The Search of the Path MTU with QUIC

Timo Völker
Michael Tüxen

Dep. of Electr. Engineering and Computer Science
FH Münster University of Applied Sciences

Erwin P. Rathgeb

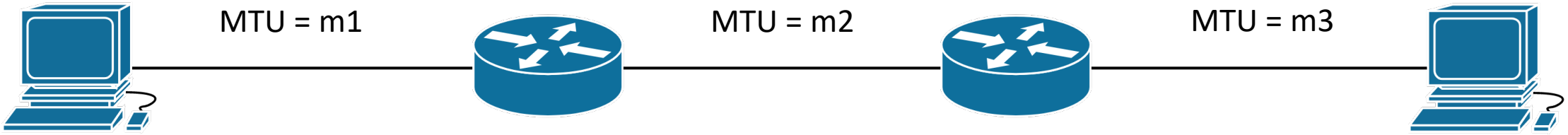
Computer Networking Technology Group
University of Duisburg-Essen

Agenda

1. Introduction
2. PMTUD Framework and QUIC
3. PMTU Search
4. Evaluation
5. Conclusion

Introduction

Path MTU



MTU = Maximum Transmission Unit

= IP packet size limit for a link

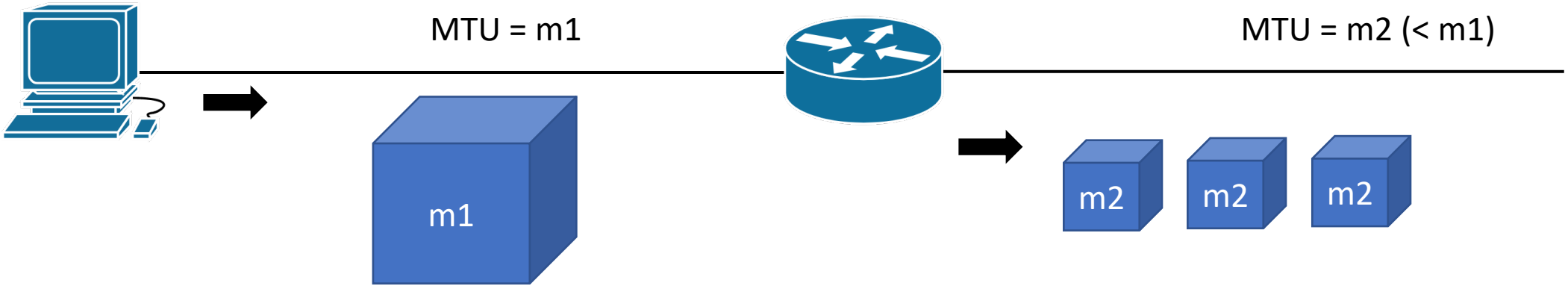
PMTU = Path MTU

= IP packet size limit for the path

= min(m1, m2, m3)

Why care about the PMTU?

- Limit size of outgoing packets only by the local link MTU (m1)
- Let router do IP fragmentation if necessary
  - Only possible with IPv4
  - IP Fragmentation considered harmful[1]

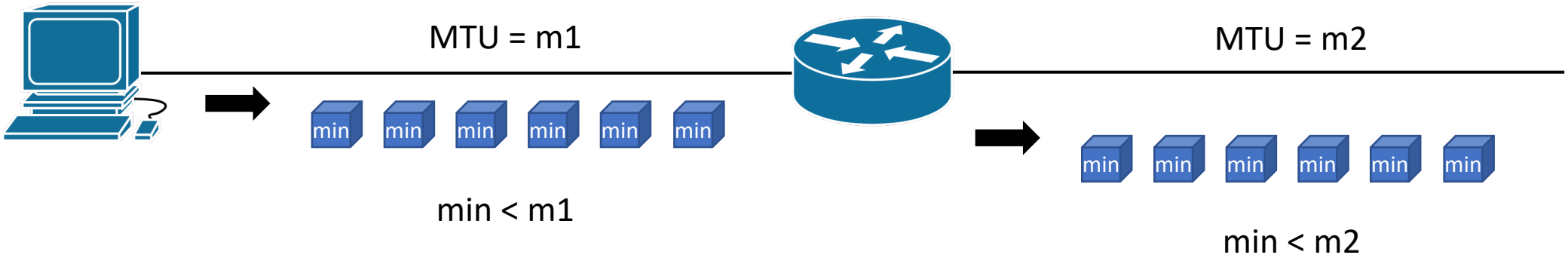


[1]Christopher A. Kent and Jeffrey C. Mogul. Fragmentation Considered Harmful. ACM SIGCOMM '87

Why care about the PMTU?

- Send IP packets small enough to avoid size constraints
  - Requires more packets
  - Increases overhead
  - Consumes more processing power

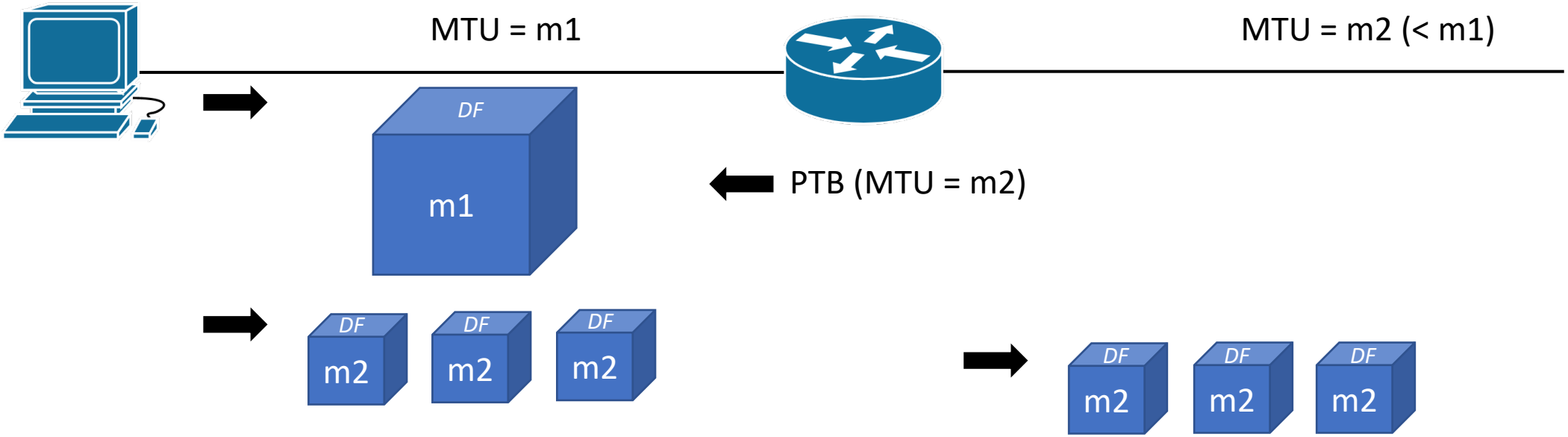
  → Reduces throughput[1]



[1]Kazuho Oku and Jana Iyengar. 2020. Can QUIC match TCP's computational efficiency? https://www.fastly.com/blog/measuring-quic-vs-tcp-computational-efficiency

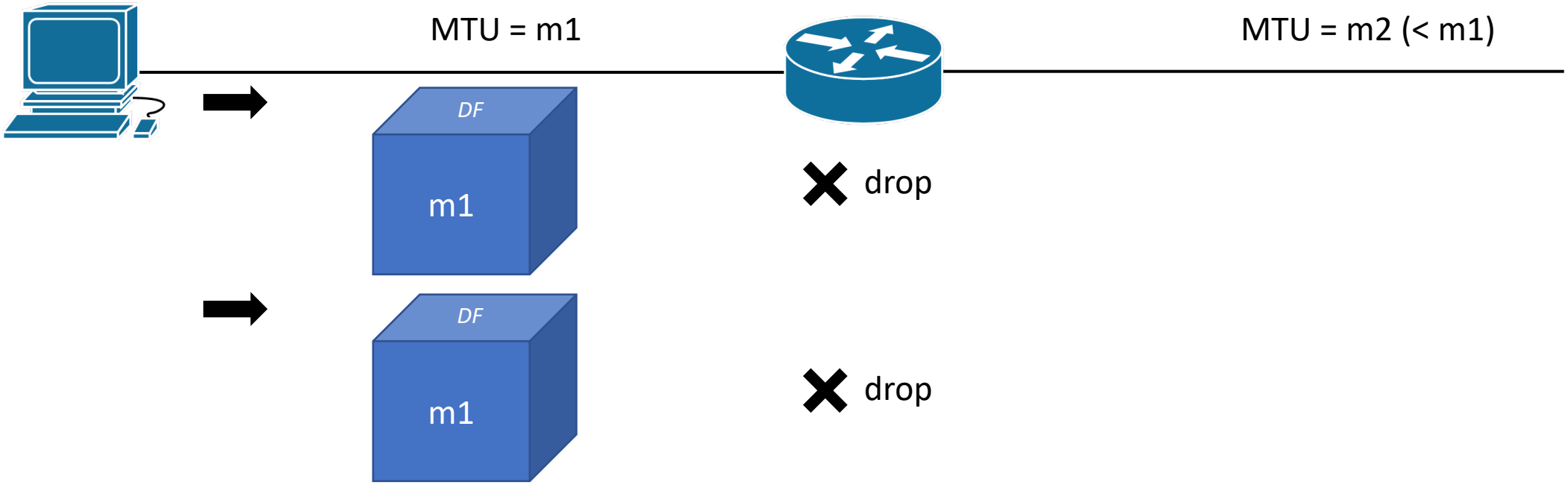
PMTU Discovery

- Limit size of outgoing packets only by link local MTU (m1)
- Set Don’t Fragment bit (implicit for IPv6)
- Rely on router to respond with Packet Too Big (PTB) message



PMTU Discovery Problem

- In the Internet, a Sender can’t rely on receiving a PTB
- E.g., because a Router silently drops packets that are too large
- Results in a Black Hole situation



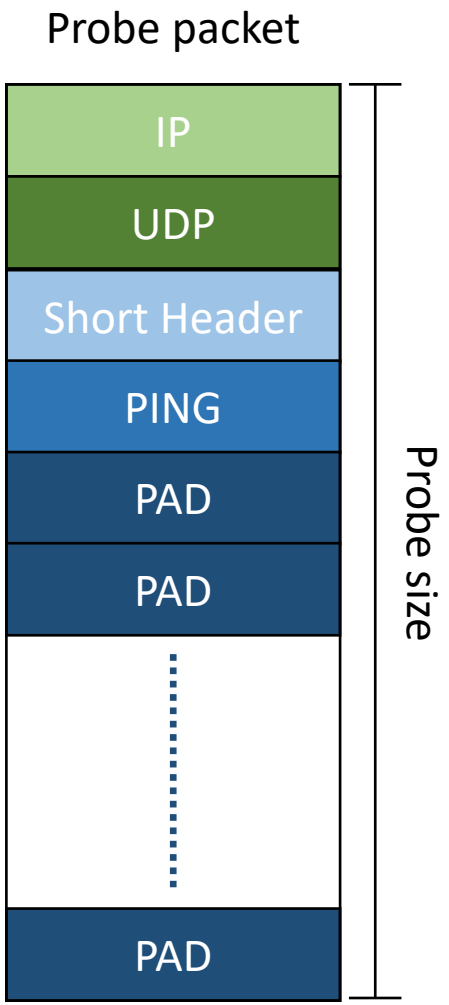
PMTU Discovery for Packetization Layer

- Don’t depend on a signal from the network
- Instead of IP, use an acknowledged Packetization Layer (PL) protocol
- IETF recently standardized this (RFC8899) for PL protocols like QUIC
- Describes the framework
- Does not contain details of how to search for the PMTU

PMTUD Framework and QUIC

PMTU Probe

- Estimate PMTU
- Begin with a small value
- Send packets to probe the network path



PMTU Probe Events

- Acknowledgement
  - Estimated PMTU = probe packet size
  - Received after approx. one RTT (+ max_ack_delay)
- Loss (use QUICs loss detection)
  - Resend probe packet
  - Consider probe as failed after MAX_PROBES lost probe packets
- PTB (valid)
  - Consider probe as failed

PMTU Search

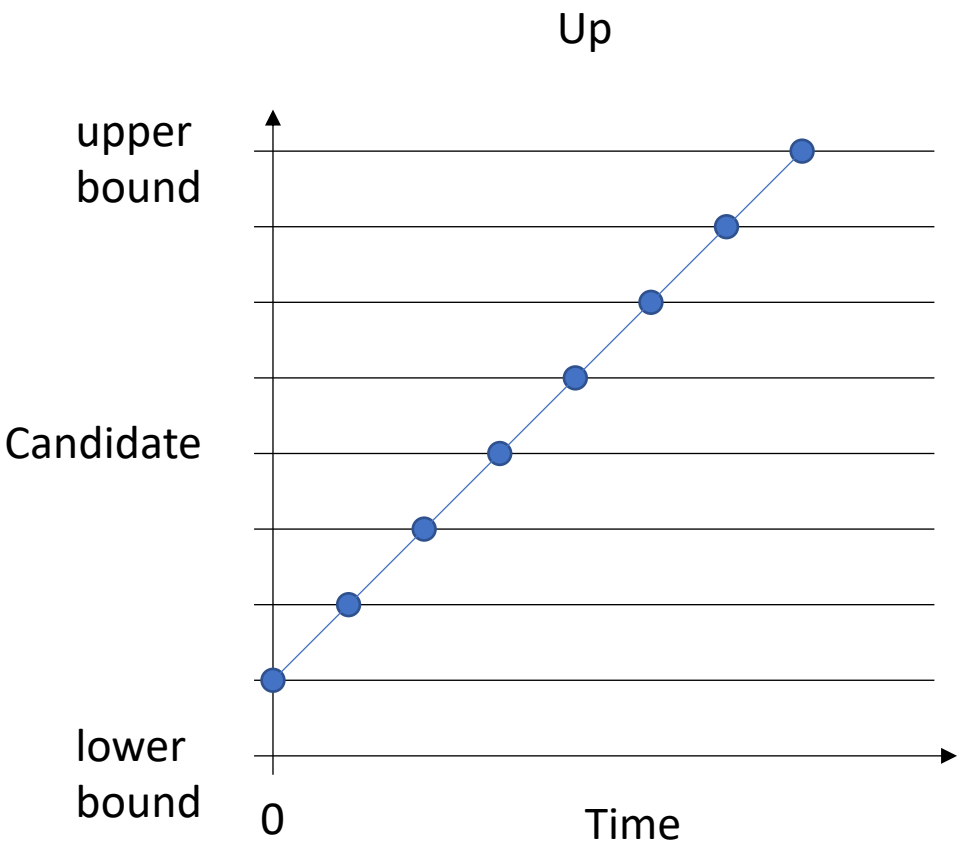
PMTU Candidates

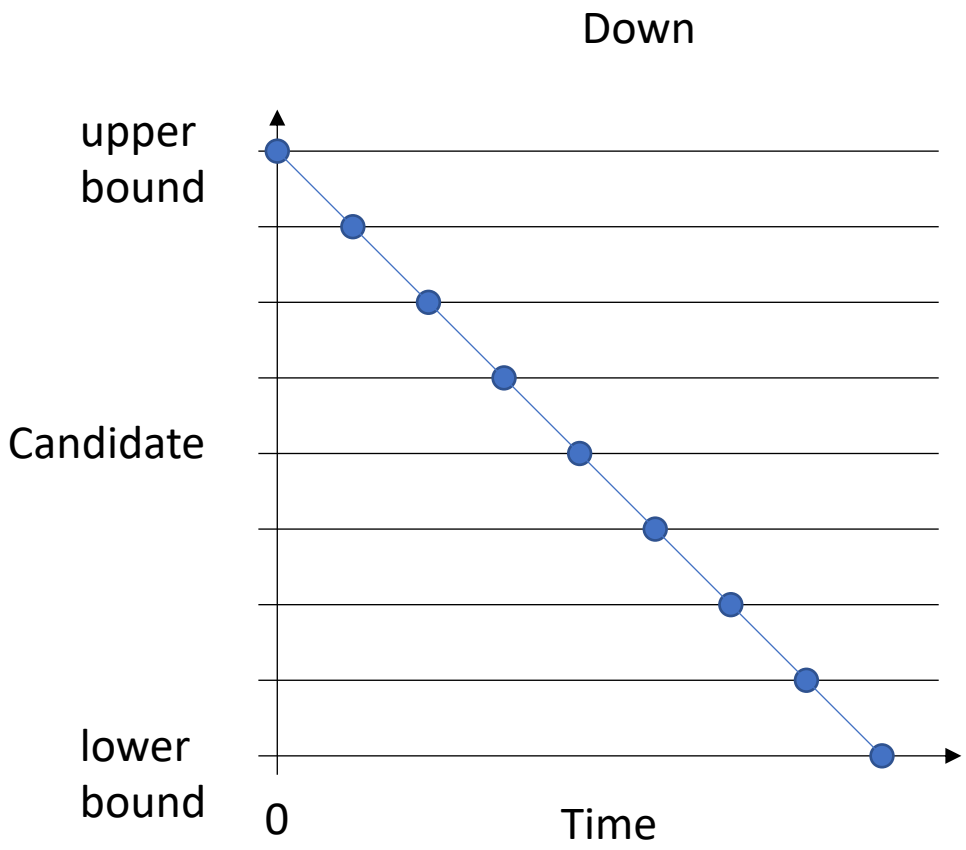
- Lower Bound: 1280 B (use this size for QUIC's path validation)
  - RFC9000: QUIC assumes each path supports this size
- Upper Bound: Minimum of these
  - Maximum IP packet size (65,535 B for IPv4 or 65,575 B for IPv6)
  - MTU specified for local network interface
  - MTU learned from another network device (Router)
  - MRU declared by the remote endpoint (max_udp_payload_size)
- Precision:
  - Consider only a subset of candidates by accepting a lower precision
  - We choose to consider only multiples of four

Candidate Sequence

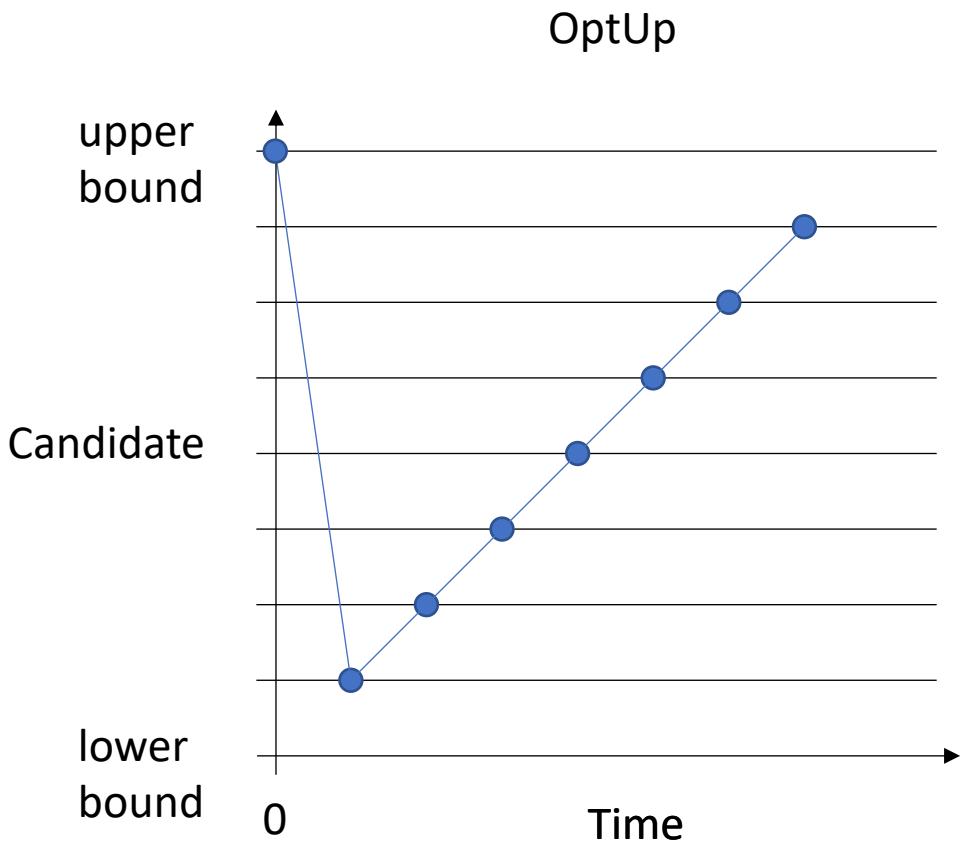
- Specifies the order of candidates to probe
- We consider 7 candidate sequences

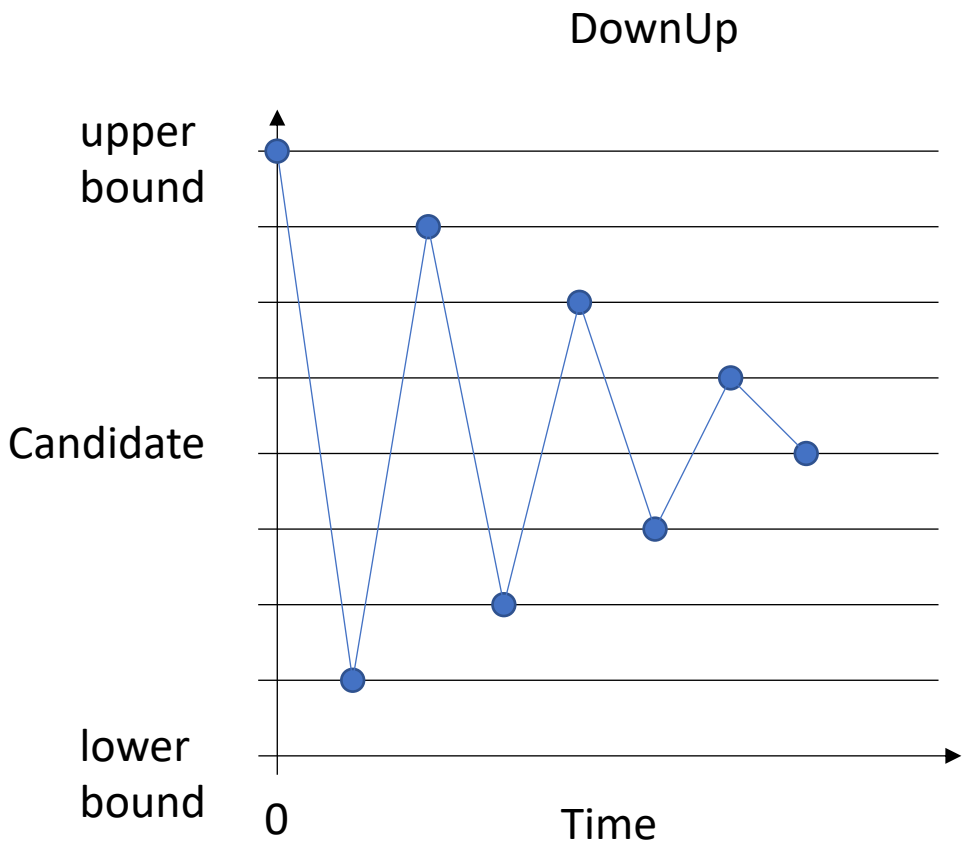
Candidate Sequences



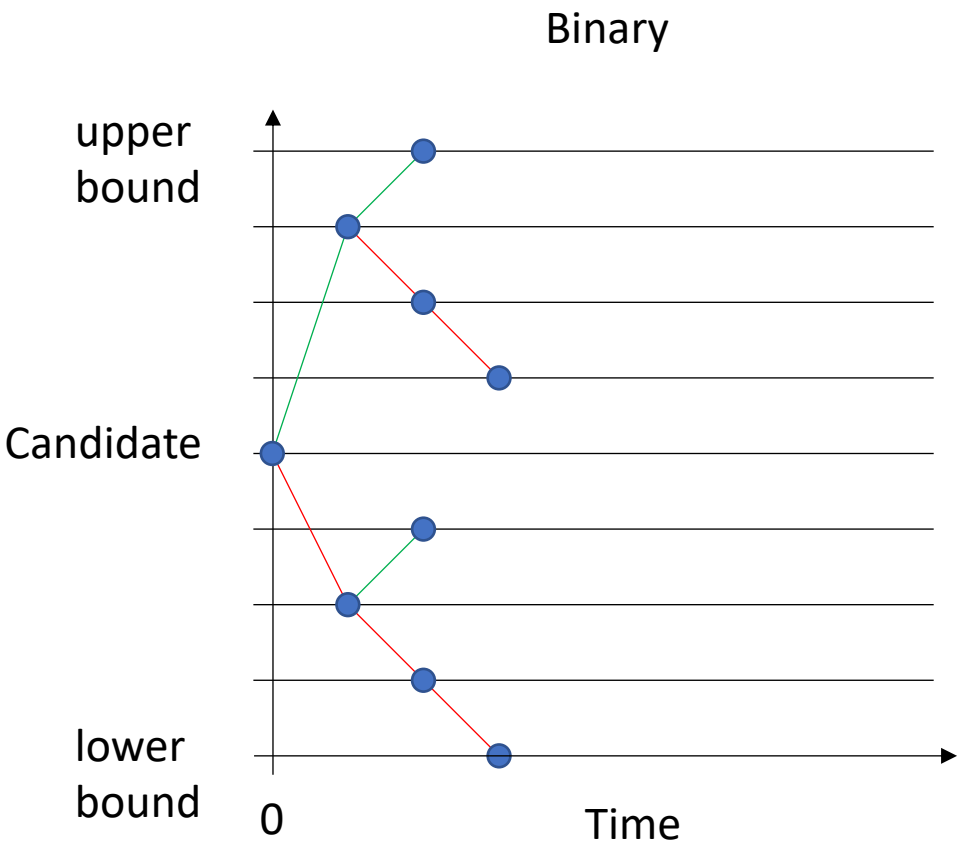


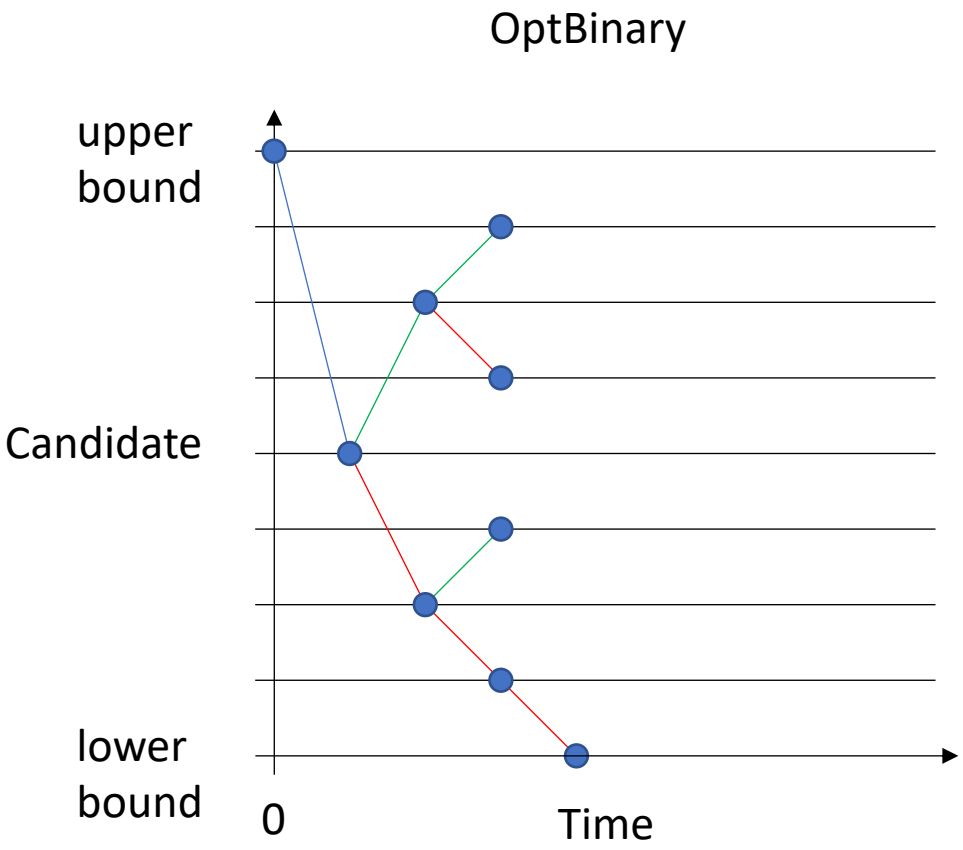
Candidate Sequences



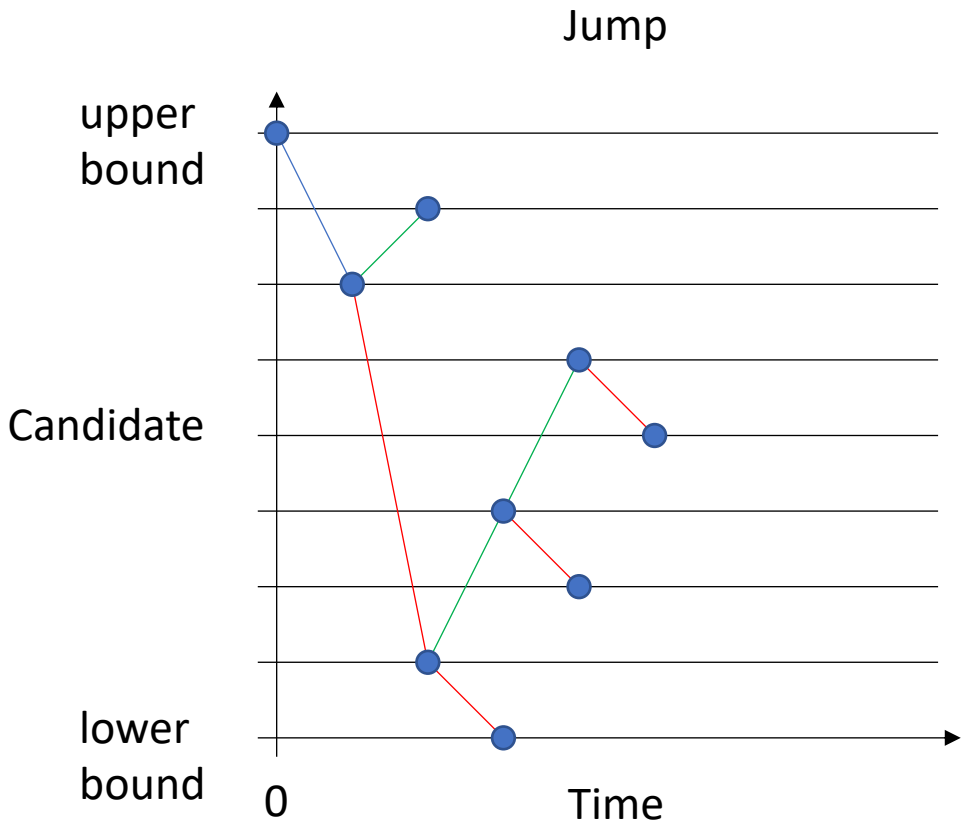


Candidate Sequences



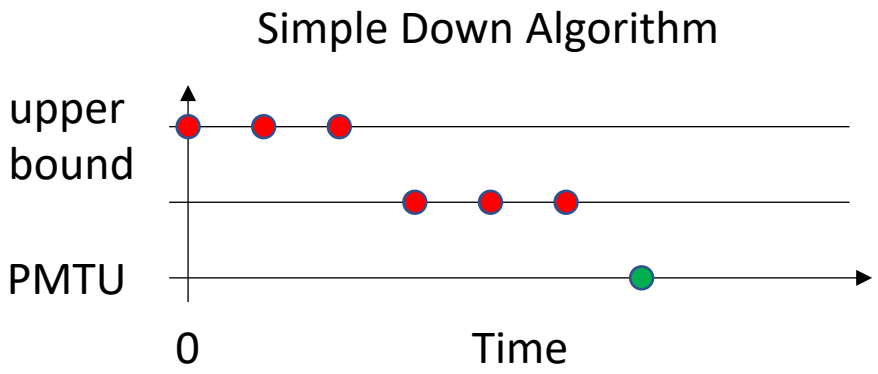


Candidate Sequences



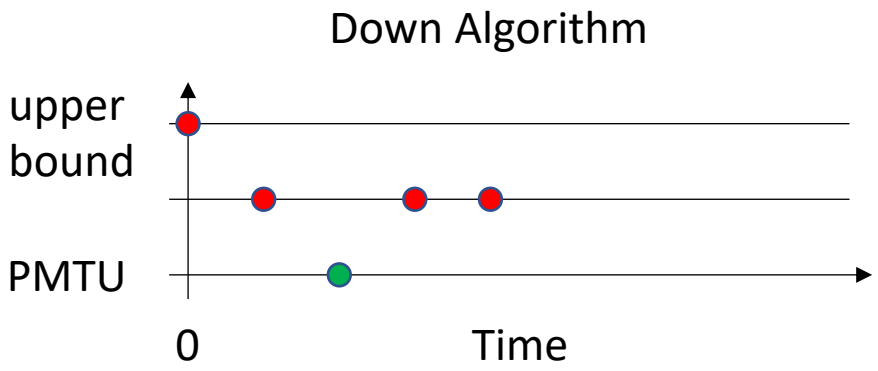
Search Algorithm

- Decides when to
  - select the next candidate
  - retransmit
  - terminate
- Simple algorithm would check one candidate after the other



Search Algorithm

- Start probe for a smaller candidate instead of rtx
- Postpones decision for previous candidate
- Successful Probe: Starts rtx if no larger candidates are available
- Failed Probe: Lets all other probes for larger candidates fail as well



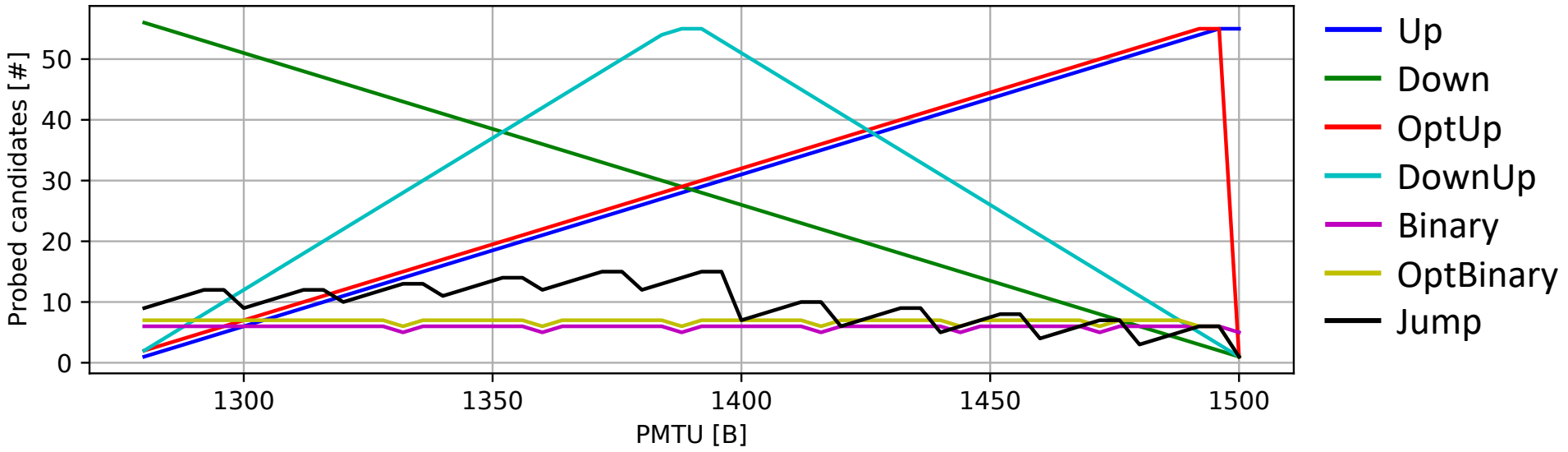
Evaluation

Analytical

- 4 Metrics
- Compare algorithm with each candidate sequence
- Consider network endpoint searches the PMTU
- Candidates: multiples of 4 between 1280 B and 1500 B
- Assume no packet loss due to another reason than packet size
- No PTB

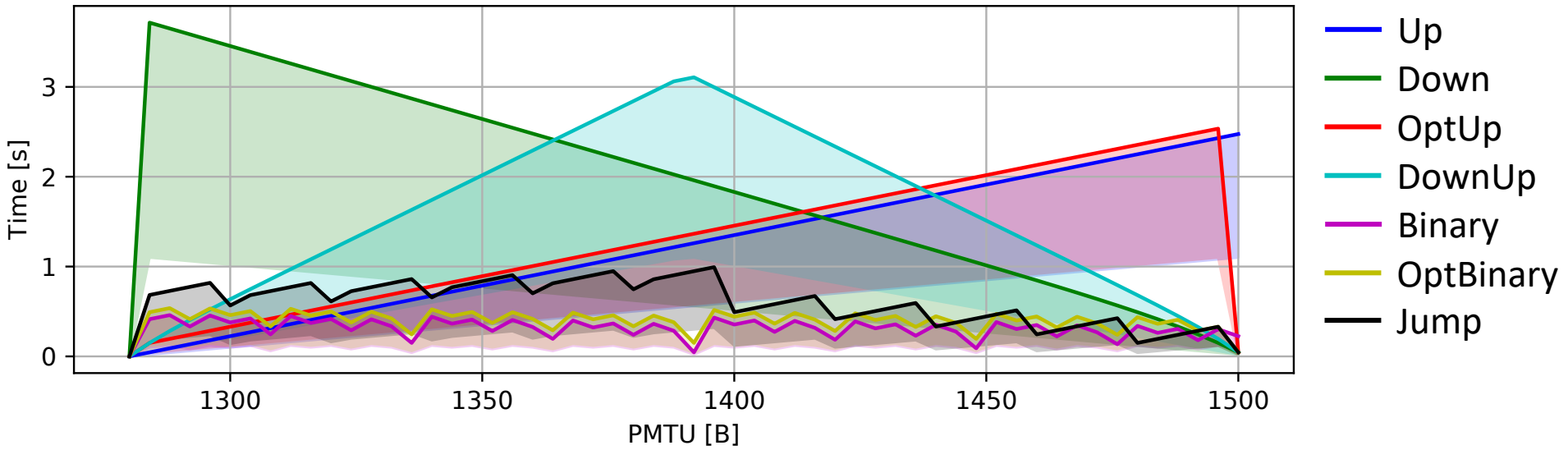
Number of probed PMTU candidates

- Number of candidates the algorithm probes
- Until it finds the PMTU and terminates



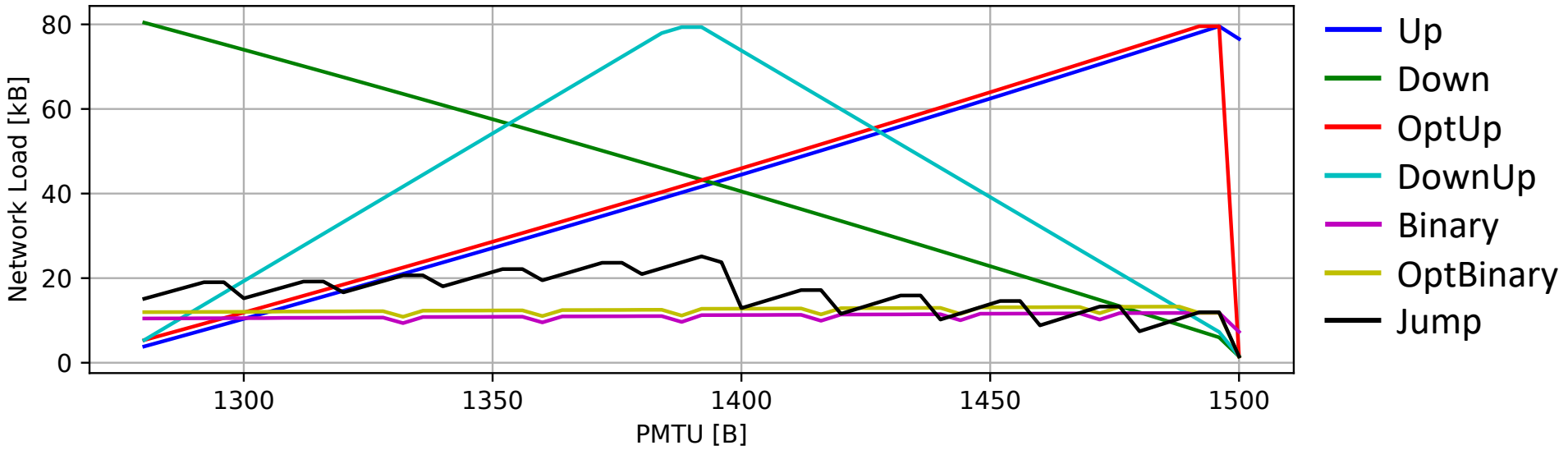
Time

- Time needed
- Until it finds the PMTU
- RTT = 20 ms, max_ack_delay = 25 ms



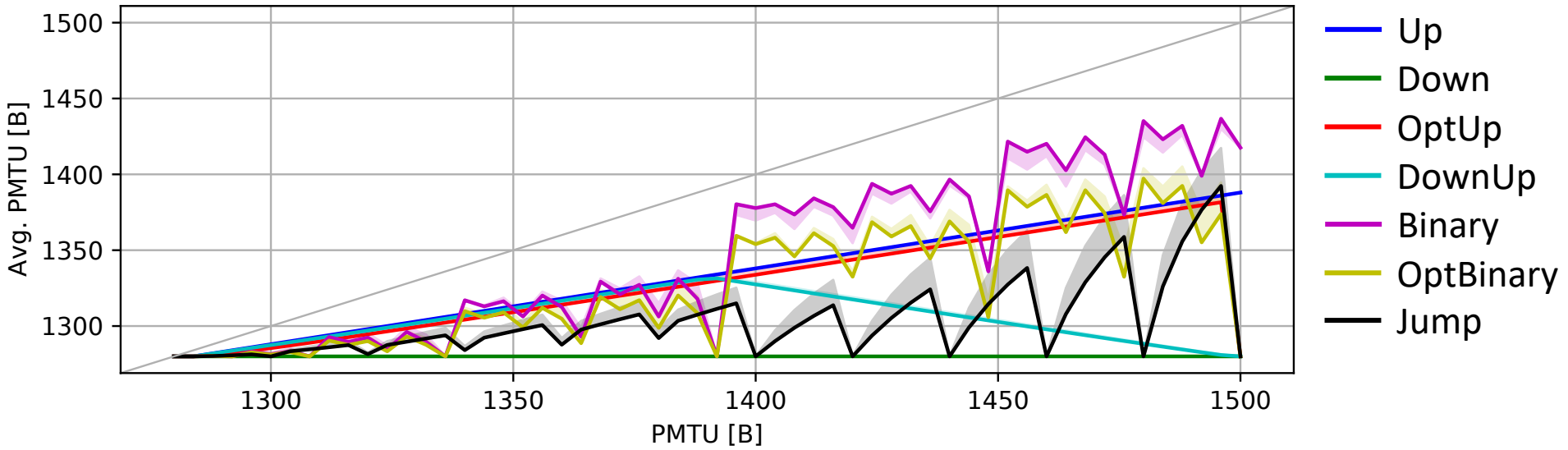
Network Load

- Load produced by probe packets
- Until it finds the PMTU and terminates
- MAX_PROBES = 3



Average PMTU Estimation

- Average PMTU estimation during search
- Before it finds the PMTU
- RTT = 20 ms, max_ack_delay = 25 ms



Result

Top 3

1. Binary
2. OptBinary
3. Jump

- Prefer OptBinary because it
  - immediately finds the PMTU or
  - triggers a PTB
- Jump depends on the initial set of candidates

Conclusion

Conclusion

- Search algorithm for PMTUD framework (RFC8899)
- Defined 4 metrics
- Analysed algorithm with different candidate sequences
- Implemented PMTUD in QUIC simulation model
- Used simulation to further evaluate search algorithm