ForCES Forwarding Element Model

<draft-ietf-forces-model-01.txt>

Steven Blake

Presentation available at:

http://www.petri-meat.com/slblake/networking/drafts/forces-model.pdf

Authors

- Lily Yang <lily.y.yang@intel.com>
- Joel Halpern <jhalpern@megisto.com>
- Ram Gopal <ram.gopal@nokia.com>
- Alan DeKok <alan.dekok@idt.com>
- Zsolt Haraszti <zsolt.haraszti@ericsson.com>
- Steven Blake <steven.blake@ericsson.com>

Changes in -01

- Last version publicly presented was `<draft-yang-forces-model-02.txt>` in Vienna.
- Merged parts of `<draft-haraszti-forces-model-00.txt>`; Haraszti & Blake added as co-authors.
- Document substantially re-organized.
- Directions taken:
  - Distinguish between FE model and LFB class model.
  - Allow packet datapaths to be represented using both topological and encoded state.
  - Allow LFBs to have multiple inputs and outputs, and do not exclusively use explicit Mux/Redirector LFBs in the general case.

Open Issues in -01

- Inter-FE topology
- Metadata
- Multiple LFB inputs/outputs
- Dynamic LFB topology reconfiguration
- LFB class partitioning
- Modeling port LFBs
- Modeling lookaside LFBs
- Modeling queueing and scheduling LFBs
- Modeling classification

Inter-FE Topology

- WG has concluded that the ForCES protocol should be capable of conveying inter-FE topology information queried by the CE from the FE:
  - FE may not have this information.
  - CE may use other means to acquire it.
- Alternatives:
  - Should inter-FE topology be part of the FE information model (i.e., part of the FE attributes)?
  - Or, should inter-FE topology be handled in the base ForCES protocol?

Metadata: Examples (1)

- ForCES protocol parameters:
  - Want to bind a next-hop to an output interface.
  - ForCES associates value *X (handle)* with output interface resource *B* when it is created in the output interface LFB.
  - Use ForCES to configure next-hop resource *A* with value *X in the next-hop LFB*.
  - *X* is the value of metadata item META_IFID conveyed between the next-hop LFB and the output interface LFB within the FE information model.
- FE implementation parameters:
  - Next-hop datastructure (*A)* has pointer *Y* to output interface datastructure (*B)*.
  - *Y* is read from a stack when the output interface routine is executed.

Metadata: Examples (2)

- Physical metadata:
  - Ex/: 16-bit value *W* is conveyed in parallel with a packet across a bus between components on a FE board.
  - Ex/: 32-bit field *Z* is appended to a packet as it is forwarded across a switching fabric between FEs (ForCES reference point Fi).
- Implicit metadata:
  - Packets arriving on a particular LFB input.

Metadata: Categories from Alan's draft

- Internal vs. External:
  - Internal metadata:
    - Not visible outside an LFB.
    - Ex/ Frame retry timer within a component implementing the Ethernet port LFB function.
  - External metadata is visible outside the component.
- Implicit vs. Explicit:
  - Implicit metadata equivalent to pre-/post-conditions on LFB input/output.
  - Explicit metadata:
    - Visible outside of an LFB.
    - Conveyed separately with a packet.

Metadata: ForCES scope

- Definitely in-scope:
  - ForCES protocol parameters, reflected as metadata in the information model.
- Definitely out-of-scope:
  - FE datastructure parameters.
  - Internal metadata.
- Possibly in-scope:
  - Implicit metadata?
    - Could be used to validate a given LFB topology.
    - Debugging
  - Physical metadata on Fi?
- Model metadata pass-through limitations in LFBs?

Multiple LFB Inputs/Outputs

- LFBs can support multiple inputs and outputs.
- LFBs can support multiple input groups and output groups.
  - Input group: Set of inputs that are expected to receive packets of the same type(s) with the same metadata elements, and are processed in the same way by an LFB. The input number within the group may be used as implicit metadata.
  - Output group: Set of outputs that produce the same type(s) of packets and the same metadata elements. Used as a redirector embedded within an LFB.
- Only use generic Mux LFBs for special cases.
- Only use Redirector LFBs when forking based on metadata not generated by the adjacent upstream LFB.

Dynamic LFB topology reconfiguration

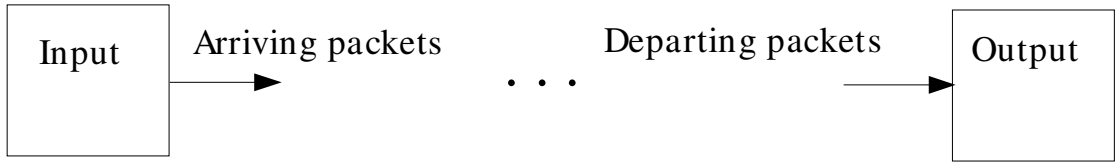
- Use cases:
  - Good for enabling support for new protocols?
    - Ex/ Turn on L2TP support.
  - Good for configuring a QoS reservation for a new RSVP flow?
    - Necessary if topological approach is used to represent the packet datapath.
    - Not (usually) necessary if encoded state approach is used instead.
- Dynamically downloading FE firmware from the CE is out-of-scope for ForCES.
- Is it safe to assume that an FE stops forwarding while the topology is reconfigured?

LFB class partitioning

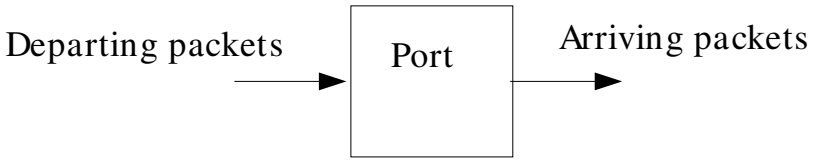
- Need to decompose common FE forwarding operations into sets of LFB classes.
- Example: IPv4 forwarding LFB vs. IPv4 header check LFB + LPM LFB + Next-hop LFB:
  - Finer-grain partitioning allows LFBs to be reused support of multiple protocols.
  - Finer-grain partitioning allows intermediate LFBs to be inserted.
  - Coarser-grain simplifies the topology.
- We need to come up with the right balance.
- Prior art available from NPF, Click, Linux TC.

Modeling port LFBs

- Can represent input and output ports in two ways:
  - Split into separate anchor nodes in the graph:



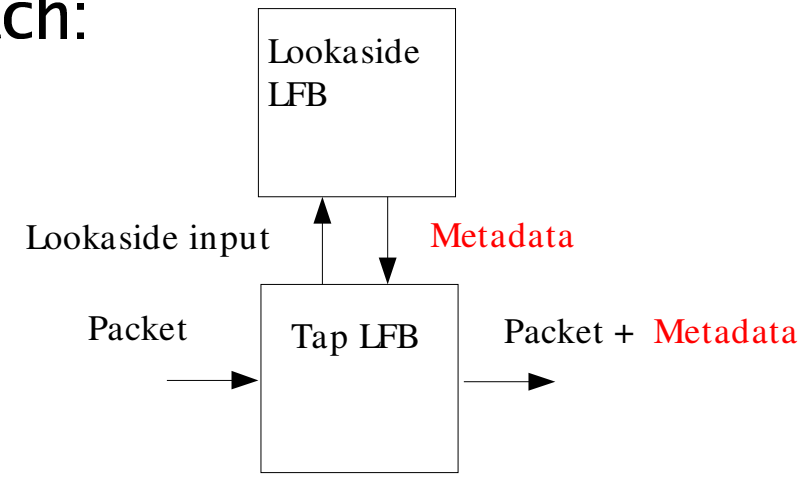
  - Combined in one graph node:



- Either case requires that port LFBs be special cased.
- Modeling issue exists for other types of LFBs:
  - L2: encapsulation/decapsulation
  - IP interface: ingress/egress

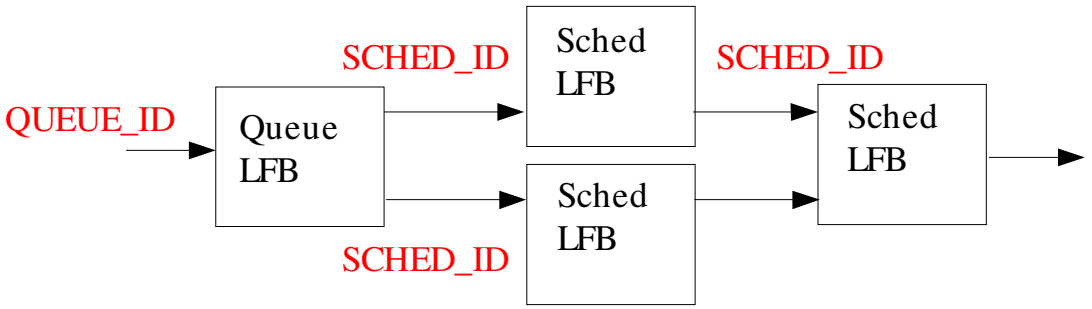
Modeling lookaside LFBs

- LFB topology models to date have assumed that packets flow through LFBs.
  - Except port LFBs
- Need to model devices that may receive part of a packet (e.g., header fields), and produce metadata, but otherwise are not in the direct datapath.
- One approach:



Modeling queueing and scheduling LFBs

- Need to model queueing/scheduling:
  - FIFO queues+ queue management:
    - AQM or Tail-drop
  - Schedulers:
    - PQ, WFQ, DRR, etc.
    - Non-work conserving (e.g., shaping)
    - Hierarchical scheduling
- Schedulers "pull" packets from queues:
  - Does this impact the model?
- One approach:



Modeling classification

- Classification LFBs
  - Generic:
    - Prefix/range/masked match on multiple header fields simultaneously
  - LPM?
  - Offset-length (e.g., fork on single header field)?
- Assume that protocol-specific LFBs look at appropriate fields within the packet:
  - Ex/ IPv4 header verify LFB

Upcoming Work Items for -02

- LFB class XML schema (Sec. 4).
- LFB topology XML schema (Sec. 5).
- FE-level attributes and capabilities XML schema (Sec. 6).
- Improve the LFB examples (Sec. 7).
- Provide examples to show that this model satisfies the FE model requirements in `<draft-ietf-forces-requirements-10.txt>` (Sec. 8).