



Wireshark for Lustre*

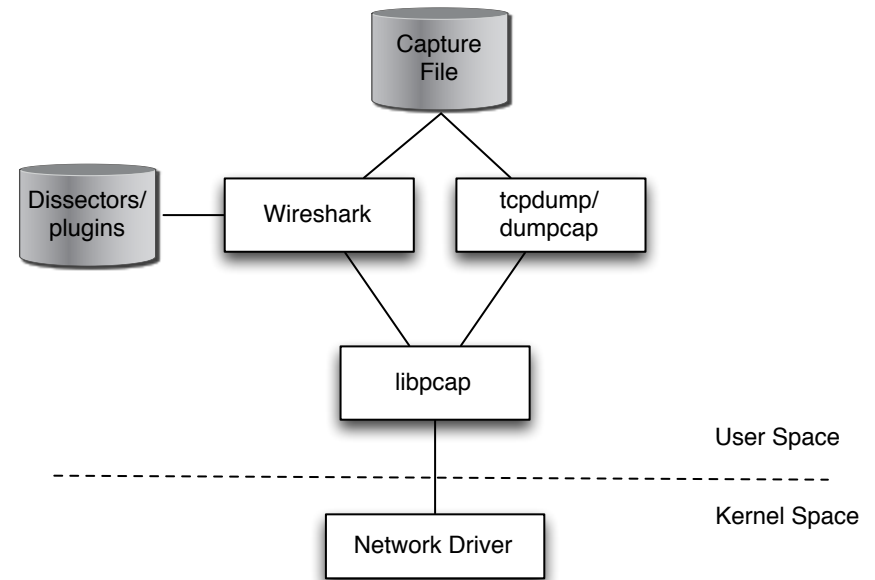
Doug Oucharek

April 17, 2013

* Other names and brands may be claimed as the property of others.

What is Wireshark?

- Protocol analyzer
- Website:
www.wireshark.org
- Powerful filtering capabilities
- Powerful analytics/statistics



Build and Install

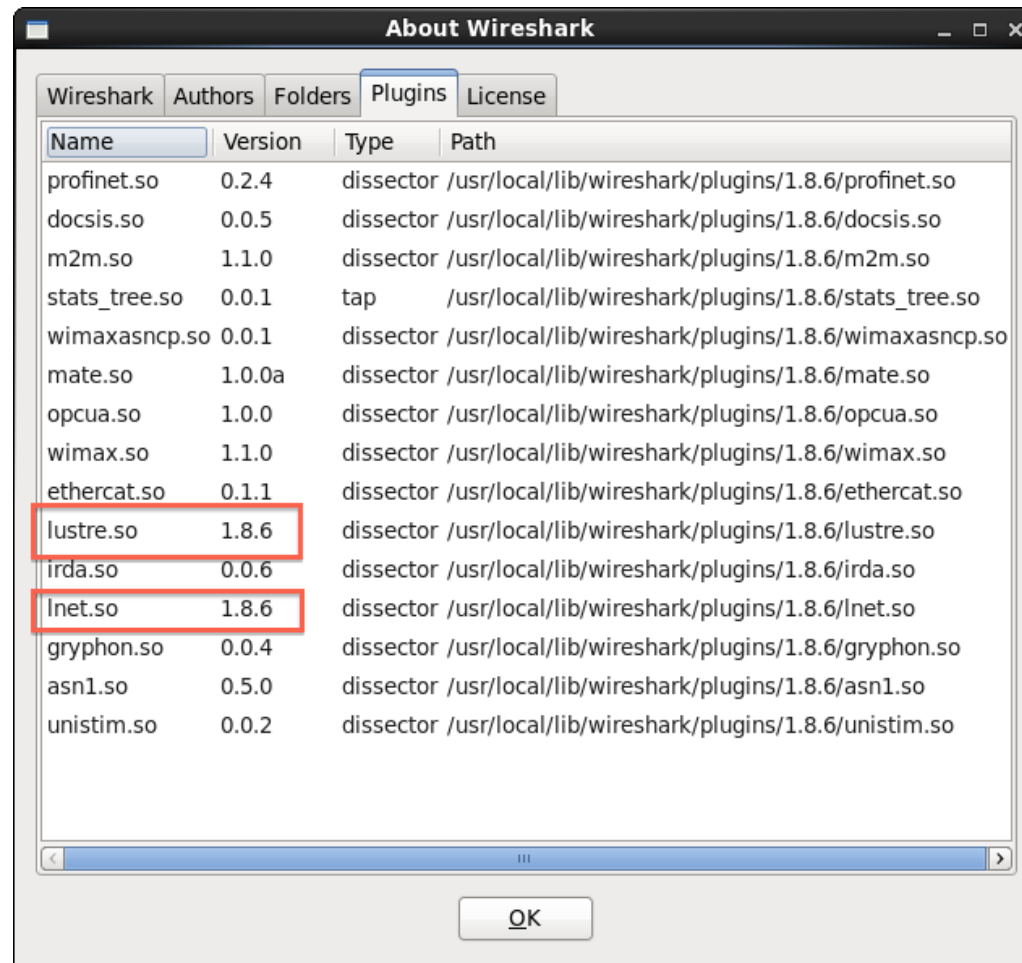
Wireshark:

- Latest Stable: 1.8.6 (recommended). Dev: 1.9.2
- Requires: gtk2-devel and libpcap-devel (CentOS 6.x)
- Usual: ./configure, make, make install
- Application: /usr/local/bin/wireshark

LNet/Lustre* Dissectors:

- Need Wireshark source tree
- Go to: lustre/contrib/wireshark (2.4 and up)
- Update Makefile if not using package manager
- Just run “make install” (“make clean” first if previously built)
- Plugins: lnet.so and lustre.so

Checking for LNet and Lustre* Plugins



Capturing Packets

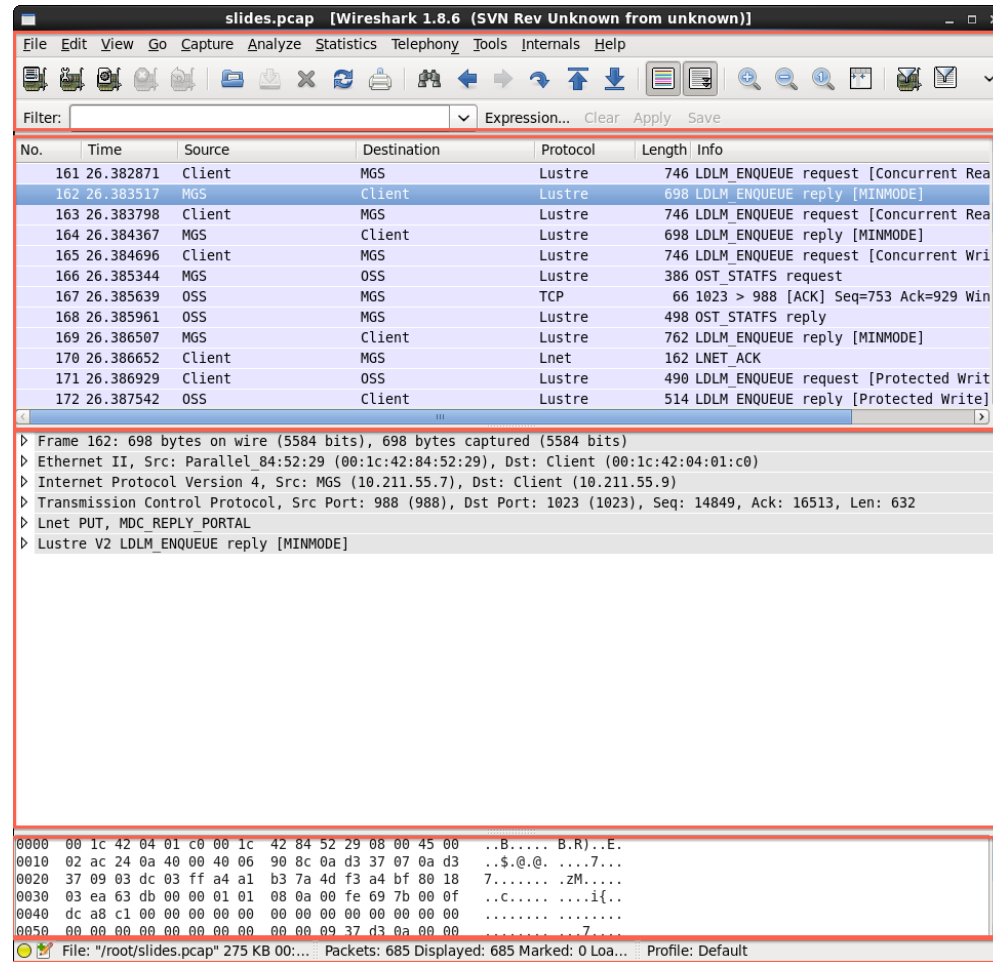
- Capture->Interfaces
- Click “Start” on interface to capture
- Resource intensive



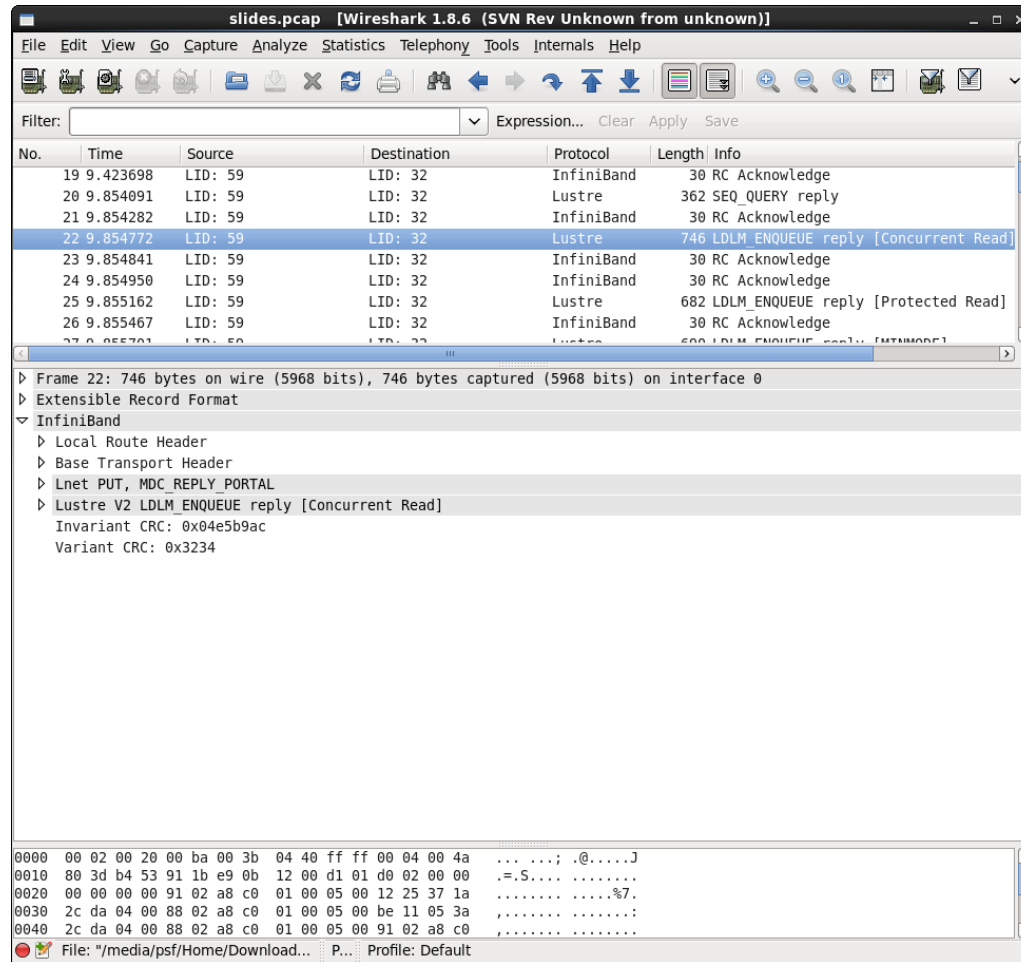
- Use “tcpdump” or “ibdump”
- (Go to:

http://www.mellanox.com/content/pages.php?pg=products_dyn&product_family=110&menu_section=34

Wireshark Interface



With Infiniband Packets



Filtering Packets

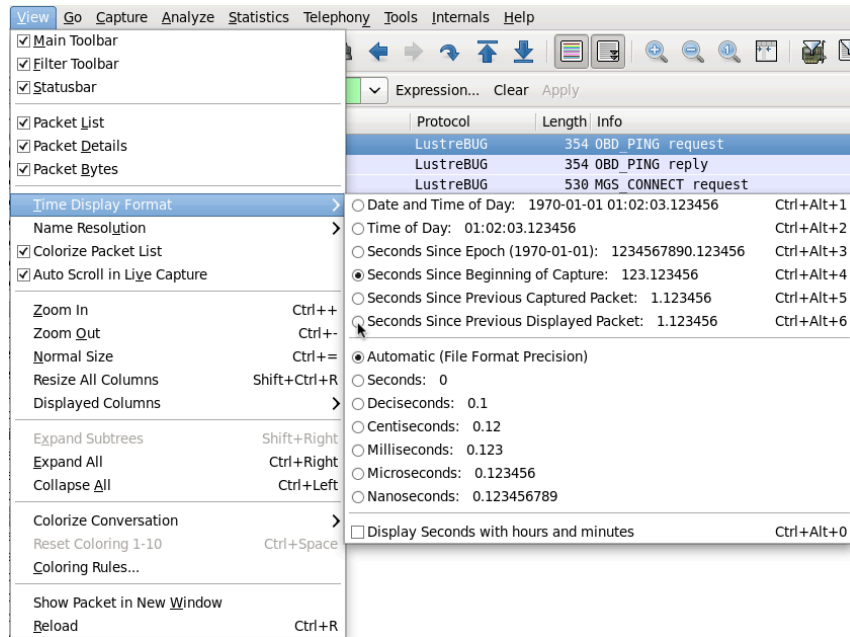
Filter: **Inet** Expression... Clear Apply Save

| No. | Time | Source | Destination | Protocol | Length | Info |
|-----|----------|--------|-------------|----------|--------|--|
| 162 | 0.000646 | MGS | Client | Lustre | 698 | LDLM_ENQUEUE reply [MINMODE] |
| 163 | 0.000281 | Client | MGS | Lustre | 746 | LDLM_ENQUEUE request [Concurrent Read][intent : getattr] filename : test3.txt |
| 164 | 0.000569 | MGS | Client | Lustre | 698 | LDLM_ENQUEUE reply [MINMODE] |
| 165 | 0.000329 | Client | MGS | Lustre | 746 | LDLM_ENQUEUE request [Concurrent Write][intent : open create][REINT_OPEN] filename : test3.txt |
| 166 | 0.000648 | MGS | OSS | Lustre | 386 | OST_STATFS request |
| 168 | 0.000322 | OSS | MGS | Lustre | 498 | OST_STATFS reply |
| 169 | 0.000546 | MGS | Client | Lustre | 762 | LDLM_ENQUEUE reply [MINMODE] |
| 170 | 0.000145 | Client | MGS | Lnet | 162 | LNET_ACK |
| 171 | 0.000277 | Client | OSS | Lustre | 490 | LDLM_ENQUEUE request [Protected Write] |
| 172 | 0.000613 | OSS | Client | Lustre | 514 | LDLM_ENQUEUE reply [Protected Write] |
| 174 | 0.000394 | Client | OSS | Lustre | 650 | OST_WRITE request |
| 175 | 0.000575 | OSS | Client | Lnet | 162 | LNET_GET, OST_BULK_PORTAL |
| 176 | 0.000088 | Client | OSS | Lnet | 211 | LNET_REPLY |

Filter: **Inet** Expression... Clear Apply Save

| No. | Time | Source | Destination | Protocol | Length | Info |
|-----|-----------|---------|-------------|----------|--------|--|
| 68 | 15.299502 | LID: 49 | LID: 32 | Lustre | 442 | LDLM_BL_CALLBACK request [Protected Write] |
| 71 | 15.299981 | LID: 49 | LID: 32 | Lustre | 338 | LDLM_CANCEL reply |
| 72 | 15.300201 | LID: 49 | LID: 32 | Lustre | 442 | LDLM_BL_CALLBACK request [Protected Write] |
| 75 | 15.300617 | LID: 49 | LID: 32 | Lustre | 338 | LDLM_CANCEL reply |
| 77 | 16.871368 | LID: 59 | LID: 32 | Lustre | 682 | LDLM_ENQUEUE reply [MINMODE] |
| 79 | 16.871742 | LID: 59 | LID: 32 | Lustre | 682 | LDLM_ENQUEUE reply [MINMODE] |
| 81 | 16.872137 | LID: 59 | LID: 32 | Lustre | 682 | LDLM_ENQUEUE reply [MINMODE] |
| 83 | 16.872750 | LID: 59 | LID: 32 | Lustre | 746 | LDLM_ENQUEUE reply [Concurrent Read] |
| 86 | 16.873245 | LID: 49 | LID: 32 | Lustre | 514 | LDLM_ENQUEUE reply [Protected Write] |

Changing Time Scale



| No. | Time | Source | Destination | Protocol | Length | Info |
|-----|----------|--------|-------------|-----------|--------|--|
| 3 | 0.000000 | OSS | MGS | LustreBUG | 354 | OBD_PING request |
| 3 | 0.004041 | MGS | OSS | LustreBUG | 354 | OBD_PING reply |
| 24 | 2.981399 | Client | MGS | LustreBUG | 538 | MGS_CONNECT request |
| 32 | 2.982412 | MGS | Client | LustreBUG | 426 | MGS_CONNECT reply |
| 35 | 2.982601 | Client | MGS | LustreBUG | 458 | LDLM_ENQUEUE request [MINMODE] |
| 41 | 2.983503 | MGS | Client | LustreBUG | 514 | LDLM_ENQUEUE reply [MINMODE] |
| 45 | 2.984624 | Client | MGS | LustreBUG | 426 | LLOG_ORIGIN_HANDLE_CREATE request filename : |
| 46 | 2.984734 | MGS | Client | LustreBUG | 482 | LLOG_ORIGIN_HANDLE_CREATE reply |

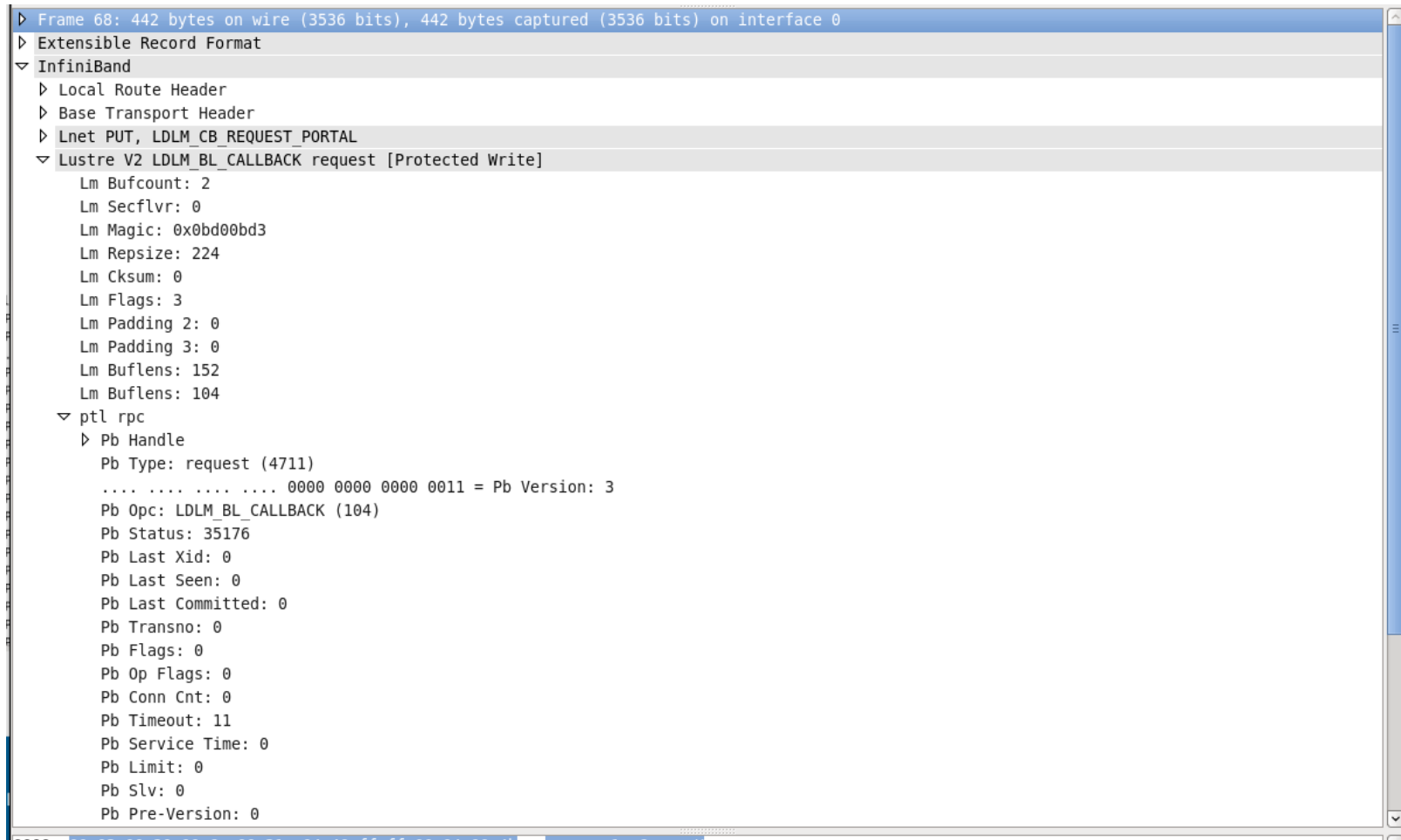
| No. | Time | Source | Destination | Protocol | Length | Info |
|-----|----------|--------|-------------|-----------|--------|--|
| 1 | 0.000000 | OSS | MGS | LustreBUG | 354 | OBD_PING request |
| 3 | 0.004041 | MGS | OSS | LustreBUG | 354 | OBD_PING reply |
| 24 | 2.897358 | Client | MGS | LustreBUG | 538 | MGS_CONNECT request |
| 32 | 0.001013 | MGS | Client | LustreBUG | 426 | MGS_CONNECT reply |
| 35 | 0.000189 | Client | MGS | LustreBUG | 458 | LDLM_ENQUEUE request [MINMODE] |
| 41 | 0.000902 | MGS | Client | LustreBUG | 514 | LDLM_ENQUEUE reply [MINMODE] |
| 45 | 0.001121 | Client | MGS | LustreBUG | 426 | LLOG_ORIGIN_HANDLE_CREATE request filename : |
| 46 | 0.000110 | MGS | Client | LustreBUG | 482 | LLOG_ORIGIN_HANDLE_CREATE reply |

LNet Header

```
▷ Frame 1: 354 bytes on wire (2832 bits), 354 bytes captured (2832 bits)
▷ Ethernet II, Src: Parallel_2a:de:5c (00:1c:42:2a:de:5c), Dst: Parallel_8e:1a:f5 (00:1c:42:8e:1a:f5)
▷ Internet Protocol Version 4, Src: OSS (10.211.55.9), Dst: MGS (10.211.55.7)
▷ Transmission Control Protocol, Src Port: 988 (988), Dst Port: 1023 (1023), Seq: 1, Ack: 1, Len: 288
▽ Lnet PUT, MGS_REQUEST_PORTAL
  Type of socklnd message: KSOCK_MSG_LNET (0x000000c1)
  checksum disabled
  ack not required
  not ack
  ▽ dest_nid = 10.211.55.7@tcp0
    Destination nid: MGS (10.211.55.7)
    lnd network interface: 0
    lnd network type: SOCKLND (2)
  ▽ src_nid = 10.211.55.9@tcp0
    Src nid: OSS (10.211.55.9)
    lnd network interface: 0
    lnd network type: SOCKLND (2)
  Src pid: 12345 (0x00003039)
  Dest pid: 12345 (0x00003039)
  Message type: PUT (1)
  Payload length: 192
  DST MD index interface: 0xffffffffffffffff (18446744073709551615)
  DST MD index object: 0xffffffffffffffff (18446744073709551615)
  Match bits: 0x0004fadac7204803 (1401717457438723)
  hdr data: 0x0000000000000000 (0)
  ptl index: MGS_REQUEST_PORTAL (26)
  offset: 0
  msg filler (padding)
  Payload
▷ Lustre V2 OBD_PING request
```



Lustre* Header



```
▶ Frame 68: 442 bytes on wire (3536 bits), 442 bytes captured (3536 bits) on interface 0
▶ Extensible Record Format
▼ InfiniBand
  ▶ Local Route Header
  ▶ Base Transport Header
  ▶ Lnet PUT, LDLM_CB_REQUEST_PORTAL
  ▼ Lustre V2 LDLM_BL_CALLBACK request [Protected Write]
    Lm Bufcount: 2
    Lm Secflvr: 0
    Lm Magic: 0x0bd00bd3
    Lm Replib: 224
    Lm Cksum: 0
    Lm Flags: 3
    Lm Padding 2: 0
    Lm Padding 3: 0
    Lm Buflens: 152
    Lm Buflens: 104
  ▼ ptl rpc
    ▶ Pb Handle
      Pb Type: request (4711)
      .... 0000 0000 0000 0011 = Pb Version: 3
      Pb Op: LDLM_BL_CALLBACK (104)
      Pb Status: 35176
      Pb Last Xid: 0
      Pb Last Seen: 0
      Pb Last Committed: 0
      Pb Transno: 0
      Pb Flags: 0
      Pb Op Flags: 0
      Pb Conn Cnt: 0
      Pb Timeout: 11
      Pb Service Time: 0
      Pb Limit: 0
      Pb Slv: 0
      Pb Pre-Version: 0
```



Interacting with Fields

- Clicking on field triggers hex to highlight
- Also shows generated name for field (if any)

```
lnd network type: SOCKLND (2)
Src pid: 12345 (0x00003039)
Dest pid: 12345 (0x00003039)
Message type: PUT (1)
Payload length: 192
DST MD index interface: 0xffffffffffffffff (18446744073709551615)
DST MD index object: 0xffffffffffffffff (18446744073709551615)
Match bits: 0x0004fadac7204803 (1401717457438723)
hdr data: 0x0000000000000000 (0)
ptl index: MGS_REQUEST_PORTAL (26)
offset: 0
```

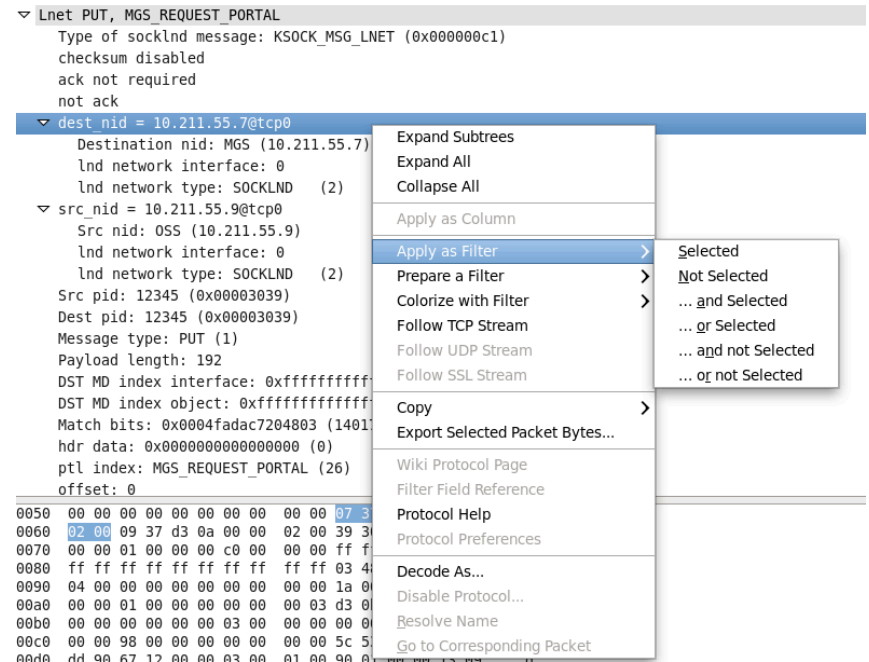
| | | | | | | | | | | | | | | | | |
|------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 0070 | 00 | 00 | 01 | 00 | 00 | 00 | c0 | 00 | 00 | 00 | ff | ff | ff | ff | ff | ff |
| 0080 | ff | ff | ff | ff | ff | ff | ff | ff | ff | ff | 03 | 48 | 20 | c7 | da | fa |
| 0090 | 04 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 1a | 00 | 00 | 00 | 00 | 00 |
| 00a0 | 00 | 00 | 01 | 00 | 00 | 00 | 00 | 00 | 00 | 03 | d3 | 0b | d0 | 0b | c0 | 00 |
| 00b0 | 00 | 00 | 00 | 00 | 00 | 00 | 03 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |
| 00c0 | 00 | 00 | 98 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 5c | 53 | d8 | 3d | 10 | c5 |
| 00d0 | dd | 90 | 67 | 12 | 00 | 00 | 03 | 00 | 01 | 00 | 90 | 01 | 00 | 00 | 13 | 09 |
| 00e0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |
| 00f0 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 |
| 0100 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 1e | 29 | 00 | 00 | 06 | 00 |

msg type (lnd.msg_type), 4 bytes

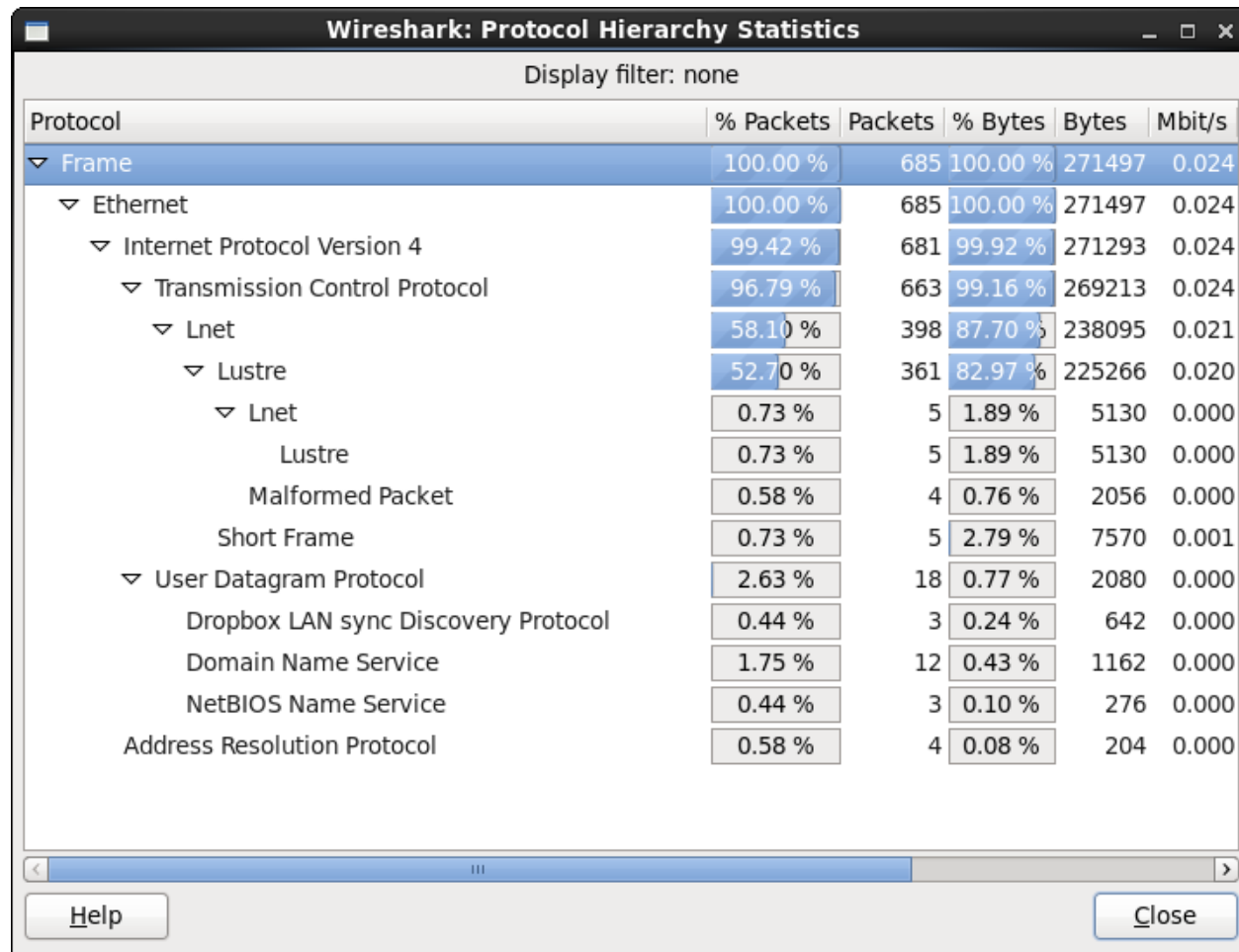


Filter Tricks – Being Lazy

- Right-Click on field
- Use “Apply as Filter” to fill in filter field for you
- Even works on text fields...which is what “dest_nid” is



Protocol Hierarchy Window



Wireshark: Protocol Hierarchy Statistics

Display filter: none

| Protocol | % Packets | Packets | % Bytes | Bytes | Mbit/s |
|-------------------------------------|-----------|---------|----------|--------|--------|
| ▼ Frame | 100.00 % | 685 | 100.00 % | 271497 | 0.024 |
| ▼ Ethernet | 100.00 % | 685 | 100.00 % | 271497 | 0.024 |
| ▼ Internet Protocol Version 4 | 99.42 % | 681 | 99.92 % | 271293 | 0.024 |
| ▼ Transmission Control Protocol | 96.79 % | 663 | 99.16 % | 269213 | 0.024 |
| ▼ Lnet | 58.10 % | 398 | 87.70 % | 238095 | 0.021 |
| ▼ Lustre | 52.70 % | 361 | 82.97 % | 225266 | 0.020 |
| ▼ Lnet | 0.73 % | 5 | 1.89 % | 5130 | 0.000 |
| Lustre | 0.73 % | 5 | 1.89 % | 5130 | 0.000 |
| Malformed Packet | 0.58 % | 4 | 0.76 % | 2056 | 0.000 |
| Short Frame | 0.73 % | 5 | 2.79 % | 7570 | 0.001 |
| ▼ User Datagram Protocol | 2.63 % | 18 | 0.77 % | 2080 | 0.000 |
| Dropbox LAN sync Discovery Protocol | 0.44 % | 3 | 0.24 % | 642 | 0.000 |
| Domain Name Service | 1.75 % | 12 | 0.43 % | 1162 | 0.000 |
| NetBIOS Name Service | 0.44 % | 3 | 0.10 % | 276 | 0.000 |
| Address Resolution Protocol | 0.58 % | 4 | 0.08 % | 204 | 0.000 |

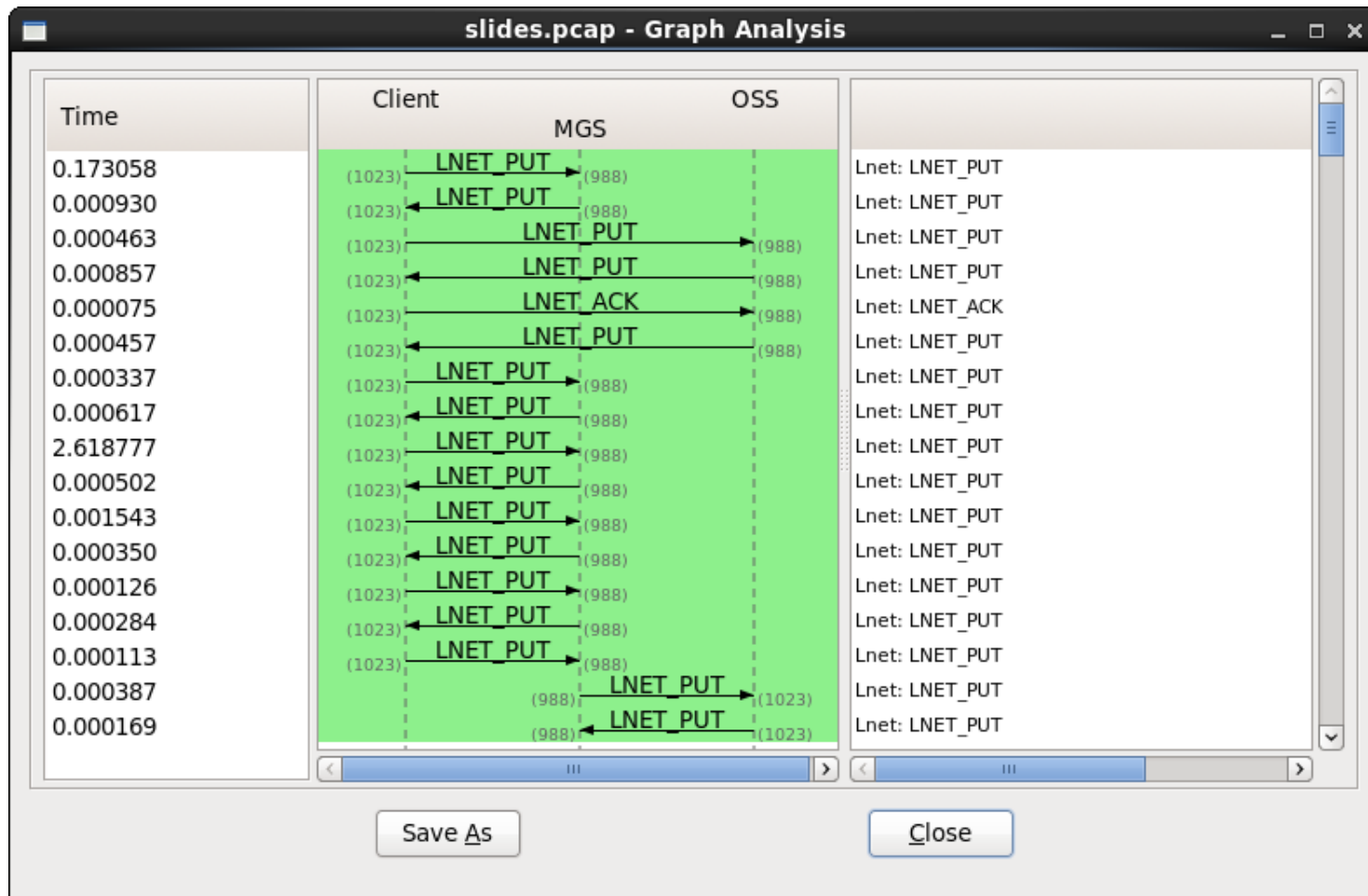
Help Close

Packet Lengths

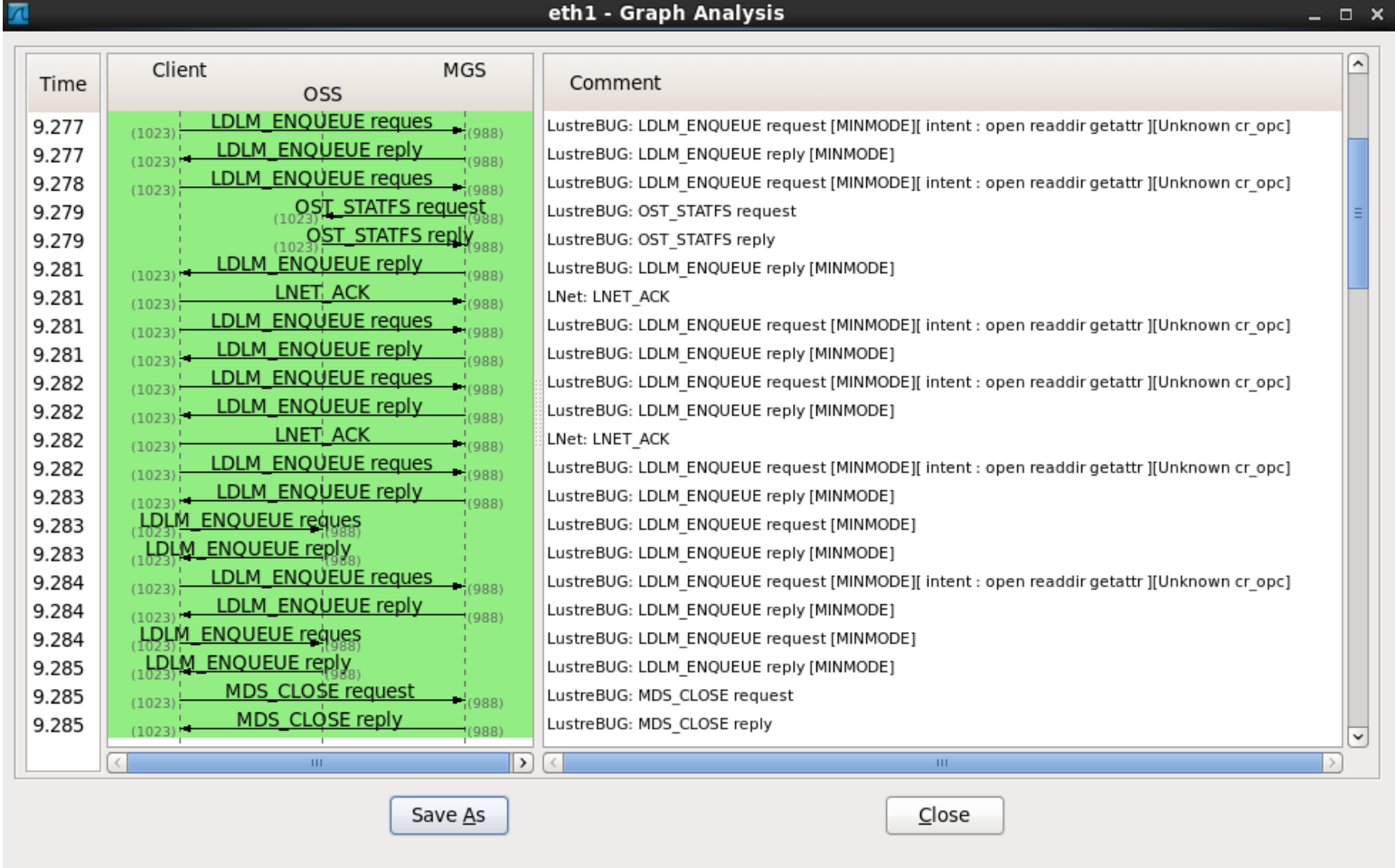
| Topic / Item | Count | Rate (ms) | Percent |
|----------------|-------|-----------|---------|
| Packet Lengths | 398 | 0.004854 | |
| 0-19 | 0 | 0.000000 | 0.00% |
| 20-39 | 0 | 0.000000 | 0.00% |
| 40-79 | 0 | 0.000000 | 0.00% |
| 80-159 | 0 | 0.000000 | 0.00% |
| 160-319 | 32 | 0.000390 | 8.04% |
| 320-639 | 177 | 0.002159 | 44.47% |
| 640-1279 | 182 | 0.002219 | 45.73% |
| 1280-2559 | 7 | 0.000085 | 1.76% |
| 2560-5119 | 0 | 0.000000 | 0.00% |
| 5120- | 0 | 0.000000 | 0.00% |

Close

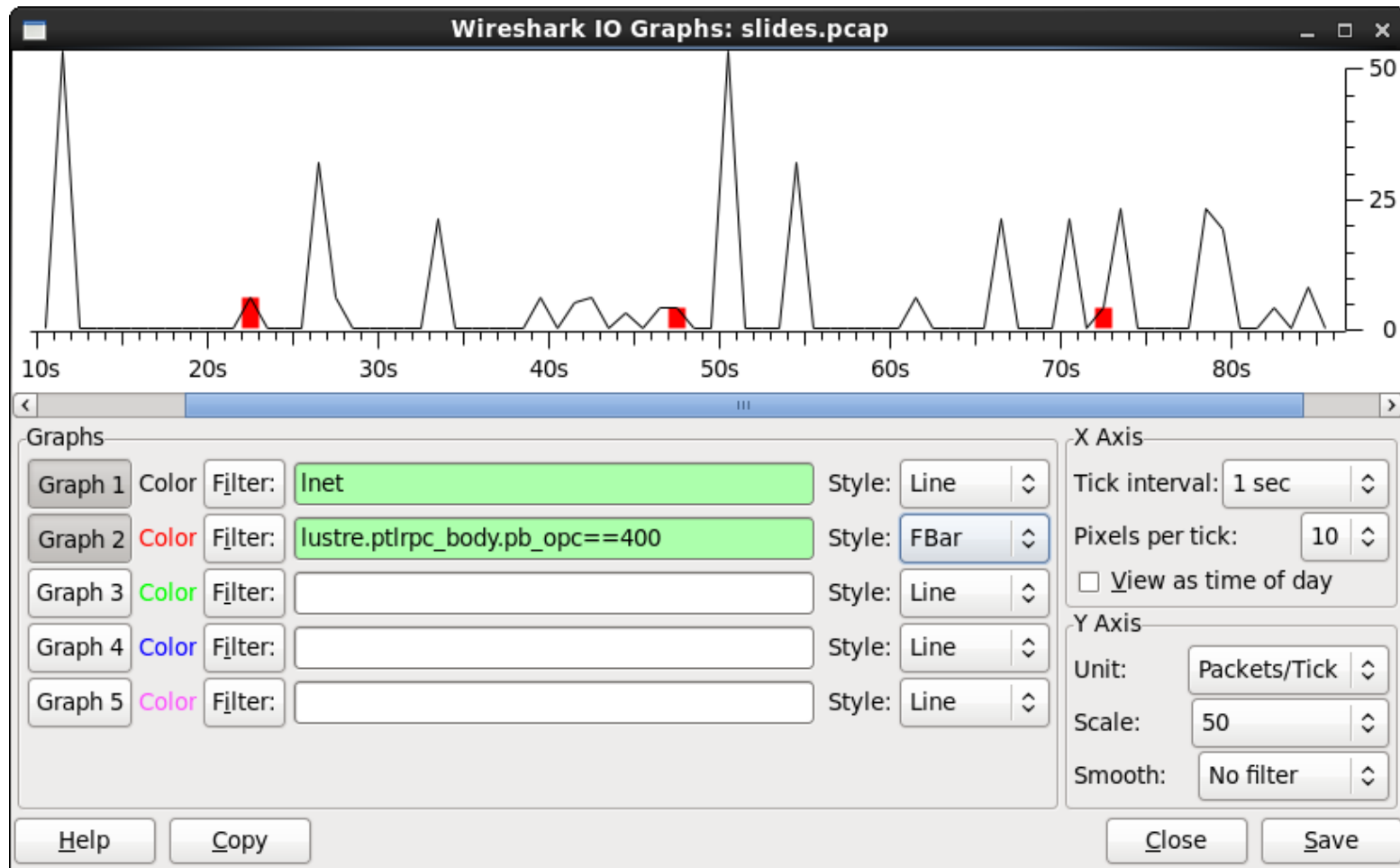
Current: Flow Graph



Future: Flow Graph



I/O Graph



I/O Graph - Advanced

