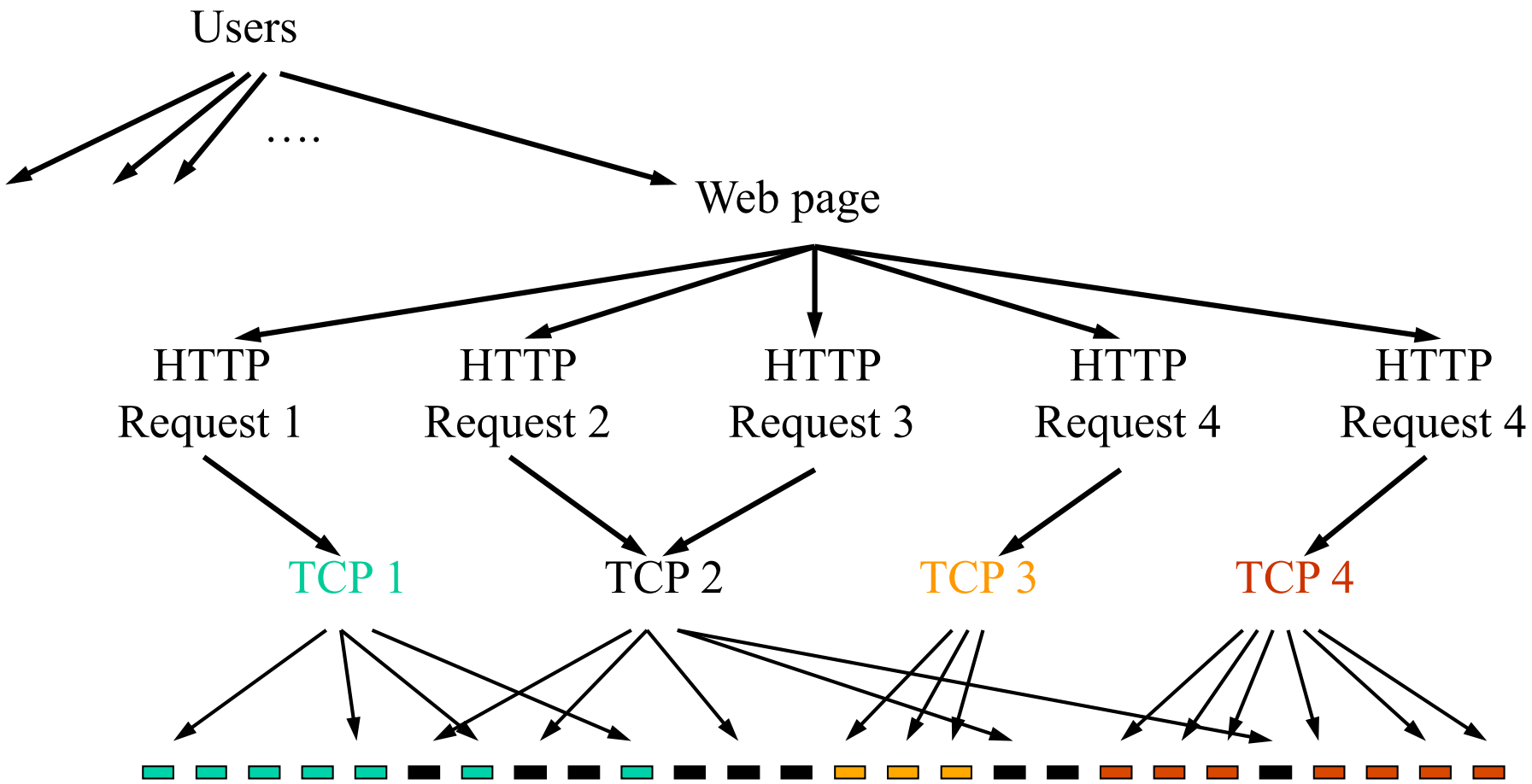
Web workloads for network and server performance evaluation

SPECweb (e.g., Specweb96)

- Method
  - Generate HTTP requests at constant rate
- Parameters
  - Target # of HTTP operations per second
  - # of threads

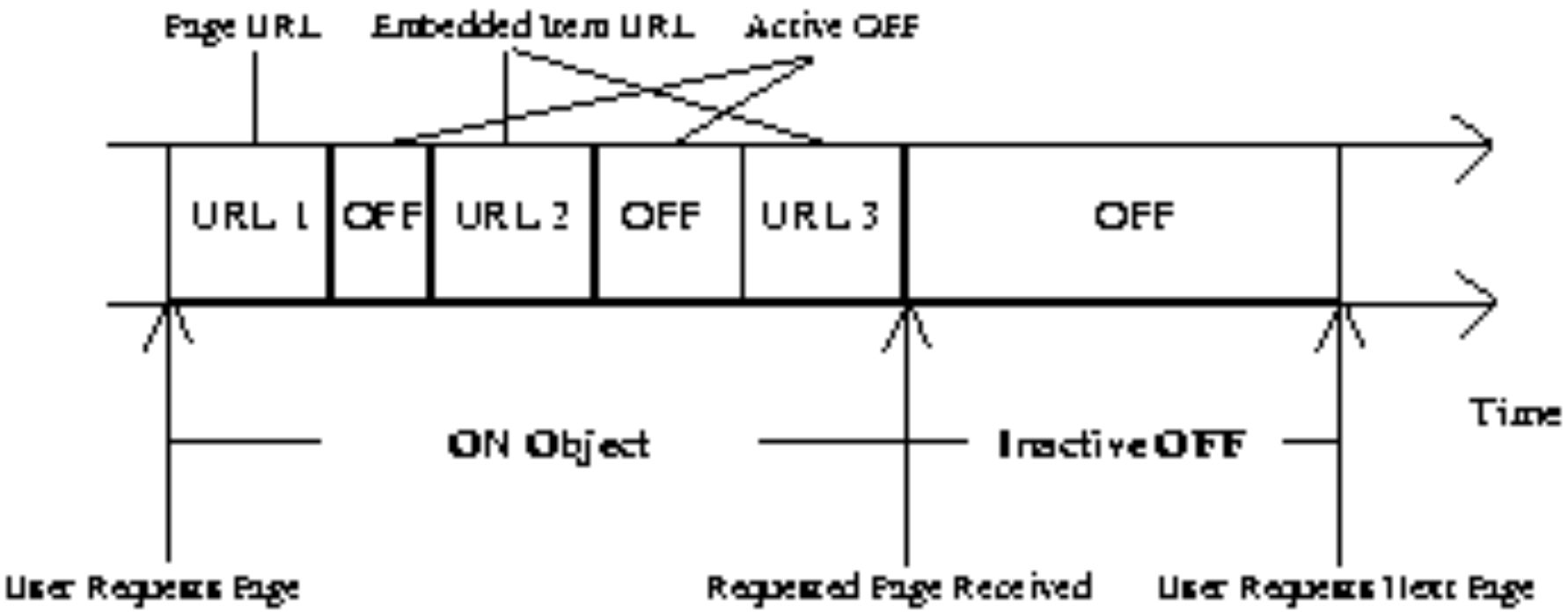
But how does Web traffic look like



SURGE: Scalable URL Ref. Generator

- Based upon empirical measurements
  - Server file size distribution
  - Request size distribution
  - Relative file popularity
  - Embedded file references
  - Temporal locality of reference
  - Idle periods of individual users
- Significant influence on Web server performance

SURGE: ON/OFF model



SURGE: Parameters

Component	Model	Probability Density Function	Parameters
File Sizes – Body	Lognormal	$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2 / 2\sigma^2}$	$\mu = 9.357; \sigma = 1.318$
File Sizes – Tail	Pareto	$p(x) = \alpha k^\alpha x^{-(\alpha+1)}$	$k = 133K; \alpha = 1.1$
Popularity	Zipf		
Temporal Locality	Lognormal	$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2 / 2\sigma^2}$	$\mu = 1.5; \sigma = 0.80$
Request Sizes	Pareto	$p(x) = \alpha k^\alpha x^{-(\alpha+1)}$	$k = 1000; \alpha = 1.0$
Active OFF Times	Weibull	$p(x) = \frac{b x^{b-1}}{a^b} e^{-(x/a)^b}$	$a = 1.46; b = 0.382$
Inactive OFF Times	Pareto	$p(x) = \alpha k^\alpha x^{-(\alpha+1)}$	$k = 1; \alpha = 1.5$
Embedded References	Pareto	$p(x) = \alpha k^\alpha x^{-(\alpha+1)}$	$k = 1; \alpha = 2.43$

Experimental setup: SURGE

- Clients
  - 6 PCs on 100 Mbps network
  - 200 Mhz Pentium Pro, 32MB RAM; Windows NT 4.0
  - Max. 50 SURGE UE
- Server
  - Apache v1.2.4; Linux 2.0
- Measures
  - Server
    - CPU utilization
    - # TCP connections
  - Network
    - # files, packets transferred

Experimental setup: SPECweb96

- Method
  - Generate HTTP requests at constant rate
- Parameters
  - Target # of HTTP operations per second
  - # of threads: 16

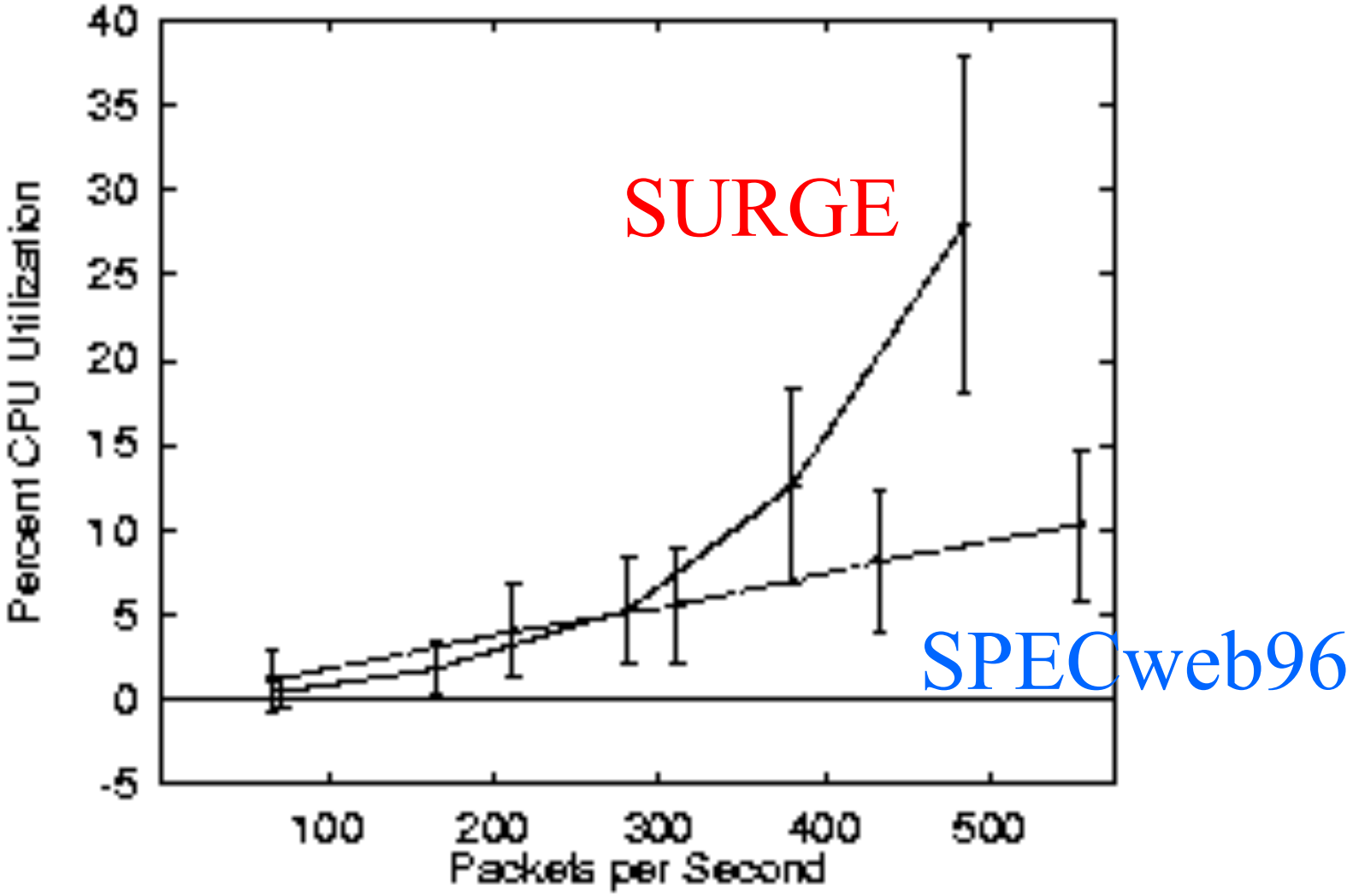
Nominal pkts/sec	SPECweb96 HTTP ops/sec	SURGE UEs
70	3	50
300	14	150
500	25	250

Table 2: Comparison SPECweb96 ops/sec and SURGE UE's.

	SPECweb96		SURGE	
Nominal pkts/sec	Requests	TCP Packets	Requests	TCP Packets
70	5901	118560	5293	131642
300	26028	560238	26055	507727
500	46520	1000289	48238	874570

Table 3: Summary of Comparison Experiments.

CPU utilization

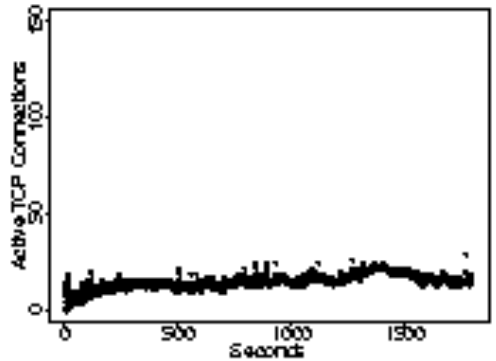


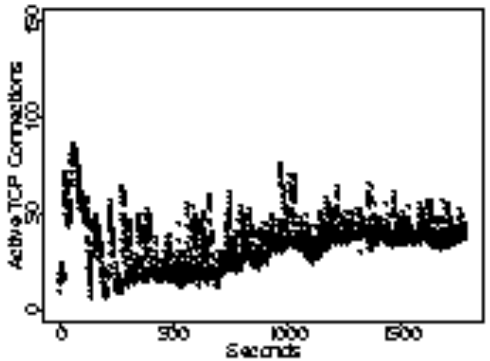
Active TCP connections

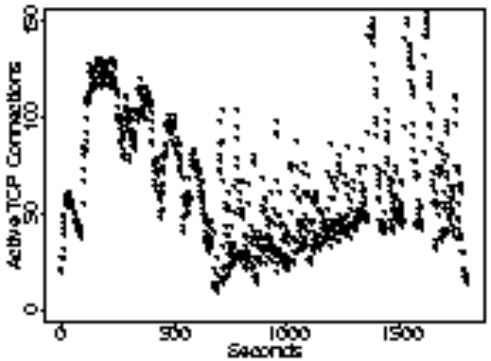
	SPECweb96		SURGE	
Nominal pps	Mean	Standard Deviation	Mean	Standard Deviation
70	0.028	0.18	13.9	3.92
300	0.37	0.69	33.2	12.1
500	0.71	1.41	67.1	35.3

Network effects

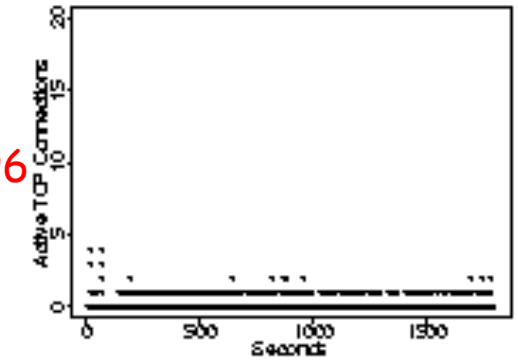
Surge

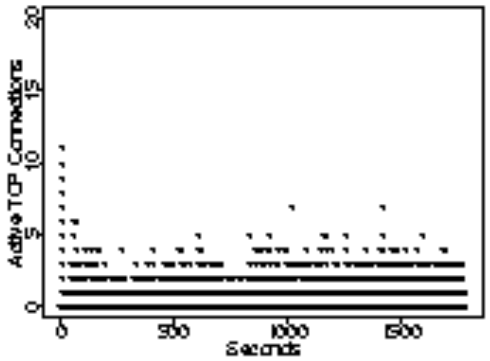


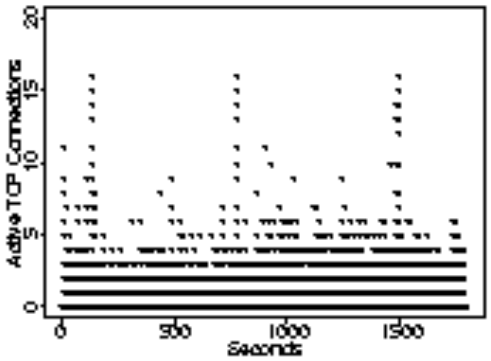




SPECweb96







70 300 500

Summary

- Workload generator
  - Based on analytical models of Web use
  - Captures important characteristics
  - Uses user equivalents to measure workload intensity
- Traffic
  - Requires larger # of open connections
  - Results in higher CPU load
  - Exercises the network differently

Realistic workload generation is crucial

otherwise optimistic assesment of system performance