Composite Link Requirements draft-ietf-rtgwg-cl-requirement-00.txt

Ning So	ning.so@verizonbusiness.com
Andrew Malis	andrew.g.malis@verizon.com
Dave McDysan	dave.mcdysan@verizon.com
Lucy Yong	lucyyong@huawei.com
Fredric Jounay	frederic.jounay@orange-ftgroup.com
Yuji Kamite	y.kamite@ntt.com

Differences between WG and Prior version

- draft-so-yong-mpls-ctg-requirement-00.txt adopted as working group draft
- Based upon mailing list discussion regarding adoption as working group draft;
  - Response to comment thread originated by Lou and wg acceptance
    - All references to Composite Transport Group and acronym CTG replaced with composite link
    - Definition of composite link (cl) from ITU-T G.800 used
  - Response to comment thread originated by Dimitri
    - Reworded 2nd bullet, section 3.2 -Motivation as partial response
    - Suggest further description of problem statement TBD

Revised 3.2 Motivation Bullet

- Advertisement of each component link into the IGP. Although this would address the problem, it has a scaling impact on IGP routing, and was an important motivation for the specification of link bundling [RFC4201]. However, there are two gaps in link bundling:
  - 1. It only supports RSVP-TE, not LDP.
  - 2. It does not support a set of component links with different characteristics (e.g., different bandwidth and/or latency).
- For example, in practice carriers commonly use link bandwidth and link latency to set link TE metrics for RSVP-TE. For RSVP-TE, limiting the component links to same TE metric has the practical effect of dis-allowing component links with different link bandwidth and latencies.

Summary of Mailing List Discussions

- Curtis Villamizar originated comments (Email was titled: Composite Link Requirements), broken down into three threads:
  - #1 Composite Link Trademark Issue
    - Appears that the conclusion of Andy's response and Curtis' reply is that "the abandoned trademark clash is in itself not relevant, only the fact that there at least was a prior implementation of an actively balanced link aggregation functionality." to be addressed in Thread #2.
  - #2. Acknowledgement of Prior Work
    - See Separate slide(s)
  - #3. Proposed Resolution of Comments
    - See Separate slide(s)

Acknowledgement of Prior Work

- When draft first submitted in October 2008, there was discussion about including IP traffic, and a decision was made to address the MPLS focused case first.
  - Only response on mailing list was from Curtis
  - Recommend asking the wg to re-evaluate this decision. (No replies in response to this Email thread).

Composite Link Terminology

- Current text in section 2.2 is a paraphrase. Could replace with the following quote from section 6.9.2 of G.800, if that is wg consensus:
  - "Multiple parallel links between the same subnetworks can be bundled together into a single composite link. Each component link of the composite link is independent in the sense that each component link is supported by a separated service layer trail. The composite link conveys communication information using different server layer trails thus the sequence of symbols cross these links may not be preserved."
- Should clarify that G.800 text related to Inverse multiplexing is one of three cases in section 6.9.2, in the appropriate section 3.1.
- Suggestion to mention abandoned “composite link” trademark as “not what we mean.”

Suggested Areas of Change

- 1. Accurately characterize what exists today, what existing CL techniques have come before this, in use or not, and accurately characterize the common use cases of existing CL.
  - Include provided references, description of OMP and other implementations and flesh out the outline provided by Curtis on the mailing list.
  - If the characterization of existing CL gets too long it could be a separate informational internet-draft that is referenced.
- 2. State as a requirement (we are at the requirement stage) that to the exent possible new CL capability will:
  - 1. Continue to accommodate common use cases today, including an ability to carry IP traffic which MAY BE omitted in an implementation but MUST be accommodated, at least as an option, by any proposed solution.
    - Related to the MPLS-only or IP and MPLS scope decision.
  - 2. ?SHOULD? Retain backward compatibility with existing MPLS/GMPLS LSR with no loss of existing capability, but possibly no gain in functionality if the legacy LSR is anywhere on the LSP path include as an LER.

Proposed Resolution of Comments

- Terminology, Section 2
  - Traffic Flow Terminology
- Motivation, Section 3.1
  - ECMP/LAG
  - Link Bundling
  - Inverse Multiplexing
- Requirements, Section 4
  - 4. Structure
  - 4.1.1.1 Traffic Flow and Connection Mapping
  - Does any of section 4.2 belong in a requirement document, at least in its current form?
  - 4.2.2.1. Signaling Protocol Extensions

Traffic Flow Terminology, Section 2

- Replace
  - "Traffic Flow: A set of packets that with common identifier characteristics that the composite link is able to use to aggregate traffic into Connections. Identifiers can be an MPLS label stack or any combination of IP addresses and protocol types for routing, signaling and management packets.
  - *Note wording of above was intended to be MPLS label stack for "data plane" and IP/ protocol types for "control plane."*
- With
  - rfc2474.txt " Microflow: a single instance of an application-to-application flow of packets which is identified by source address, destination address, protocol id, and source port, destination port (where applicable).
  - rfc2475.txt:
    - "Microflow: a single instance of an application-to-application flow of packets which is identified by source address, source port, destination address, destination port and protocol id.
    - Traffic stream: an administratively significant set of one or more microflows which traverse a path segment. A traffic stream may consist ofthe set of active microflows which are selected by a particular classifier.
  - Proposed: MPLS microflow is properly identified by its entire label stack and if the payload is IP, by the IP source and destination address. If not IP as indicated by a PW CW, then only the label stack is used.
  - Reference the "Avoiding ECMP" section in rfc4385 which explains the reason to avoid certain values in the first nibble of payload and rfc4928 which provides the following terms
    - IP ECMP: A forwarding behavior in which the selection of the next-hop between equal cost routes is based on the header(s) of an IP packet.
    - Label ECMP A forwarding behavior in which the selection of the next-hop between equal cost routes is based on the label stack of an MPLS packet

Link Bundling, Section 3.1

- Drop "Link Bundle Bashing" bullet item and stick to the requirement, with proposed wording
  - "This document proposes that a link metrics allow for grouping together a set of parallel data links that may have different characteristics, and for advertising and operating them as a single link."

Section 3.1, ECMP/LAG Bullet

o ECMP/Hashing/LAG: IP traffic composed of a large number of flows with bandwidth that is small with respect to the individual link capacity can be handled relatively well using ECMP/LAG approaches.

- However, these approaches do not make use of MPLS control plane
- information nor traffic volume information.

+ While nothing precludes using traffic volume information, and
+ some implementations have done so, in practice few if any
+ implementations today make use of MPLS control plane information
+ or traffic volume information. Implementations commonly use the
+ entire MPLS label stack for non-IP MPLS traffic.

Distribution techniques applied only within the data plane can result in less than ideal load balancing across component links of a composite link.

Section 3.1, Inverse Multiplexing

- Inverse Multiplexing: Making multiple parallel links to appear as a single link using inverse multiplexing techniques, such as proposals under discussion in the [PWBONDING]. However, the inverse multiplexed link will have a latency of the link with the largest latency. When there is a mix of latency sensitive traffic with other traffic that is less sensitive to latency, having all traffic experience the latency of the worst link is not acceptable to operators.
- Add text to describe maximum packet processor and/or link bandwidth rates impact on the largest supportable "traffic flow."
- Also a comment to summarize issues with current use of metrics to approximate TE, but this appears broader than just Imux.

Section 4, Requirements Structure

As defined in the terminology section a (traffic) flow is the smallest unit of traffic assignable to a connection, and connections are assigned to a component link that is part of a composite link. The composite link has routing information, normal IGP that has no TE information and IGP with TE extensions (IGP-TE) and signaling information with TE information (RSVP-TE) and without TE information. The following table summarizes the three cases covered in this requirements section.

Traffic Flows	IGP	IGP-TE	RSVP-TE	LDP
With TE Info	Y	Y	Y	N
Without TE Info	Y	N	N	Y
With & Without TE Info	Y	Y	Y	Y

•Meaning of Y, N needs to be stated/ clarified.

Furthermore, if a requirement would be repeated for each of the above three cases (e.g., IGP related routing information) it is described in a section common to all flows.

Therefore, the outline of this section is structured as follows:

- Management/Measurement of Interior Functions
  - Functions common to all LSP flows
  - Functions specific to LSP flows with TE information
  - Functions specific to LSP flows without TE information
  - Sets of LSP flows with and without TE information
- Exterior Functions
  - Functions common to all LSP flows
  - Functions specific to LSP flows with TE information
  - Functions specific to LSP flows without TE information
  - Sets of LSP flows with and without TE information

•Does organization of document need to be clarified?

Traffic Flow and Connection Mapping

- "4.1.1.1. Traffic Flow and Connection Mapping
  - The solution SHALL support operator assignment of traffic flows to specific connections.
  - The solution SHALL support operator assignment of connections to specific component links.
  - The solution SHALL support IP packet transport across a composite link for control plane (signaling, routing) and management plane functions.
  - In order to prevent packet loss, the solution must employ make- before-break when a change in the mapping of a connection to a component link mapping change has to occur."
- Need further description of how this could work for signaled and statically configured LSPs.
- Scalability requirements/objectives, from (at least) a configuration standpoint should be described.
- Need to clarify what is meant by make-before-break is local to the composite link (i.e., not the current RFC 3209 RSVP definition that has an end-end tunnel semantic and not exactly like MPLS FRR either).

Signaling Protocol Extensions

- "4.2.2.1. Signaling Protocol Extensions Requirements
  - The solution SHALL support per LSP signaling of at least the following additional parameters for an LSP
    - Maximum (estimated or measured) acceptable latency
    - Actual (estimated or measured) accumulated latency based upon the actual component link assigned by the composite link
    - Bandwidth of the highest and lowest speed"
- Author proposed clarificaiton of last sub-bullet
  - Maximum and minimum acceptable bandwidth of the LSP
- Semantics needs to be described. Describe whether this would work with existing protocols, and if so, how.

Next Steps

- WG agreement on scope
  - Continue with current MPLS-only scope, or expand to IP and MPLS scope
- Agree on terminology:
  - Composite Link
  - Traffic Flow
- Augment problem statement section
  - Need more discussion on mailing list
- Augment description of prior work section
  - Break out as separate draft and reference if this gets too long
- Focus on requirements section
  - Organization/ structure
  - General direction of separating requirement from solution
  - Specific text changes

Acknowledgements

Co-Authors like to thank Adrian F., Lou B., Kireeti K., Eric G., Dmitri P., Curtis V., and other for reviews and suggestions