Data Applications Traffic Characteristics

Internet Data Applications

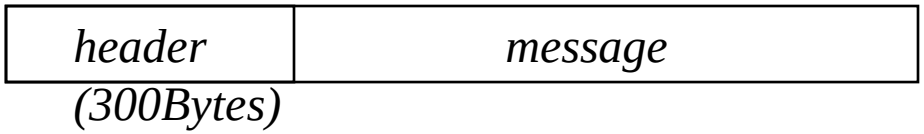
- *telnet* Remote Login
- *smtp* Simple Mail Transfer Protocol
- *nntp* Network News Transfer Protocol
- *ftp* File Transfer Protocol
- *http* Hypertext Transfer Protocol

Data Source Traffic Characteristics

- Extensive measurements of Internet traffic have been performed in recent years
  - *Vern Paxson, 1994*
  - *Mark Crovella, 1997*
- Measurements categorized by applications capturing
  - session duration
  - amount of data sent in a “transaction”
- Analytical models are created to fit measurements

Measurements of Interest (1)

- *telnet*
  - number of bytes sent by the responder in response to a command
- *smtp*
  - number of bytes in a mail message

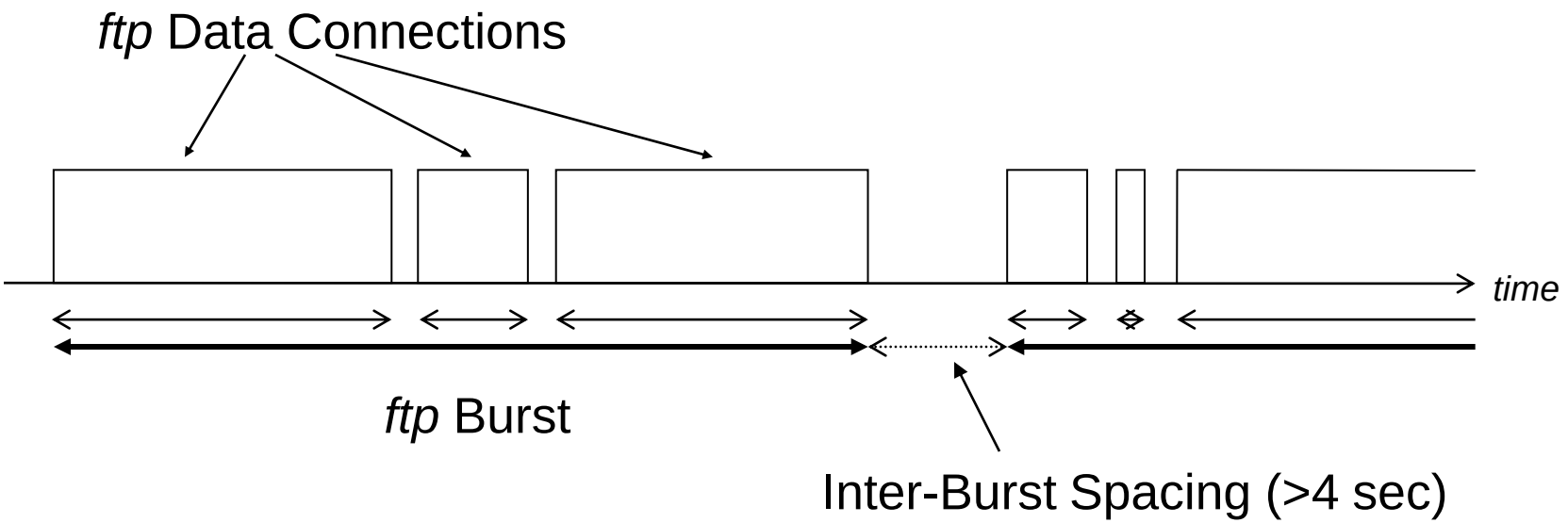


- *nntp*
  - number of bytes in a news message posted

Measurements of Interest (2)

- *ftp*:
  - *ftp* data connection:
    - number of bytes in a file transferred by *ftp*.
  - *ftp* session
    - total number of bytes transferred in *ftp* data connections during the same period of an *ftp* control connection
  - *ftp* burst (as *ftp* data connections tend to occur in bursts)
    - sequence of *ftp* data connections spaced out less than 4 s.
- *http*
  - *http* file size distribution found to be similar to *ftp* bursts size distribution

ftp Data Connection Burst



- Note:
  - model focuses on the upper 5% of the bursts (ranked by burst data volume) as they contribute 90% of the data

Analytic Models of Data Application Traffic Characterisitcs

Application	Variable	Model	Parameters	Mean
telnet	Responder (bytes)	Log_2-normal	$x = 4500$, $\sigma_x = 7.2$	31582
	Duration (sec.)	Log_2-normal	$X = 240$, $\sigma_x = 7.8$	1979
nntp	Originator (bytes)	Log_2-normal	$X = 2^{11.5}$, $\sigma_x = 3$	5295
smtp	Originator (bytes)	Log_2-normal	$X = 2^{10}$, $\sigma_x = 2.75$	1708
ftp	Connection (bytes)	Log_2-normal	$X = 3000$, $\sigma_x = 16$	140083
	Burst size (bytes)	Pareto	$\alpha = 1$, $k = 10^{5.5}$	∞
http	File size (bytes)	Pareto	$\alpha = 1.1$, $k = 10^{5.5}$	3478505

(Reference: Vern Paxson, "Empirically-Derived Analytic Models of Wide-Area TCP Connections")

The Log-Normal Distribution

- $Y \sim N(y_m, \sigma_y^2)$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\cdot\sigma_y} e^{\frac{-1}{2\sigma_y^2}\cdot(y-y_m)^2}$$

- X is $\log_2$-normal iff $X = 2^Y$; i.e., $Y = \log_2 X$
- Let

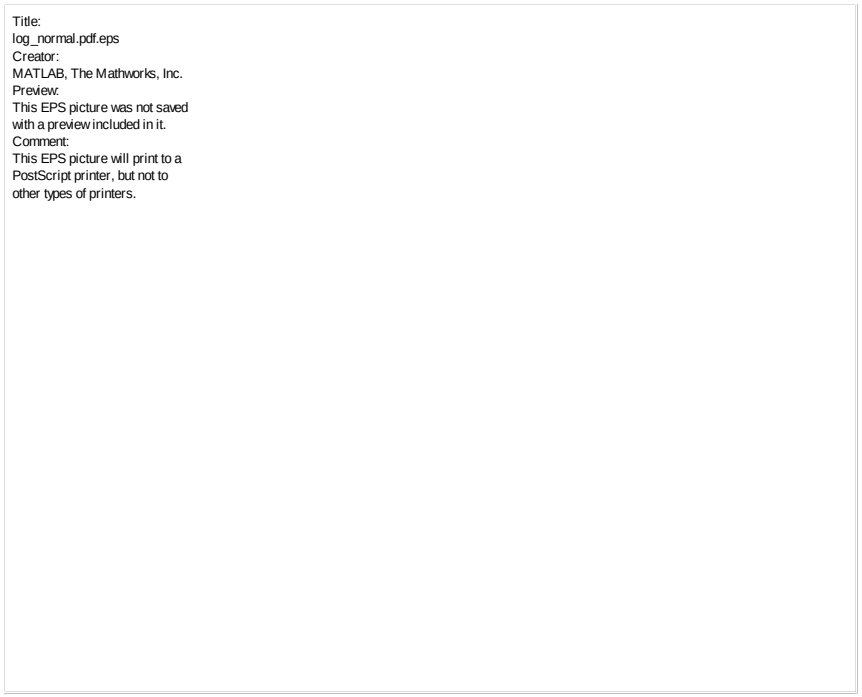
$x_m = 2^{y_m}$ geometric mean of X

$\sigma_X = 2^{\sigma_Y}$ geometric std deviation of Y

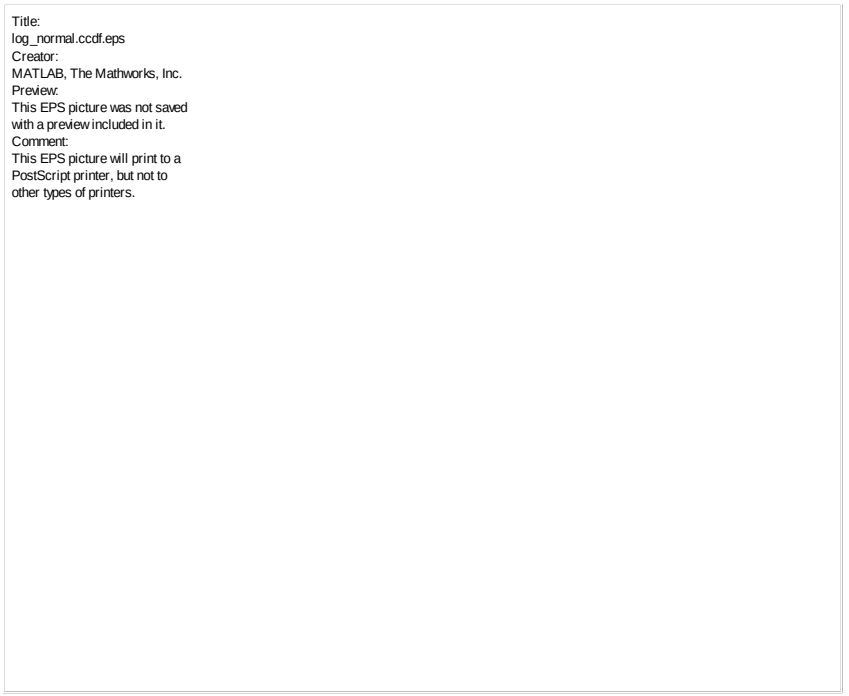
$$f_X(x) = \frac{1}{\sqrt{2\pi}\cdot\ln\sigma_x}\left(\frac{x}{x_m}\right)^{\frac{-1}{2\ln\sigma_x}\cdot\frac{\ln(x/x_m)}{\ln\sigma_x}-1}$$

The Log-Normal Distribution (2)

x_m=3000 Bytes, σ_X = 16



Probability density function (pdf) $f_X(x)$



Complementary distribution function (CDF) $Pr\{X \geq x\}$

The Pareto Distribution

- CDF and pdf:

$$F_c(x) = (k/x)^{\alpha}, \quad x \geq k > 0, \quad 1 < \alpha < 2$$

$$f(x) = \alpha k^{\alpha} x^{-\alpha-1}$$

- Heavy-tailed distribution

$$(a) \quad P[X \geq x] \to x^{-\alpha} h(x), \quad 0 < \alpha < 2$$

$$x \to +\infty$$

$$(b) \qquad Var(X) \to \infty$$

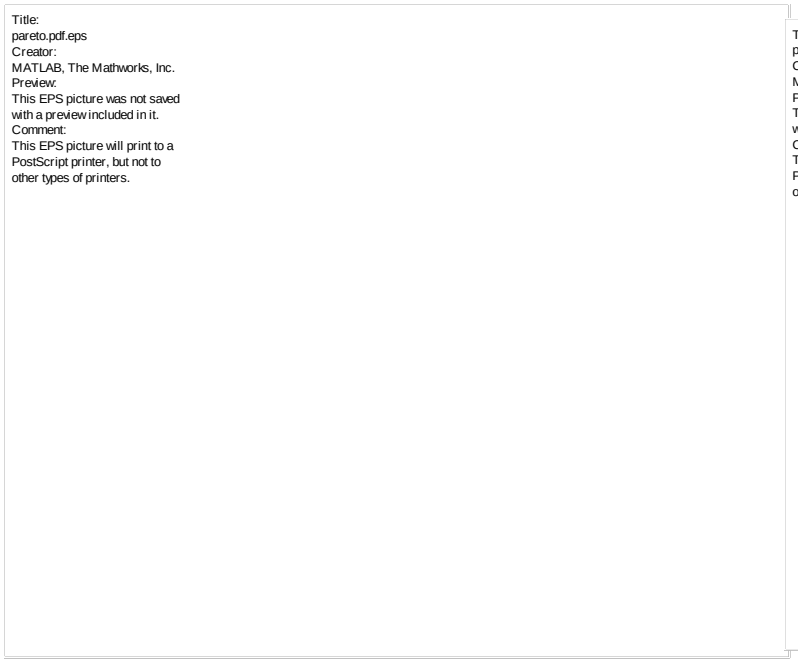
– where h is a slow varying function at ∞, i.e.

$$\frac{h(tx)}{h(t)} \to 1, \quad x > 0$$

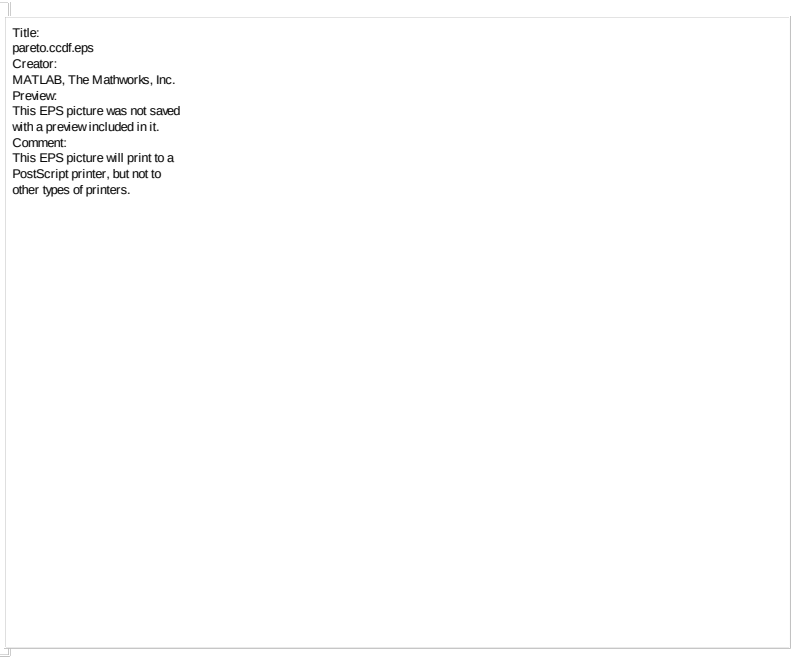
$$t \to \infty$$

The Pareto Distribution

$\alpha=1, k=10^{5.5}$

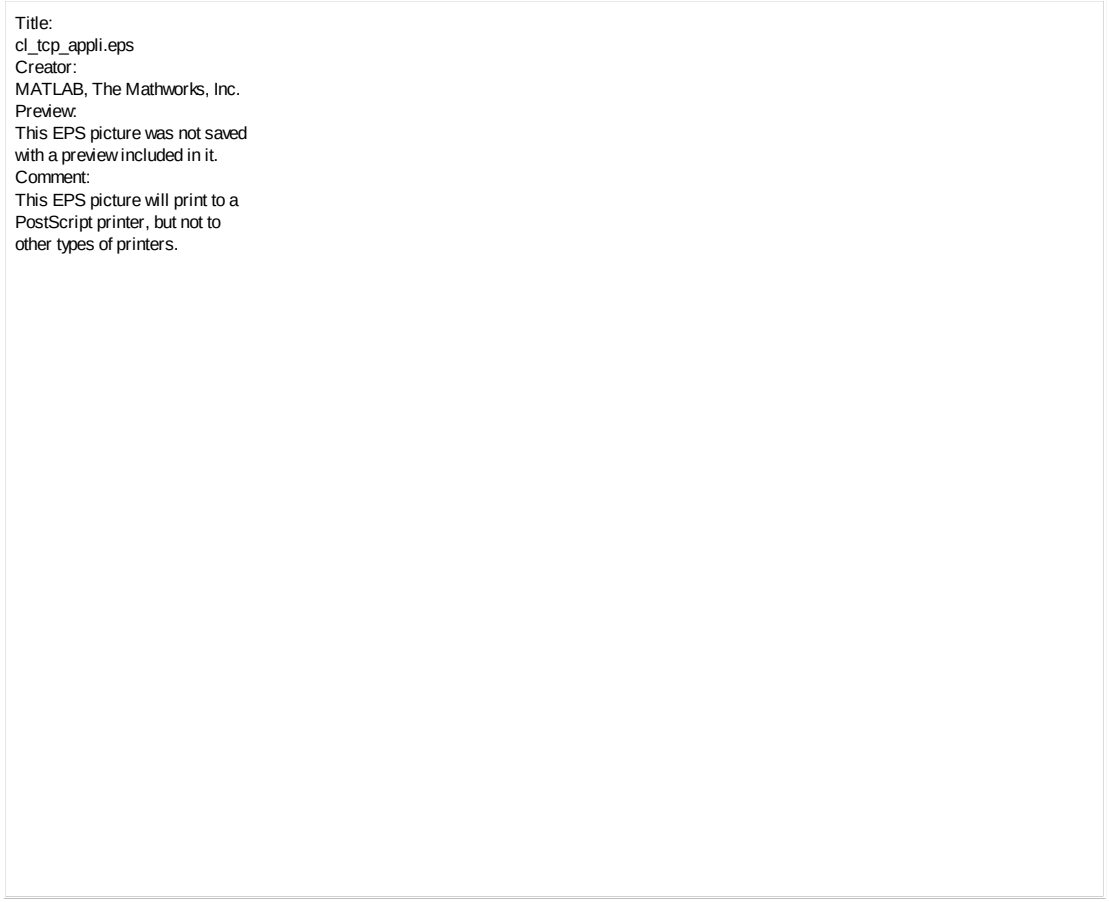


Probability density function (pdf) $f_X(x)$



Complementary distribution function (CDF) $Pr\{X \geq x\}$

Complementary Distribution Functions For Data Applications



Network Data Traffic Characteristics

- Shaped by TCP
  Transport Protocol used by all internet data applications

- Self-similarity characteristics
  Exhibits a high degree of burstiness with long range dependence

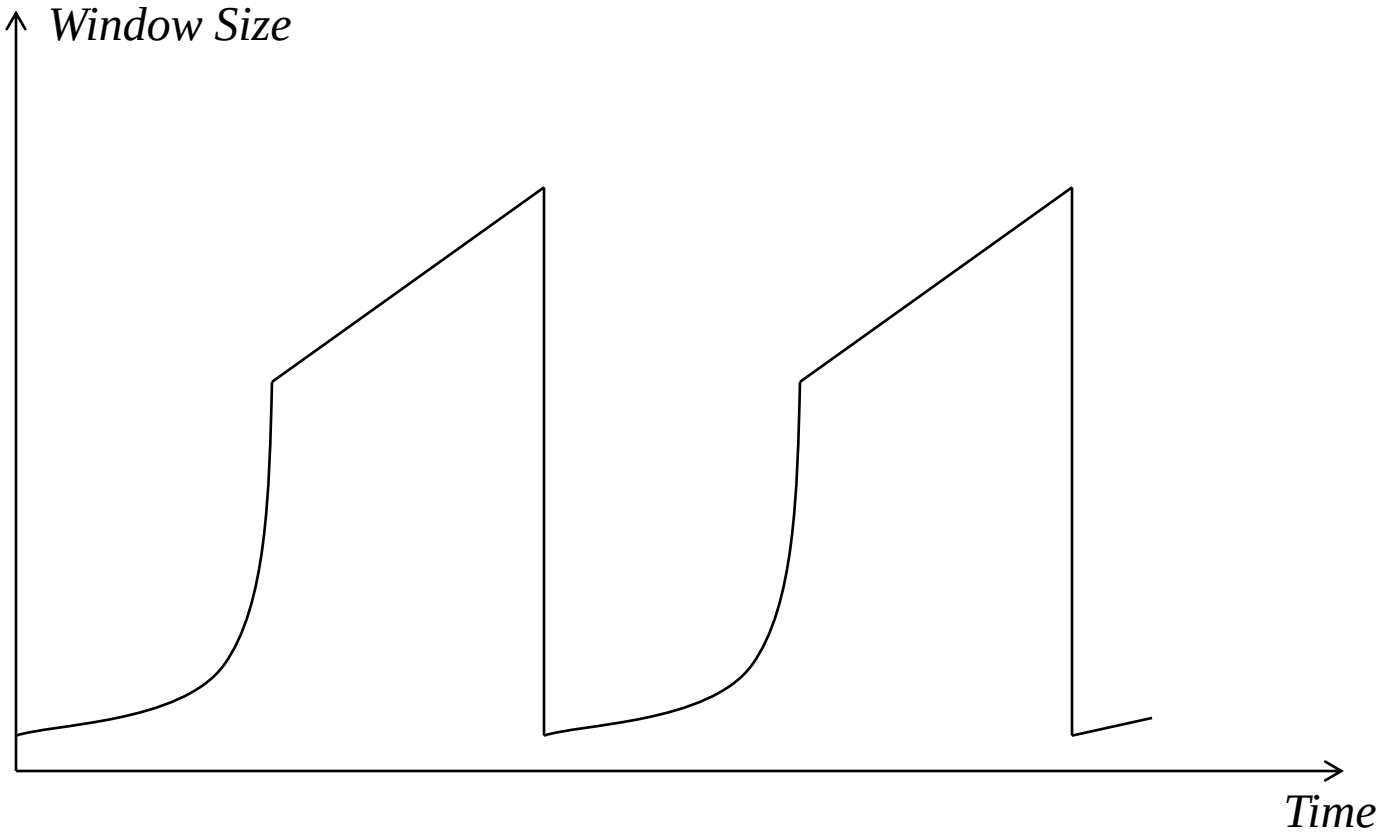
TCP

- Window based
- Provides
  - end-to-end reliability
  - end-to-end flow control
    - with explicit feedback from the destination
  - network congestion control
    - without feedback from the network
    - Congestion window size found empirically

TCP Behavior

- "Slow start"
  - start with a window size of 1
  - double window size with each ack received up to a threshold
  - beyond threshold, increase window size by 1 with each ack received up to a prespecified maximum window size
  - loss of a packet restarts the process
- Many variants (tahoe, Reno, …)
- Aggressive behavior:
  - tries to maximize throughput, and thus causes congestion
- Generates bursty traffic on the network

TCP Behavior (2)



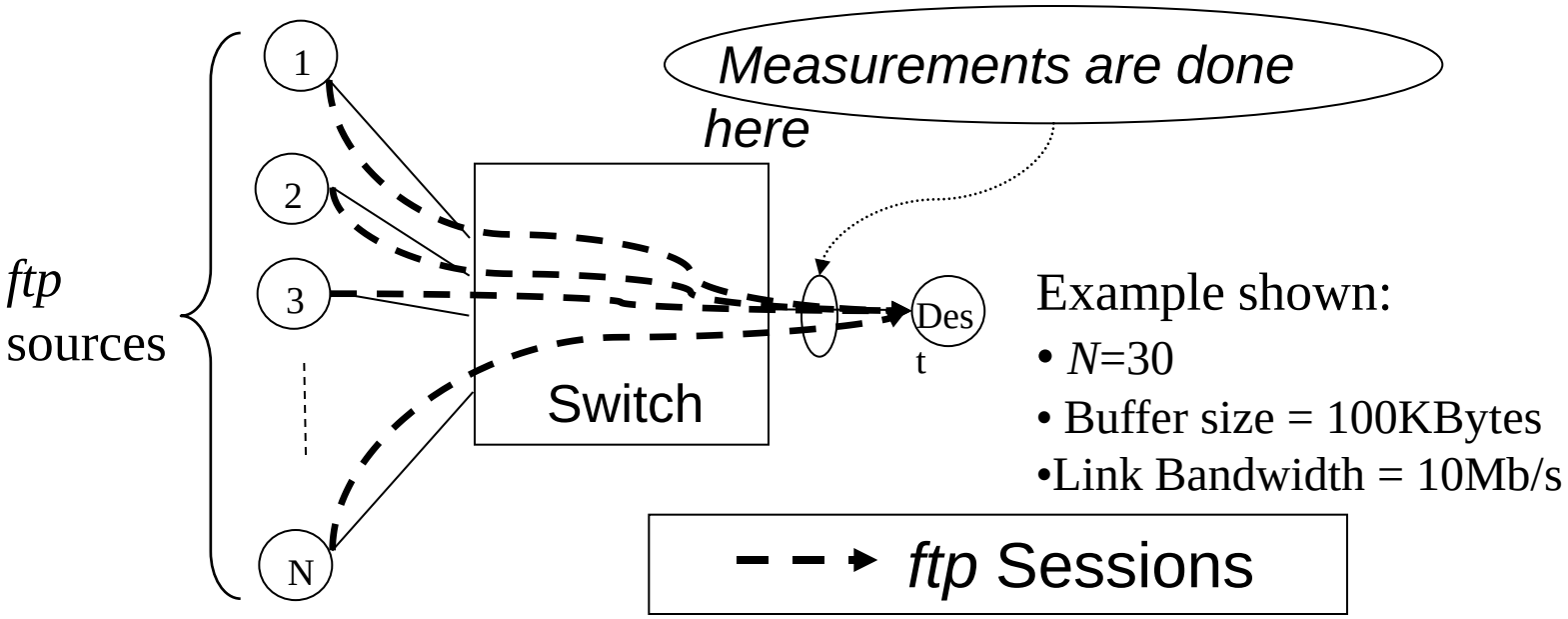
"Self-Similarity" characteristic

- Highly bursty
- Same degree of burstiness observed at all time scales
- Long range dependence
  - slow decay in autocorrelation function

Experienced in both LANs and WANs

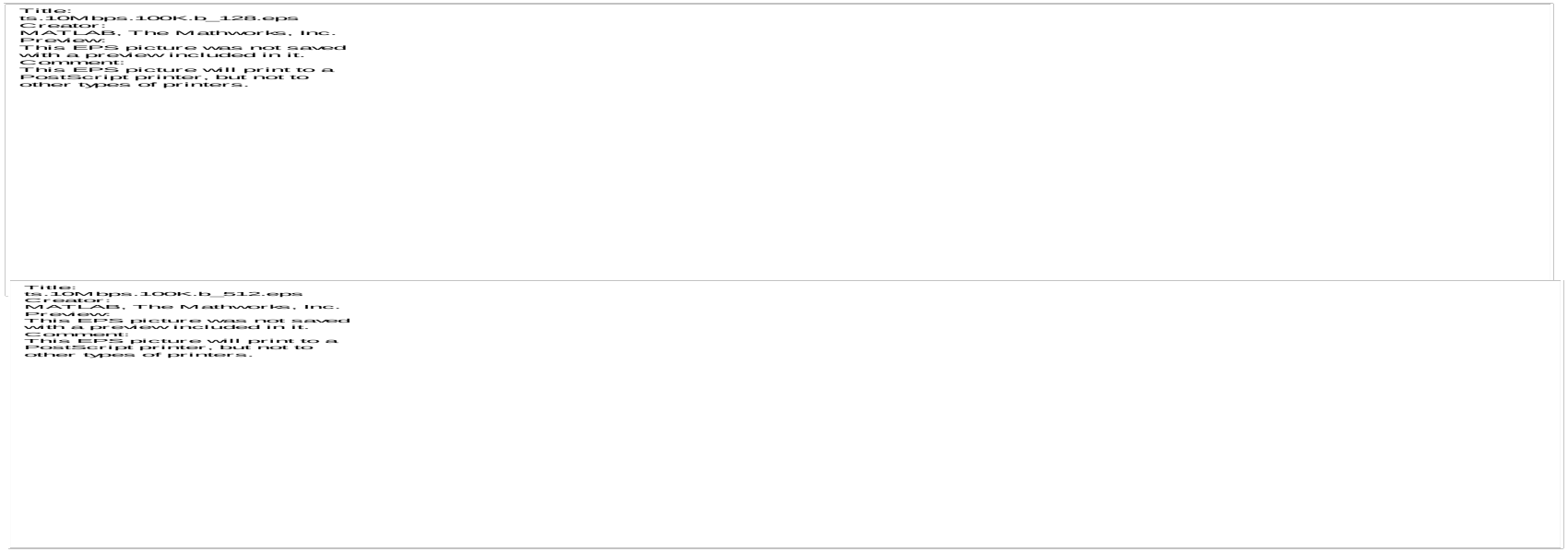
Example of Self-Similar Traffic

- Aggregation of traffic from many *ftp* users on a network link



- Traffic "self-similarity" varies as a function of the number of sources, the switch buffer size and the link bandwidth

Example of Self-Similar Network Traffic (2)



- Traffic shown at two time scales
  - top: 1.28 sec. intervals
  - bottom: 5.12 sec. intervals

Causes of Self-Similarity

- Aggregation of traffic generated by a large number of bursty TCP sources
- Very wide-tailed and heavy-tailed distributions of data application level characteristics
- Human-computer interactions occurring over wide range of time scales

Consequences of Self-Similarity

- No traffic smoothing possible
- Large buffers and high bandwidth required to prevent congestion and keep packet loss low
  - hence, low average utilization
- Incompatibility of data traffic and real-time multimedia traffic