

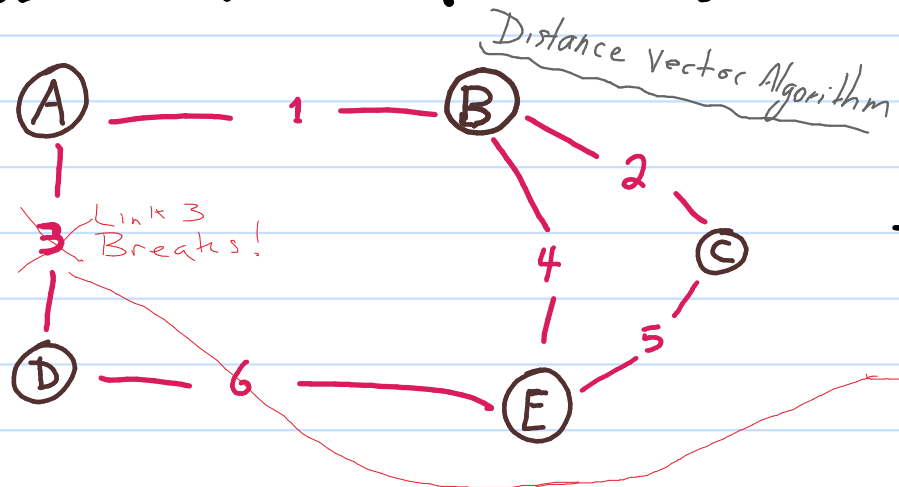
Oct 22, 2013

EXAM
 • DNS
 • Routing
 • What are zone files

The first node to recognize ∞ is the last to remove it!

ASSIGNMENT 4 - Pt #1

1



A	Link	Cost
A	local	∞
B	1	1
C	1	2
D	3	∞
E	1	2

Either A or D could have changed first.

B		C	
A	1	1	2
B	local	∞	1
C	2	1	L 0
D	1	2	5
E	4	1	5

3

B & D now calculate new routes.

A	B	D
Same	Same	A 6 3
		...
		4 2

2

A sends to adjacent nodes, B & D.

A	B	C	D
Same	Same	Same	A 3 ∞
			B 3 ∞
			C 6 2
			D L 0
			E 6 1

4

Finally, A calculates the updated route.

A
A
B
C
D
E

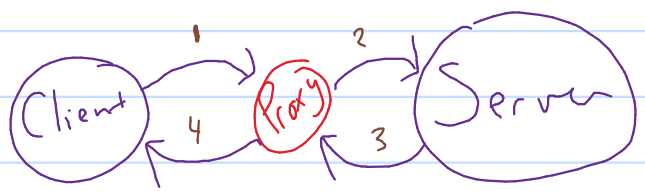
1 3 ✓
Now we're done!

Oct 23 - Proxies

Concerns:

- Scalability
- caching
- proxy caching

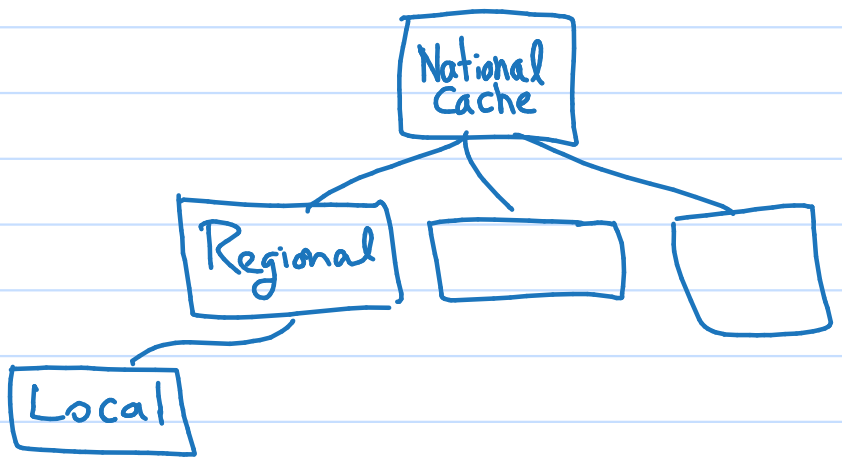
- Traffic
- Server Load
- Latency



- 2,3 will not be needed if already cached.

* proxy → DNS server to know the source server.

"NLNR"



Hit vs Miss

- Hit → Proxy has the doc cached
- Byte-hit ratio - shows reduction of traffic.

Oct 29

Exam

- 100 min
- Pt A - 32 Q
- Pt B - 5 Q

SQUID - slow

Part 1

- create DIRs
- config file
- /usr/sbin/squid...

username,
etc...

} just getting setup

Part 2

- log cleaning ①
 - Want only ~200" responses; filter Query String URLs; ②
 - ③ No HTTPS
- timestamp, Size, Code, URL

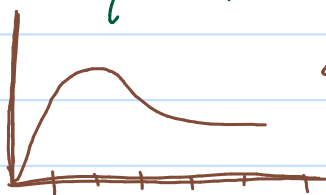
~ 2100 resulting URLs

Part 3

wget [url file] >/dev/null

Part 4

~/squid/... access log



URLs

← gnu plot * (or Excel)

Midterm Study Notes

Sockets

(JAVA) ServerSocket s = new ServerSocket(6528)
Socket s2 = s.accept();

(C) int s = socket(AF_INET, SOCK_STREAM, 0);

Threads

• Light-weight process

• Java:

① Thread subclass

② Runnable Interface

MyThread x = new MyThread();
x.run();

Semaphores:

~~int semget~~

Sem = createSem(1)

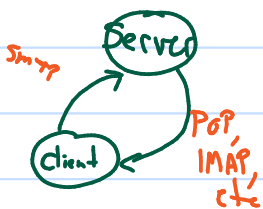
down(sem); up(sem);

Semaphore sem = new Semaphore(100, true);
sem.acquire();
sem.release();

HTTP

- 501 - not implemented

Mail



IMAP - usually doesn't delete emails from the server.
better for multiple devices.

SMTP

telnet localhost 25

HELO wuhu

MAIL FROM

RCPT TO

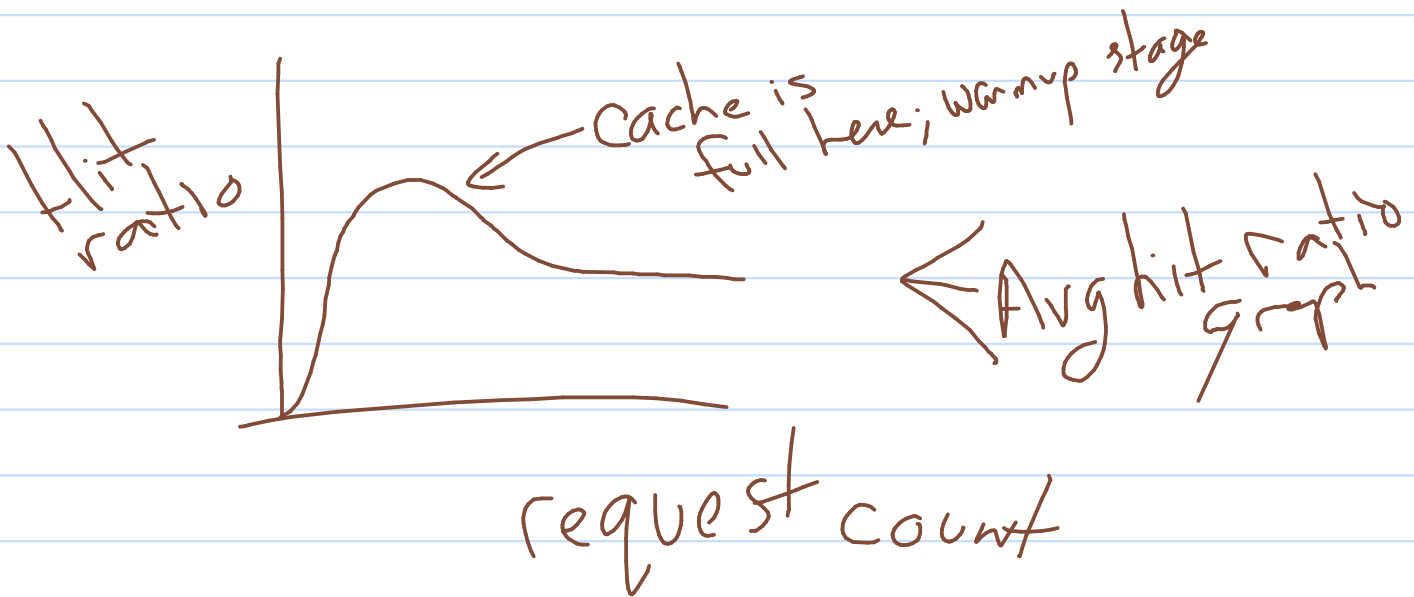
DATA

telnet wuhu 110

USER j2mair
PASS **
LIST

telnet localhost 25 ← years!!
MAIL FROM j2mair@sss...
RCPT TO
DATA
...
telnet wuhu 110
USER -
PASS -
LIST -

Need Squid log format for later lab (?)



* number of hits graph is more linear.

Cache Replacement Policy

Important Aspects

- ① - Popularity
- ② - Temporal Locality - if a request has been sent to Res A recently, what is the chance it will be requested again soon.

N.B. Metrics

- ① Hit ratio
- ② Byte hit ratio

Temporal Loc → better hit ratio

Example: employees looking at common online resources.
→ Analyze ① Locality
② Popularity

Greedy Dual Size Policy

$$h = \frac{\text{cost}}{\text{size}} \text{ for ea object}$$

latency of bringing the obj into the cache

Network cost -- # of hops

$$A \text{ obj: } h = \frac{4 \text{ hops}}{4M} = 1$$

$$B \text{ obj: } h = \frac{12 \text{ hops}}{4M} = 3 \leftarrow \text{Keep this one b/c more costly to get}$$

Nov 12/13

AG

• Graphs of diff policies should look about same
-- if the cache never becomes full!

- Need multiple stats: mean, mode, etc
→ because: can tell variance.

Sequence:

4100, 4700, 4200, 20000, 4000

mean: 7400, median: 4200

Seq 2:

4100, 4700, 4200, 4800, 4000

mean: 4300, median: 4200


Same

Workload Characterization

Workload \leadsto Inputs

- Number of requests
- Size of req' resources

Characterization \leadsto Modelling

- Graphs 

PDF
CDF

Nov 13

Web Workload

- 1 - # of requests
- 2 - file size
- 3 - # of embedded resources
- 4 - inter-arrival times

From yesterday

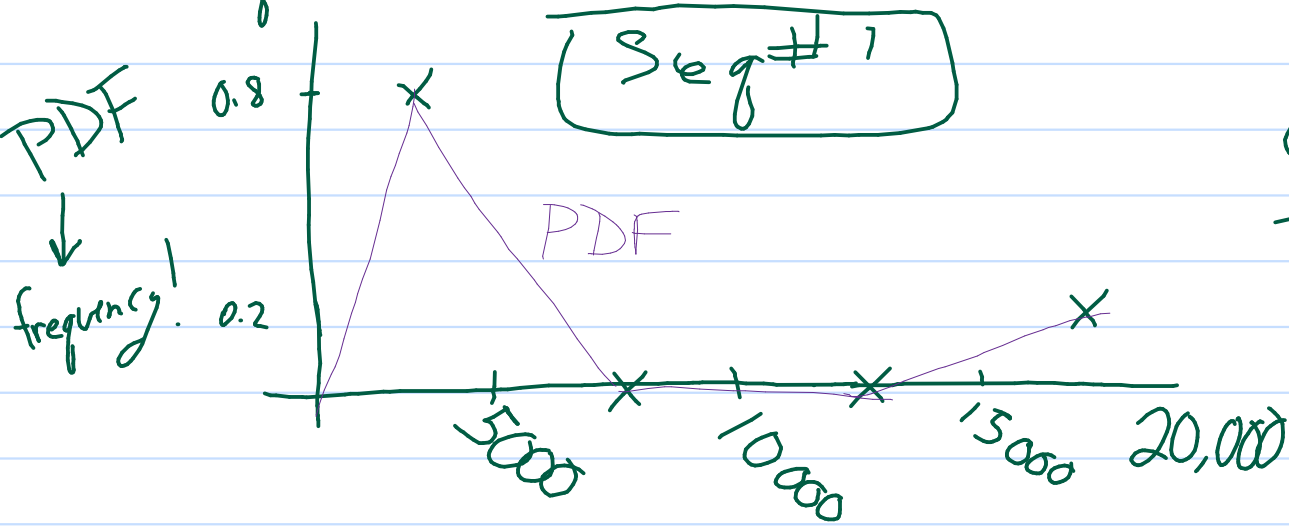
Sequence:

4100, 4700 4200 20000 4000

mean: 7400, median: 4200

Seq 2: 4100 4700 4200 4800 4000

mean: 4360 median: 4200



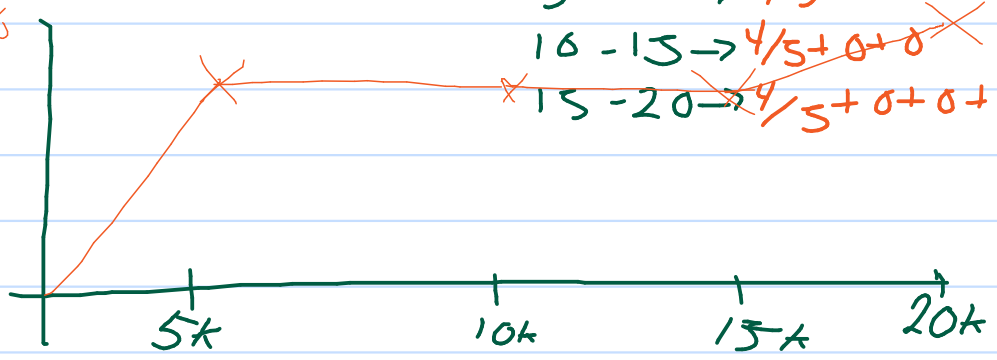
Excel
Data Analysis
Pack
"bin"

$0-5000 = 4 \rightarrow 4/5 = 0.8$
 $5001-10000 = 0 \rightarrow 0$
 $10k-15k = 0 \rightarrow 0$
 $15k-20k = 1 \rightarrow 1/5 = 0.2$

PDF \rightarrow connect points

CDF - Cumulative

1.0



$0-5k \rightarrow 4/5 = 0.8$
 $5-10 \rightarrow 4/5 + 0$
 $10-15 \rightarrow 4/5 + 0 + 0$
 $15-20 \rightarrow 4/5 + 0 + 0 + 1/5 = 1.0$

Nov 19/13

AG - explanation

#4: PDF & CDF - Interarrival Size

#5: URL Size == file Size

Interarrival Time \rightarrow diff in timestamps of arrivals

lognormal = ? \cdot X

\rightarrow See "Sample Graphs" \rightarrow real numbers.

"imperial" \rightarrow PDF, CDF

table 10.2 of the Notes \rightarrow Distributions

Excel = logNormal(X, mean, StdDev)

PDF = f_{bin} / Total
normal(x_{bin} , Mean, StdDev)
 \uparrow
20, 40, 60, etc...

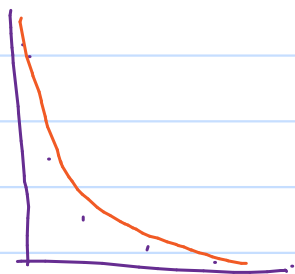
Interarrival
 \downarrow

expon.* should be

file size
 \downarrow

Nov 20

Zipf's Law: $p(r) = Kr^{-c}$



- Zipf applies to the popularity of things

Assignment #7

- Need to create Access Log \rightarrow copy 1000 timestamps from Squid Log
 - Use "weibul 2" in C for video file sizes (random)
-

Nov 26

C-program: "a", "b", vars; you can choose.

- remove negatives; mean should be around 10M
- can use Excel \rightarrow Weibul
- Vid View Count \rightarrow should be sim to Zipf.
 - * can be simple histogram
 - * view count \rightarrow log-logistic

File Sizes: need 10M as mean
So convert #s to KB

Final Exam

- Same format; MC & Q/A
- Calculator allowed
- 80-90% → After Midterm
 - 10-20% → Protocols
 - 90%: Web Proxies, Caching, Performance, Serving Multimedia in Web
 - Hit/Byte Ratio
 - Distributions → how to decide which is a good model
 - Assignments
 - MAYBE Q's from Midterm.

KHE118 Thurs 5th
3-6 pm.

Multimedia

