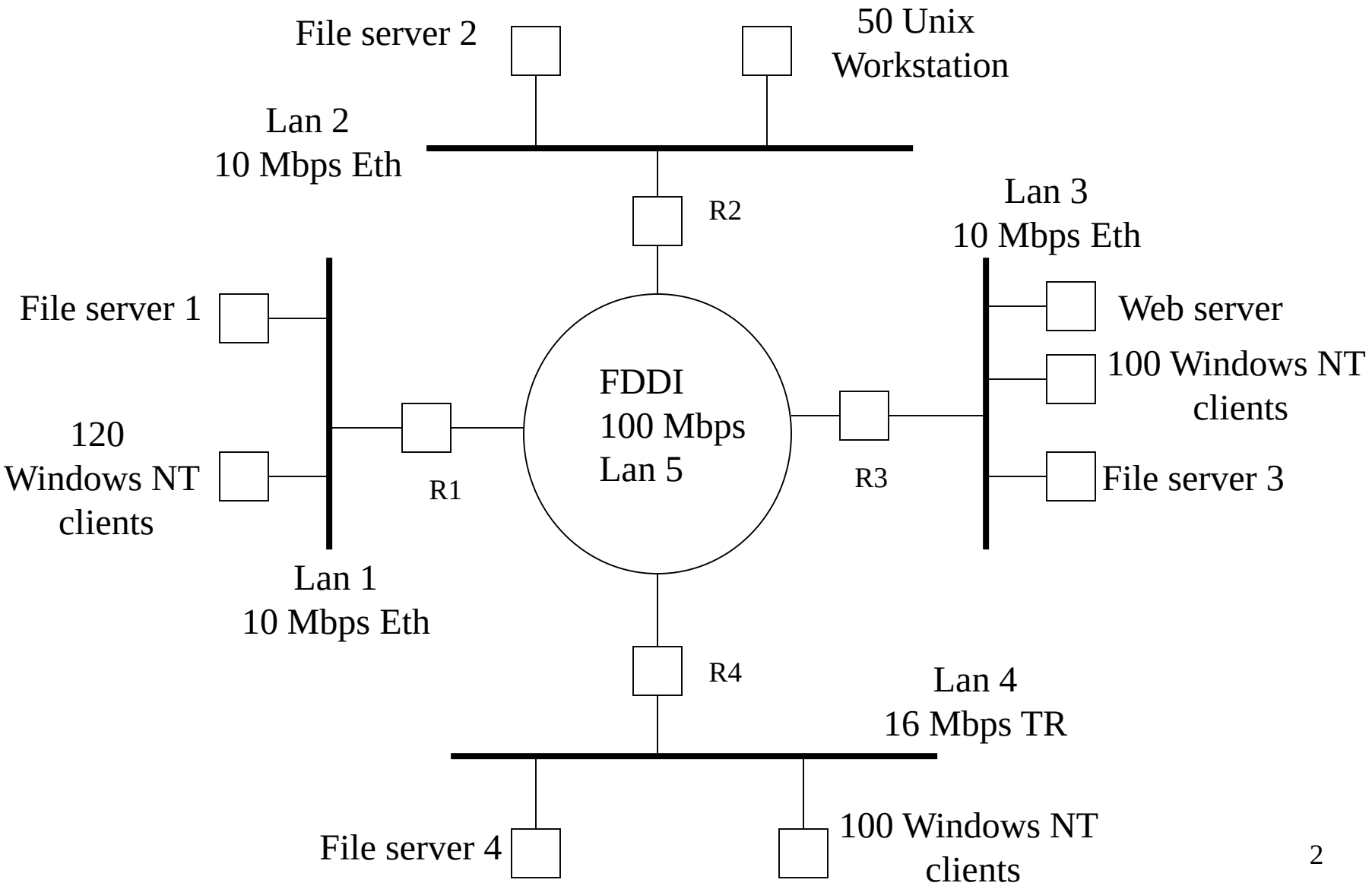
Evaluation of applications over an intranet

Intranet architecture



Workload analysis

Intranet applications (there are different classes):

1) Corporate training: web based application

2) Access to local file system

Infrastructure hypothesis

- All file server and web server have a single CPU and a single disk
- The FDDI and the routers are very fast versus the Lan Ethernet, then they can be modelled as a simple delay, therefore they are server without queue
- All the other components can be modelled as queues with service time independent by the load.

User hypotheses

- Finite number of users
- An user generates a new request, after having received an answer to the previous request, after a time equal to its thinking time
- 85% of users is working with the local file system
- 15% of users is working with the Web Server

Queueing network

Multiclass closed queue model, each class is charaterized by:
client group, application, server

- client group: CLi: clients in Lan i (i: 1 to 4)
- application:
  - FS for local file server access,
  - TR for Training
- server:
  - FSi: i-th NFS server (i: 1 to 4)
  - WebS: Web Server

Class types and number of users

(CL1, FS, FS1)	120 x 0.85 = 102
(CL2, FS, FS2)	50 X 0.85 = 43
(CL3, FS, FS3)	100 x 0.85 = 85
(CL4, FS, FS4)	100 x 0.85 = 85
(CL1, TR, WebS)	120 x 0.15 = 18
(CL2, TR, WebS)	50 x 0,15 = 7
(CL3, TR, WebS)	100 x 0,15 = 15
(CL4, TR, WebS)	100 x 0,15 = 15

Server types

Routers: low delay, given their low latency

FDDI ring: low delay, given high bandwidth

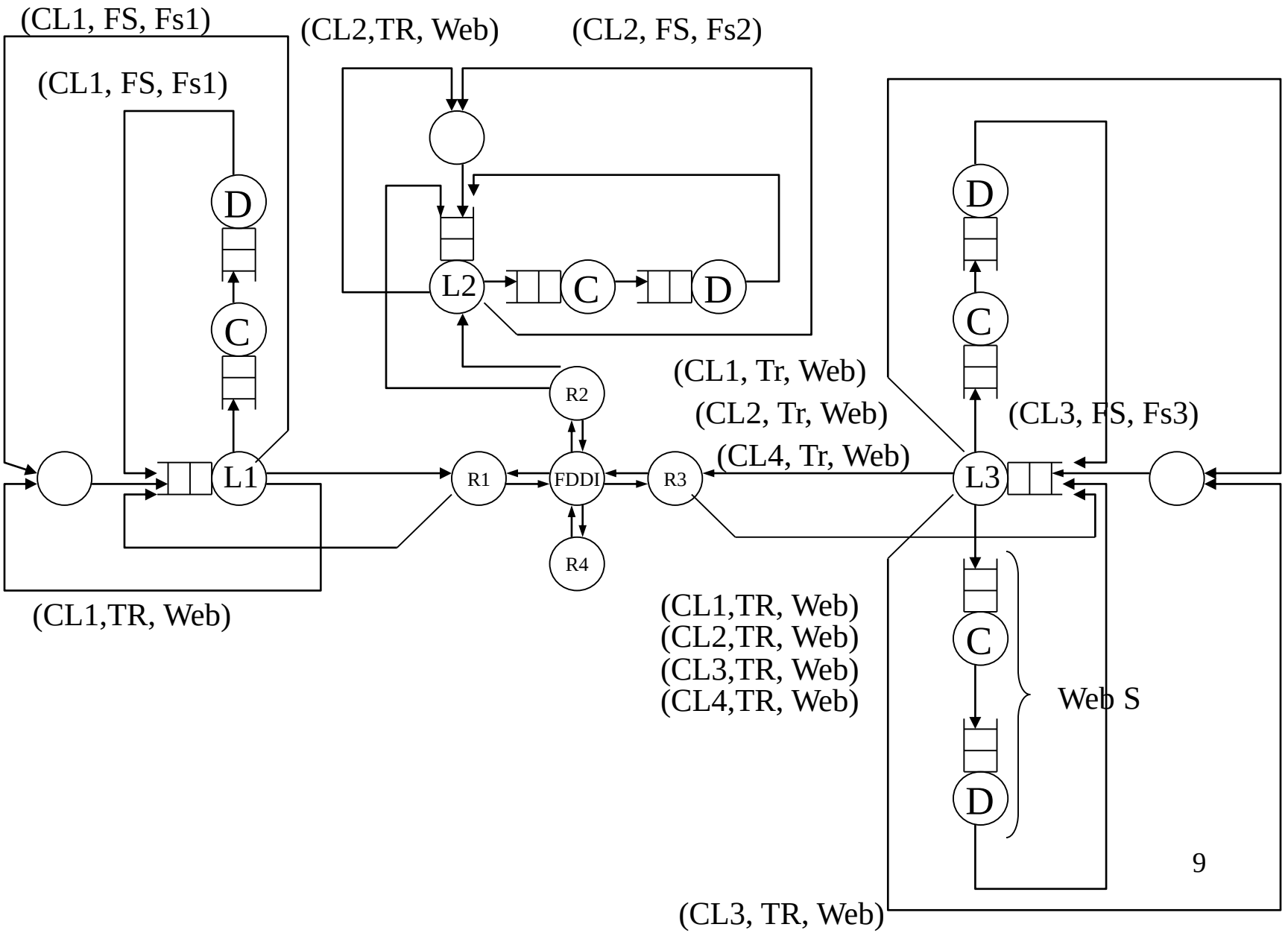
CPU, Disks, LANs } servers with load independent service time

Service Demands: $D_{i,r} = V_{i,r} \times S_{i,r}$

where $\mathbf{V_{i,r}}$ = Visit Ratio

$\mathbf{S_{i,r}}$= Service Time

QN model


(CL1, FS, Fs1)
(CL2,TR, Web)
(CL2, FS, Fs2)
(CL1, FS, Fs1)
D
C
L2
C
D
D
C
(CL1, Tr, Web)
R2
(CL2, Tr, Web)
(CL3, FS, Fs3)
(CL4, Tr, Web)
L1
R1
FDDI
R3
L3
R4
(CL1,TR, Web)
(CL1,TR, Web)
(CL2,TR, Web)
(CL3,TR, Web)
(CL4,TR, Web)
C
Web S
D
(CL3, TR, Web)

Web server workload characterization
(for a training session)

- Avg request document size per HTTP request:
  - 20 rqs for txt documents (2.000 bytes per doc)
  - 100 rqs for inline images (50.000 bytes each)
    - (20 text pages x 5 inline/text pages)
  - 15 rqs for other multi-media (mm) obj (2.000.000 bytes each)

Web server workload characterization

- % request for:
  - txt documents = 20/(20+100+15) = 15 %
  - inline images = 100/(20+100+15) = 74 %
  - other mm obj = 15/(20+100+15) = 11 %

Web server workload characterization

- Average document size

  0.15 x 2.000 + 0.74 x 50.000 + 0.11 x 2.000.000 =

  = 257.300 bytes

Note it is an average, the overhead associated to the three document sizes is different

Web server workload characterization

- Document request arrival rate is function of the think time and of the number of users in the thinking phase

(CLi, TR, Web) ⇒ $\frac{\#Users_i \text{ in thinking phase}}{\text{think time}}$

45 sec

Max # $Users_i$

18 per Lan 1
7 per Lan 2
15 per Lan 3
15 per Lan 4

Web server workload characterization

- Device service time
  - CPU: 1 msec processing time x HTTP request
  - Disk:

    We need to consider
    - Seekrand= avg time to position at a random cylinder
    - DiskRevTIme = time for a complete disk revolution
    - TransferTime = BlockSize/ 10^6 x TransferRate
    - ControllerTime = time spent at the controller for an I/O req.

$$S_d = \text{ControllerTime} + \text{Pmiss} \times (\text{SeekRand} + \text{DiskRevolutionTime}/2 + \text{TransferTime})$$

Web server workload characterization

- Lan hp: no fragmentation i.e. max data area 1500 bytes; hp no data overhead for HTTP request

$$NDatagrams = \frac{MessageSize + TCPOvhd}{\min_n MTU_n - IPOvhd}$$

$$Overhead_n = TCPOvhd + Ndatagrams \times (IPOvhd + FrameOvhd_n)$$

$$ServiceTime_n = \frac{8 \times (MessageSize + Overhead_n)}{10^6 \times Bandwidth}$$

Web server workload characterization

- Lan hp: no fragmentation i.e. max data area 1500 bytes; hp no data overhead for HTTP request
- **Ethernet**

$$NDatagrams = \frac{257300 + 20}{1500 - 20} = 174$$

$$Overhead_n = 20 + Ndatagrams \times (20 + 18) = 6632$$

$$ServiceTime_n = \frac{8 \times (257300 + 6632)}{10^6 \times Bandwidth}$$

Web server workload characterization

- Lan hp: no fragmentation i.e. max data area 1500 bytes; hp no data overhead for HTTP request
- **Token ring**

$$NDatagrams = \frac{257300 + 20}{1500 - 20} = 174$$

$$Overhead_n = 20 + Ndatagrams \times (20 + 28) = 8372$$

$$ServiceTime_n = \frac{8 \times (257300 + 8372)}{10^6 \times Bandwidth}$$

Web server workload characterization

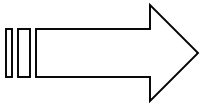
- Router
  - delay 134 **μsec** x packet (approximated in total to 1 msec)

- FDDI
  - delay with

$$\text{ServiceTime}_n = \frac{8 \times (\text{MessageSize} + \text{Overhead}_n)}{10^6 \times \text{Bandwidth}}$$

Local file system workload characterization

- File dimension 8192 bytes
- avg NFS request arrival rate is function of the think time and number od users in the thinking phase

(CLi, FS, FSi) 

$$\frac{\#Users_i \text{ in thinking phase}}{\text{think time}}$$

10 sec

Max $\#Users_i$

102 per Lan 1
43 per Lan 2
85 per Lan 3
85 per Lan 4

Local file system workload characterization

- Device service time
  - CPU: 1 msec per file request
  - Disk:

    We need to consider
    - Seekrand= avg time to position at a random cylinder
    - DiskRevTIme = time for a complete disk revolution
    - TransferTime = BlockSize/ 10^6 x TransferRate
    - ControllerTime = time spent at the controller for an I/O req.
    - N blocks to read = 8192/2048 = 4

  - Lan i with 8192 bytes

Throughput & response time

Class	Throughput (req/sec)	Response time (sec)
• CL1, FS, FS1	10,12	0,08
• CL2, FS, FS2	4,23	0,06
• CL3, FS, FS3	8,44	0,08
• CL4, FS, FS4	8,44	0,07
• CL1, TR, WebS	0,34	8,58
• CL2, TR, WebS	0,14	8,55
• CL3, TR, WebS	0,28	7,96
• CL4, TR, WebS	0,28	8,35