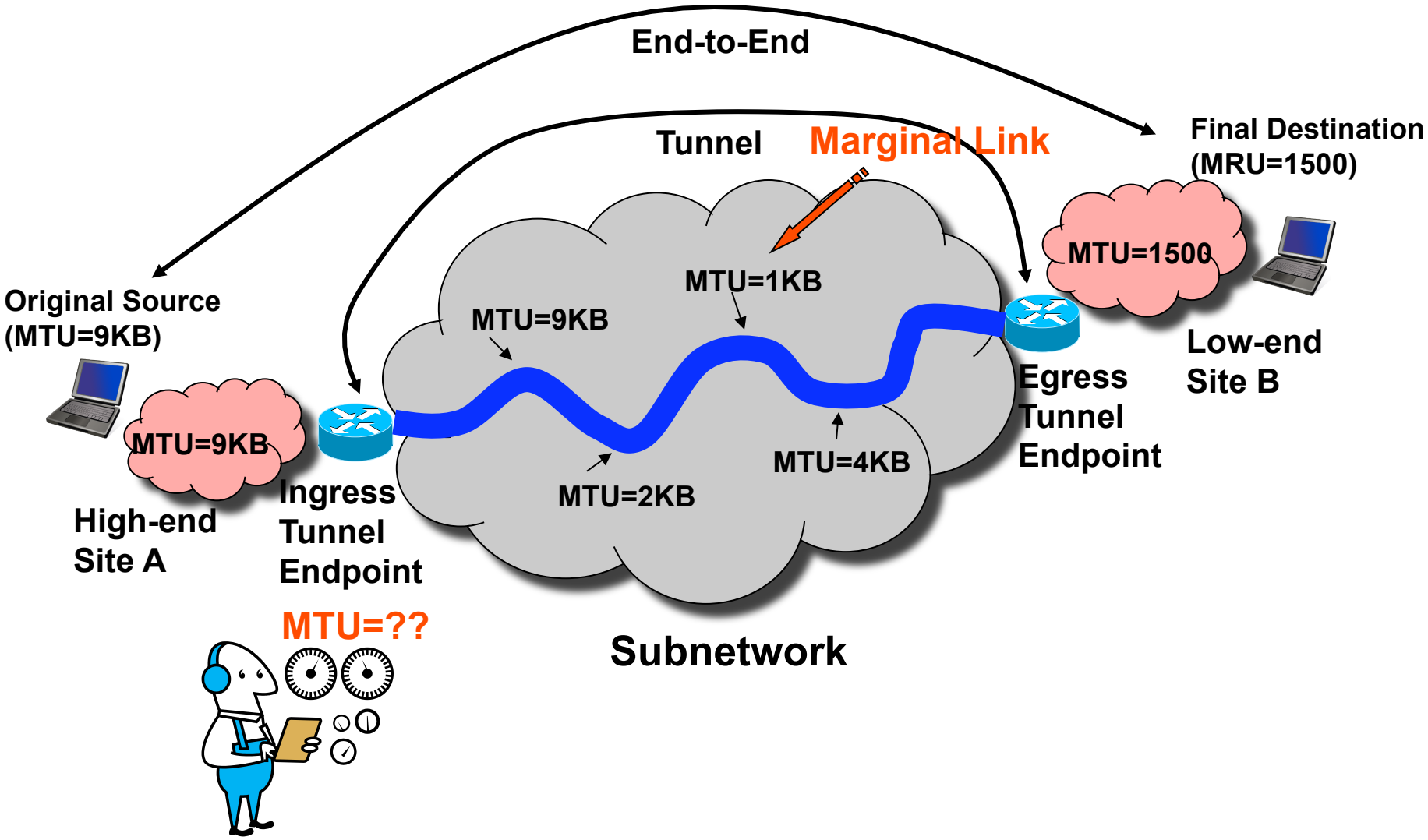
Subnetwork Encapsulation and Adaptation Layer (SEAL)

IETF75 INTAREA Meeting

Fred L. Templin

fred.l.templin@boeing.com

Tunnel Maximum Transmission Unit (MTU)



SEAL Approach

- **4Byte encapsulation sublayer** with **32 bit SEAL-ID**
- Track MTU **w/o classical path MTU discovery**
- **Detect** and **tune out** in-the-network IPv4 fragmentation
- Segmentation to mitigate **misconfigured MTUs** and **marginal links**
- Promotes desired end-state of **MTU-robust subnetworks**

Draft Status

- New draft name - ***draft-templin-intarea-seal***
- Updated based on review input and list discussions
- New approach since IETF74
- Standards-track submission through INTAREA
- Now two distinct “modes” of operation:
  - SEAL-FS (SEAL with Fragmentation Sensing)
  - SEAL-TE (SEAL with Traffic Engineerig)

SEAL With Fragmentation Sensing (SEAL-FS)

- Minimal mechanism for discovering tunnel MTU
- Egress Tunnel Endpoint (ETE):
  - listens for IP fragmentation
  - drops all IP fragments
  - sends “Fragmentation Reports” to Ingress Tunnel Endpoint (ITE)
- ITE adjusts tunnel MTU based on fragmentation reports
- ITE never has to segment and ETE never has to reassemble
- Use cases:
  - performance-intensive core routers that support many tunnels over paths containing robust links (MTU >> 1500)

SEAL With Traffic Engineering (SEAL-TE)

- Same features as SEAL-FS, but includes segmentation and reassembly at a layer below IP
- **MTU based on maximum size the ETE can reassemble**; NOT on the link with the smallest MTU in the path
- **End systems see a solid 1500 MTU at a minimum,** and can often send packets that are MUCH larger than the path MTU
- **IPv6 jumbograms supported even if not all links in the path support jumbograms**
  - **Uses segmentation at a layer below IP**
  - **Does not reduce the integrity of L2 CRC checks**
- Adapts to loss based on reassembly reports
- SEAL-TE tunnels can be configured over SEAL-FS tunnels or even over other SEAL-TE tunnels
- Use cases:
  - Enterprise routers connecting high-performance data centers
  - CPE routers
  - MANET routers

Observations

- **“Unmitigated** Fragmentation Considered Harmful”
- **“Carefully-managed** Fragmentation Considered **Useful”**
- **In-the-network fragmentation is NOT a misfeature!**

For more information:

http://tools.ietf.org/html/draft-templin-intarea-seal (specification)

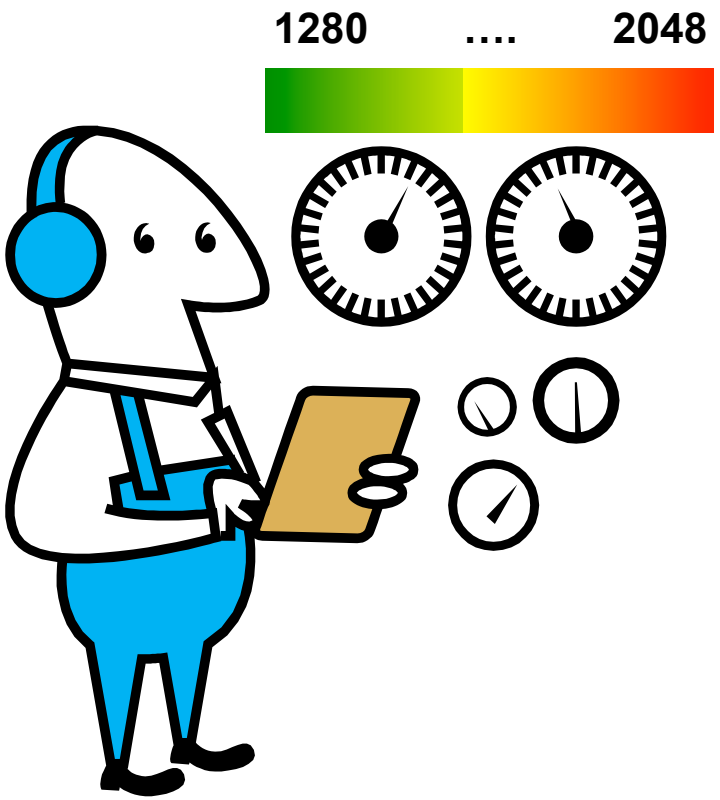
http://osprey67.com/seal (linux source code)

BACKUPS

Problems with Classical Path MTU Discovery

- ICMPs may be lost, erroneous, fabricated
- ICMPs may have insufficient information for relaying
- ALWAYS drops packets when MTU insufficient
- In-the-network tunnels may have 1000's of packets in-flight when a routing change hits an MTU restriction:
  - all packets are dropped
  - flood of ICMPs returned to ITR
  - resources wasted

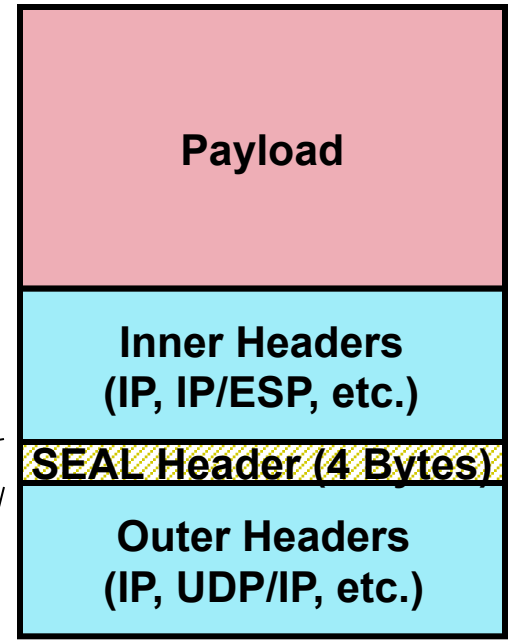
MTU Configuration Knob



- < 1280: MinMTU underflow
- < 1400: fragmentation unlikely
- < 2048: fragmentation managed
- 2048 – 64KB: best-effort
- > 64KB: jumbogram

SEAL Encapsulation

- Extends IP-ID to 32 bits
- Report Fragmentation mechanism
- Tunnel segmentation and reassembly
- Nonce-protected error feedback
- Compatible with wide variety of tunnels




```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          ID Extension         |A|R|M|RSV| SEG |  Next Header  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

ID Extension (16 bits)

A – Acknowledgement Requested (1 bit)

R – Report Fragmentation (1 bit)

M – More Segments (1 bit)

RSV – Reserved (2 bits)

SEG – Segment number (3 bits)

Next Header (8 bits)