

# Path Maximum Transmission Unit Discovery

draft-ietf-pmtud-method-01.txt

4-Mar-2004

Matt Mathis <mathis@psc.edu>  
John Heffner <jheffner@psc.edu>  
Kevin Lahey <kml@patheticgeek.net>

Comments to the list:  
pmtud {-request} @ietf.org

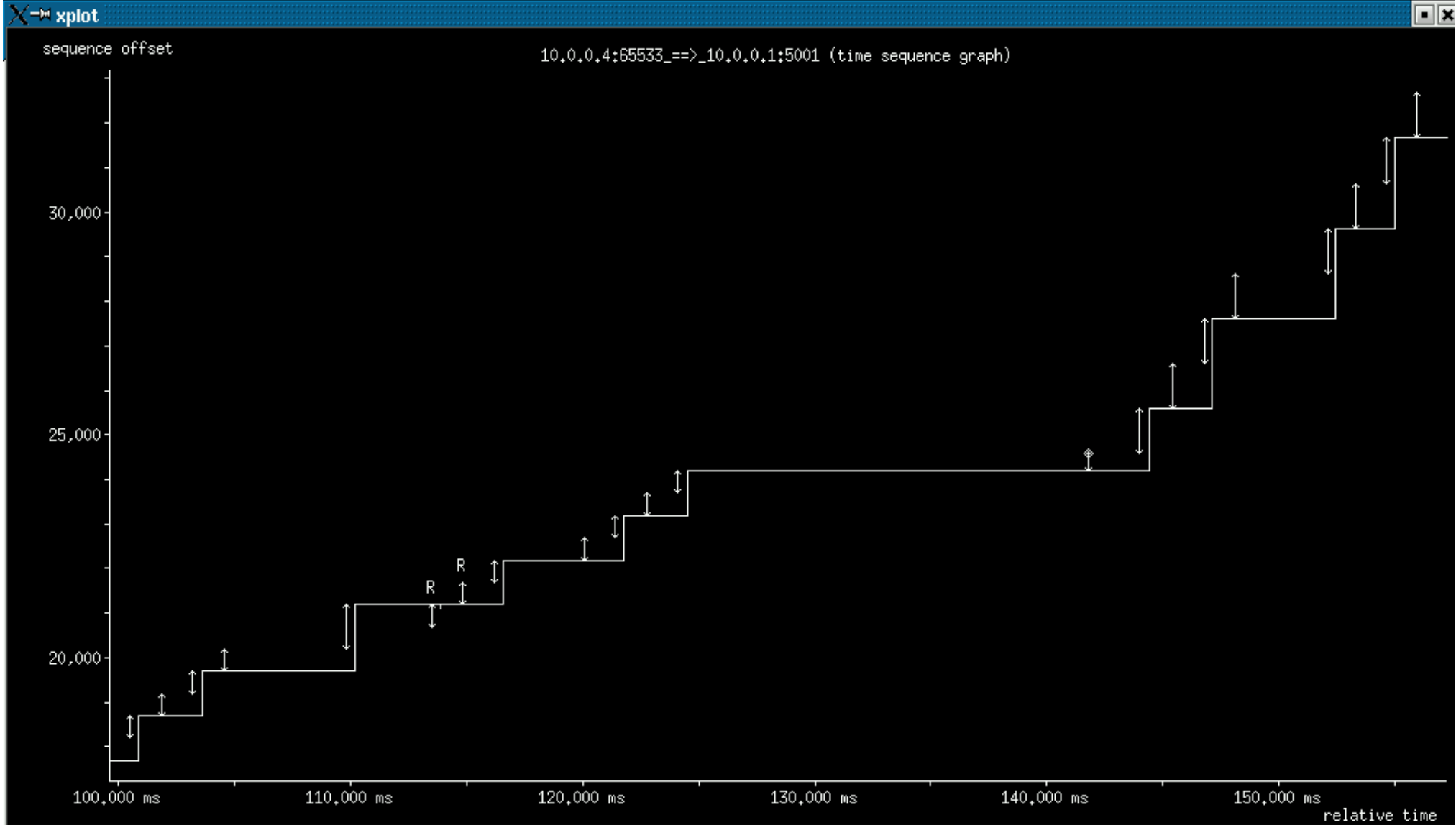
# Outline

- Current status
- Robustness Issues
- Next steps

# Algorithm Review

- Start with a "moderate" MTU (1k?)
- Test larger MTUs by probing
  - Raise MTU if successful
  - (Optional) process any RFC1191/1981 ICMP
  - Do not reduce TCP window on lost (unsuccessful) probes
- Most of the algorithm runs in the transport layer
  - TCP, SCTP, or higher layer (e.g. NFS)
- Keep cached/shared state in the IP layer
  - IP Maximum Payload Size (MPS)

# Running Code



# Key Point

- We are not defining a protocol
  - A method using existing protocols
- Implementation differences do not affect interoperability
- Careful thought to maximize robustness
  - Cover bugs in other parts of the stack

# Editorial Updates

- Recast as an extension to standard pmtud
  - Does not respecify ICMP based pmtud
  - (Prior restructure made it parallel to RFC1981)
  
- No longer TCP centric
  
- Ongoing terminology cleanups

# Algorithm Updates

- Main algorithm description is now in layer 3 terminology
  - (IP packet size, MTU, etc)
  - Much clearer
  
- Added a full RTT verification phase
  - Address the situations where raising the MTU causes excess loss
  
- Added extensive discussion of failure modes and recovery
  - (not quite complete yet)

# Call for implementations

- Core ideas are stable
- Implementation experience is the best way to clarify some details



# Open Robustness Issues

## ■ Two old items

- Routers or tunnels not honoring DF
- Repeated Timeouts

## ■ Resolved

- Raising MTU raises loss rate

# Not honoring IPv4 DF

- THIS IS A SERIOUS BUG in other gear
  - Routers and tunnels that violate DF
- Becoming MORE common
- Write "Fragmentation is worse than we thought"
  - What happens when we wrap the (16 bit) IP ID field?
  - Miss-associated fragments might have IP checksum errors

# Hard (Repeated) Timeouts

- On double(?) timeouts, reset MTU to 512(?)
  - and trigger timeout actions at other layers...?
    - ▶ router discovery, server pooling, etc
    - ▶ comprehensive language would be useful
- Do we need to consider the recurrent failure case?

# Plans for the Next Draft

- Still some obsolete and missing text
- Seeking contributors

# Comments please

pmtud {-request} @ietf.org