VPFC Planning Scheme Based on VPFP in scaling Detnet

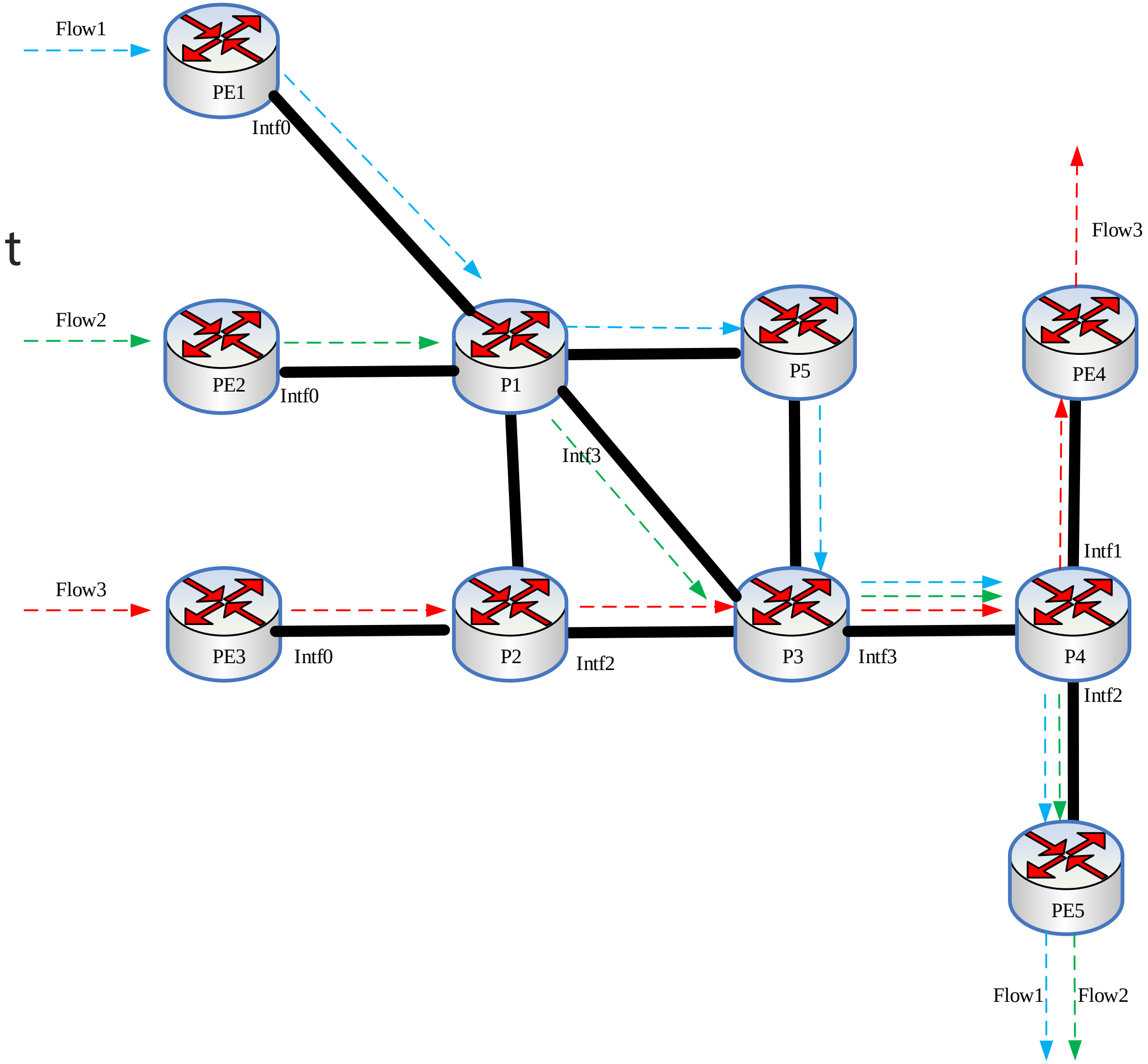
draft-guo-detnet-vpfc-planning-01

D. Guo, G. Wen, New H3C
K. Yao, China Mobile(Presenter)
G. Peng, BUPT

IETF 116 Detnet

Introduction

- CSQF is effective for bounded jitter and latency
- Prerequisite of CSQF is that queuing resources don’ t conflict
- In Scaling Detnet, non-linear topology will make queuing resource collision problem more difficult to solve
- When conflict happens on the converged point, change the planning cycle of the head node and perform cycle calculation to eliminate the collision.



VPFP

Virtual Periodic Forwarding Path(VPFP) : In the forwarding path, the virtual path forwarding based on the cycle and the mapping relationship between cycles is called a VPFP. The VPFP has the following characteristics:

- The outbound interface of each node in the forwarding path supports cycle-based forwarding;
- In each segment link of the path, there is a mapping relationship between the scheduling cycle of the outbound interface of the upstream node and the scheduling cycle of the outbound interface of the downstream node;

```
+-- uint16 vpfpid # virtual periodic forwarding path identifier
+-- uint8 cycles # is the number of cycles used across all
                 # interfaces in the CSQF/TCQF domain.
+-- policy_info [policy] # Policy information, e.g. SRv6 policy
+-- pipe_info[0..cycles-1] # The scheduling cycle pipeline
                           # corresponding to each scheduling cycle
                           # on the head node
    +-- uint8 hops # Number of hops
    +-- map_info[0..hops-1] # The mapping target in each pipeline
                            # is a specific scheduling cycle
        +--uint8 out_cycle #output cycle
```

Figure 5: VPFP configuration data structure

VPFC

Virtual Periodic Forwarding Channel (VPFC) : A forwarding channel established on VPFP. The basic elements of a VPFC are:

- VPFCID(VPFC Identifier). VPFCID is an integer that uniquely identifies a VPFC within the same deterministic periodic forwarding domain;
- VPFP. VPFP is the path that carries the VPFC;
- Cycle Info. Cycle Info contains the scheduling cycle and the resources corresponding to the scheduling cycle, describes the bandwidth and periodicity characteristics of the VPFC, and is the result of resources reservation;

```
+-- uint16 vpfcid # Virtual Periodic Forwarding Channel Identifier
+-- uint16 vpfpid # virtual periodic forwarding path identifier
+-- if_config[oif] # Outgoing InterFace configuration
    +-- uint16 cycles # Number of cycles involved in resources
                      # reservation
    +-- cycleinfo[0..cycles-1] #Cycle Info
        +--uint16 cycleid #Cycle ID
        +--uint16 res #Number of Resources
```

Figure 4: VPFC configuration data structure

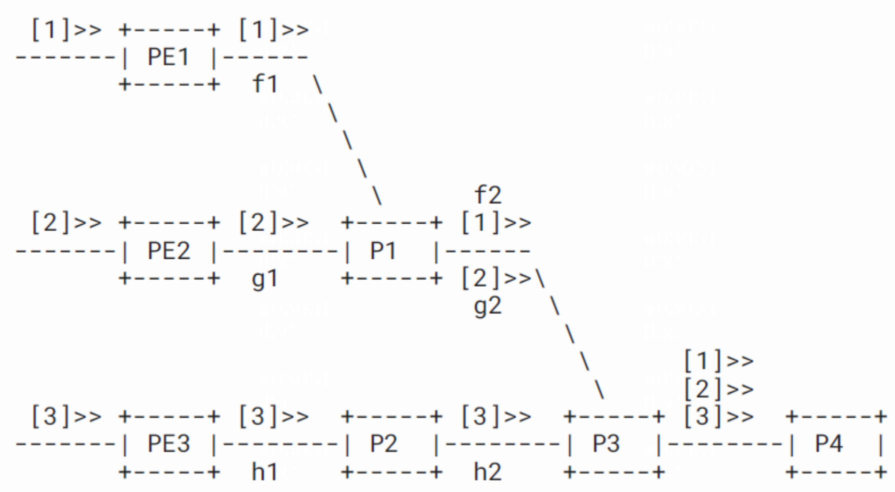
Mapping Function

- After the calibration is completed, a definite mapping relationship (mapping function) is established between the scheduling cycles of the outbound interface of the two adjacent nodes.

  Some constraints can be imposed on the mapping relationship:

  - During calibration, any scheduling cycle in the downstream node has one and only one scheduling cycle calibrated with it in the upstream node. Under this constraint, the function f becomes an injective function.
  - In the same CSQF domain, all interface plan the same number of scheduling cycles. Under this constraint, all mapping functions have the same domain and range.

- A composite function formed from a set of injective functions along a forwarding path is still an injective function. In the figure on the right, f = f2(f1), g = g2(g1), h = h2(h1) are all injective functions.



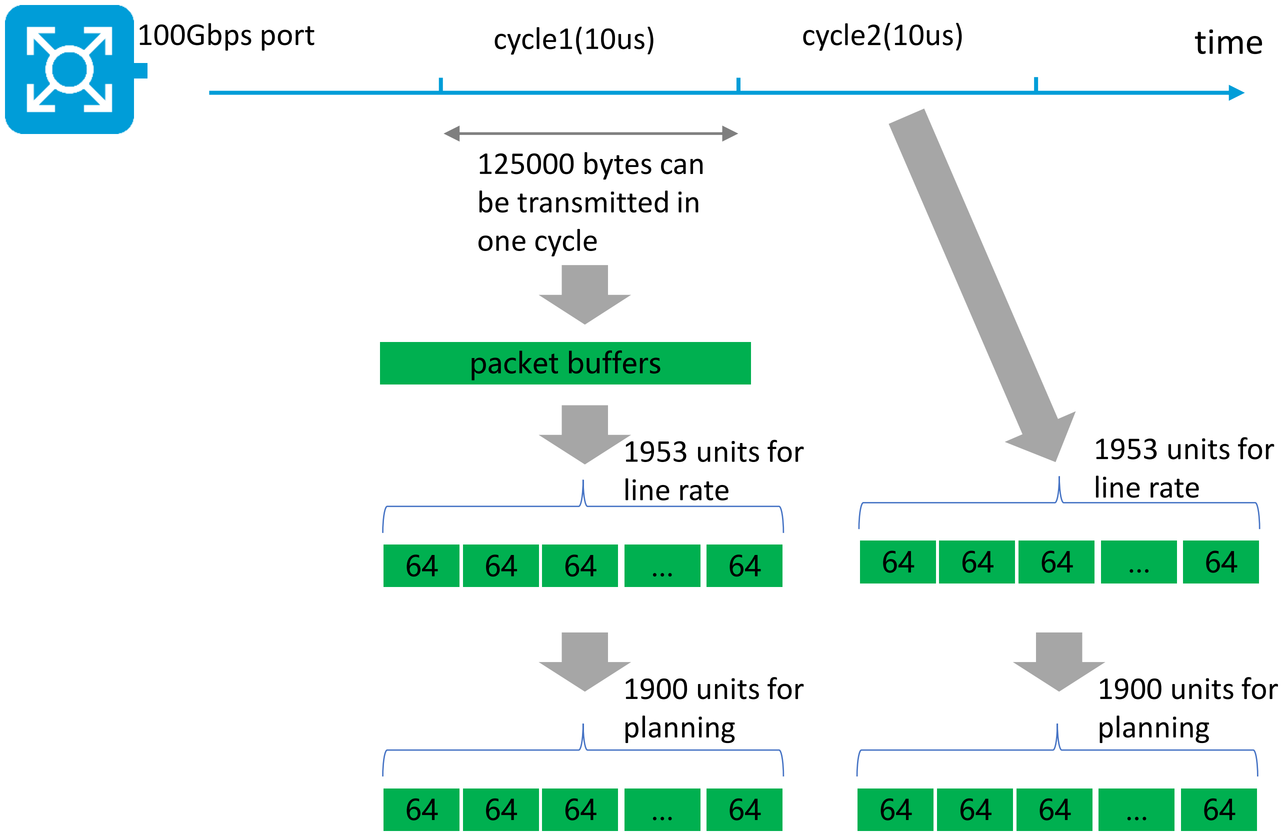
- Changing the cycle selection on the ingress PE will make all the cycles selected along the path change, which gives a easy way to solve cycles confliction.

Resource Quantification

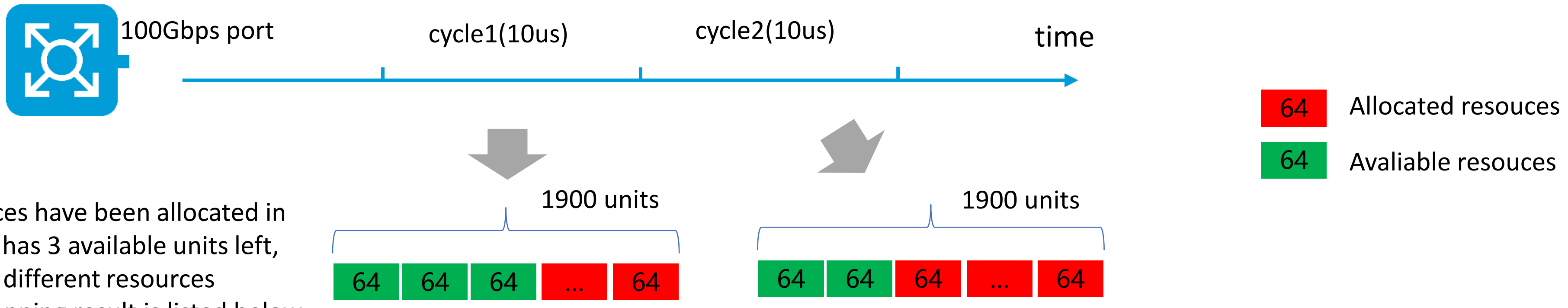
- Each cycle has one or more queues for buffering packets data. The amount of buffering resources is critical for cycle based forwarding capabilities. For a 100Gbps port, about 125000 bytes buffers is required to achieve full line rate.

- In our scheme, the amount of buffer resources is expressed in resource units to simplify the resources planning. If a resource unit is 64 bytes long, a 100Gbps port requires 1953 resources units in one cycle to achieve full line rate.

- For factors that have not been considered, some capabilities need to be reserved. For example, the available planning number of resources can be initialized as 1900 units for a 100Gpbs port.



Resource Planning



Assume most resources have been allocated in cycle 1 and 2, cycle 1 has 3 available units left, cycle 2 has 2 left. For different resources requirements, the planning result is listed below.

Requirements	Enough in cycle 1 ?	Enough in cycle 2 ?	Allocating result
1 100-bytes-long data	Yes	Yes	2 units from cycle 1 or 2
1 150-bytes-long data	Yes	No	3 units from cycle 1
1 100-bytes-long data ＋ 1 100-bytes-long data	Yes for 1 100-bytes-long data	Yes for 1 100-bytes-long data	2 units from cycle 1 2 units from cycle 2
1 250-bytes-long data	No	No	Fail

** Note : For simplicity, we use 2 cycles to illustrate the process of resource planning. In actual scenarios, more cycles will be used.*

Overall Processing

1. The data plane generates topology information. IGP (OSPF or IS-IS) collects the network topology information, BGP-LS summarizes the topology information discovered by IGP protocol and sends it to the MCPE (management/control plane entity). The MCPE stores this information in the topology database of the control plane.
2. MCPE measures the transmission delay between nodes through NQA or TWAMP. This delay is a relatively coarse granularity in accuracy, usually in milliseconds.
3. MCPE obtains the link transmission delay measurement results through NETCONF [RFC6241]/YANG [RFC6020] and uses the results with less accurate delay info to update the topology data.
4. User provides flow information required to establish a session.
5. MCPE uses CSPF to calculate the end-to-end path to obtain an optimal path, or multiple paths with close propagation delays for PREOF.
6. The MCPE performs accurate segmentation measurement on the forwarding path. According to the measurement results and the resident delay in the node, the correlation mapping is established to form VPFP. VPFP is delivered to the data plane and integrated into the forwarding table to direct data forwarding.
7. MCPE uses the resources planning scheme to reserve resources. After the planning is successful, the result forms VPFC.
8. MCPE delivers the planned resources to the head node (e.g., Ingress PE) of VPFP, and creates VPFC.
9. The head node of VPFP schedules the data of the DetNet flow according to the cycle resources owned by the VPFC to which the DetNet flow belongs.
10. MCPE maintains the states of VPFP and VPFC, if some path becomes unavailable, new planning is performed.
11. MCPE recycles the resources when needed.

Thank you!