Architectural Considerations for Load

(draft-campbell-dime-load-considerations-00)

Ben Campbell
IETF91, Honolulu

Draft is mostly questions, not answers

- How are Load and Overload Different?
- How is Load used?
- Piggyback vs dedicated application?
- Which nodes exchange load?
- What is the scope of a load indication?
- Do we need to negotiate?
- Usual security questions

Differences between Load and Overload

- How is Load different from Overload?

- Proposal:
  - Overload is an explicit request for action
    - Contract between Reporting and Reacting Nodes
    - Exceptional condition
  - Load is informational
    - Hint to the recipient
    - Always have a load value
  - No standardized link between Load and Overload

How is Load Used

- Input to multiple decisions
  - Target selection for diverted requests
  - General purpose load balancing?
  - Inferred overload?

Who uses Load

- Next Hop Selection
  - Always immediate peer

- Server Selection
  - Peer (same as next hop selection)
  - Client (Destination-Host)
  - Sometimes In-between nodes (Destination-Host)

Who Uses Load (cont)

- Do we need to support load for server-selection?

  - Potential for more complexity (c.f. DOIC)

  - Is a non-peer node that inserts D-H likely to be able to do anything useful with load?

Next Steps

- More discussion of these and the remaining questions on list

- Do we need an architecture draft?
    - Could expand this draft into one.

- Do we need requirements beyond what is in RFC 7068