Moving Forward with Existing Proposals

Anna Charny

reflecting work of other authors of
draft-charny-pcn-performance comparison and many other people

History

- Summary of well-defined proposals as of IETF70
  - CL, 3SM, SM
- Functional comparison
- Summary of simulation efforts
- Pseudocode of the core marking behaviors encompassing all possible options
- Tabled to IETF71

New Developments

- Edge-based Marked Flow Termination
  - Approximates 3SM behavior by moving 3SM’s slow-down logic from core to edge
- LC-PCN draft clarification
- Reduction of encoding options due to uncovered tunneling issues
- AD feedback: scarcity of assigned DSCP codepoints define scope of viable solutions
  - Strong incentive to pursue 2 codepoint solution

This Presentation:

- Will try to reflect emerging consensus on how to move forward assuming a 2-codepoint solution (at least initially)
- Will clarify what may be lost by that (using key relevant points from draft-charny-pcn-comparison)
- Will NOT explicitly compare various proposals
  - see draft-charny-pcn-comparison and individual proposal drafts for details on that

Attempt to Summarize Emerging Consensus

- Ask for one global DSCP at this point
  - Tentative: can reuse the admitted-EF of draft-baker?
  - Use this DSCP and standardize core behaviour and PCN info message format to work with a 2-codepoint solution
- That means: either admission only OR termination only Or some schemes that does both with 2 codepoints
    - SM is the only proposal on the table today that does both with 2 codepoints
- Implication: if allow both admission and termination, then proposed behaviours must work with SM
  - But keep the door open to other options to the extent possible
- Use experimental DSCP for 3 codepoint solutions
  - Experiment = Understand whether/when 3 code point solutions needed/wanted by operators
- Describe boundary node behaviour that allows SM (informational)
  - Make it as general as possible without breaking SM
  - Keep the door open for other boundary node behaviours

Options With 2 Codepoints

- The Options:
  - Allow just admission
  - Allow just termination
  - Allow both admission and termination (SM)
  - Allow any of the 3 options and make operator choose/configure?
- NOTE: if allow SM with two codepoints, CL can be done by adding threshold marking when/if extra codepoint becomes available (also need minor changes to boundary behaviors that could be pre-built with SM)

What is Lost in the Only Admission or Only Termination Case?

- Need to configure which one you are using in the domain
- Don’t get the other one…
  - Is it acceptable to force either just admission or just the termination, but not both?
    - This presentation assumes must allow to have both
  - If must allow both, then solution must support SM
    - Unless and until another/better solution found and tested
- Anything else???

Assuming SM Must be Supported…

- **Core MUST do Excess Rate Metering and Marking**
  - A token bucket, which is sized in bits. It has a configured bit rate. Tokens MUST be added at the configured rate, to a maximum value TB.max
  - Tokens MUST be removed equal to the size of the metered-packet, to a minimum TB.size=0
  - If the token bucket is within an MTU of being empty, then the meter SHOULD indicate "excess-rate mark" to the Mark function. MTU means the maximum size of PCN-packets on the link.
  - If the token bucket is empty (TB.size = 0), then the meter MUST indicate "excess-rate mark" to the Mark function.

Other Things Core Node Should Do (if it must support SM):

- When doing excess-rate marking) SHOULD:
  - If the metered-packet is already “excess-rate marked”, then the Excess Rate Meter function SHOULD NOT be performed.
  - If the PCN-traffic level on the link is such that PCN-packets need to be dropped, then excess-rate marked packets SHOULD be preferentially dropped
  - If the PCN-traffic level on the link is such that the metered-packet is dropped, then the Excess Rate Meter function SHOULD NOT be performed on this packet

Other Things that Must be Defined

- PCN information exchange messages will contain (some of):
  - To be used to communicate PCN info from egress, to ingress and possibly PDP (wherever that is)
    - CLE
    - Sustainable Rate
    - Rate to terminate (optional: may be useful for PDP)
    - Ingress sending rate (optional: may be useful for PDP)
- Boundary Node Behaviors to be specified
  - Informational
  - Not in this presentation
  - Assumption: SM will be the initial one (assuming both admission and termination is needed)
  - Assumption: may define more that one boundary behavior

Limitations and Sacrifices (1)

- Core behavior definition:
  - The “SHOULD preferentially drop excess-marked packet condition” is problematic for 3SM and EMFT proposals in the presence of heavy loss
    - Limits the possibility of defining simpler edge behaviors
    - No edge behaviors that provably work with 2 codepoints are described as of today
  - Does not allow optimizations proposed in LC-PCN
    - Could be useful if termination decision made at the edge
    - Require additional implementation complexity at the core
    - Not fully understood at this time

Limitations and Sacrifices (2)

- Have only SM as two-function, 2-codepoint solution
- BUT SM has a number of known performance limitations compared to some of the 3-code point solutions:
    - when there are a small number of flows in ingress-egress aggregates
        - Not an infrequent case at all!
    - Some performance degradation in the presence of multiple simultaneously congested bottlenecks
    - Discussed in draft-charny-single-marking presentation later today
- SM is suboptimal for ECMP support
    - need 3 code-points to fix
- SM is suboptimal for support of probing
    - Need threshold-marking for admission to fix
- Does not allow simpler edge implementations possibly afforded by 3SM and EMFT solutions

What about Threshold Marking?

- Core MUST do Threshold Metering and Marking if:
  - want to experiment with 3 codepoints
  - want to allow just admission
- Threshold marking defined by Phil on Tuesday
  - Not changed and not discussed in this presentation in detail

What Needs to Happen to Move On?

- Reach consensus on which 2-codepoint solution to pursue
  - Assuming there is consensus that 2 codepoint is what we must do
- Agree on specific encoding
  - two choices (but not discussed in this presentation)
- Turn slides into appropriate core behavior draft
- Specify any signaling requirements
- Specify (informational) boundary behaviors

That is it

Thank you!

BACKUP

- The following slides summarize some of the draft-charny-pcn-comparison conclusions

Marking and Encoding

	SM	3SM	CL	EMFR	LC-PCN
# encoding states	2	3	3	3	2 (3 with AfM)
# metering mechanisms in forwarding path	1	2	2	2	1
Type of marking for admission	excess	threshold	threshold	threshol d	Excess or rate msremnt with proportional marking
Type of marking for termination	not required	Excess with slowdown	excess	excess	Not required

- All existing proposals except 3SM and LC-PCN can be supported with threshold and/or excess rate marking
- 3SM and LC-PCN need additional core functionality
  - But EMFR can approximate 3SM without this additional core functionality. However, performance results are preliminary

Caveats: other differences

	SM	3SM	CL	EMFR	LC-PCN
Look at marking prior to metering?	yes (do not meter excess-marked packets)	yes (put token buckets in if packet excess-rate marked;	Yes (do not meter excess–marked packets;)	Yes (do not meter excess-marked packets	yes (do not meter excess-marked packets)
Re-Mark a previously marked packet	n/a	Do not remark excess to threshold	Do not remark excess to threshold	Do not remark excess rate to threshold	n/a
Drop preference in case of packet loss	Drop excess marked pkts first	Prefer not to drop excess-rate marks but OK if some dropped	Drop excess-rate marks first	Prefere not to drop excess rate marks but OK if some dropped	depends thres. set. Typically, prefer not to drop ex. rate

- **Choice of algorithm defines "red" behaviors (CL, SM, LC-PCN vs 3SM or EMFR)**
- **Orange behaviors might be OK?**
- **Green the same for all**

Other Differences: Decision Location

- Admission Decisions
  - At ingress for CL and SM as described
    - But OK to do at egress
  - At egress for 3SM, EMFT and LC-PCN
- Termination decisions
  - At ingress of CL and SM
    - Could do at egress with performance degradation
  - At egress for 3SM, EMFT and LC-PCN
  - Note: if ingress decides termination, can police/drop packets while signaling deals with teardown (could be substantial delay); egress cannot do it

Other differences: what is signaled

- CL and 3SM:
  - CLE and Sustainable Rate as described
    - The meaning of these are slightly different between CL and 3SM, but the format is the same
    - Note: if admission decision moved to egress, then just Sustainable Rate will need to be signaled
  - 3SM, EMFT and LC-PCN
    - Nothing for admission
    - Set of flows to terminate for termination

Performance Comparisons

- Extensive apples-to-apples CL to SM comparison
- Substantial 3SM simulation study
- Some amount of simulations of EMFT
  - Conjecture: close to 3SM?
- No simulations of LC-PCN as of today
- Across-the-board performance comparisons difficult due to lack of apples-to-apples simulations

Other Comparisons

- Probing
  - Out of scope now but:
    - SM and LC-PCN need many probes to reliably decide admission
      - Router alert options has been suggested
        - » Performance impact a serious concern
    - CL and 3SM need just one probe
- ECMP
  - No direct support for admission other than by probing for any proposals
  - For termination
    - Good support for 3SM and EMFT
    - CL can support at the expense of signalling set of flows to ingress
    - SM is not accurate even if signals set of flows to ingress