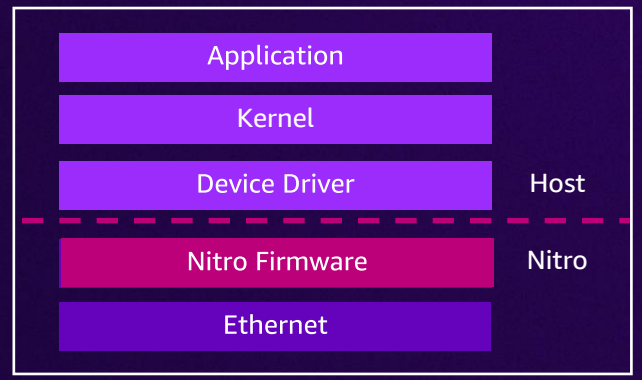
CLIENT: iperf3 -c <receiver> -M <segment_size> -P 2
RECEIVER: iperf3 -s

Multi-flow Analysis

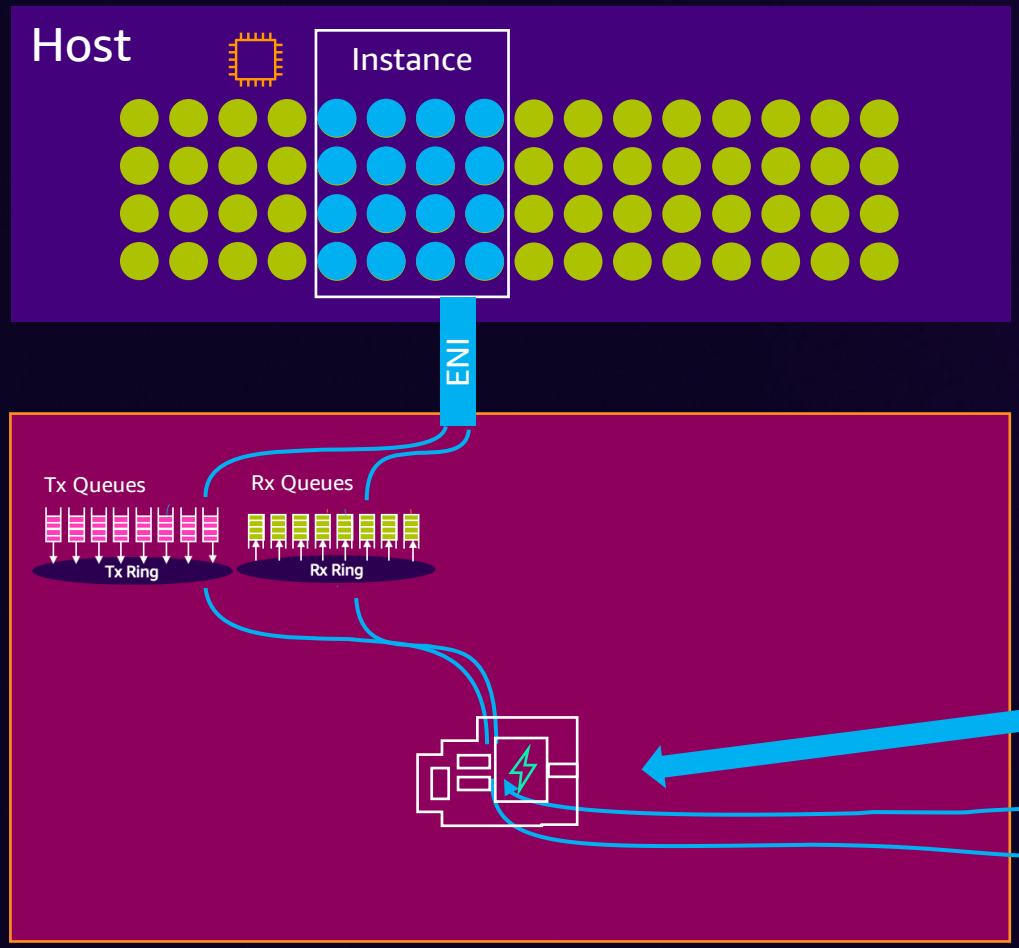
- Number of Flows
- Baseline Bandwidth
- Burst Bandwidth
- Packets Per Second
- Shared Nitro Resources

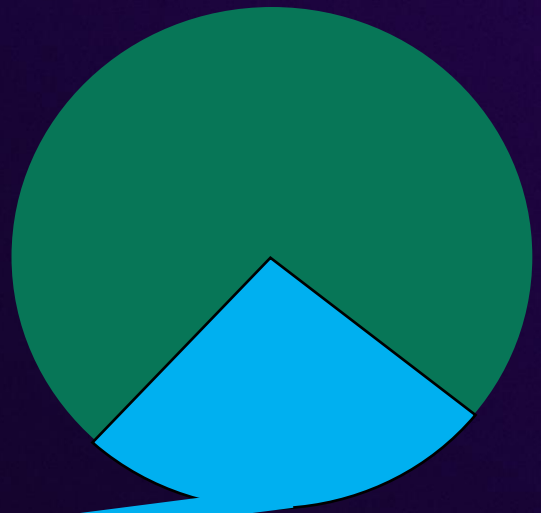


Opportunistic Allocation



Resource Allocation Per Instance

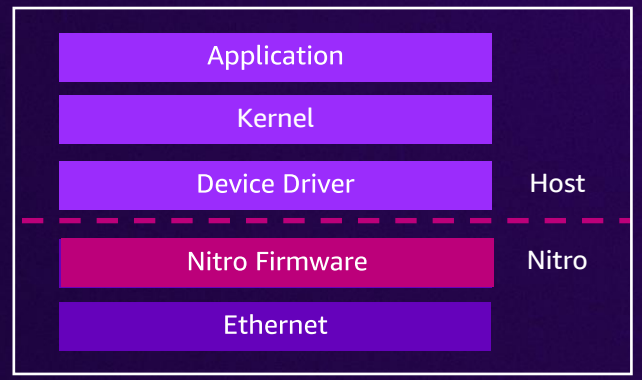




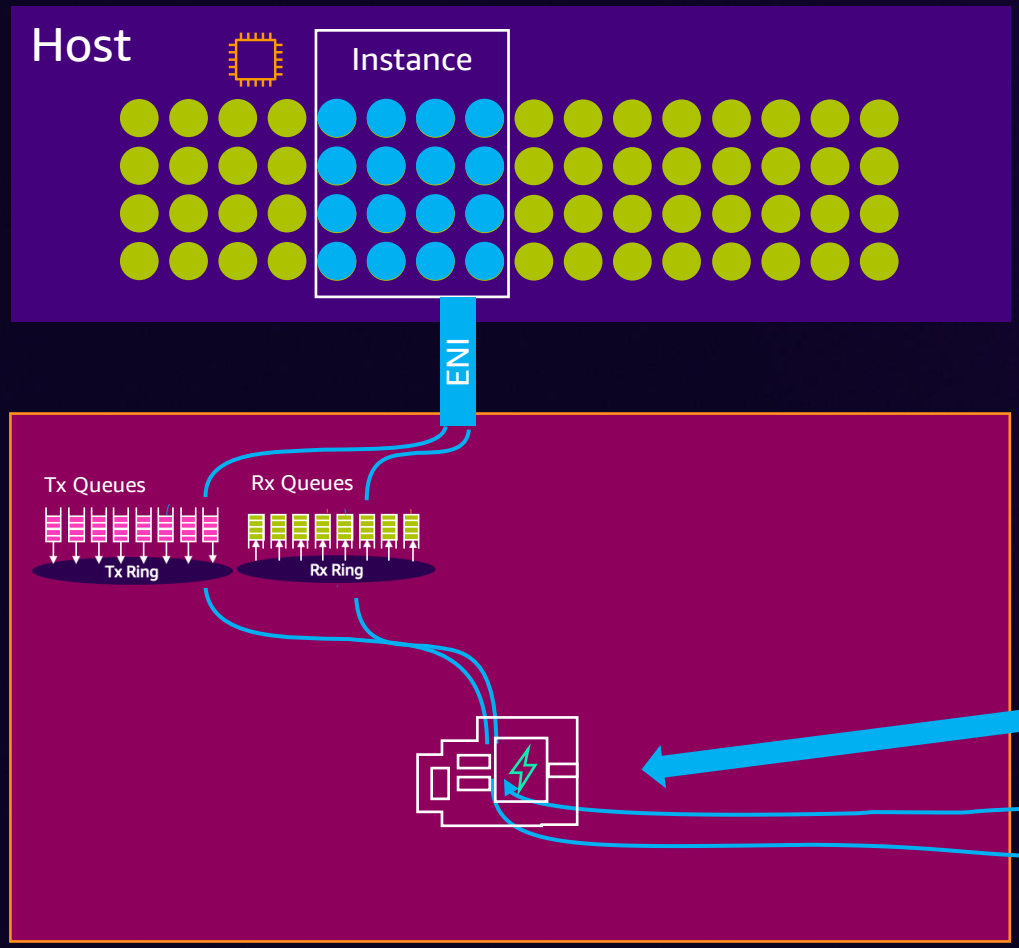
Cycle Time Allotted for Each Nitro Processor

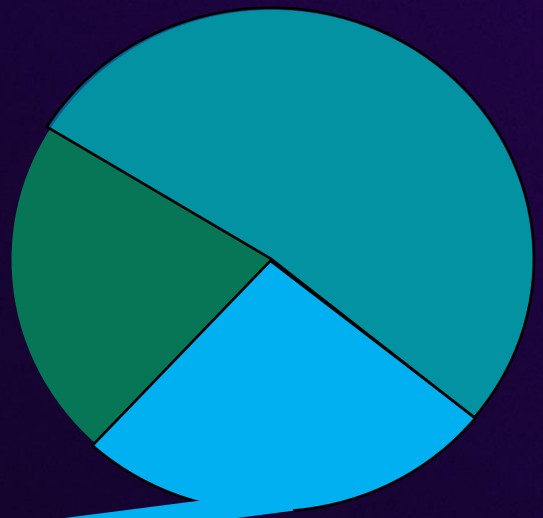
Instance = Percent of Cycle Time

Opportunistic Allocation



Resource Allocation Per Instance

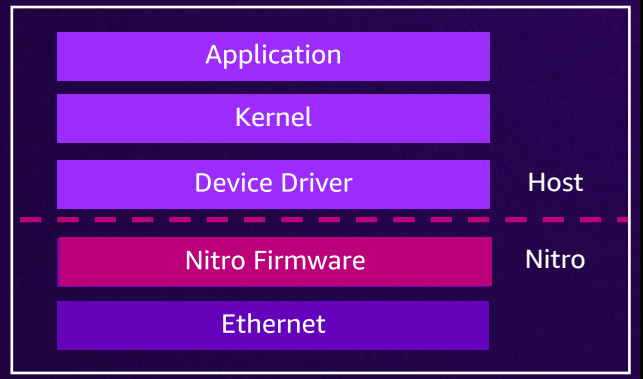




Cycle Time Allotted for Each Nitro Processor

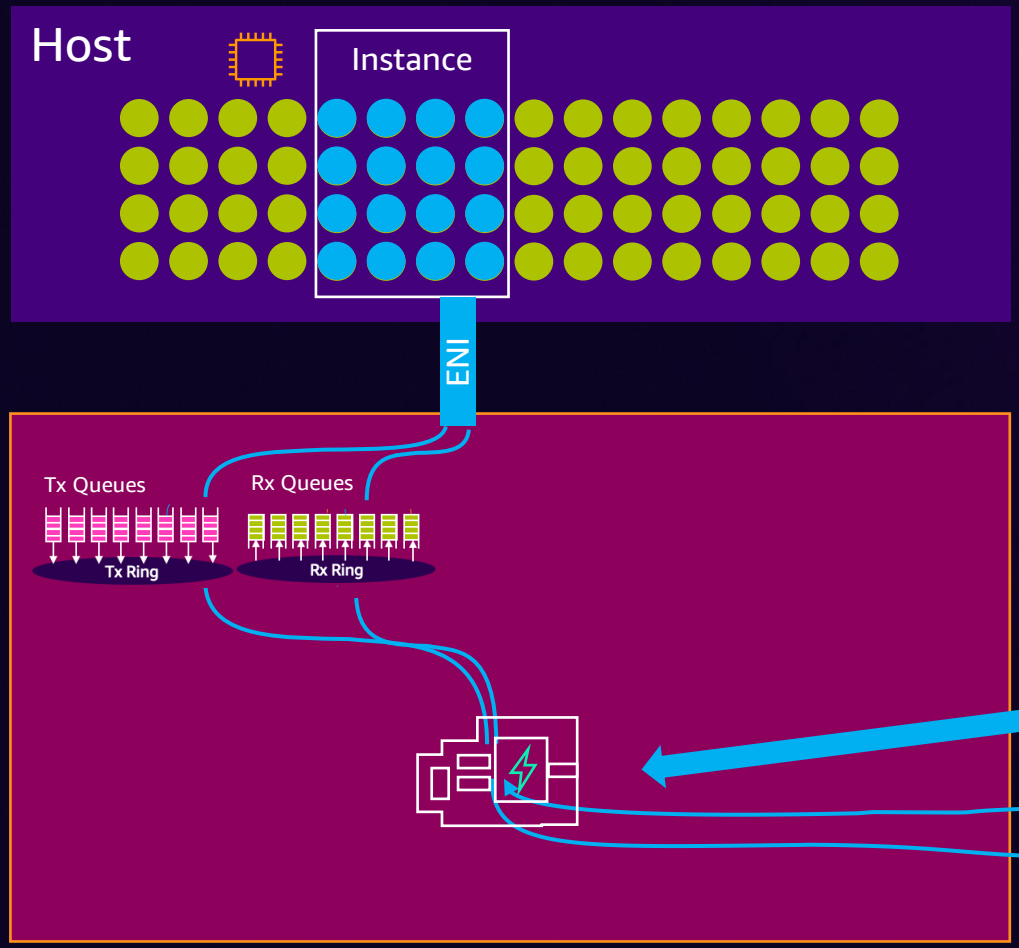
Instance = Percent of Cycle Time

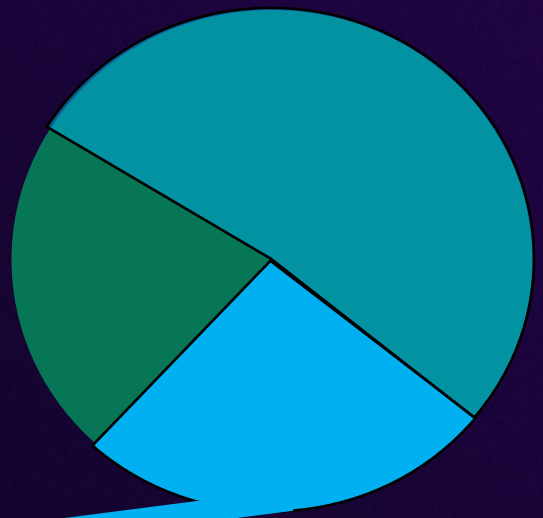
Opportunistic Allocation



Enhanced ENA Metrics:
- pps_allowance_exceeded

Resource Allocation Per Instance



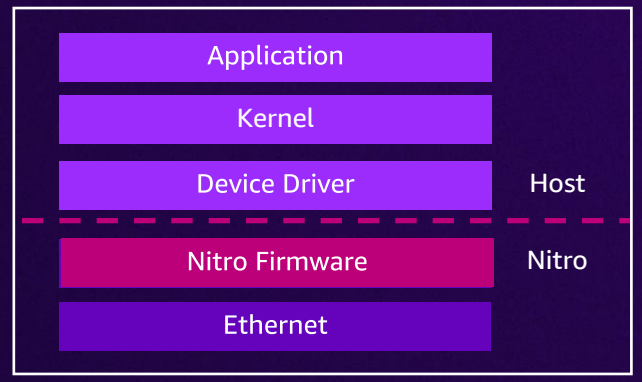


Cycle Time Allotted for Each Nitro Processor

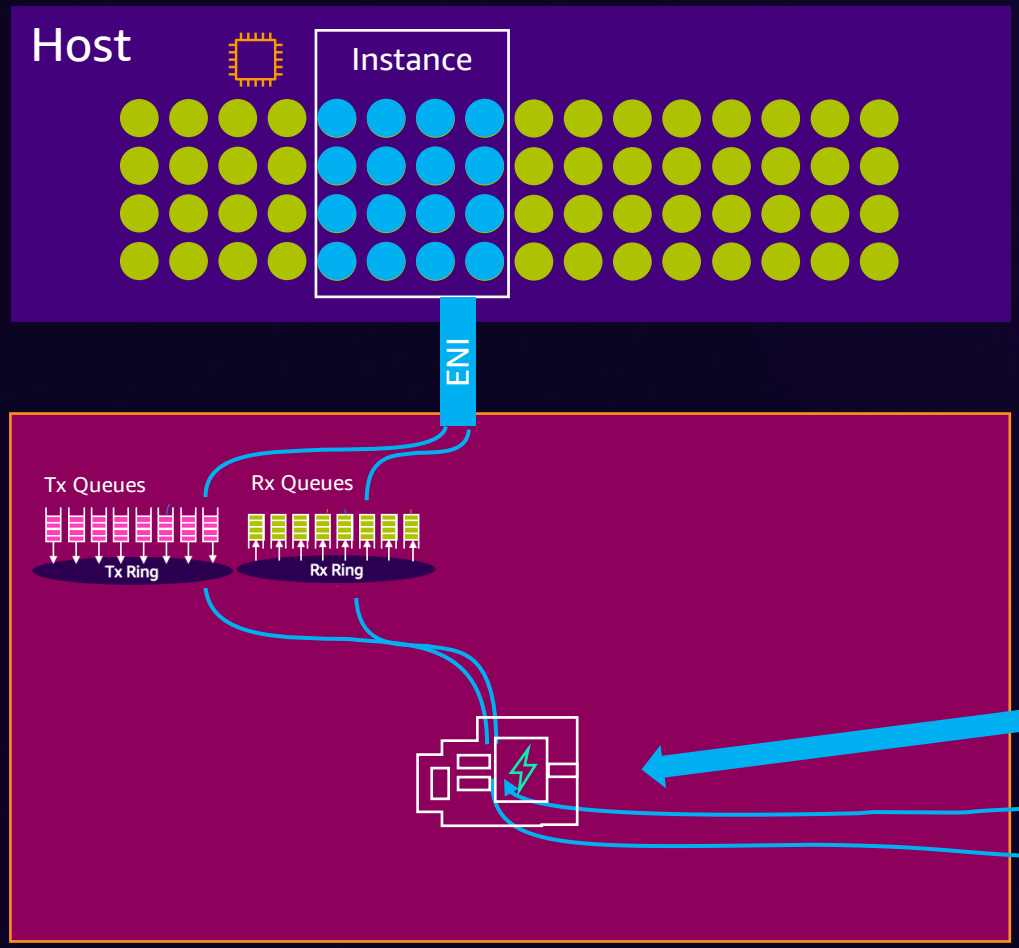
Instance = Percent of Cycle Time

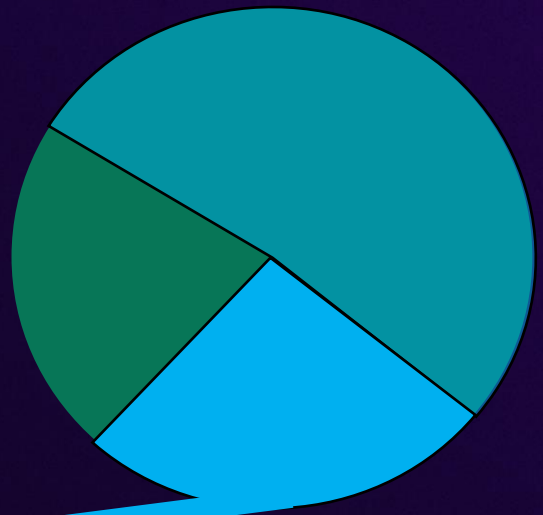
Maximize your throughput:
- Accelerated Flows
- Larger Packets
- Long Lived Flows

Opportunistic Allocation



Resource Allocation Per Instance

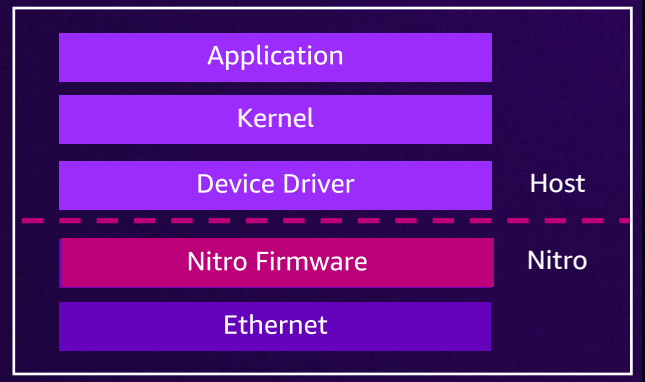




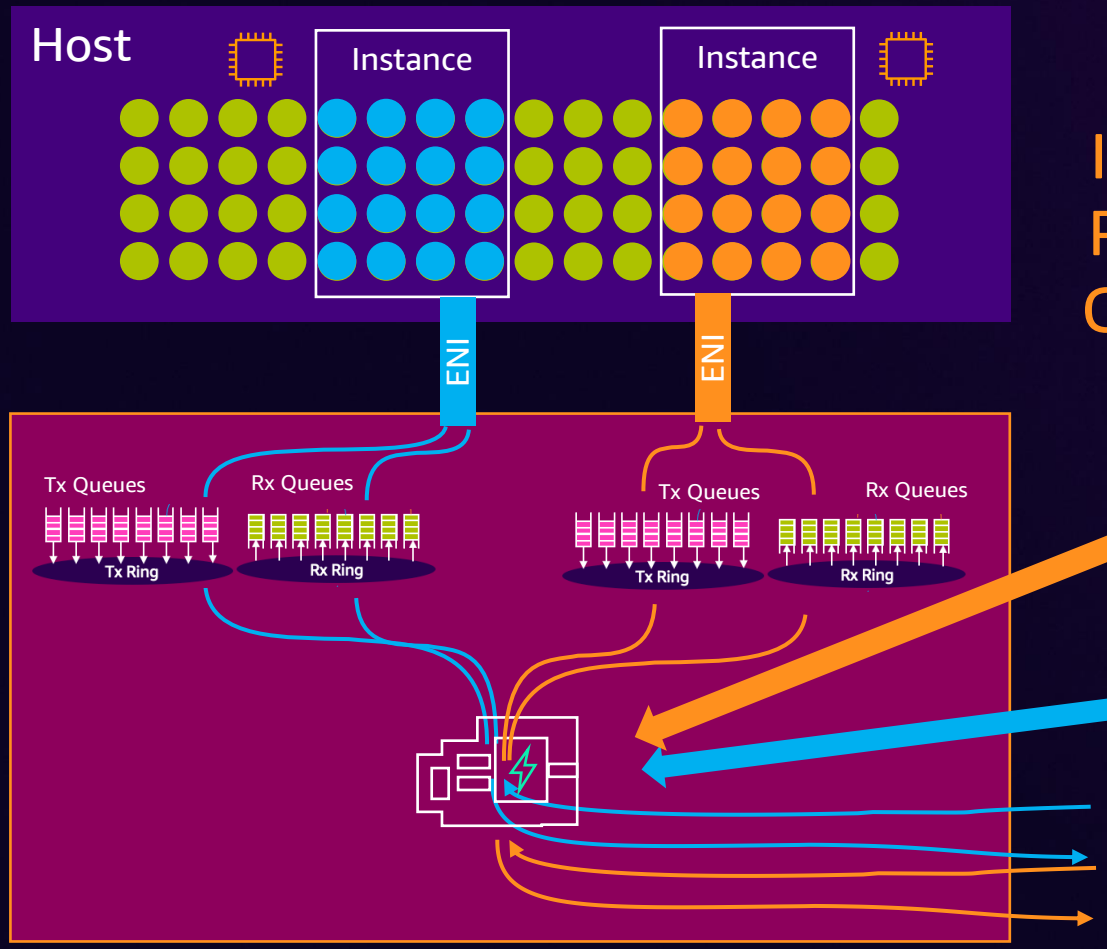
Cycle Time Allotted for Each Nitro Processor

Instance = Percent of Cycle Time

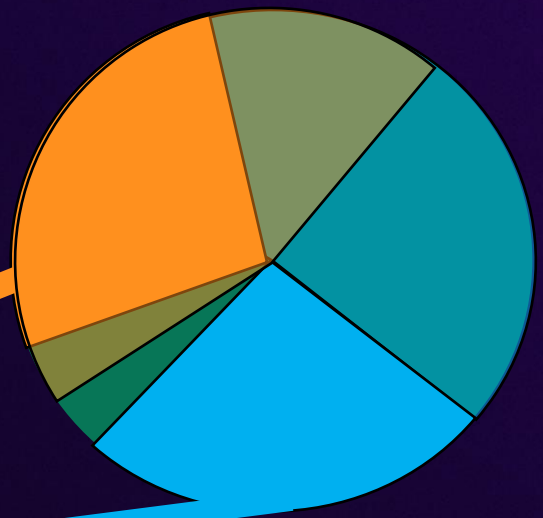
Opportunistic Allocation



Resource Allocation Per Instance



Instance = Percent of Cycle Time

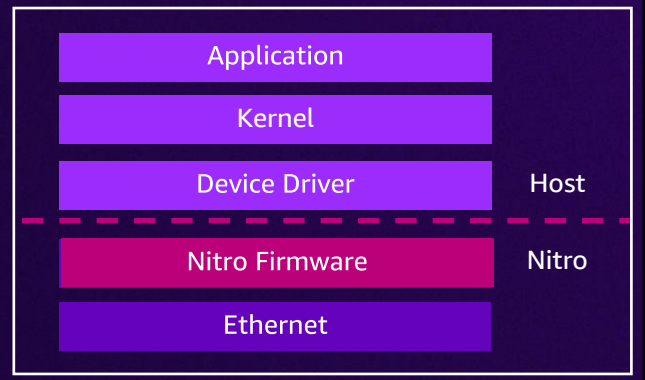


Cycle Time Allotted for Each Nitro Processor

Instance = Percent of Cycle Time

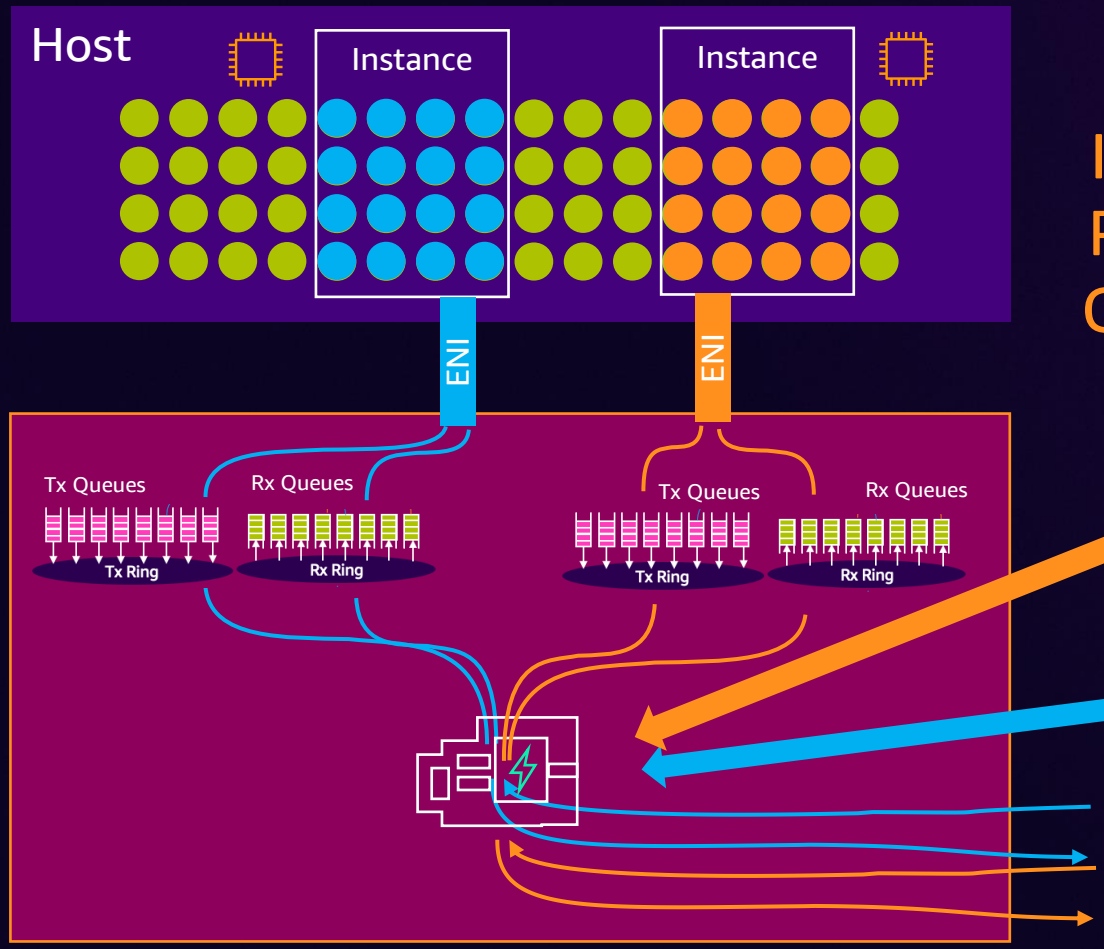
Shared Tenancy
- Resource Isolation
- Contention Mitigation
- Maximized Throughput

Opportunistic Allocation

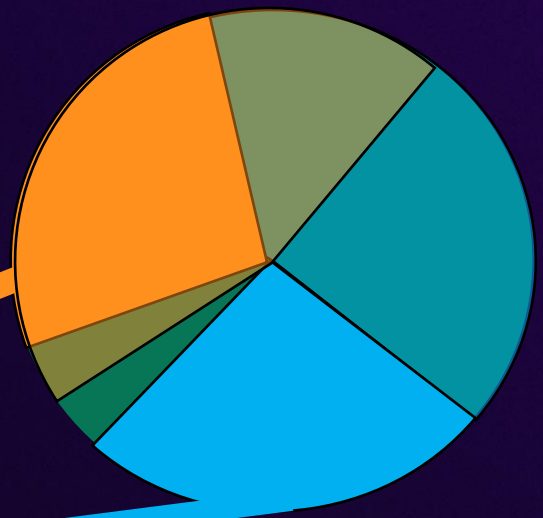


Enhanced ENA Metrics:
- pps_allowance_exceeded

Resource Allocation Per Instance



Instance = Percent of Cycle Time



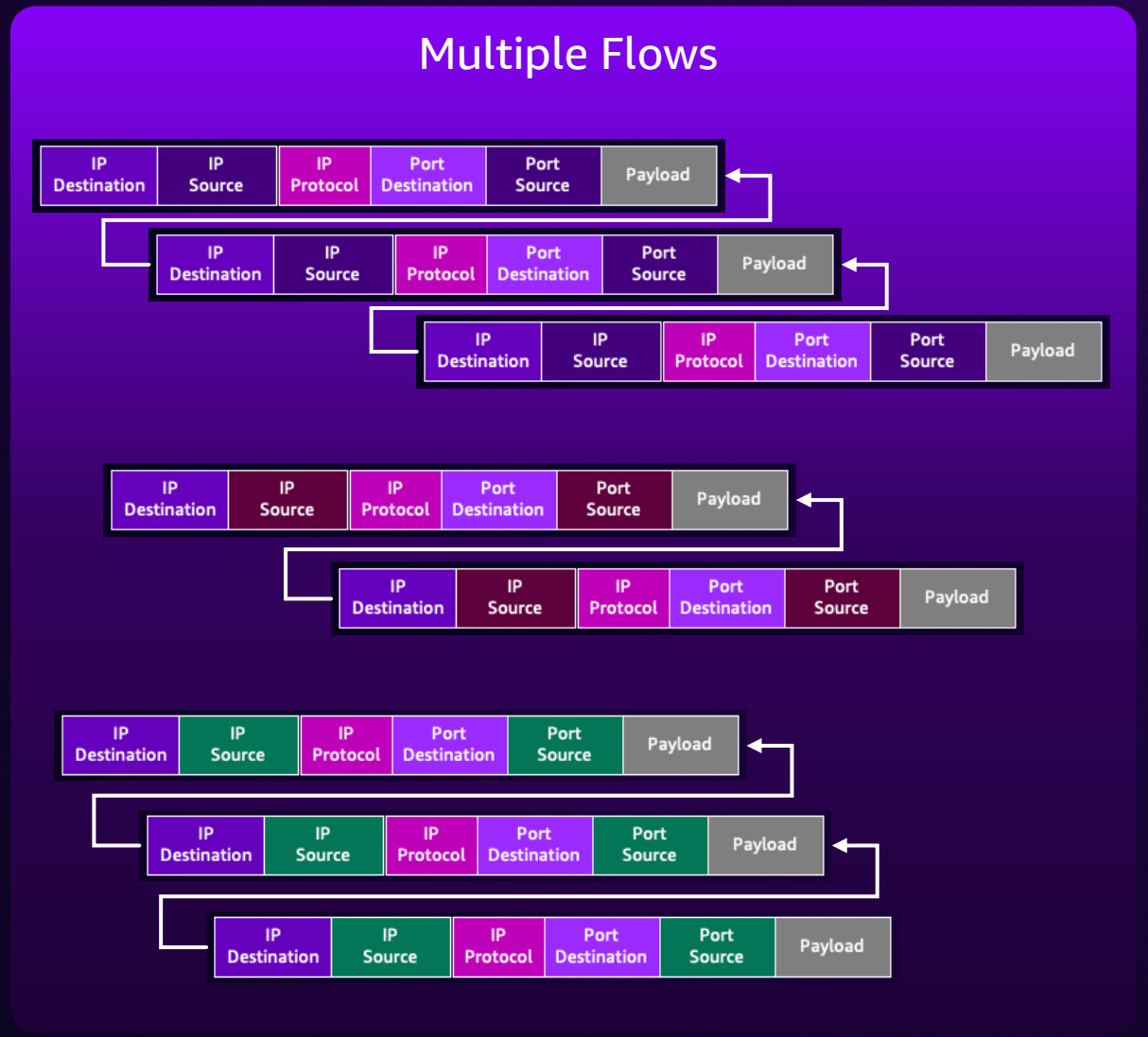
Cycle Time Allotted for Each Nitro Processor

Instance = Percent of Cycle Time

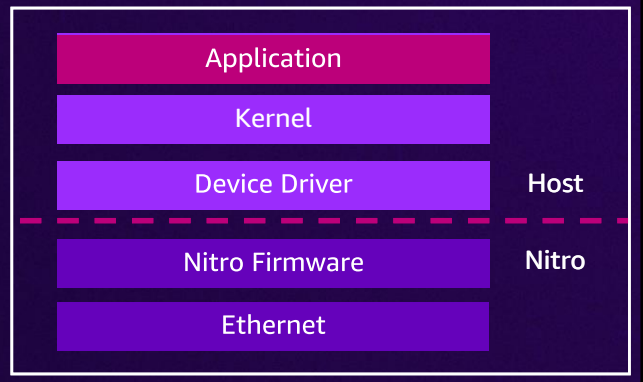
Shared Tenancy
- Resource Isolation
- Contention Mitigation
- Maximized Throughput

Tuning and Monitoring Tools

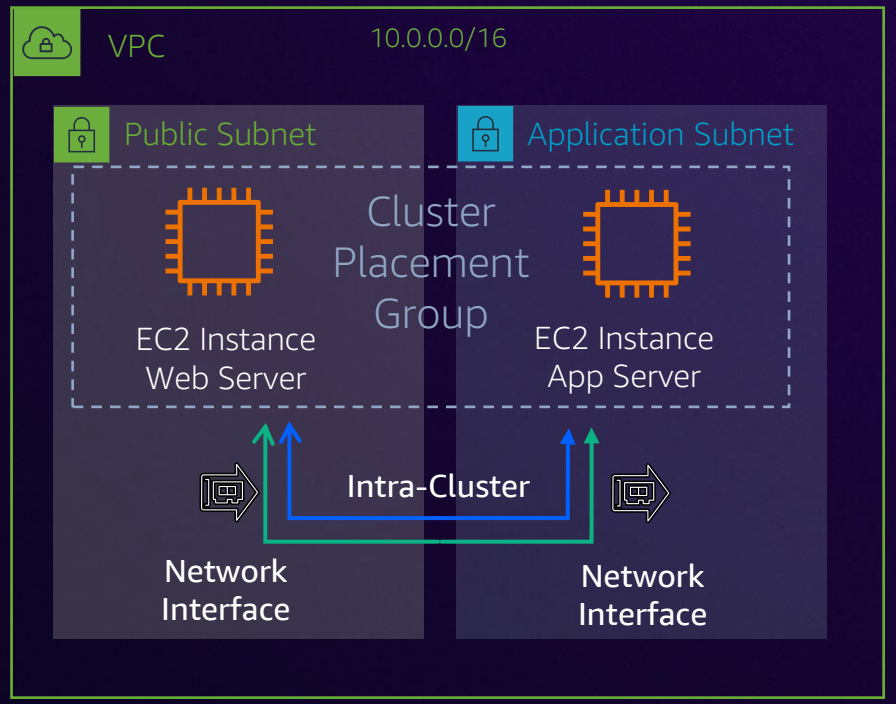
- Latency Tuning
- Queue Management Tools
- Monitoring Tools



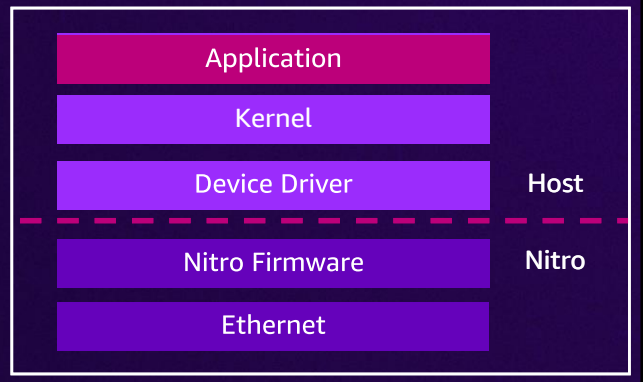
Latency Considerations



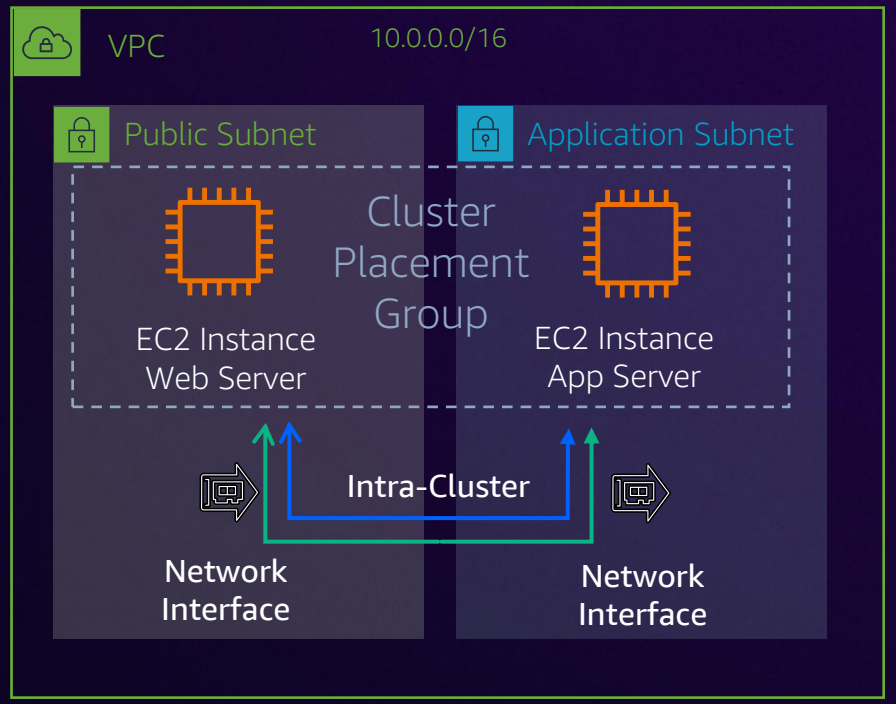
- Placement – CPGs



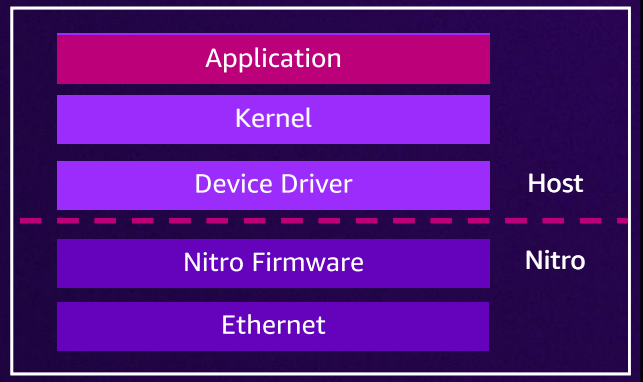
Latency Considerations



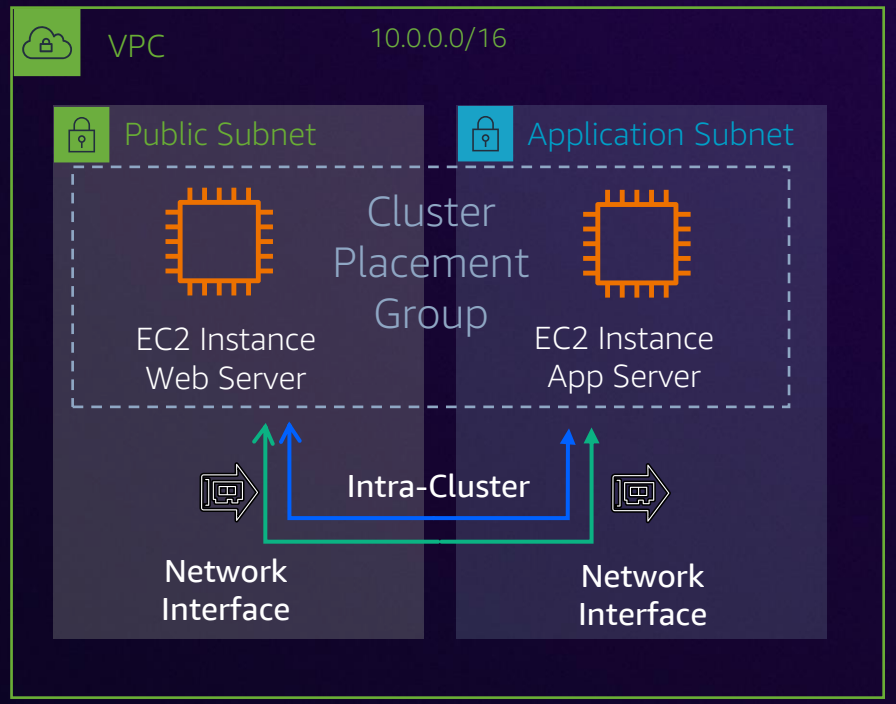
- Placement – CPGs
- Drivers – DPDK



Latency Considerations

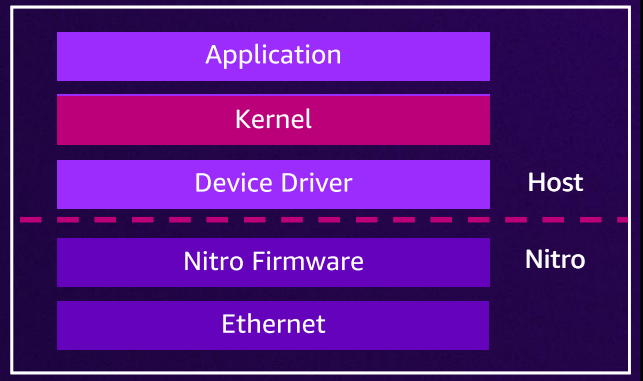


- Placement – CPGs
- Drivers – DPDK
- Protocol – UDP



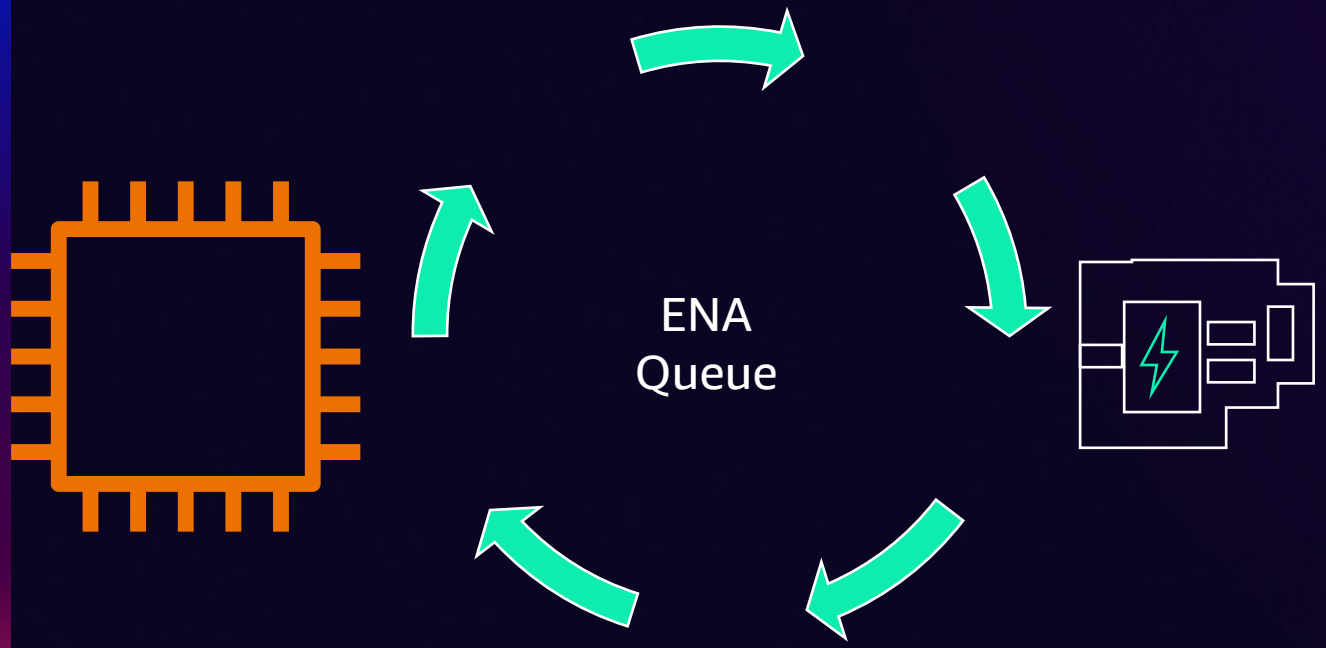
Latency Considerations

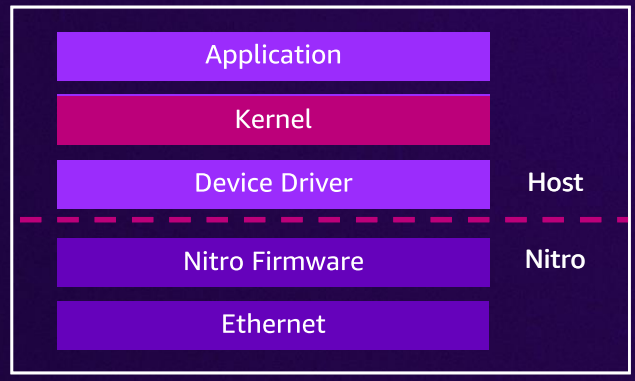
- Poll Mode Operation



Latency Considerations

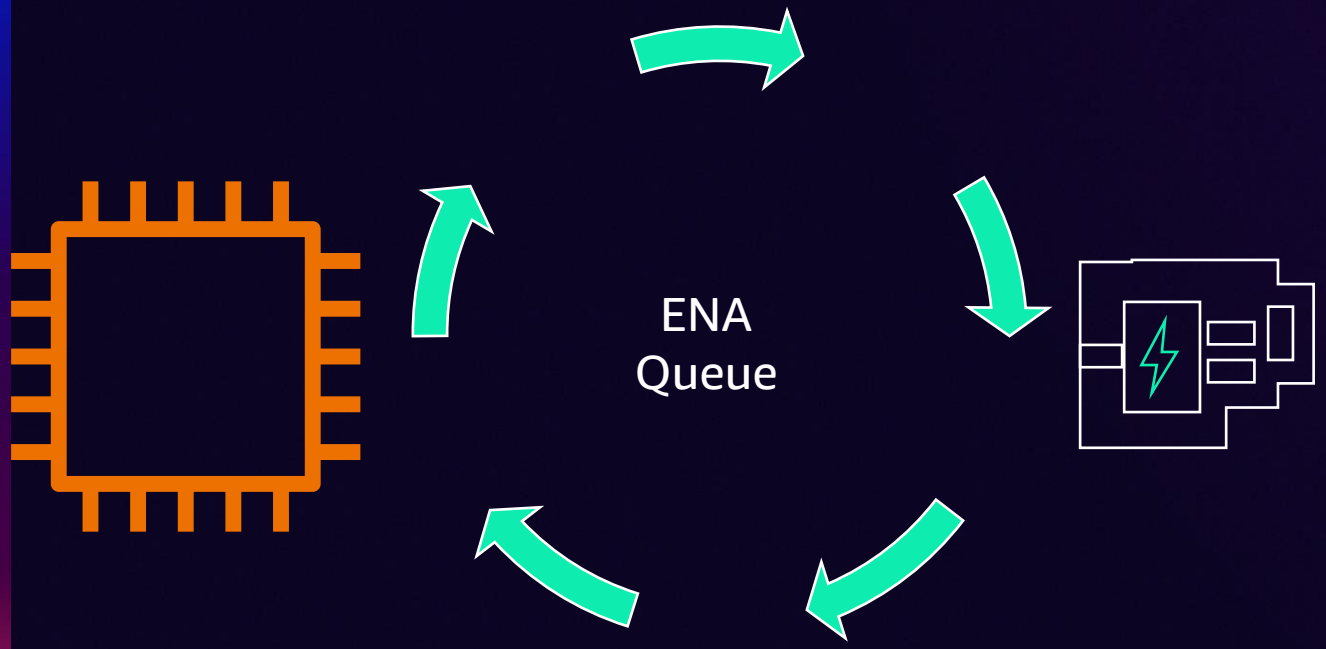
- Poll Mode Operation

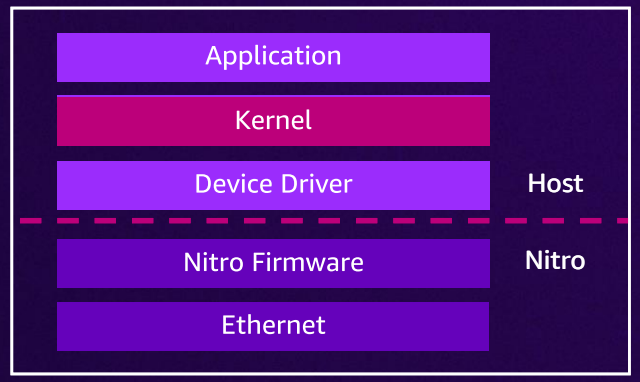


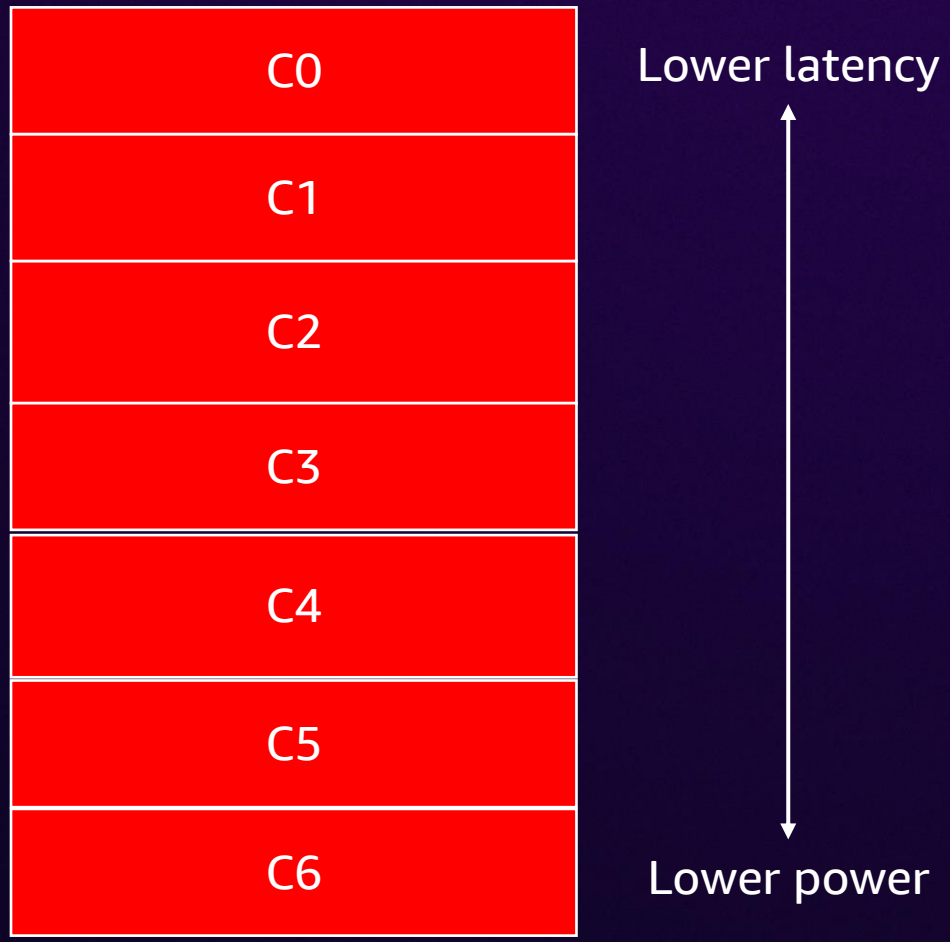


Latency Considerations

- Poll Mode Operation
- CPU State

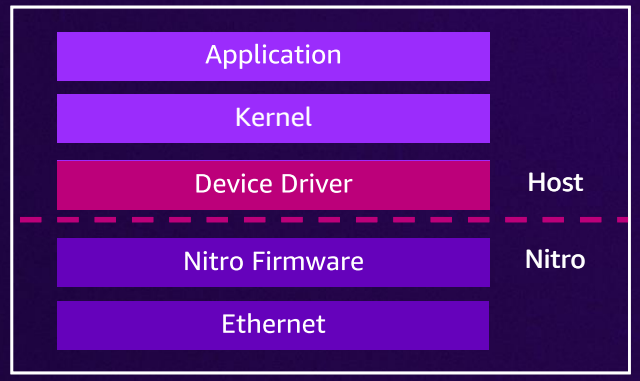


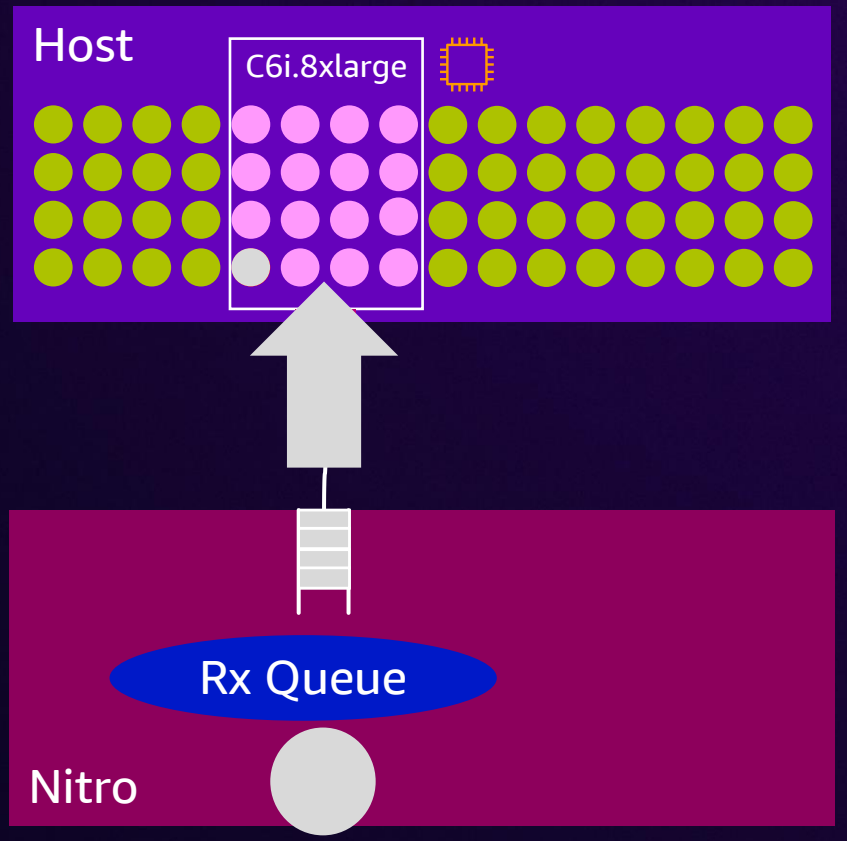




Latency Considerations

- Interrupts

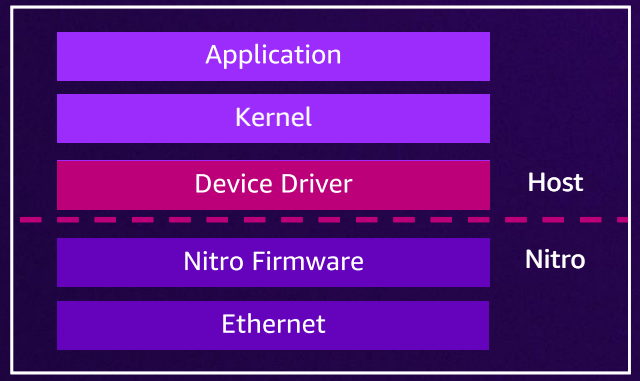


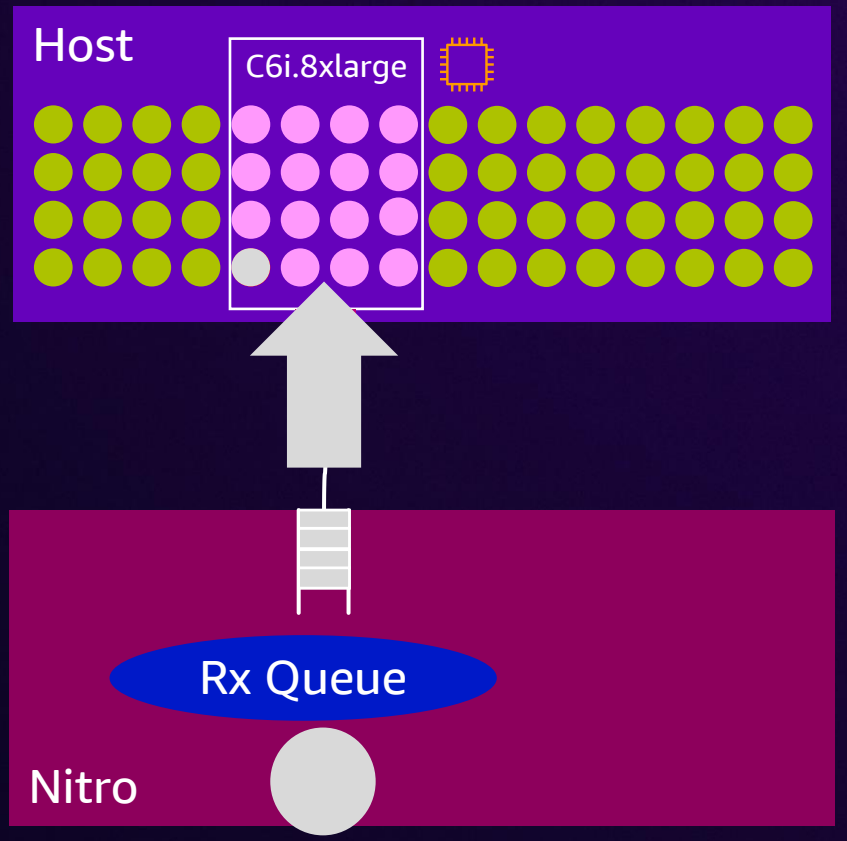


Latency Considerations

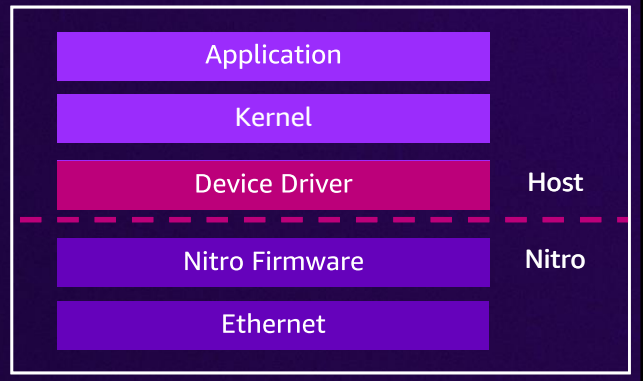
- Interrupts
- Interrupt Moderation

  [ec2-user ~]$ ethtool -c interface | egrep "rx-usecs:|tx-usecs:|Adaptive RX" Adaptive RX: off TX: off rx-usecs: 20 tx-usecs: 64





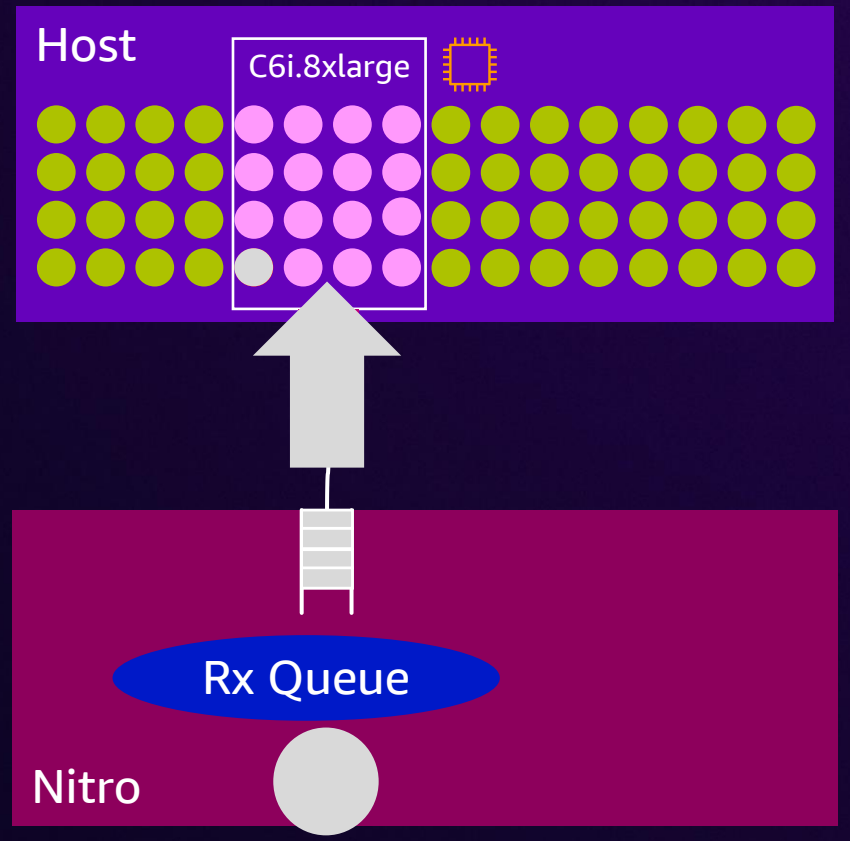
Latency Considerations



- Interrupts
- Interrupt Moderation

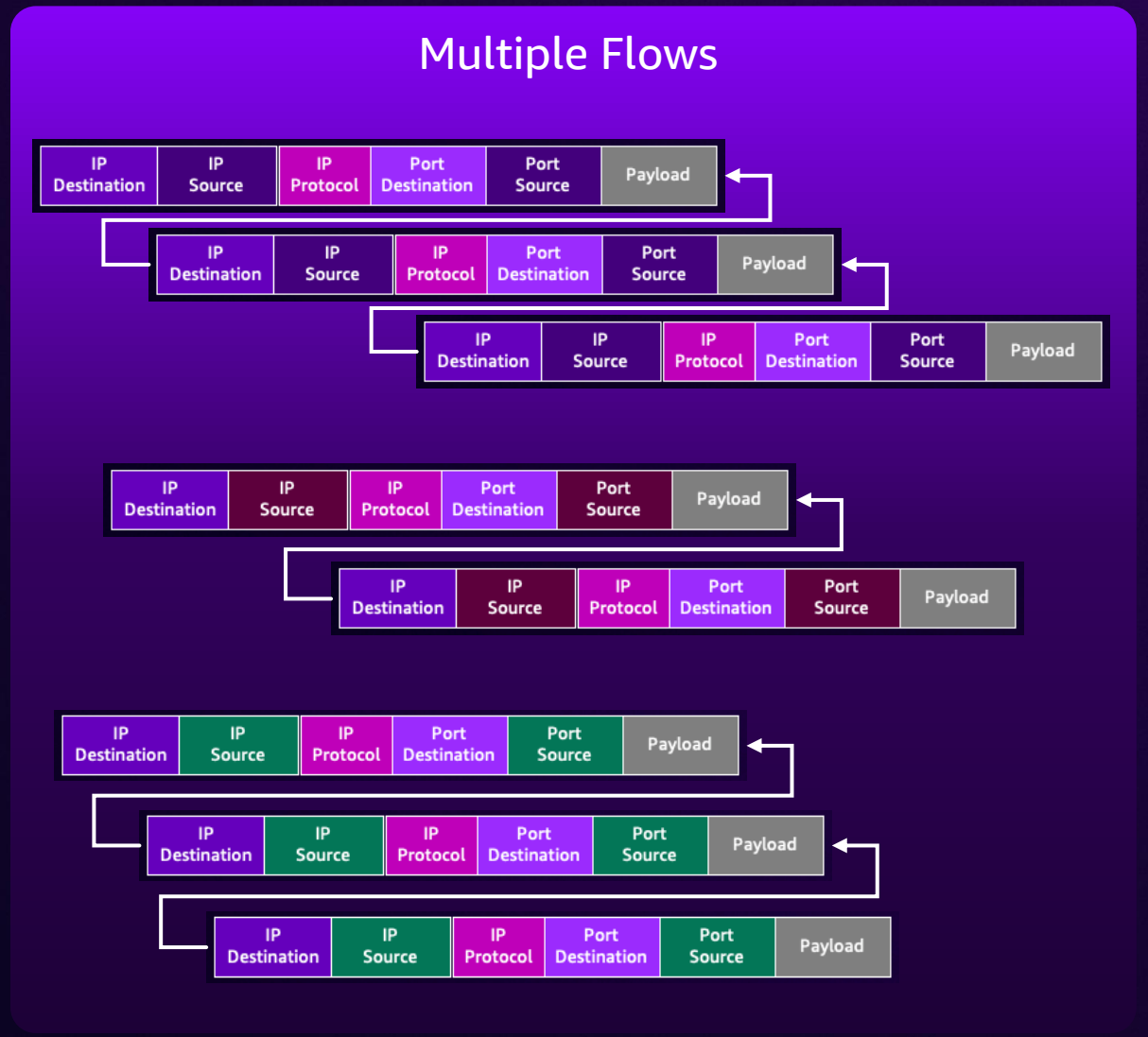
  [ec2-user ~]$ ethtool -c interface | egrep "rx-usecs:|tx-usecs:|Adaptive RX" Adaptive RX: off TX: off rx-usecs: 20 tx-usecs: 64
- Dynamic Interrupt Moderation

  [ec2-user ~]$ ethtool -c interface | egrep "rx-usecs:|tx-usecs:|Adaptive RX" Adaptive RX: on TX: on rx-usecs: 20 tx-usecs: 64

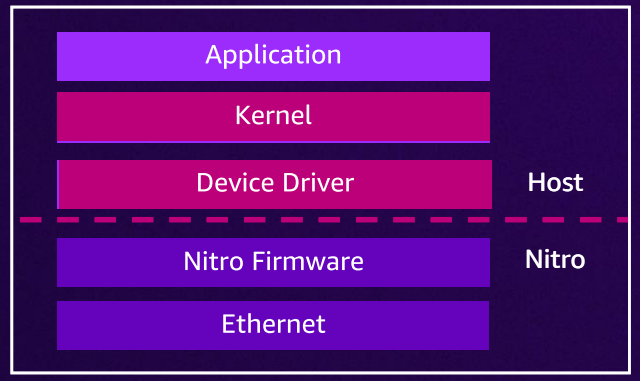


Tuning and Monitoring Tools

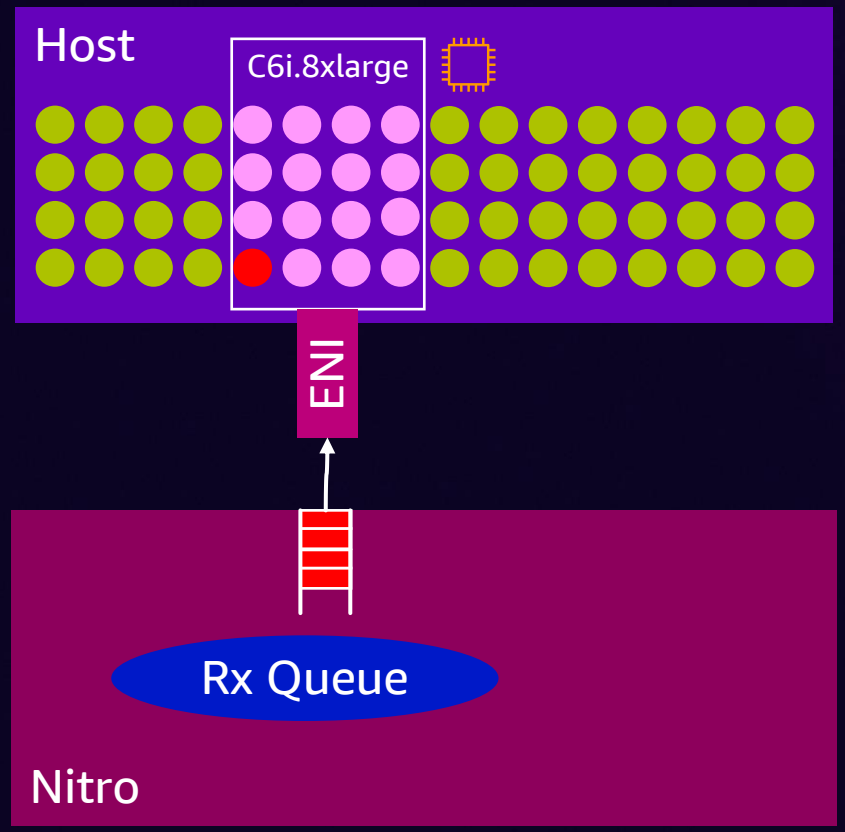
- Latency Tuning
- Queue Management Tools
- Monitoring Tools



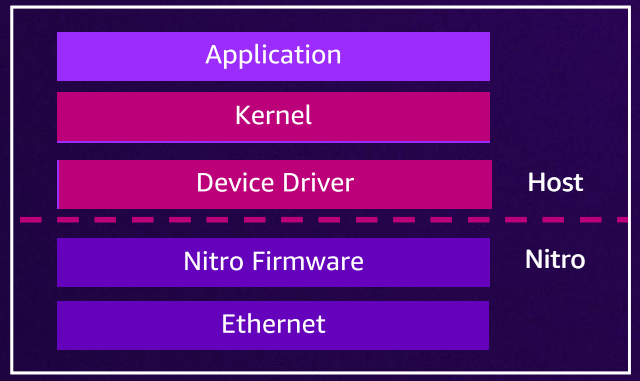
ENA Receive Side Scaling (RSS)



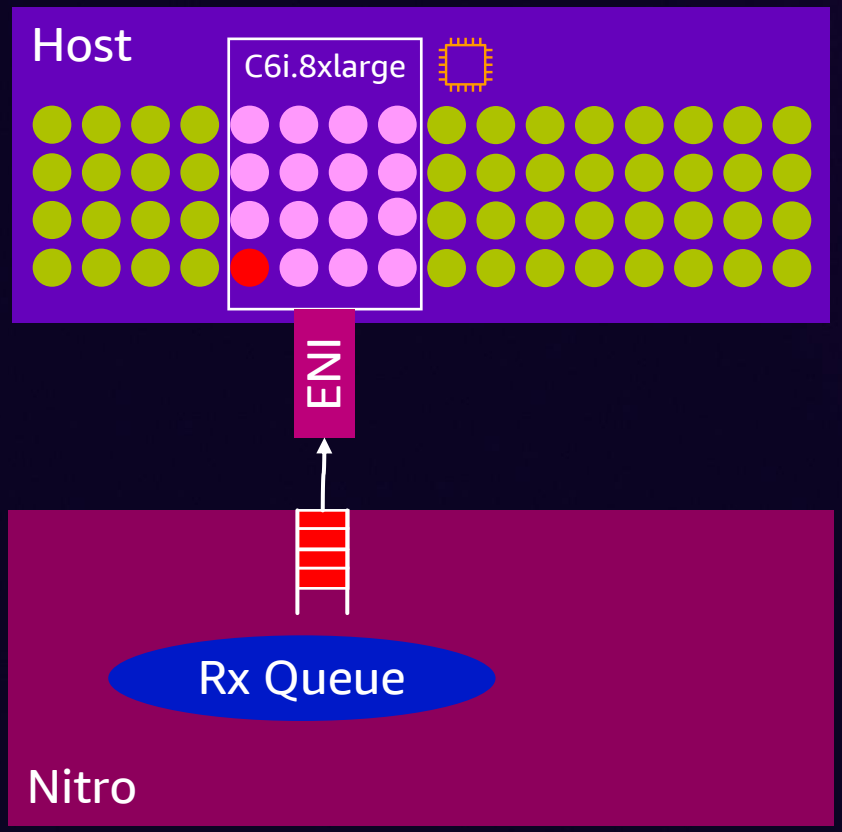
C6i.8xlarge prior to RSS



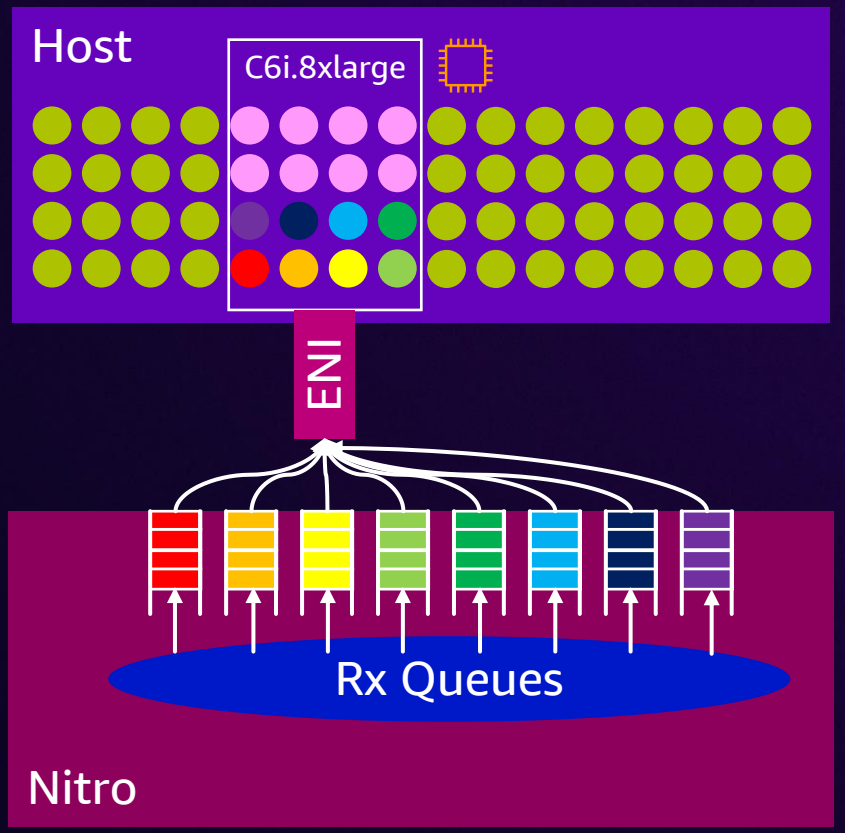
ENA Receive Side Scaling (RSS)



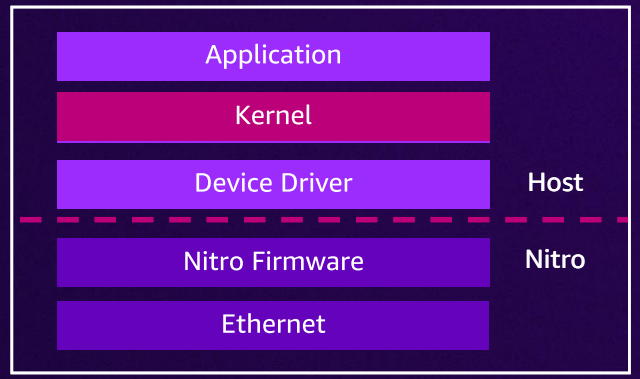
C6i.8xlarge prior to RSS



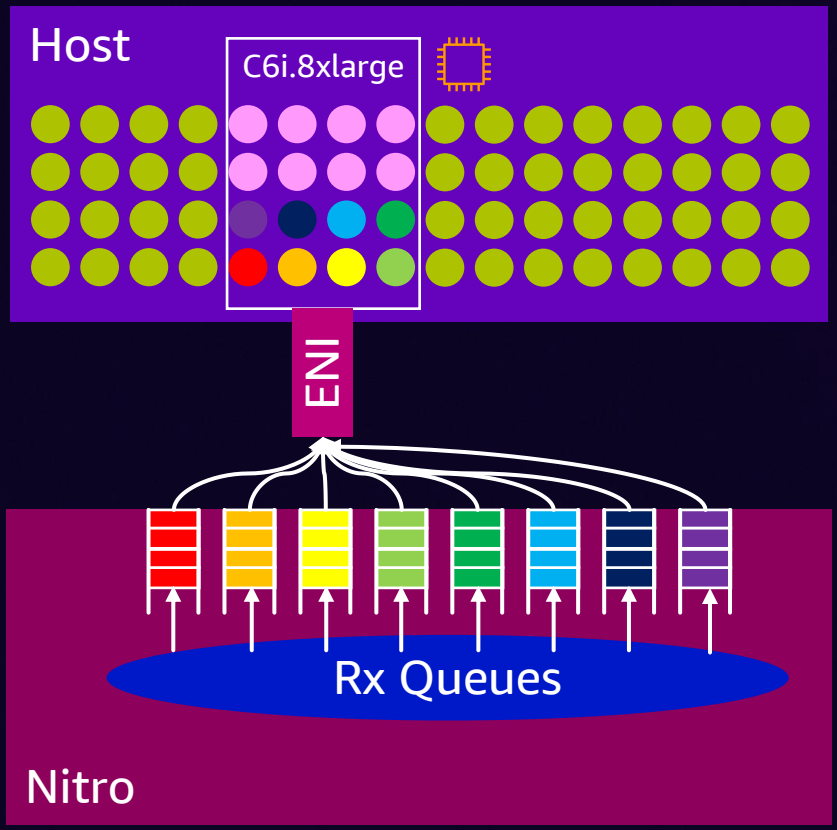
C6i.8xlarge with RSS (default)



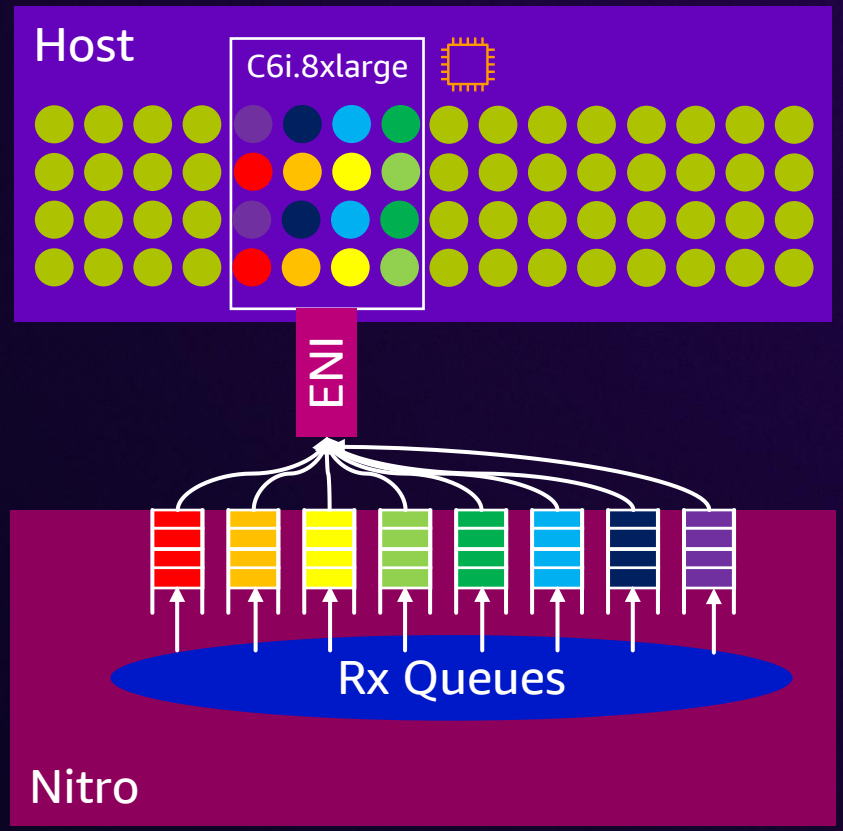
Nitro Receive Packet Steering (RPS)



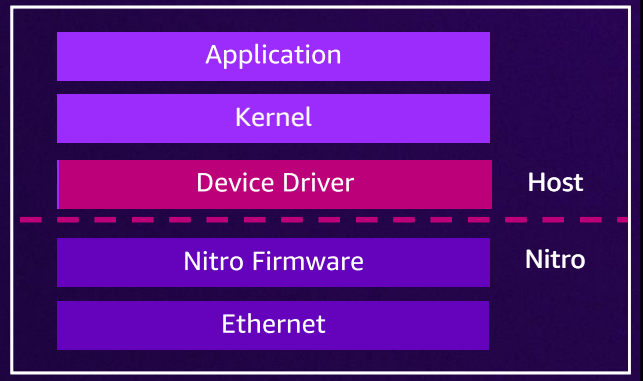
C6i.8xlarge with RSS



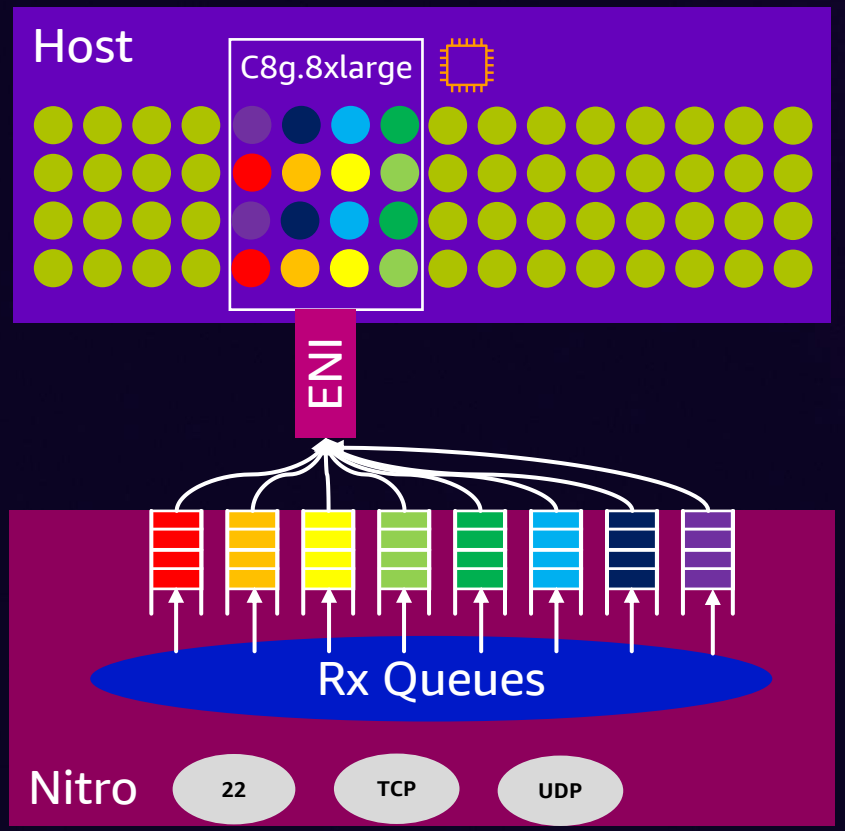
C6i.8xlarge with RPS (Linux)



ENA Flow Steering - N-tuple Filtering



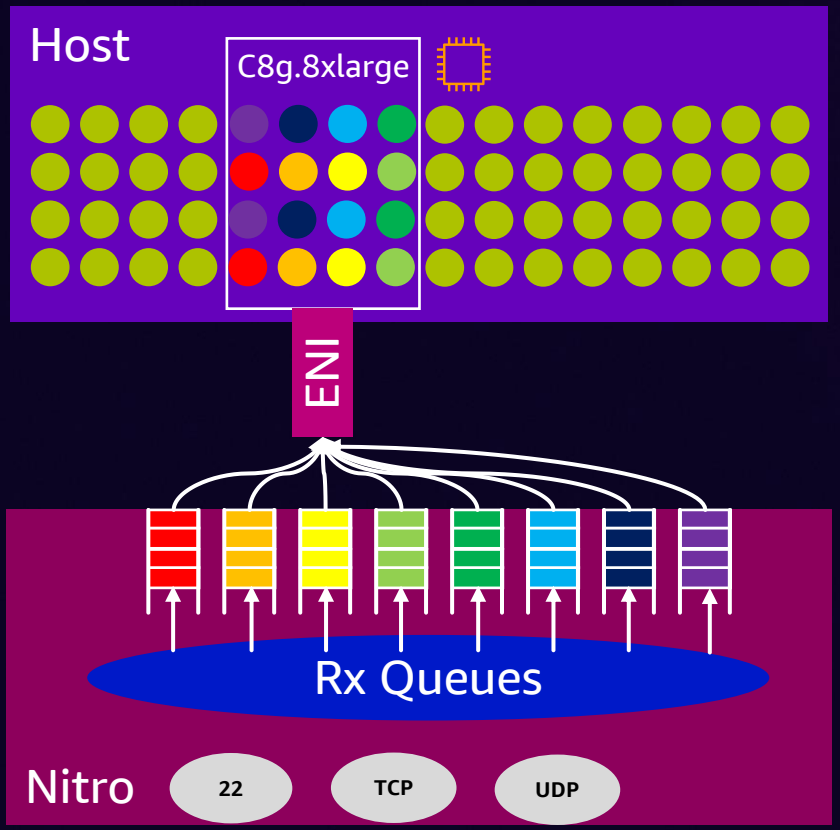
C8g.8xlarge with RPS (Linux)

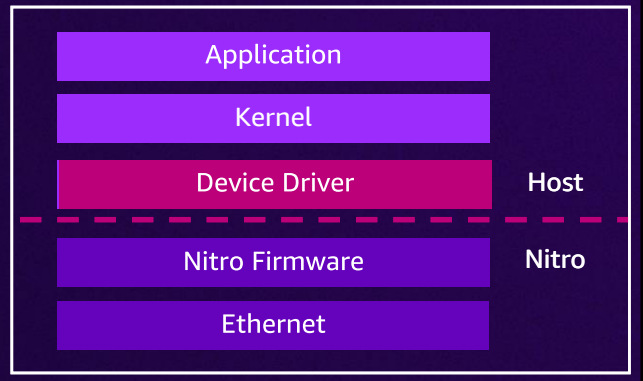


ENA Flow Steering - N-tuple Filtering

ethtool -N eth1 flow-type tcp4 dst-port 22 action 1 loc 0
ethtool -N eth1 flow-type tcp4 action 2 loc 1
ethtool -N eth1 flow-type udp4 action 3 loc 2

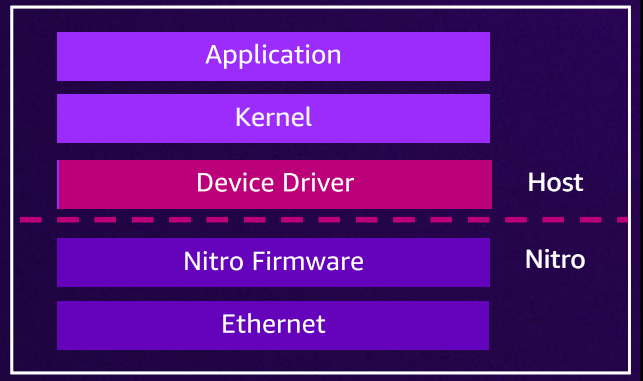
C8g.8xlarge with RPS (Linux)



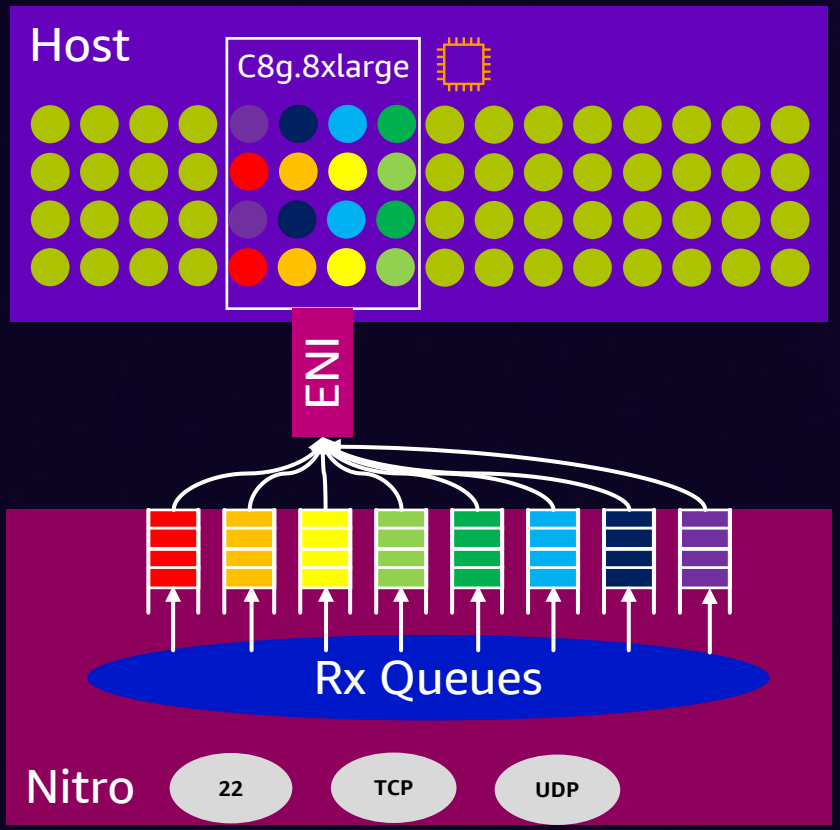


ENA Flow Steering - N-tuple Filtering

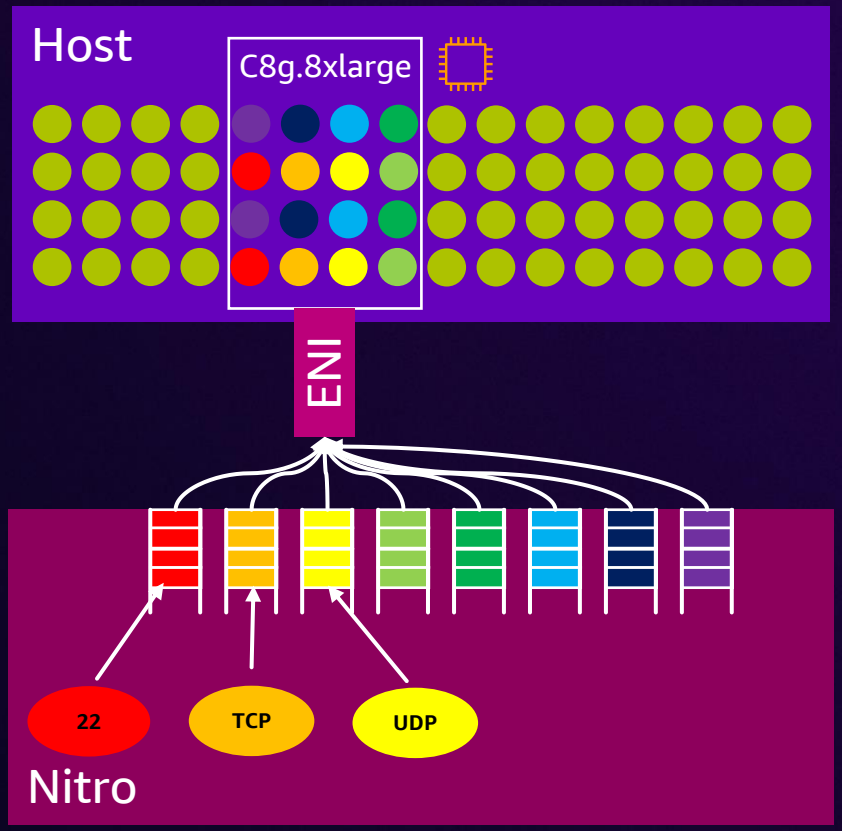
```
ethtool -N eth1 flow-type tcp4 dst-port 22 action 1 loc 0
ethtool -N eth1 flow-type tcp4 action 2 loc 1
ethtool -N eth1 flow-type udp4 action 3 loc 2
```



C8g.8xlarge with RPS (Linux)

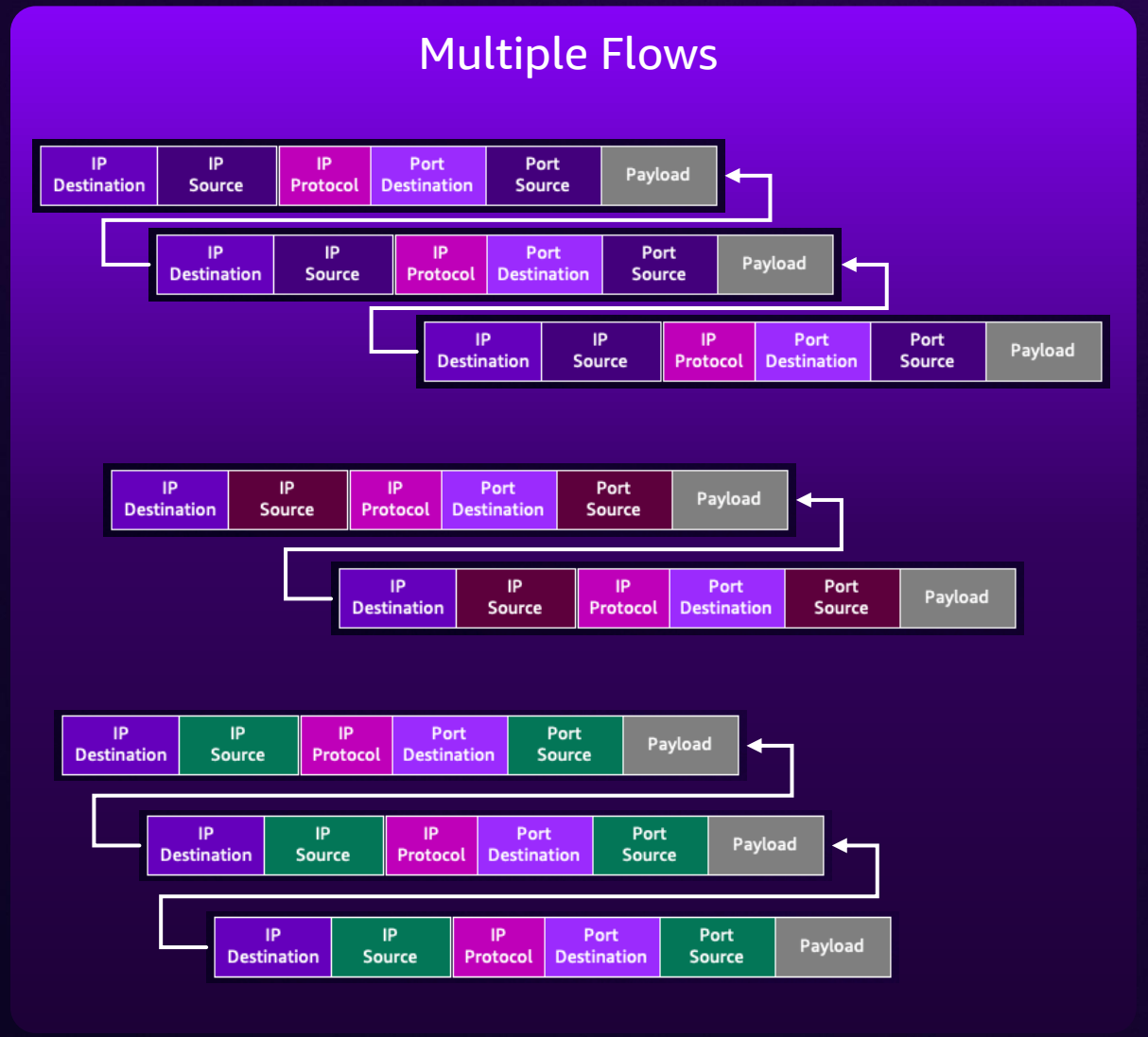


With Flow Steering



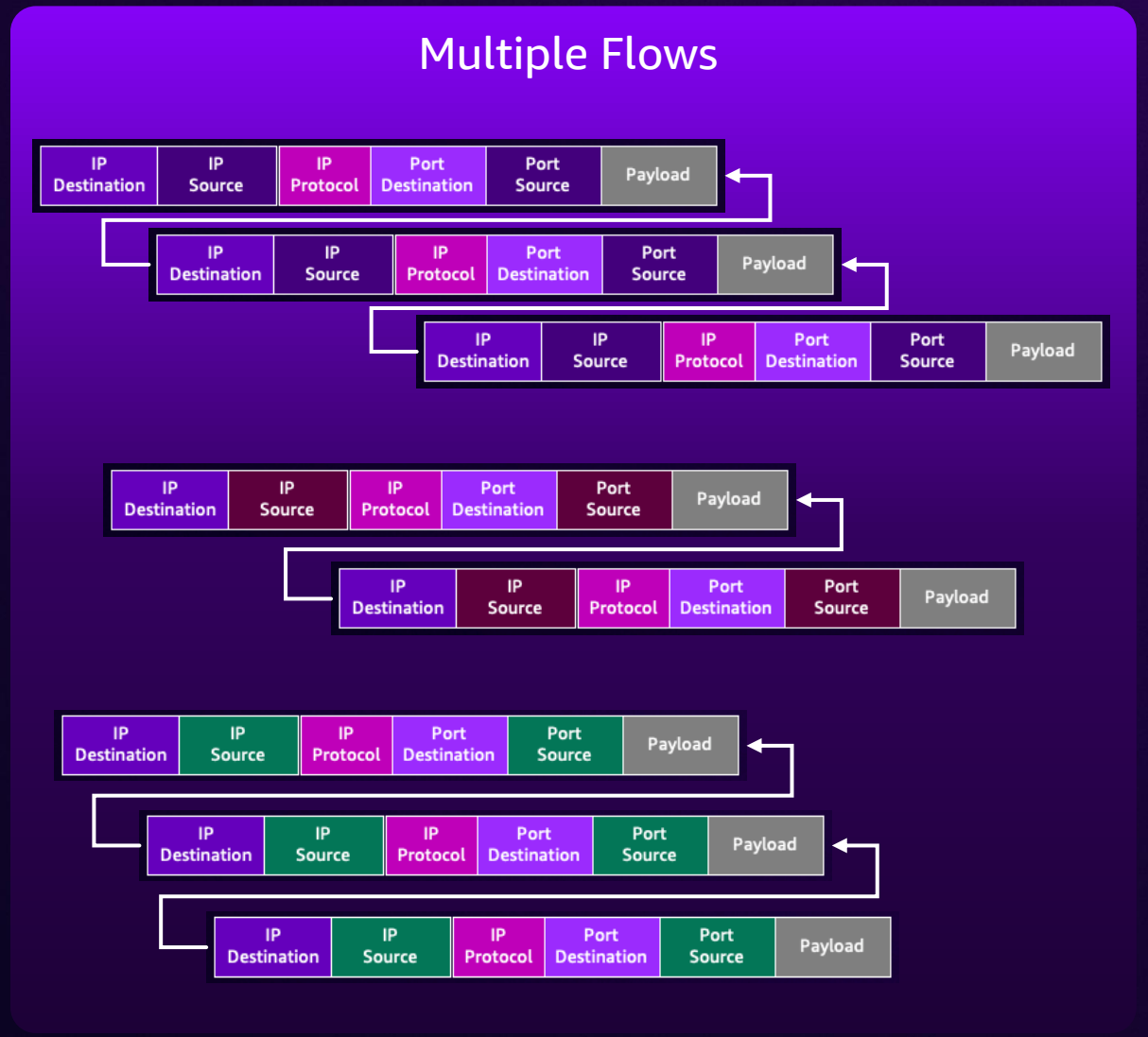
Tuning and Monitoring Tools

- Latency Tuning
- Queue Management Tools
- Monitoring Tools



Flow Analysis – Performance and Scaling Metrics

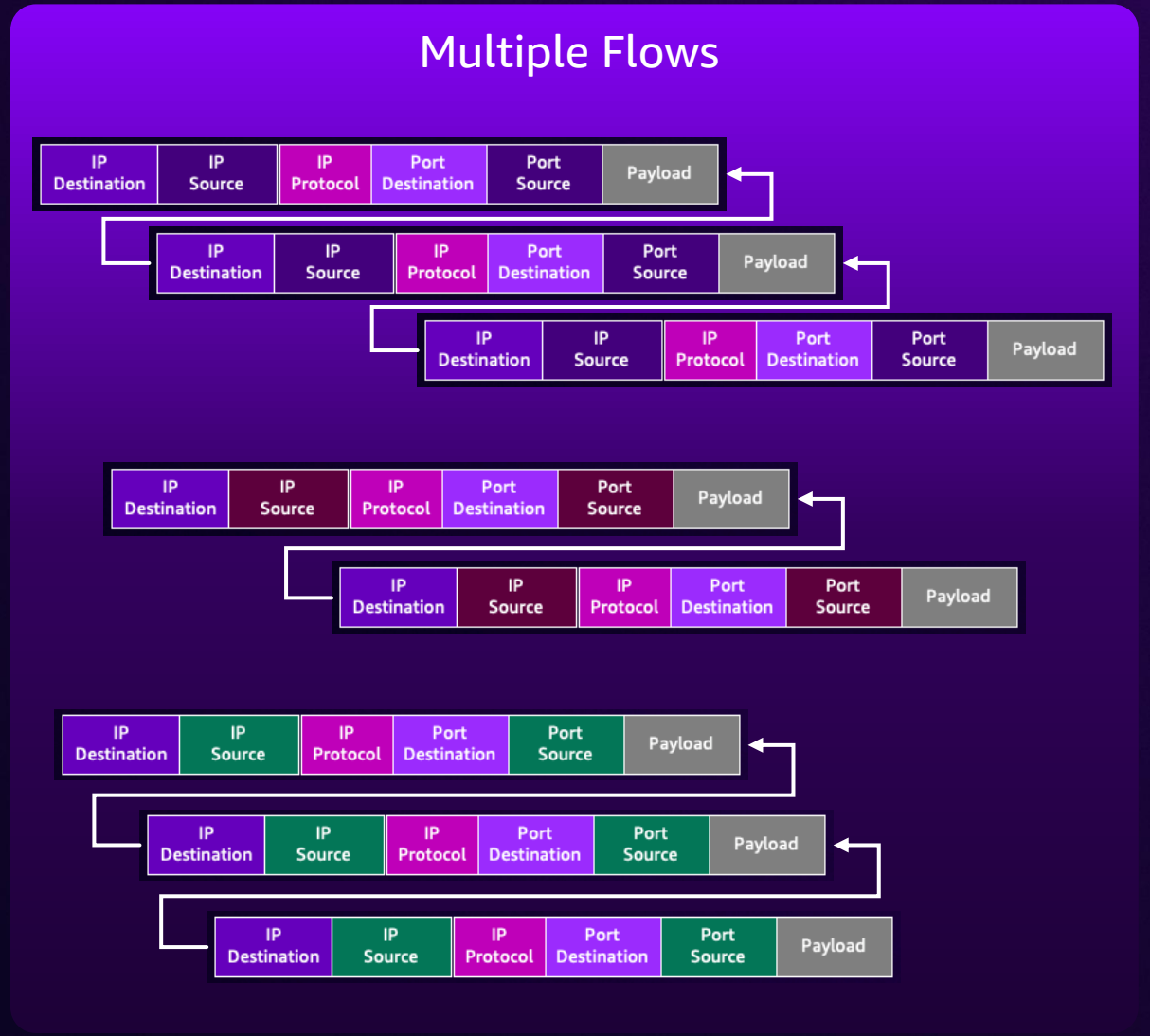
```
[ec2-user ~]$ ethtool -S eth0
     bw_in_allowance_exceeded: 0
     bw_out_allowance_exceeded: 0
     pps_allowance_exceeded: 0
     conntrack_allowance_exceeded: 0
     linklocal_allowance_exceeded: 0
     conntrack_allowance_available: 136812
```



Flow Analysis – Performance and Scaling Metrics

```
[ec2-user ~]$ ethtool -S eth0
     bw_in_allowance_exceeded: 0
     bw_out_allowance_exceeded: 0
     pps_allowance_exceeded: 0
     conntrack_allowance_exceeded: 0
     linklocal_allowance_exceeded: 0
     conntrack_allowance_available: 136812
```

CloudWatch Agent

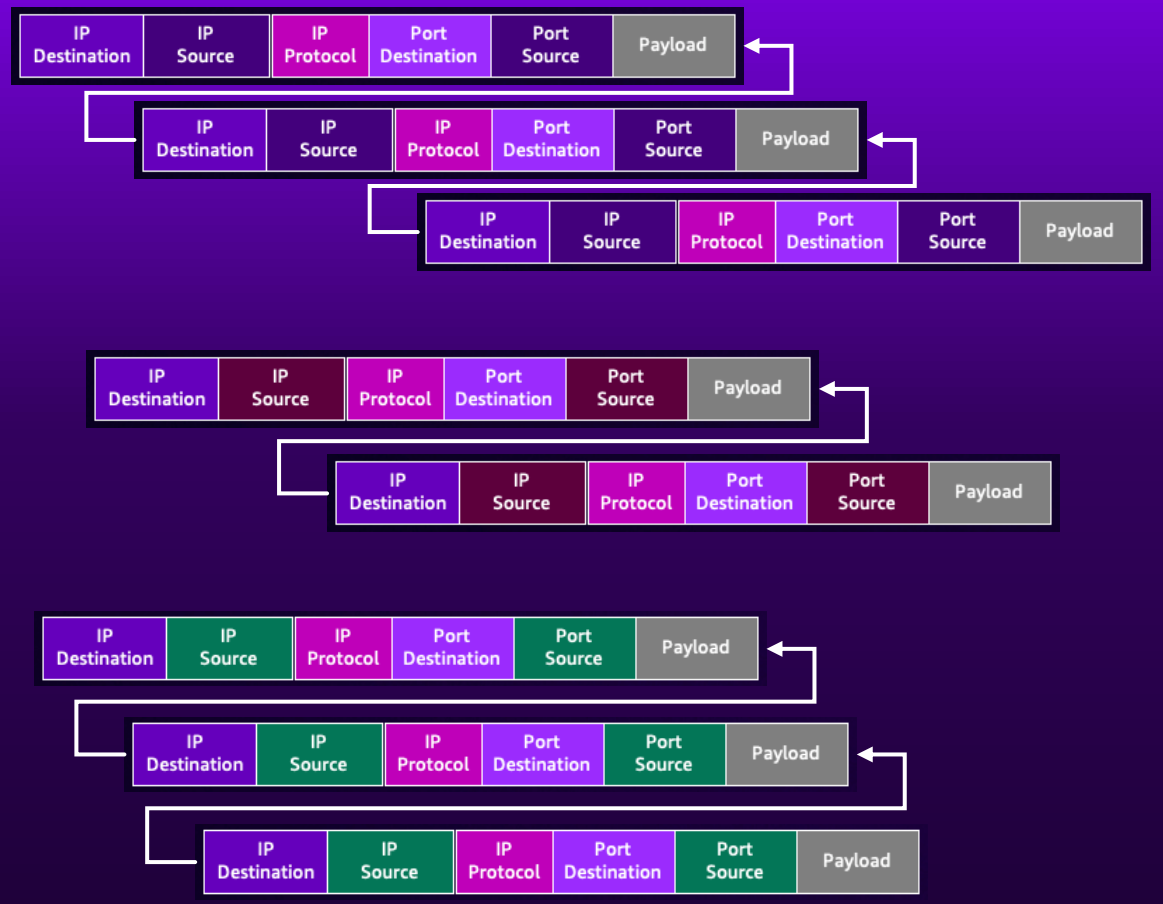


Flow Analysis – Performance and Scaling Metrics

```
[ec2-user ~]$ ethtool -S eth0
     bw_in_allowance_exceeded: 0
     bw_out_allowance_exceeded: 0
     pps_allowance_exceeded: 0
     conntrack_allowance_exceeded: 0
     linklocal_allowance_exceeded: 0
     conntrack_allowance_available: 136812
```

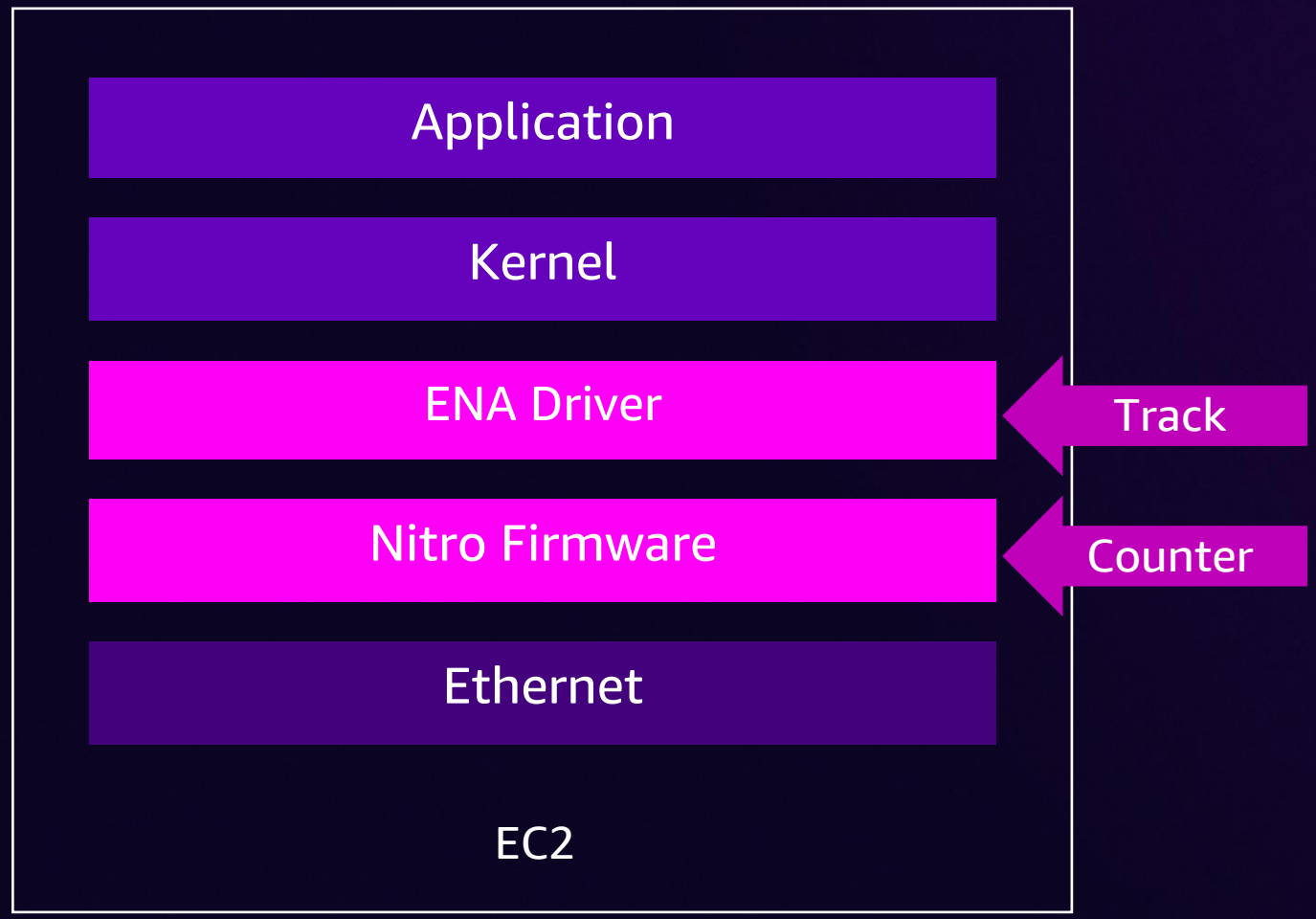
CloudWatch Agent

Multiple Flows

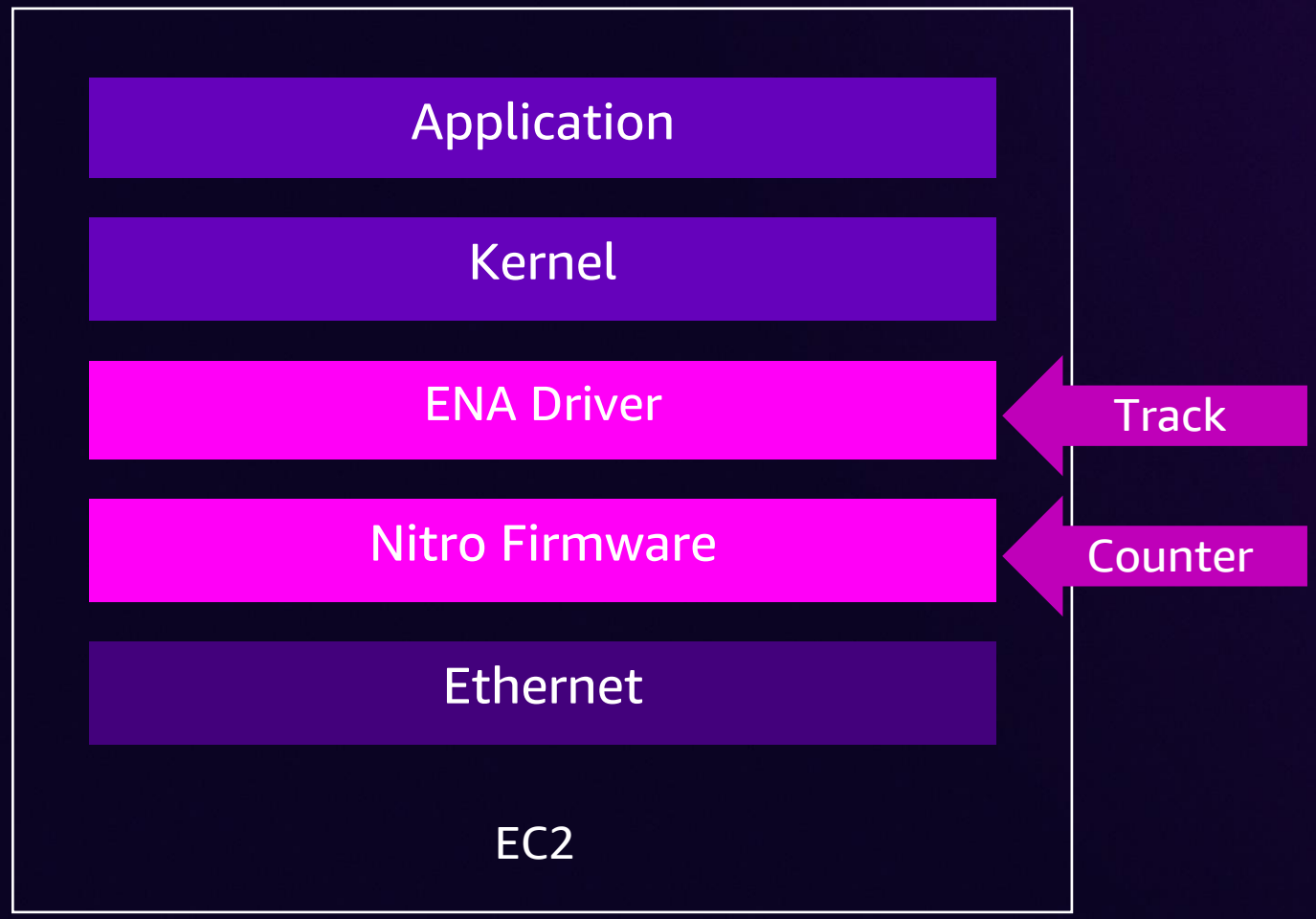


Connection Tracking

Connection Tracking



Connection Tracking

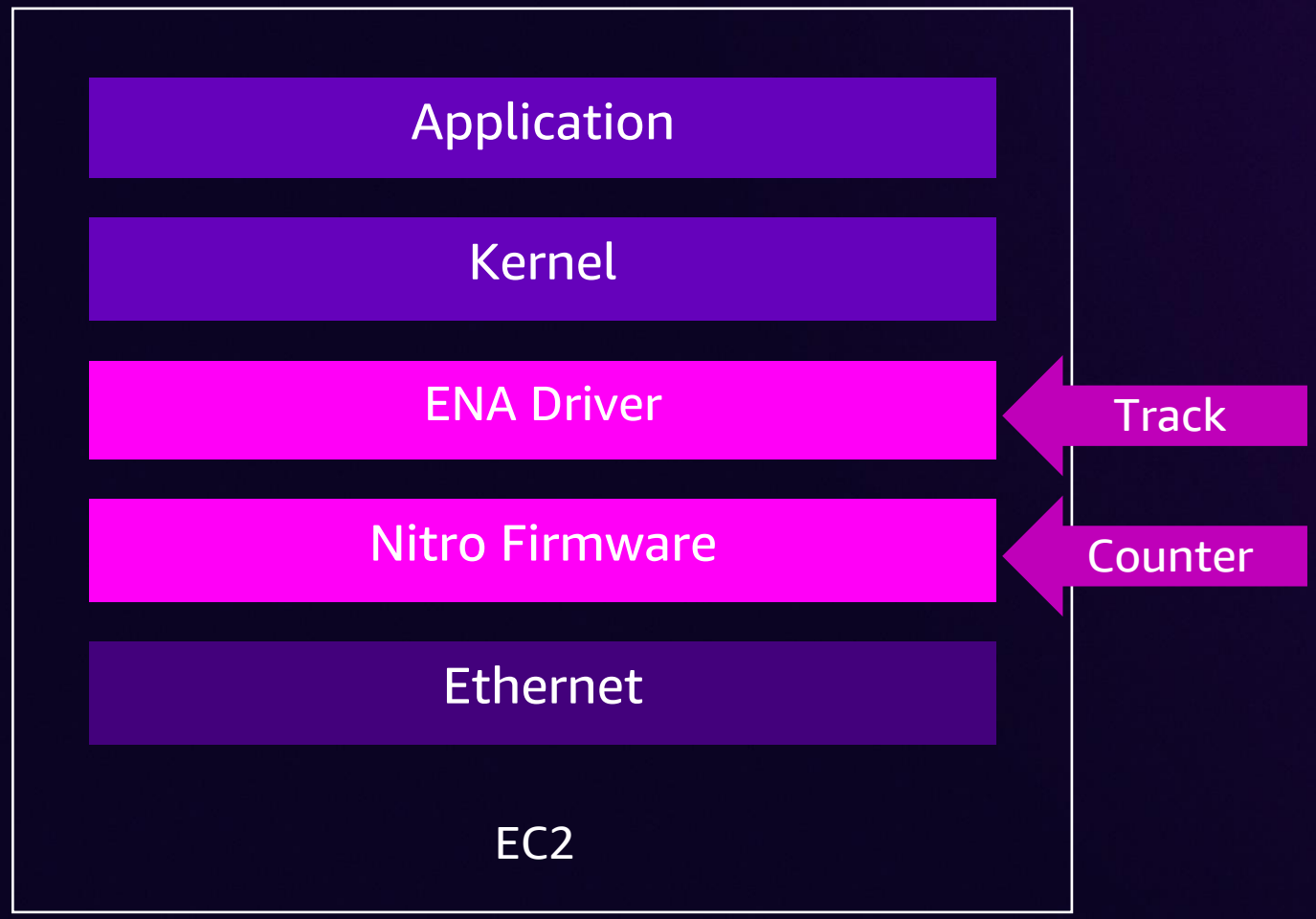


Configurable Idle conntrack timeout

Proto	Default	Min	Max
TCP	5 Days	60 sec	5 days
UDP 1-direction	30 sec	30 sec	60 secs
UDP stream	180 sec	60 secs	60 secs

Connection Tracking

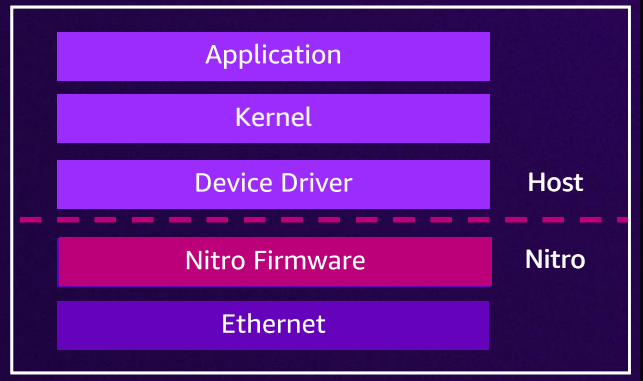
- Conntrack_allowance_exceeded
- Conntrack_allowance_available

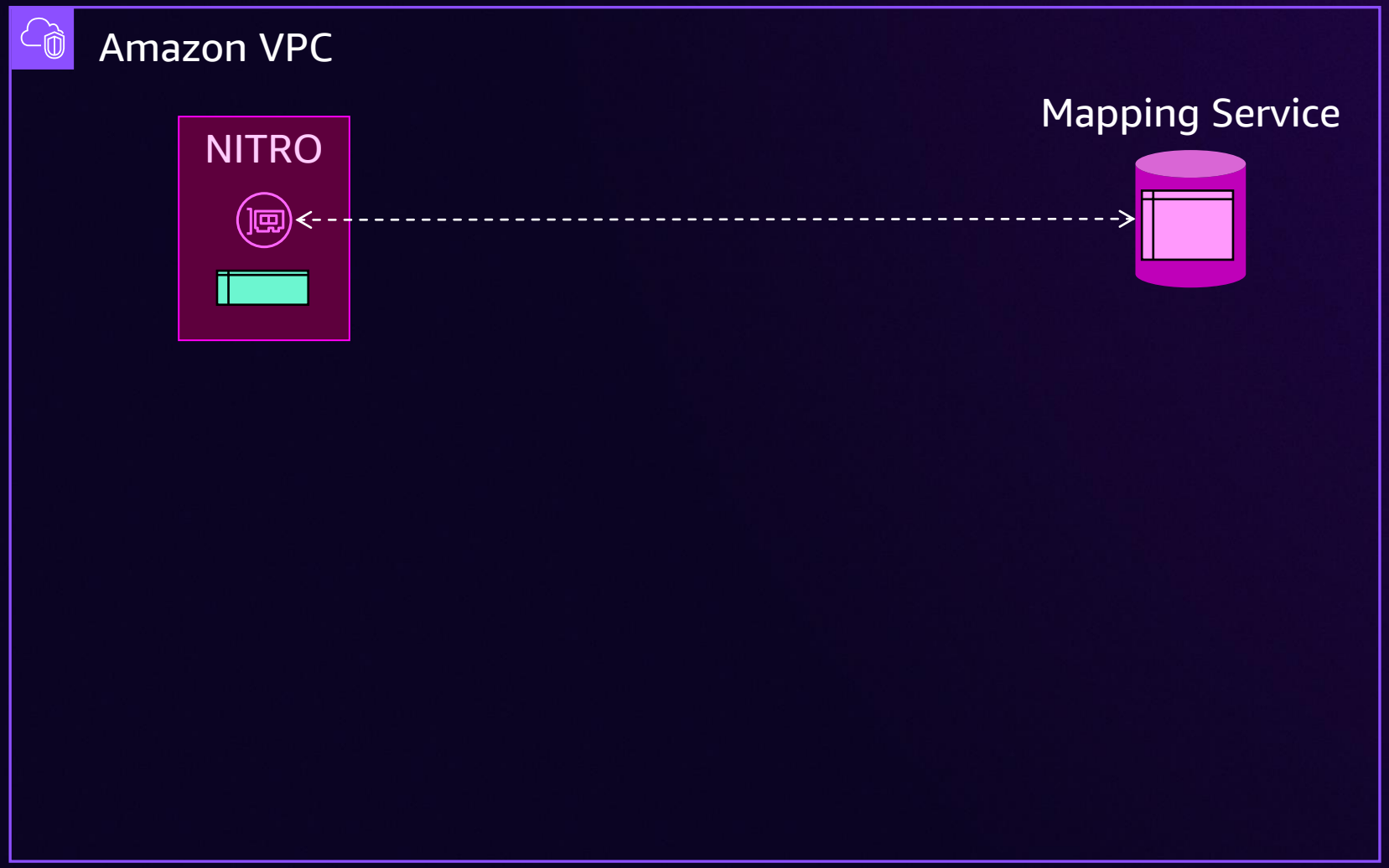


Configurable Idle conntrack timeout

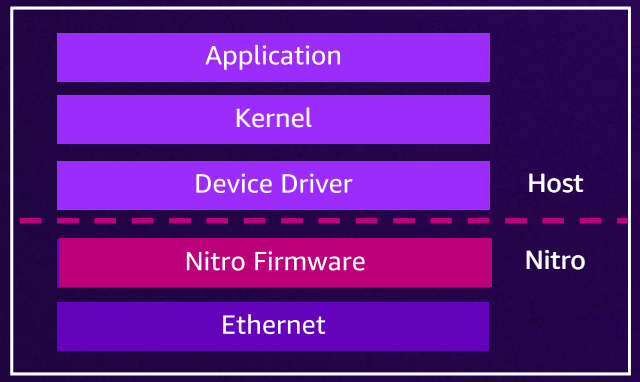
Proto	Default	Min	Max
TCP	5 Days	60 sec	5 days
UDP 1-direction	30 sec	30 sec	60 secs
UDP stream	180 sec	60 secs	60 secs

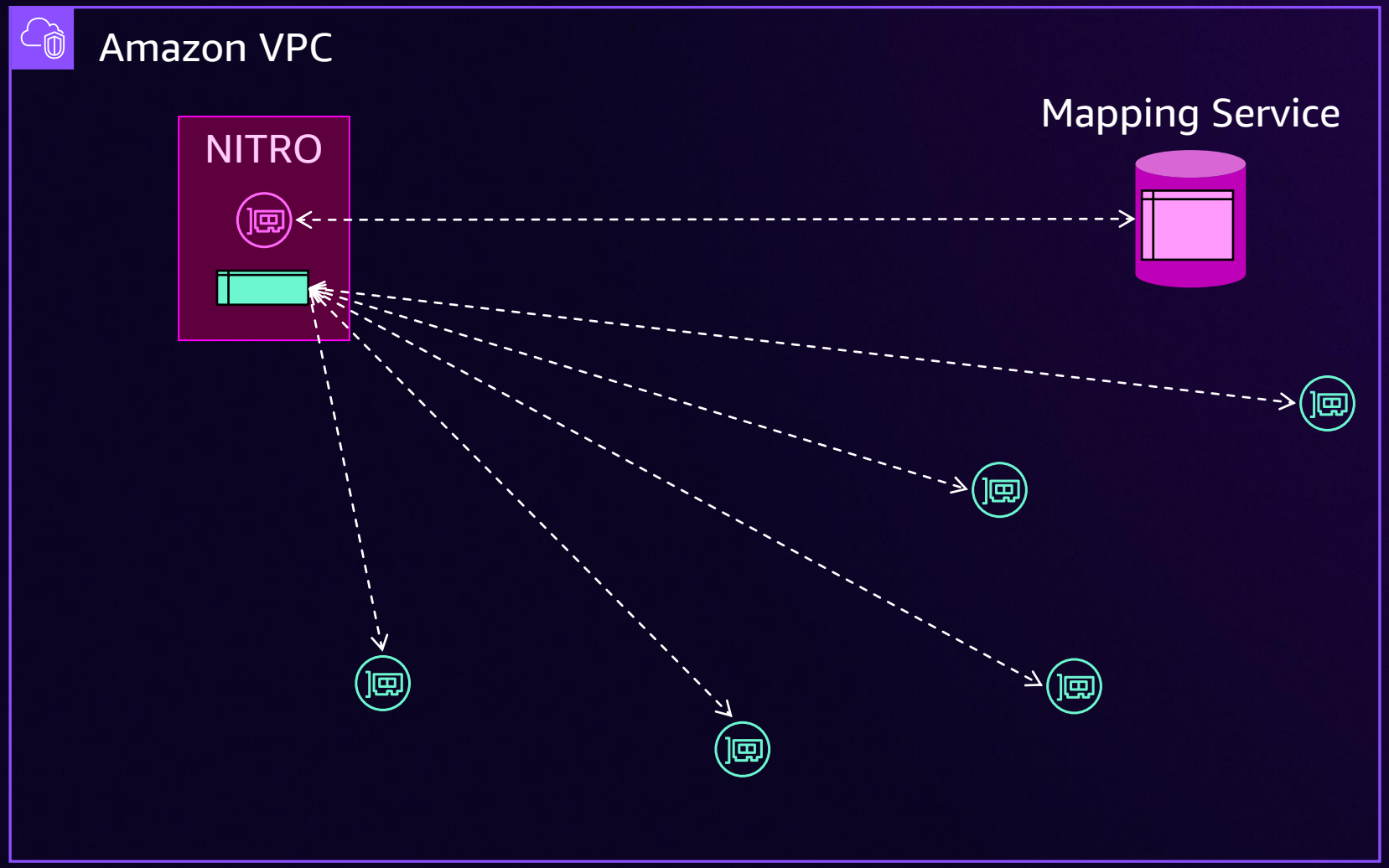
VPC Network Address Units



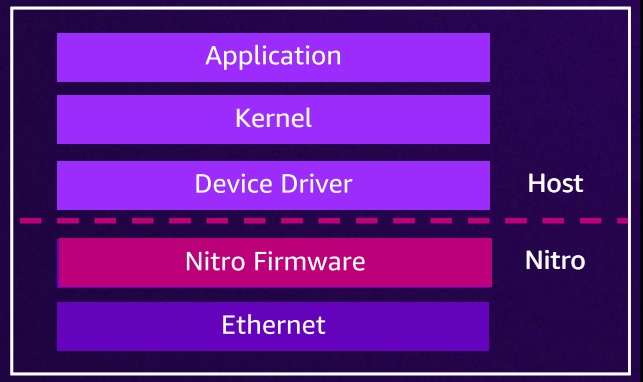


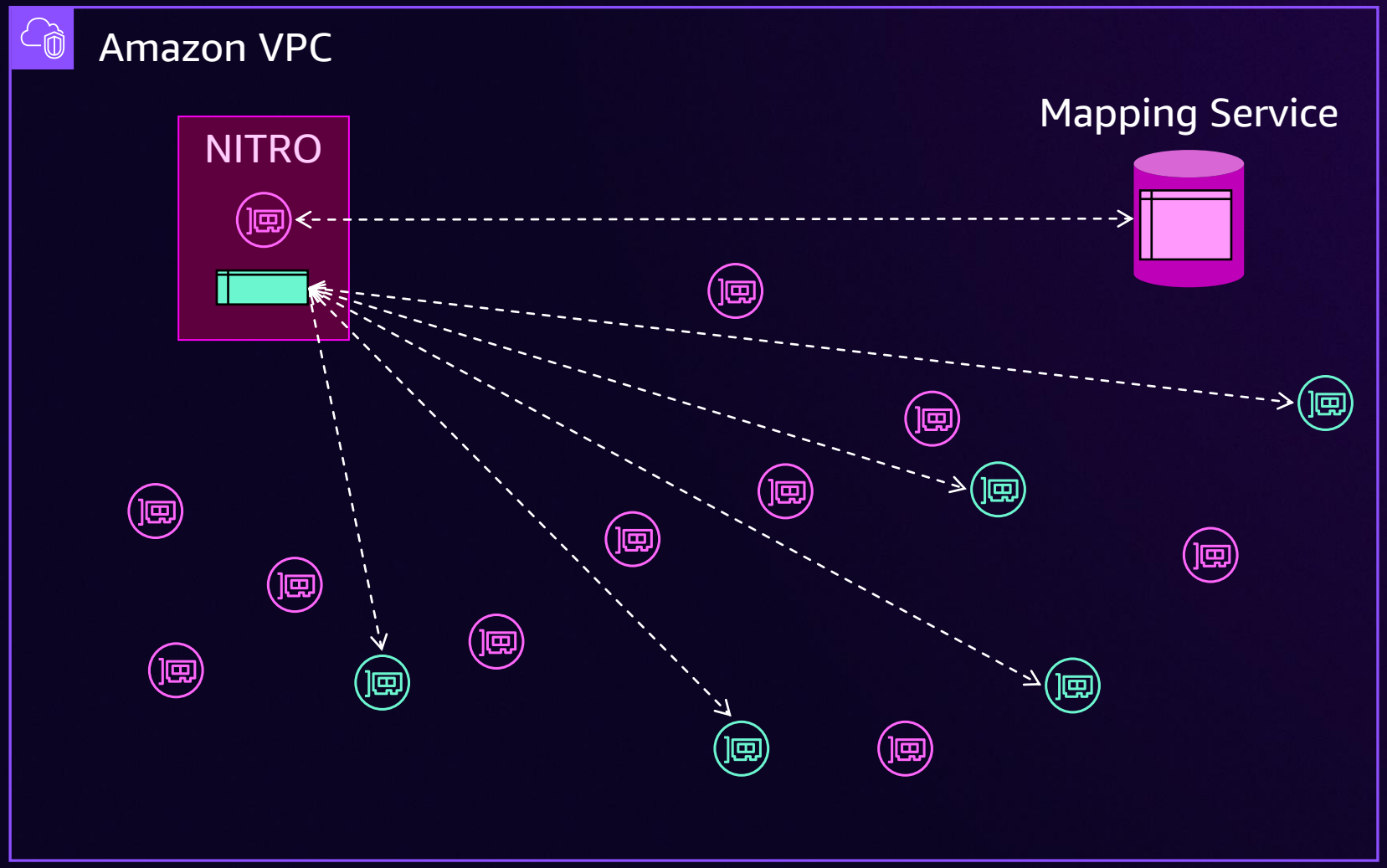
VPC Network Address Units



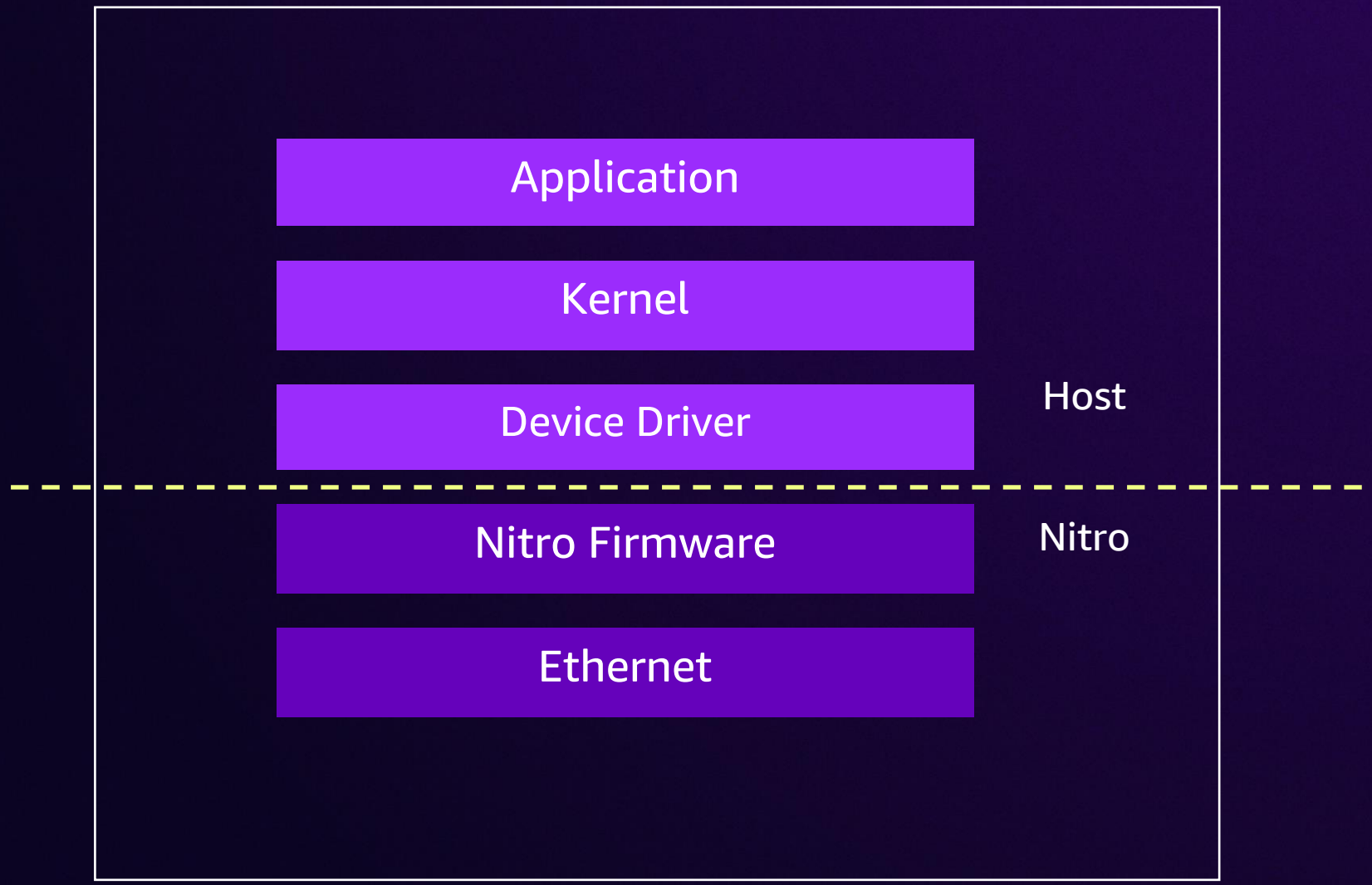


VPC Network Address Units

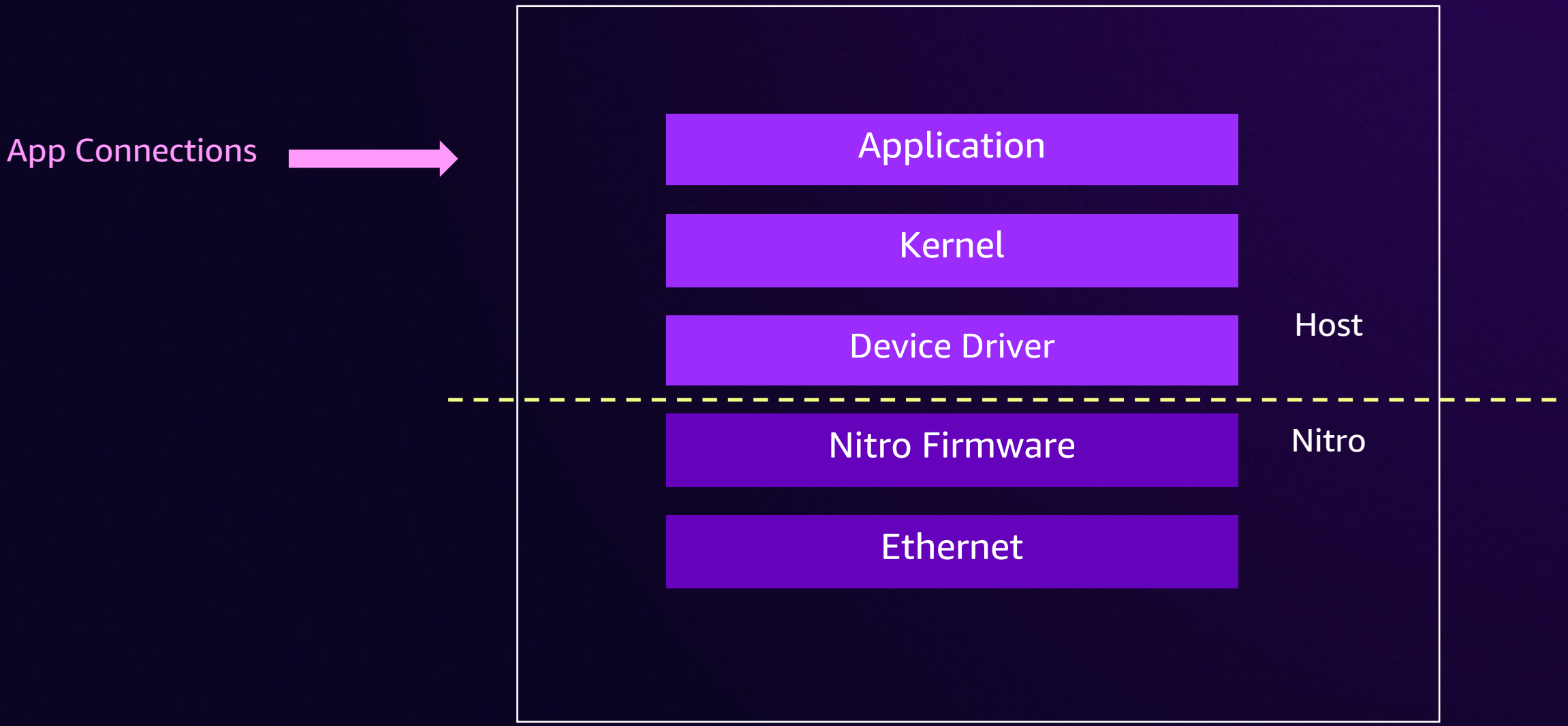




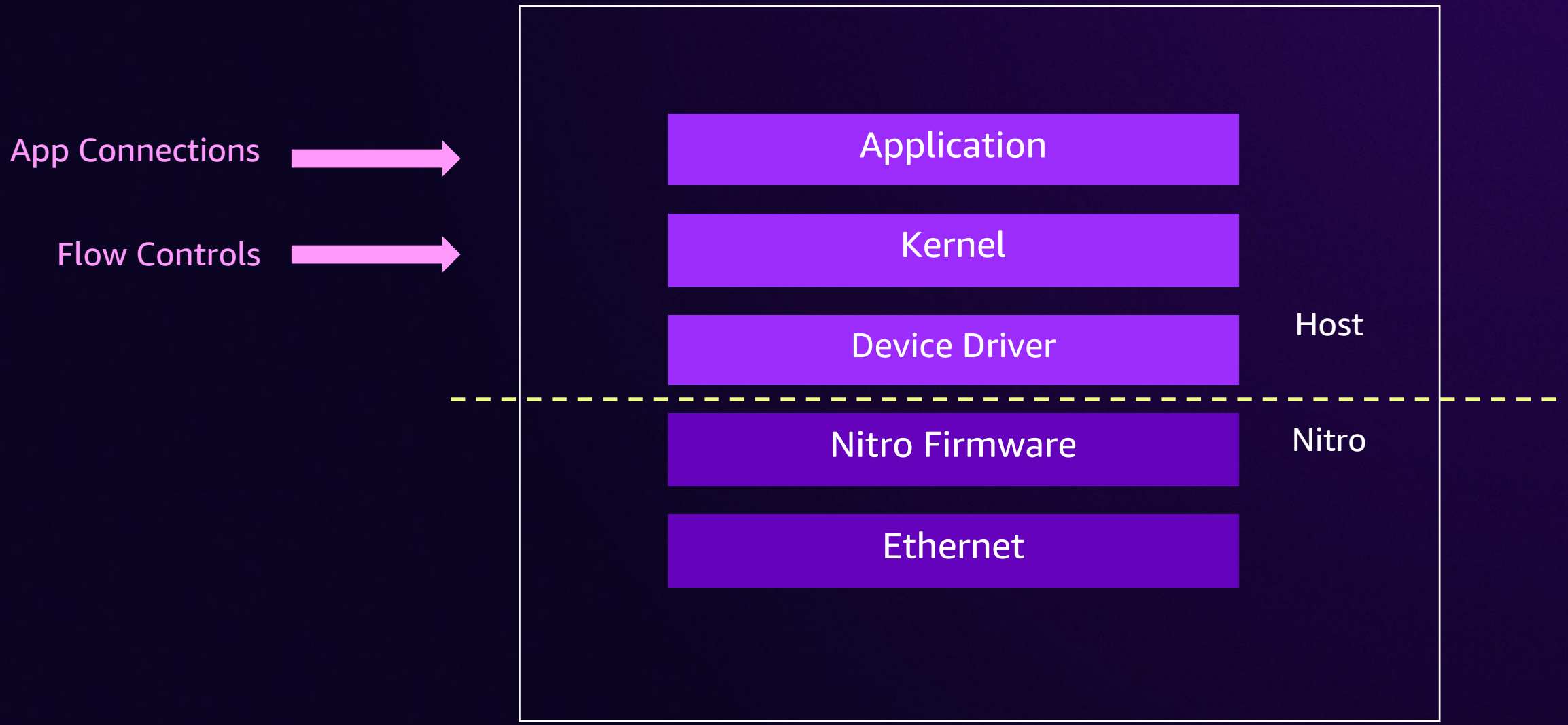
Elastic Network Adapter (ENA) Stack



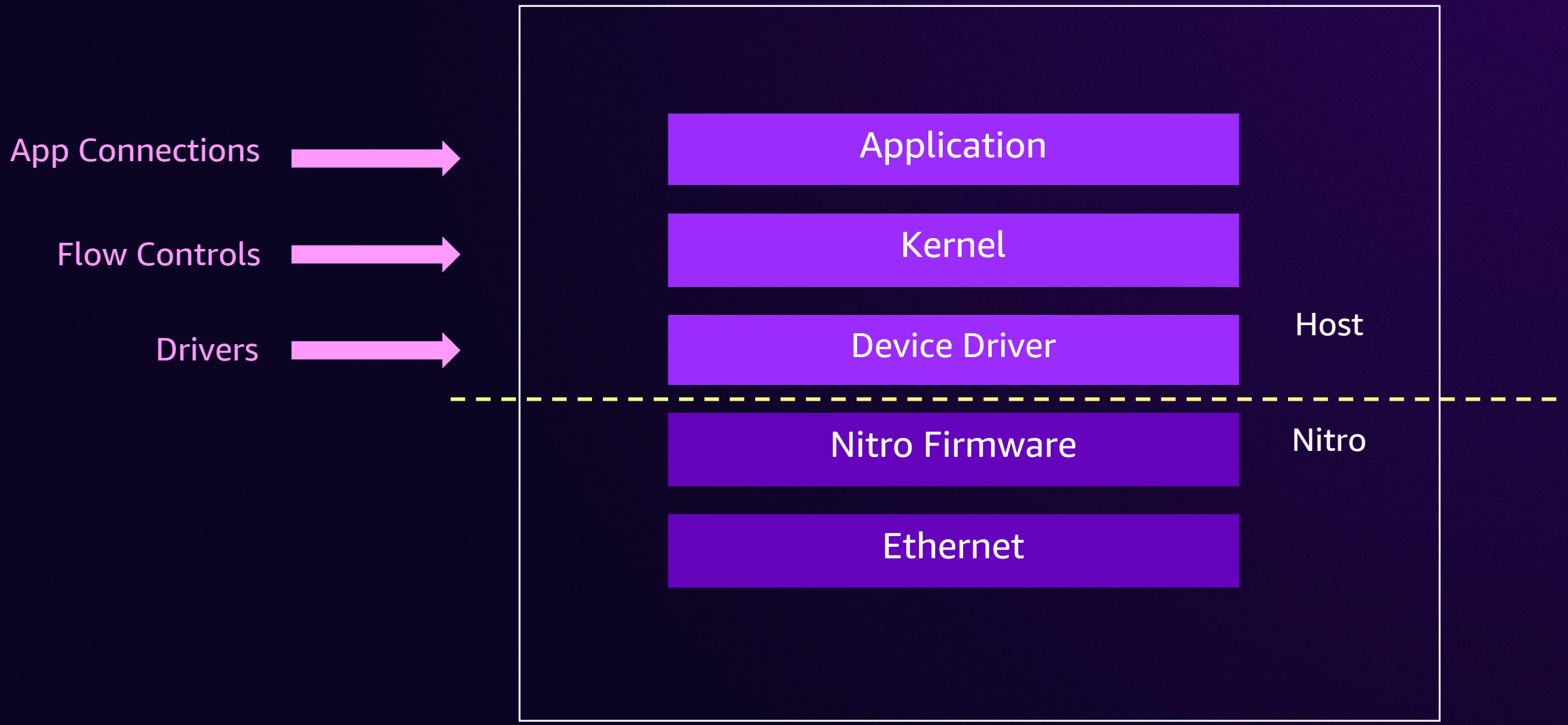
Elastic Network Adapter (ENA) Stack



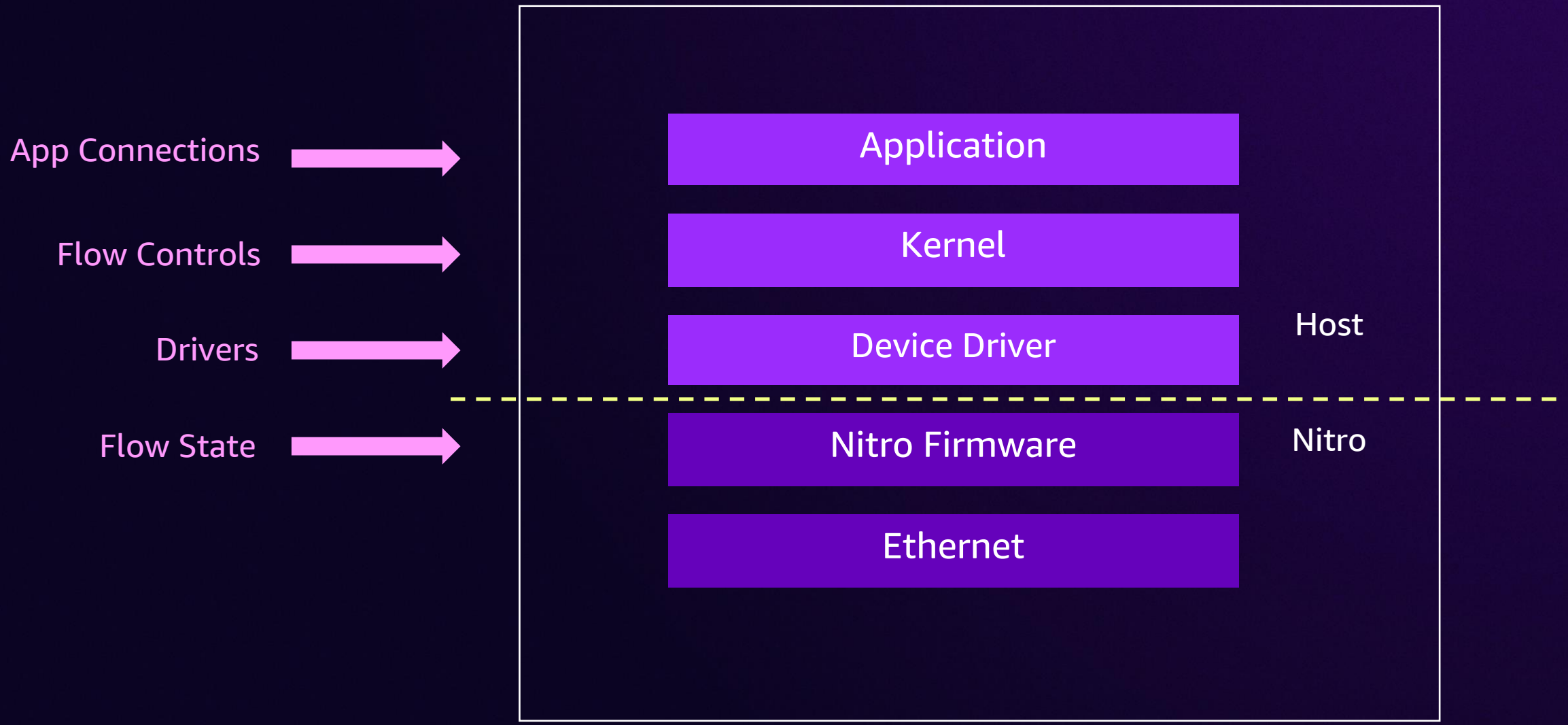
Elastic Network Adapter (ENA) Stack



Elastic Network Adapter (ENA) Stack

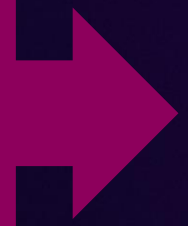


Elastic Network Adapter (ENA) Stack



Action Plan for Optimizing Nitro Throughput

1. Know your traffic types
   a. ‘Whale flows’
   b. TCP / UDP Flow Entropy
   c. Single vs. Many flows



Action Plan for Optimizing Nitro Throughput

1. Know your traffic types
   a. ‘Whale flows’
   b. TCP / UDP Flow Entropy
   c. Single vs. Many flows

2. Know the traffic type’s profile
   a. Packets Per Second
   b. Bits Per Second
   c. Connections Per Second

Action Plan for Optimizing Nitro Throughput

1. Know your traffic types
   a. ‘Whale flows’
   b. TCP / UDP Flow Entropy
   c. Single vs. Many flows

2. Know the traffic type’s profile
   a. Packets Per Second
   b. Bits Per Second
   c. Connections Per Second

3. Choose your Instance Type
   a. Baseline and Burst Specifications
   b. Nitro Performance Capabilities
   c. Density of Elastic Network Interfaces

Action Plan for Optimizing Nitro Throughput

1. Know your traffic types
   a. ‘Whale flows’
   b. TCP / UDP Flow Entropy
   c. Single vs. Many flows

2. Know the traffic type’s profile
   a. Packets Per Second
   b. Bits Per Second
   c. Connections Per Second

3. Choose your Instance Type
   a. Baseline and Burst Specifications
   b. Nitro Performance Capabilities
   c. Density of Elastic Network Interfaces

4. Tune and Monitor your Instances
   a. Driver Tuning
   b. Ethtool Metrics
   c. VPC Metrics

Additional Resources

Nitro Tuning Performance User Guide

Nitro Generation User Guide

ENA Driver Best Practice

Connection Tracking

Network Address Units

Cloudwatch Agent

ENA Express

Microbursts

Networking Specs

Additional Presentations Recommended

NET201-INT: The power of cloud network innovation
CMP215-INT: Compute innovation for any application, anywhere
NET301: Amazon VPC: Advanced design and what’s new
NET317: AWS Application Networking: Build Simple, secure, and reliable apps
CMP301: Dive Deep into the AWS Nitro System
CMP320: AWS Graviton: The best price performance for your AWS workloads

Thank you!

Please complete the session survey in the mobile app

John Pangle
(he/him)
Senior PMT
AWS

Scott Wainner
(he/him)
Principal SA
AWS