Lecture 9

Labwork

Antonio Cianfrani

DIET Department – Networking Group

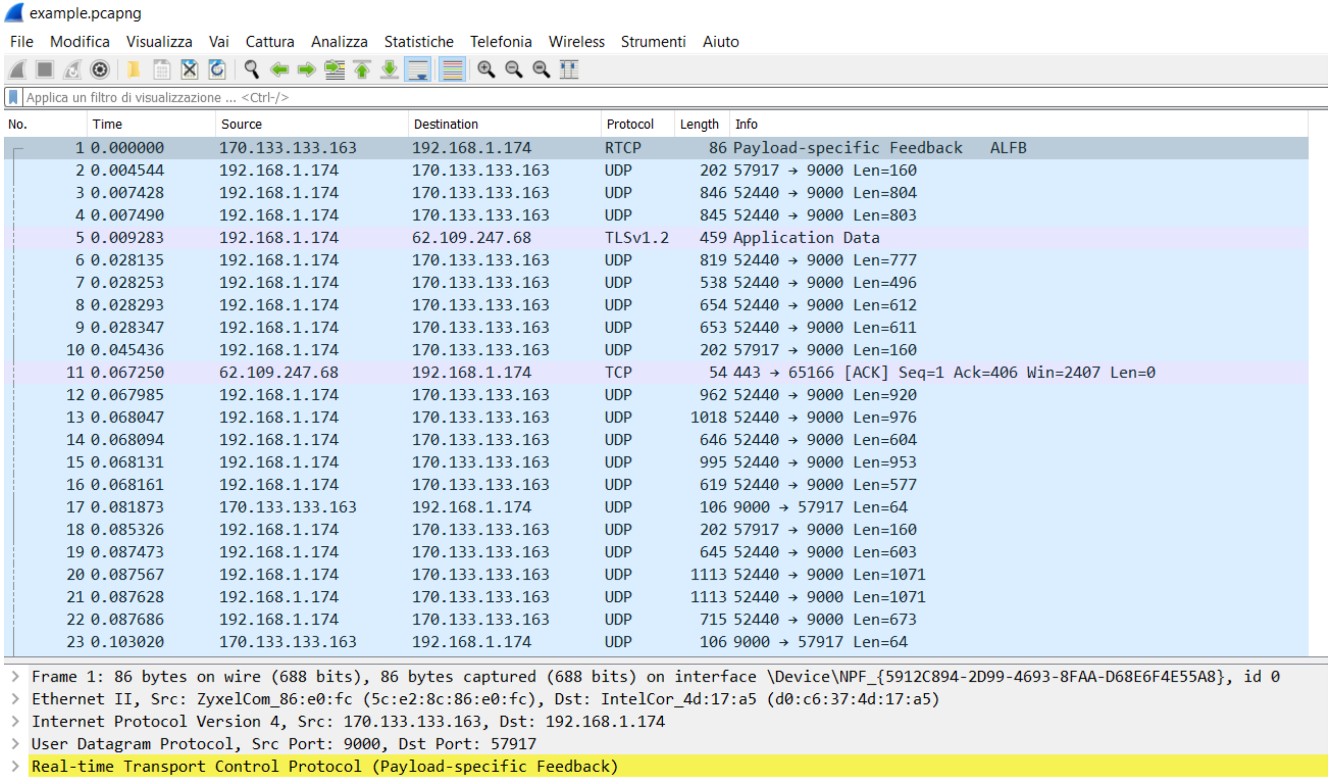
web.uniroma1.it/netlab

DIPARTIMENTO DI SCIENZA
E TECNICA DELL'INFORMAZIONE
E DELLA COMUNICAZIONE INFOCOM

SAPIENZA
UNIVERSITÀ DI ROMA

Internet traffic analysis

- The aim of the labwork is to investigate Internet traffic features, starting from raw packets captures.
- Raw packets captures are available in .pcap format (Wireshark traces).

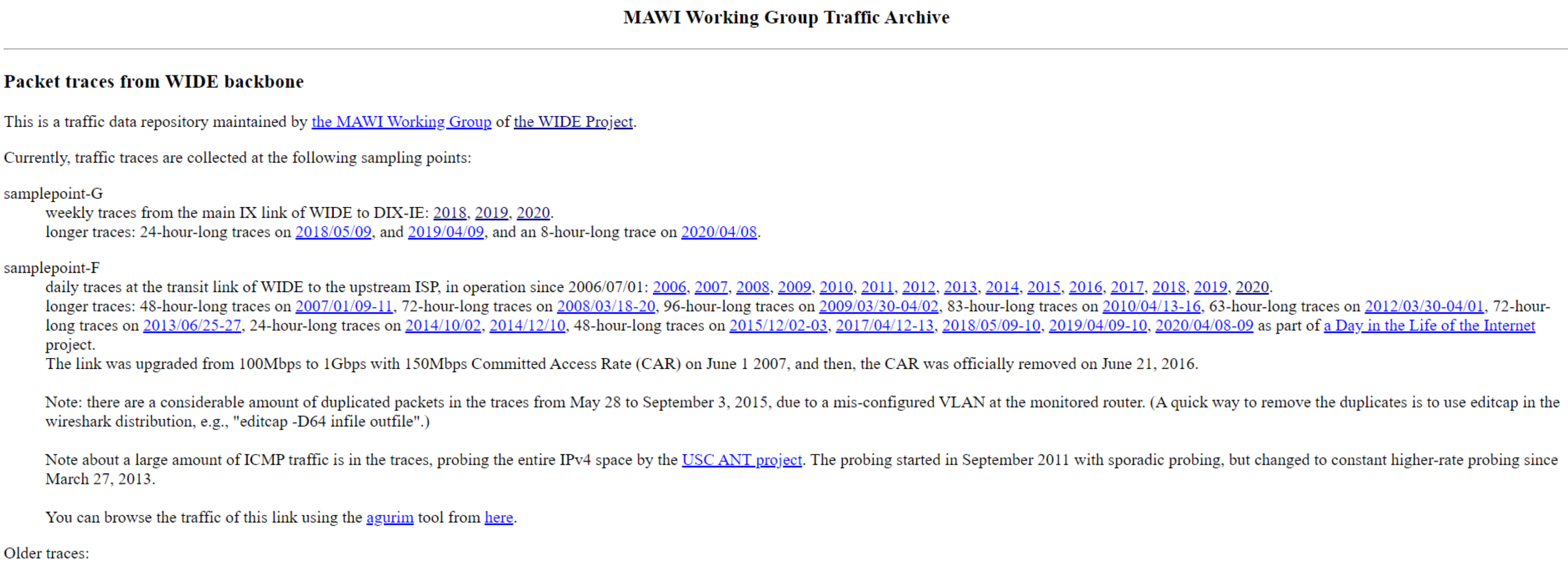

example.pcapng
File Modifica Visualizza Vai Cattura Analizza Statistiche Telefonia Wireless Strumenti Aiuto
Applica un filtro di visualizzazione ... <Ctrl-/>

```
No.   Time        Source            Destination       Protocol  Length  Info
    1 0.000000    170.133.133.163   192.168.1.174     RTCP          86 Payload-specific Feedback   ALFB
    2 0.004544    192.168.1.174     170.133.133.163   UDP          202 57917 → 9000 Len=160
    3 0.007428    192.168.1.174     170.133.133.163   UDP          846 52440 → 9000 Len=804
    4 0.007490    192.168.1.174     170.133.133.163   UDP          845 52440 → 9000 Len=803
    5 0.009283    192.168.1.174     62.109.247.68     TLSv1.2      459 Application Data
    6 0.028135    192.168.1.174     170.133.133.163   UDP          819 52440 → 9000 Len=777
    7 0.028253    192.168.1.174     170.133.133.163   UDP          538 52440 → 9000 Len=496
    8 0.028293    192.168.1.174     170.133.133.163   UDP          654 52440 → 9000 Len=612
    9 0.028347    192.168.1.174     170.133.133.163   UDP          653 52440 → 9000 Len=611
   10 0.045436    192.168.1.174     170.133.133.163   UDP          202 57917 → 9000 Len=160
   11 0.067250    62.109.247.68     192.168.1.174     TCP           54 443 → 65166 [ACK] Seq=1 Ack=406 Win=2407 Len=0
   12 0.067985    192.168.1.174     170.133.133.163   UDP          962 52440 → 9000 Len=920
   13 0.068047    192.168.1.174     170.133.133.163   UDP         1018 52440 → 9000 Len=976
   14 0.068094    192.168.1.174     170.133.133.163   UDP          646 52440 → 9000 Len=604
   15 0.068131    192.168.1.174     170.133.133.163   UDP          995 52440 → 9000 Len=953
   16 0.068161    192.168.1.174     170.133.133.163   UDP          619 52440 → 9000 Len=577
   17 0.081873    170.133.133.163   192.168.1.174     UDP          106 9000 → 57917 Len=64
   18 0.085326    192.168.1.174     170.133.133.163   UDP          202 57917 → 9000 Len=160
   19 0.087473    192.168.1.174     170.133.133.163   UDP          645 52440 → 9000 Len=603
   20 0.087567    192.168.1.174     170.133.133.163   UDP         1113 52440 → 9000 Len=1071
   21 0.087628    192.168.1.174     170.133.133.163   UDP         1113 52440 → 9000 Len=1071
   22 0.087686    192.168.1.174     170.133.133.163   UDP          715 52440 → 9000 Len=673
   23 0.103020    170.133.133.163   192.168.1.174     UDP          106 9000 → 57917 Len=64

> Frame 1: 86 bytes on wire (688 bits), 86 bytes captured (688 bits) on interface \Device\NPF_{5912C894-2D99-4693-8FAA-D68E6F4E55A8}, id 0
> Ethernet II, Src: ZyxelCom_86:e0:fc (5c:e2:8c:86:e0:fc), Dst: IntelCor_4d:17:a5 (d0:c6:37:4d:17:a5)
> Internet Protocol Version 4, Src: 170.133.133.163, Dst: 192.168.1.174
> User Datagram Protocol, Src Port: 9000, Dst Port: 57917
> Real-time Transport Control Protocol (Payload-specific Feedback)
```

WIDE Dataset

- The dataset to be used is the WIDE backbone network one.
- Available at http://mawi.wide.ad.jp/mawi/



MAWI Working Group Traffic Archive

Packet traces from WIDE backbone

This is a traffic data repository maintained by the MAWI Working Group of the WIDE Project.

Currently, traffic traces are collected at the following sampling points:

samplepoint-G
weekly traces from the main IX link of WIDE to DIX-IE: 2018, 2019, 2020.
longer traces: 24-hour-long traces on 2018/05/09, and 2019/04/09, and an 8-hour-long trace on 2020/04/08.

samplepoint-F
daily traces at the transit link of WIDE to the upstream ISP, in operation since 2006/07/01: 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020.
longer traces: 48-hour-long traces on 2007/01/09-11, 72-hour-long traces on 2008/03/18-20, 96-hour-long traces on 2009/03/30-04/02, 83-hour-long traces on 2010/04/13-16, 63-hour-long traces on 2012/03/30-04/01, 72-hour-long traces on 2013/06/25-27, 24-hour-long traces on 2014/10/02, 2014/12/10, 48-hour-long traces on 2015/12/02-03, 2017/04/12-13, 2018/05/09-10, 2019/04/09-10, 2020/04/08-09 as part of a Day in the Life of the Internet project.
The link was upgraded from 100Mbps to 1Gbps with 150Mbps Committed Access Rate (CAR) on June 1 2007, and then, the CAR was officially removed on June 21, 2016.

Note: there are a considerable amount of duplicated packets in the traces from May 28 to September 3, 2015, due to a mis-configured VLAN at the monitored router. (A quick way to remove the duplicates is to use editcap in the wireshark distribution, e.g., "editcap -D64 infile outfile".)

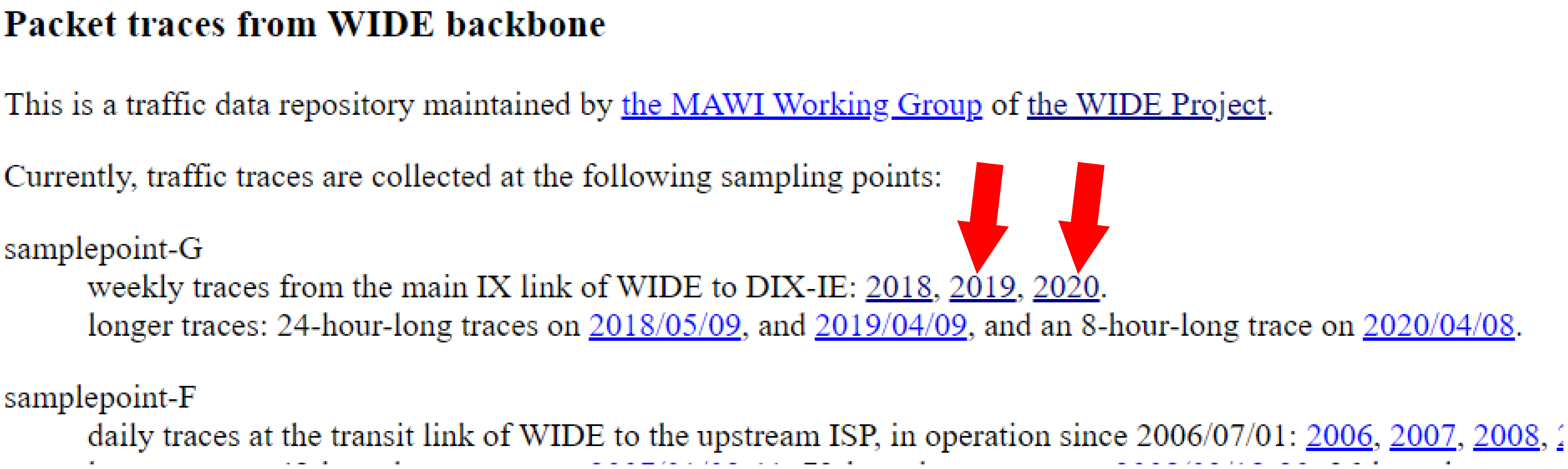
Note about a large amount of ICMP traffic is in the traces, probing the entire IPv4 space by the USC ANT project. The probing started in September 2011 with sporadic probing, but changed to constant higher-rate probing since March 27, 2013.

You can browse the traffic of this link using the agurim tool from here.

Older traces:

WIDE Dataset

- Samplepoint-G
- Each group will have a different dataset



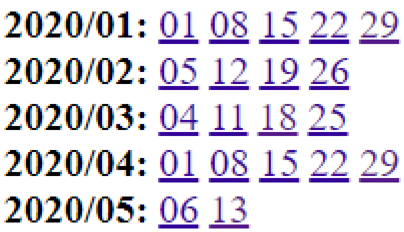
Packet traces from WIDE backbone

This is a traffic data repository maintained by the MAWI Working Group of the WIDE Project.

Currently, traffic traces are collected at the following sampling points:

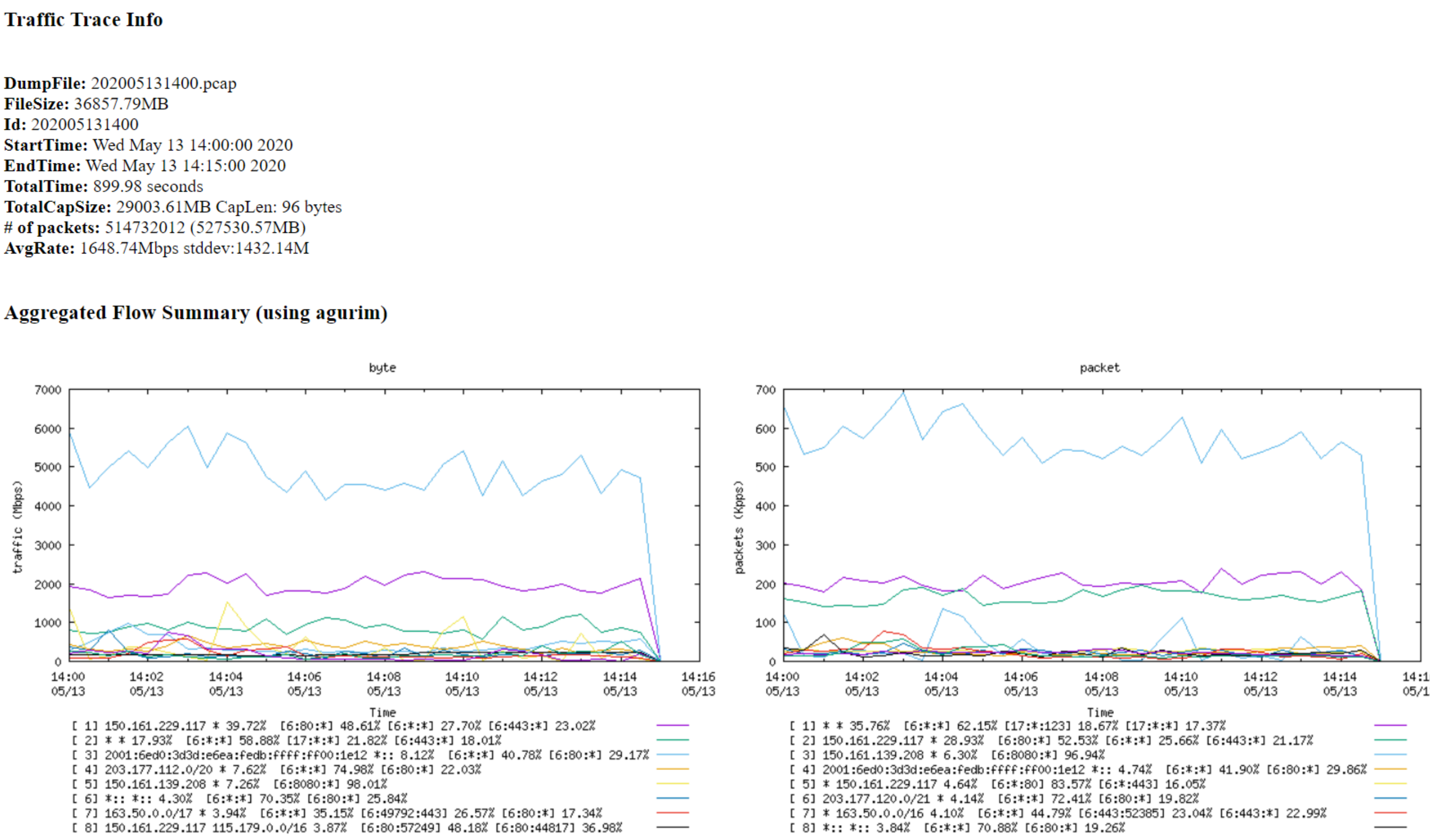
samplepoint-G
weekly traces from the main IX link of WIDE to DIX-IE: 2018, 2019, 2020.
longer traces: 24-hour-long traces on 2018/05/09, and 2019/04/09, and an 8-hour-long trace on 2020/04/08.

samplepoint-F
daily traces at the transit link of WIDE to the upstream ISP, in operation since 2006/07/01: 2006, 2007, 2008,



2020/01: 01 08 15 22 29
2020/02: 05 12 19 26
2020/03: 04 11 18 25
2020/04: 01 08 15 22 29
2020/05: 06 13

WIDE Dataset



Traffic Trace Info

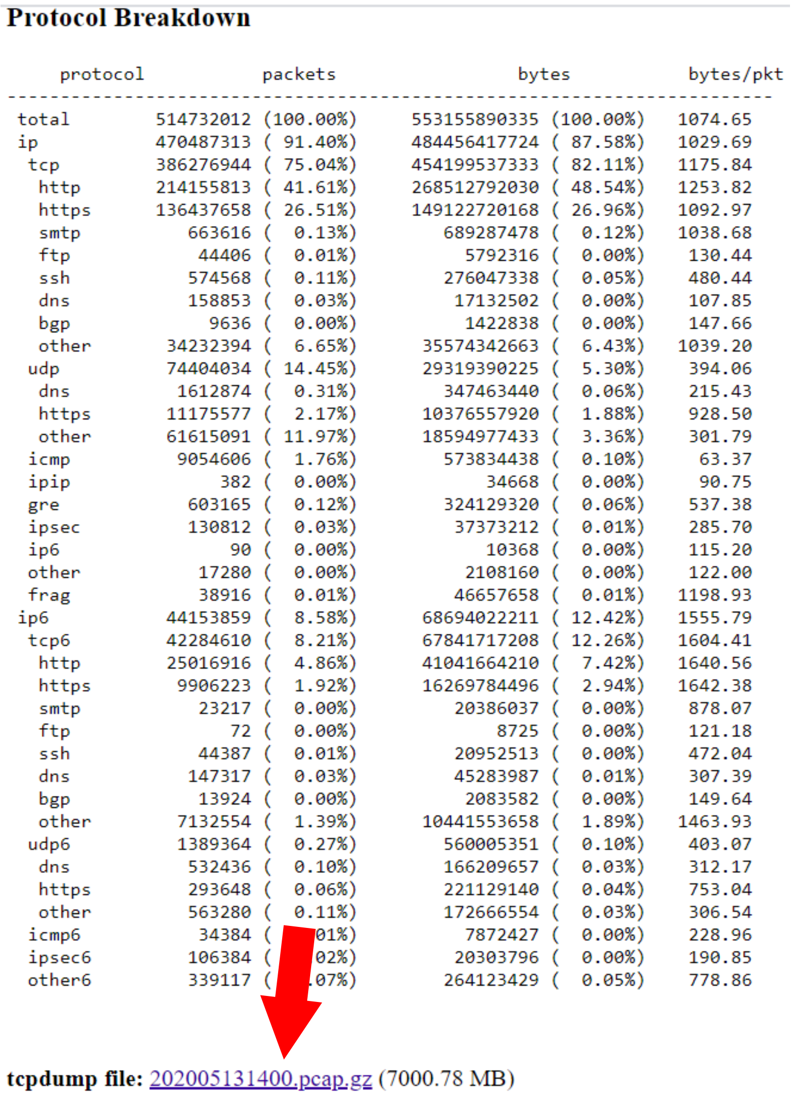
DumpFile: 202005131400.pcap
FileSize: 36857.79MB
Id: 202005131400
StartTime: Wed May 13 14:00:00 2020
EndTime: Wed May 13 14:15:00 2020
TotalTime: 899.98 seconds
TotalCapSize: 29003.61MB CapLen: 96 bytes
# of packets: 514732012 (527530.57MB)
AvgRate: 1648.74Mbps stddev:1432.14M

Aggregated Flow Summary (using agurim)

```
[ 1] 150.161.229.117 * 39.72%  [6:80:*] 48.61% [6:*:*] 27.70% [6:443:*] 23.02%
[ 2] * * 17.93%  [6:*:*] 58.88% [17:*:*] 21.82% [6:443:*] 18.01%
[ 3] 2001:6ed0:3d3d:e6ea:fedb:ffff:ff00:1e12 *:: 8.12%  [6:*:*] 40.78% [6:80:*] 29.17%
[ 4] 203.177.112.0/20 * 7.62%  [6:*:*] 74.98% [6:80:*] 22.03%
[ 5] 150.161.139.208 * 7.26%  [6:8080:*] 98.01%
[ 6] *:: *:: 4.30%  [6:*:*] 70.35% [6:80:*] 25.84%
[ 7] 163.50.0.0/17 * 3.94%  [6:*:*] 35.15% [6:49792:443] 26.57% [6:80:*] 17.34%
[ 8] 150.161.229.117 115.179.0.0/16 3.87%  [6:80:57249] 48.18% [6:80:44817] 36.98%
```

```
[ 1] * * 35.76%  [6:*:*] 62.15% [17:*:123] 18.67% [17:*:*] 17.37%
[ 2] 150.161.229.117 * 28.93%  [6:80:*] 52.53% [6:*:*] 25.66% [6:443:*] 21.17%
[ 3] 150.161.139.208 * 6.30%  [6:8080:*] 96.94%
[ 4] 2001:6ed0:3d3d:e6ea:fedb:ffff:ff00:1e12 *:: 4.74%  [6:*:*] 41.90% [6:80:*] 29.86%
[ 5] * 150.161.229.117 4.64%  [6:*:80] 83.57% [6:*:443] 16.05%
[ 6] 203.177.120.0/21 * 4.14%  [6:*:*] 72.41% [6:80:*] 19.82%
[ 7] * 163.50.0.0/16 4.10%  [6:*:*] 44.79% [6:443:52385] 23.04% [6:443:*] 22.99%
[ 8] *:: *:: 3.84%  [6:*:*] 70.88% [6:80:*] 19.26%
```

WIDE Dataset



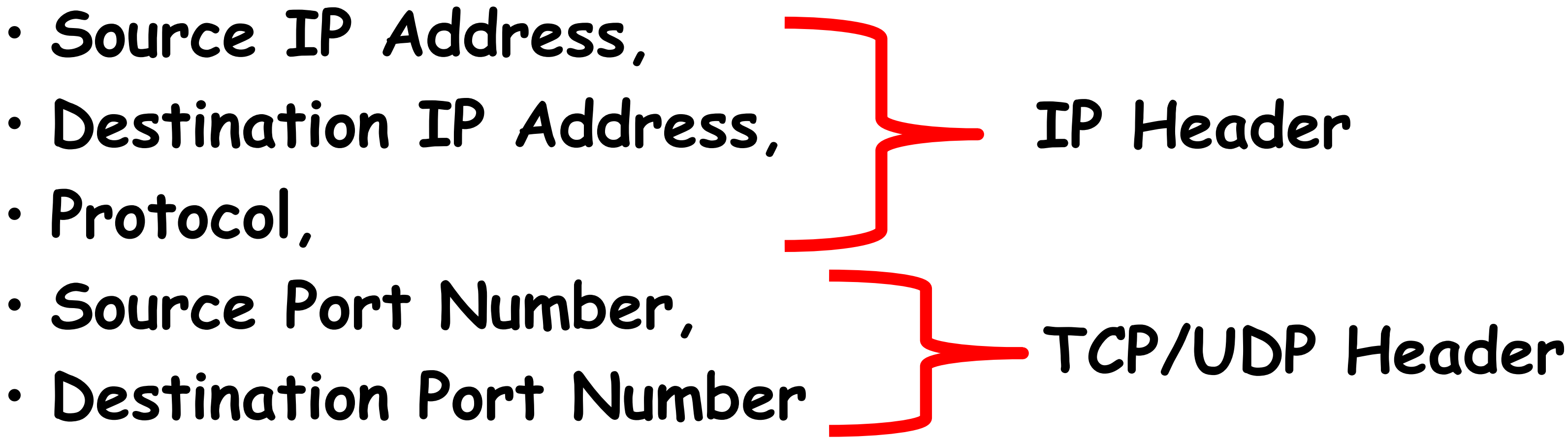
Protocol Breakdown

```
     protocol            packets                 bytes           bytes/pkt
------------------------------------------------------------------------
 total          514732012 (100.00%)     553155890335 (100.00%)   1074.65
 ip             470487313 ( 91.40%)     484456417724 ( 87.58%)   1029.69
  tcp           386276944 ( 75.04%)     454199537333 ( 82.11%)   1175.84
   http         214155813 ( 41.61%)     268512792030 ( 48.54%)   1253.82
   https        136437658 ( 26.51%)     149122720168 ( 26.96%)   1092.97
   smtp            663616 (  0.13%)        689287478 (  0.12%)   1038.68
   ftp              44406 (  0.01%)          5792316 (  0.00%)    130.44
   ssh             574568 (  0.11%)        276047338 (  0.05%)    480.44
   dns             158853 (  0.03%)         17132502 (  0.00%)    107.85
   bgp               9636 (  0.00%)          1422838 (  0.00%)    147.66
   other         34232394 (  6.65%)      35574342663 (  6.43%)   1039.20
  udp            74404034 ( 14.45%)      29319390225 (  5.30%)    394.06
   dns            1612874 (  0.31%)        347463440 (  0.06%)    215.43
   https         11175577 (  2.17%)      10376557920 (  1.88%)    928.50
   other         61615091 ( 11.97%)      18594977433 (  3.36%)    301.79
  icmp            9054606 (  1.76%)        573834438 (  0.10%)     63.37
  ipip                382 (  0.00%)            34668 (  0.00%)     90.75
  gre              603165 (  0.12%)        324129320 (  0.06%)    537.38
  ipsec            130812 (  0.03%)         37373212 (  0.01%)    285.70
  ip6                  90 (  0.00%)            10368 (  0.00%)    115.20
  other             17280 (  0.00%)          2108160 (  0.00%)    122.00
  frag              38916 (  0.01%)         46657658 (  0.01%)   1198.93
 ip6             44153859 (  8.58%)      68694022211 ( 12.42%)   1555.79
  tcp6           42284610 (  8.21%)      67841717208 ( 12.26%)   1604.41
   http          25016916 (  4.86%)      41041664210 (  7.42%)   1640.56
   https          9906223 (  1.92%)      16269784496 (  2.94%)   1642.38
   smtp             23217 (  0.00%)         20386037 (  0.00%)    878.07
   ftp                 72 (  0.00%)             8725 (  0.00%)    121.18
   ssh              44387 (  0.01%)         20952513 (  0.00%)    472.04
   dns             147317 (  0.03%)         45283987 (  0.01%)    307.39
   bgp              13924 (  0.00%)          2083582 (  0.00%)    149.64
   other          7132554 (  1.39%)      10441553658 (  1.89%)   1463.93
  udp6            1389364 (  0.27%)        560005351 (  0.10%)    403.07
   dns             532436 (  0.10%)        166209657 (  0.03%)    312.17
   https           293648 (  0.06%)        221129140 (  0.04%)    753.04
   other           563280 (  0.11%)        172666554 (  0.03%)    306.54
  icmp6             34384 (    01%)          7872427 (  0.00%)    228.96
  ipsec6           106384 (    02%)         20303796 (  0.00%)    190.85
  other6           339117 (   .07%)        264123429 (  0.05%)    778.86
```

tcpdump file: 202005131400.pcap.gz (7000.78 MB)

Labwork

- Data from .pcap format to a new format (.txt?) needed for further analysis
- First analysis (mandatory): from packets to flows
  - A **flow** is a set of packets related to the same traffic relationship;
  - All the packets of a flow will have the same:



- **Source IP Address,**
- **Destination IP Address,**
- **Protocol,**

IP Header

- **Source Port Number,**
- **Destination Port Number**

TCP/UDP Header

- Additional analysis (facultative).

Labwork

- The Labwork will be presented by the group.
- Presentation time: 10 minutes
- Before the presentation, the group must send to me the code (don't share it among groups!).
- Presentation before 20 July.

Labwork grade

- My part: up to **21**
- Prof. Baiocchi's part: up to **11**

- Midterm : up to **18**
- Labwork
  - Mandatory part: up to **2**
  - Facultative part: up to **2**

- I know that 18 + 2 + 2 = 22 ……… let's say that 22=21 ☺

Groups

N	Group name	Trace
1	Abramson	2020/01: 01
2	Baran	2020/01: 08
3	Cerf	2020/01: 15
4	Dijkstra	2020/01: 22
5	Erlang	2020/01: 29
6	Floyd	2020/02: 05
7	Gray	2020/02: 12
8	Huffman	2020/02: 19
9	Iverson	2020/02: 26
10	Jacobson	2020/03: 04
11	Kleinrock	2020/03: 11
12	Little	2020/03: 18
13	Markov	2020/03: 25
14	Metcalfe	2020/04: 01
15	Nyquist	2020/04: 08
16	Ohm	2020/04: 15
17	Perlman	2020/04: 22
18	Quimby	2020/04: 29
19	Rivest	2020/05: 06
20	Shannon	2020/05: 13
21	Tesla	2019/12: 18
22	Turing	2019/12: 11
23	Umeda	2019/12: 04
24	Viterbi	2019/11: 27
25	Wiener	2019/11: 20
26	Young	2019/11: 13
27	Zipf	2019/11: 06