Traffic Modeling(4)

Analysis of traffic characteristics(Self-similar)

<Results of Joo and Ribeiro's study>

: Traffic characteristics are also different because the probability distribution of the transport file size is different for each application.

- Web traffic: Small TCP transmission multiple times, each TCP connection starts slowly.

  → High utilization of available bandwidth, small variance

Analysis of traffic characteristics(Self-similar)

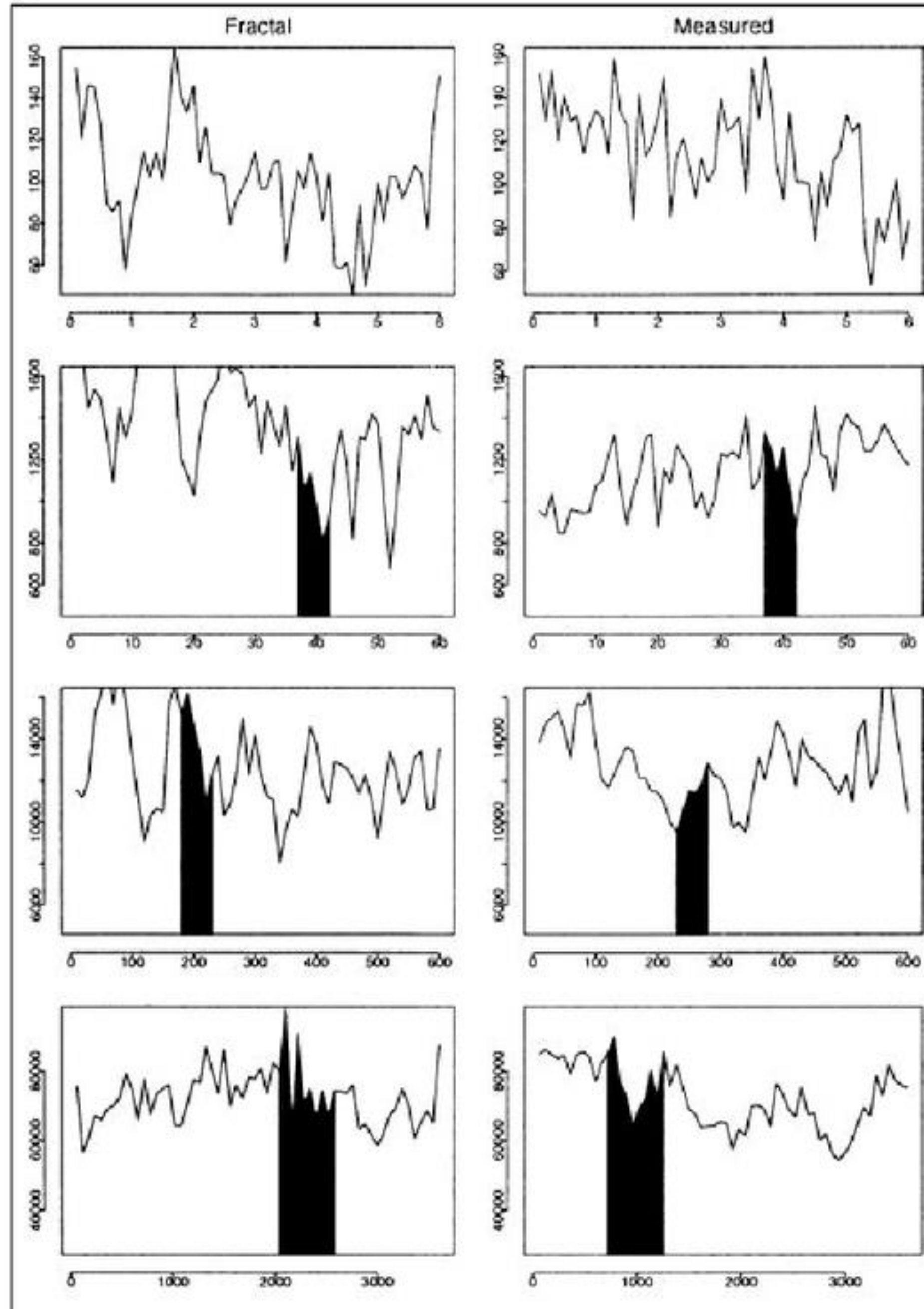
- FTP Traffic: Large-capacity file transfer
  → High possibility of avoiding congestion
  → Low bandwidth utilization
  → Self-similarity

- HTTP Traffic: → Self-similarity

- E-mail applications: → No Self-similarity

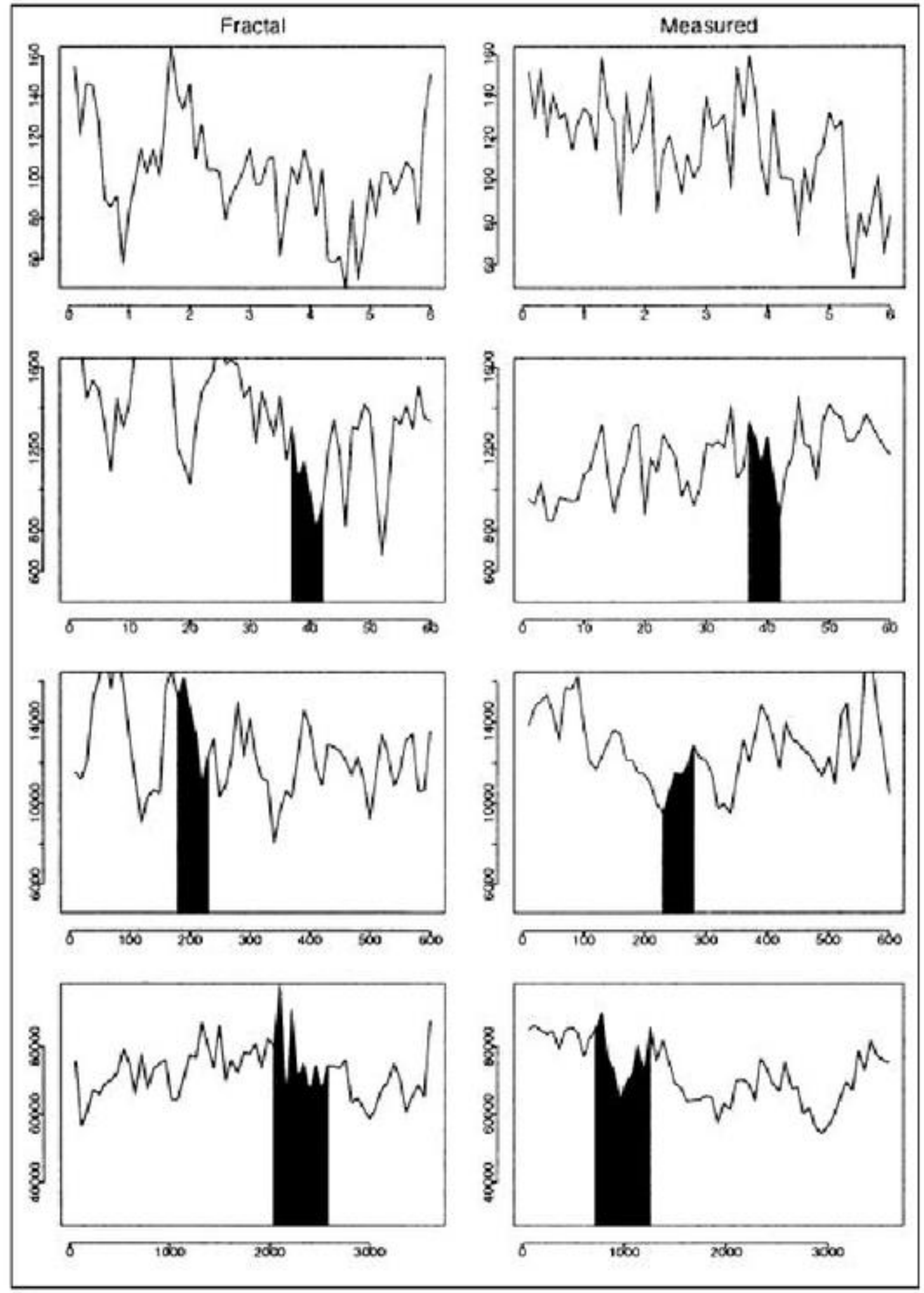
Fractal

- It is another term to describe the self-similarity of traffic
- A fractal process is characterized by significant long bursts



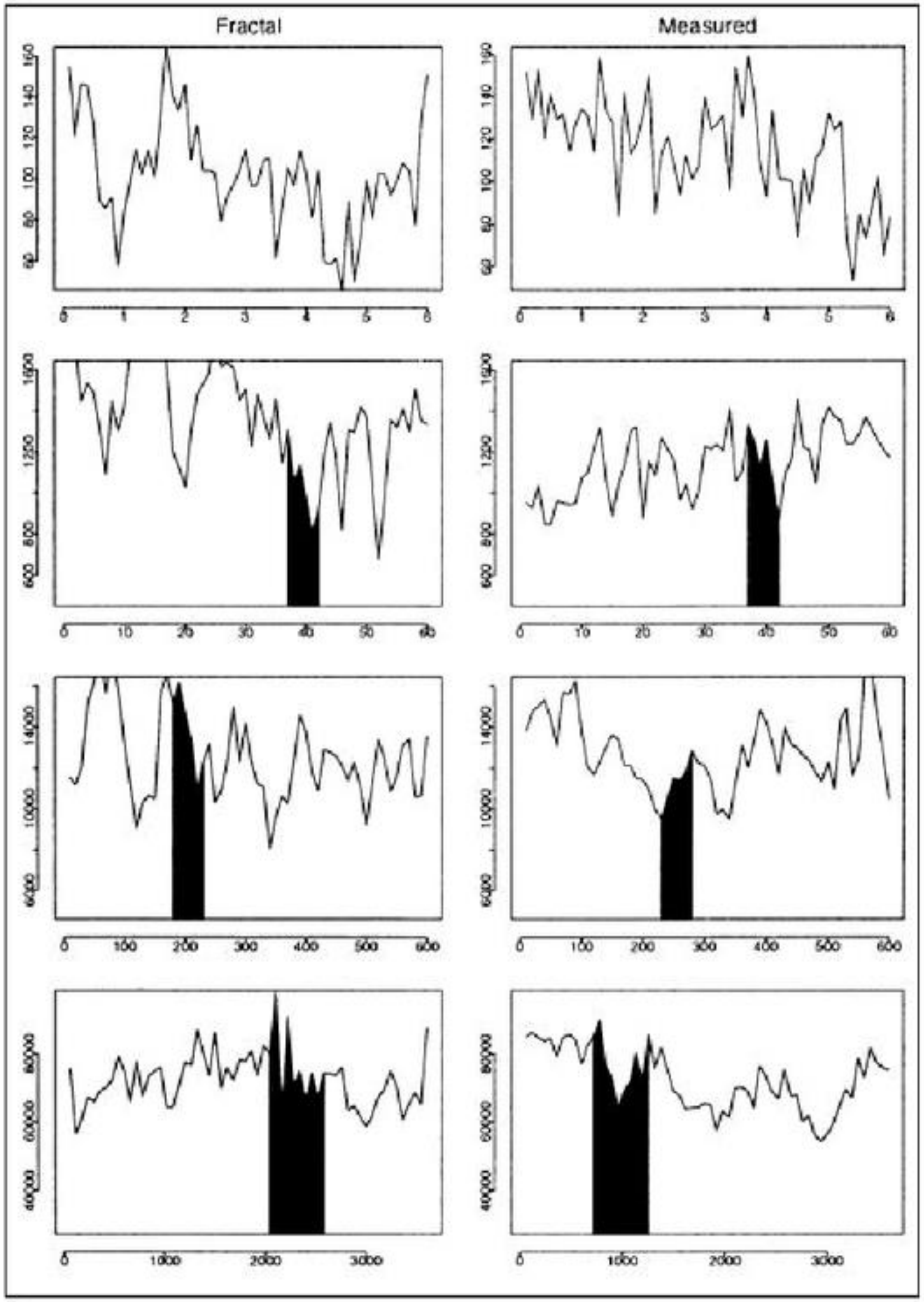
Fractal

- Downloading large files
  - e.g. video files, long periods of high levels of VBR video, intensive bursts of database activities



Fractal

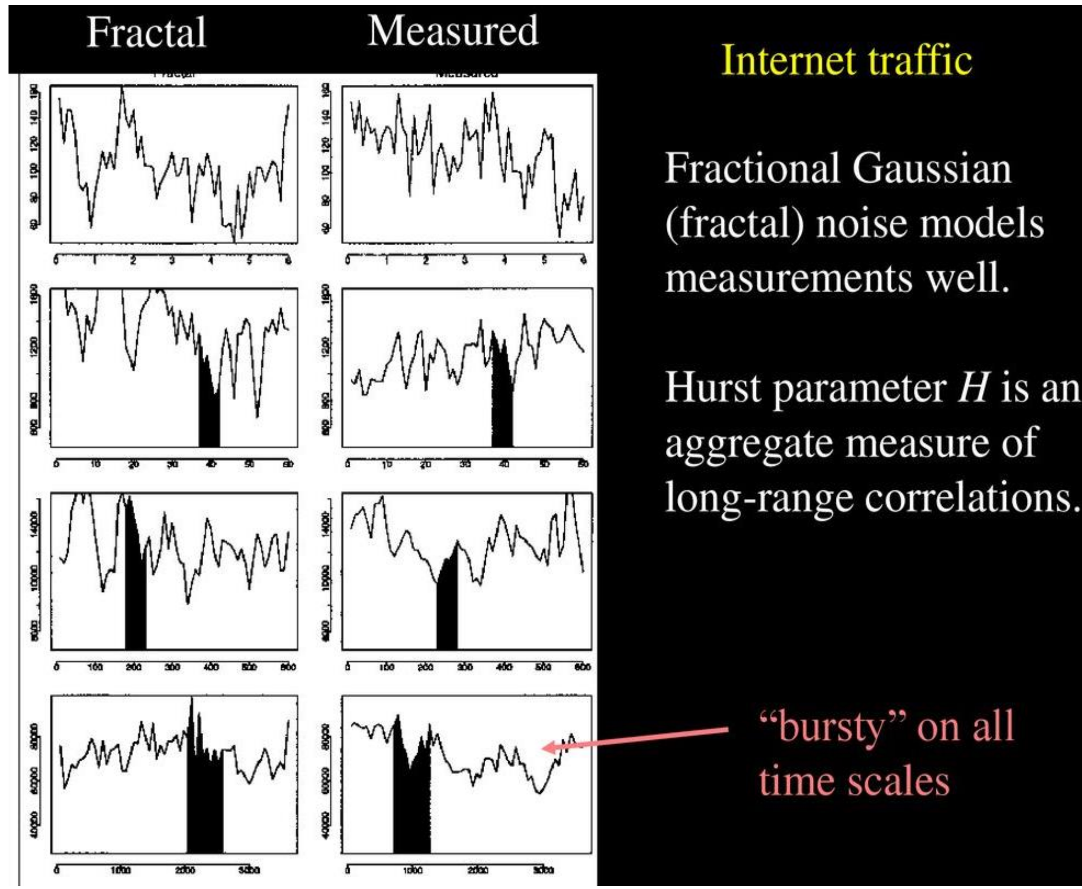
- The fractal (or self-similar) model obtained by fitting the traffic mean, variance and Hurst parameter is displayed.
- As can be seen, this model exhibits burstiness at all time scales as the original traffic.



Fractal

- Current WAN(Wide Area Networks) traffic is often described as multi-fractal.

- Multi-fractal traffic can be considered as an extension of self-similar traffic, it can capture more irregularities in the distribution.



This table summarizes some typical traffic types and associated traffic distributions and models

Traffic types	Traffic distribution	Frequently used traffic models
Individual source traffic	Heavy-tailed ON/OFF distribution	• Pareto • Weibull
Individual application traffic or LAN	Self-similar	• FGN • FARIMA
Aggregate traffic	LRD Multifractal	• Fractional Brownian motion(FBm) model • M/G/∞ • M/Pareto

Traffic distributions and frequently used traffic models

Traffic distributions and frequently used traffic models

Suitability of self-similar and LRD

- "Why does the traffic display these characteristics?"
- It is pointed out that heavy-tailed nature of ON and OFF periods has more to do with basic properties of information storage and processing.
- It is not a result of the network protocols or user preference.
- Therefore, changes in protocol processing and document display cannot remove the self-similarity of the web traffic.

Suitability of self-similar and LRD

- Also, it is shown that both the user's thinking or reading times and the file-size distributions are strongly heavy-tailed.

- In addition, Internet provides explicit support for multimedia formats; the file distribution is strongly heavy-tailed.

- Often, self-similarity in today's network traffic is explained in terms of application traffic.

Suitability of self-similar and LRD

- The burst data traffic and VBR real-time applications such as compressed video and audio display a certain degree of correlation between arrivals and slow LRD in time.
- As a result, the aggregate traffic is self-similar.
- Or, it could be the high variety of individual connections (i.e. infinite variance) that contributes to the aggregate traffic.

Suitability of self-similar and LRD

- Overall, the factors, apart from application traffic itself, that contribute to the self-similar nature and the LRD behavior of the emerging network traffic are
  - User behavior- user-reading time and user-induced delay
  - File- size distribution
  - Set of files available in the server

Traffic distribution	Description
Poisson	Session arrival process
Exponential	Session duration
Heavy-tailed	Suitable for burst individual source traffic with ON/OFF patterns
• Pareto	File-transfer time distribution, user-reading(thinking) time, user-induced delay
• Weibull	Machine-processing time, file downloading time

Traffic distributions and suitable applications

Traffic distributions and suitable applications

Current self-similar models

- Current and future models: fluid traffic model

  In this model, traffic is considered as volume and is characterized by a flow rate.

- Suitable to model the traffic where the individual traffic unit is insignificant

  e.g. individual cells in broadband ISDN(B-ISDN) ATM networks

Current self-similar models

- Here, larger traffic units provide a simpler and better analysis of the network performance as well as saving, simulation, and computing resources.

- Suitable for modeling burst traffic with ON/OFF patterns

Current self-similar models

- Following assumptions are made

  i) The ON-state traffic arrives deterministically at a constant rate

  ii) Traffic is switched off during the OFF state

  iii) The ON and OFF periods are exponentially distributed and mutually independent

Current self-similar models

- Fractional ARIMA
  - Most commonly used model
  - Can model both LRD and SRD processes simultaneously
    - particularly useful to simulate the queuing performance of SRD and LRD traffic simultaneously

Current self-similar models

- Fractional ARIMA
  - Provides quick simulation
  - By changing the parameters that affect the degree of SRD and LRD, we can identify the parameters that are more or less sensitive to SRD or LRD

Current self-similar models

- Fractional Gaussian Noise (FGN)
  - Most frequently used stochastic model for self-similar traffic modeling
  - Suitable for burst data and multimedia application traffic modeling with a prevalence of LRD
  - Provides a good estimation of queuing performance for aggregate traffic.

Current self-similar models

- Transform-Expand-Sample (TES)
  - Can capture both the marginal distributions and the autocorrelations of the measured traffic
  - Should satisfy the following three requirements

    i) The histogram of measured traffic matches the model's marginal distribution

    ii) The model's autocorrelations should match the measured traffic up to a reasonable lag

    iii) Good correspondence exists between the sample paths of the simulated and the measured data

LRD traffic models

- Fractional Brownian motion (FBm)
  - Gaussian process with a mean zero and stationary increments
  - Should satisfy the following three requirements simultaneously:

LRD traffic models

- M/G/∞
  - Is chosen to generate self-similar arrivals
  - Introduces multifractal behavior at small/medium timescales without affecting the asymptotic self similarity
  - More conservative than FBm as it predicts a stricter queuing performance

LRD traffic models

- M/Pareto
  - A particular type of the general M/G/∞ model
  - Simple and useful to estimate the queuing performance of a variety of realistic multimedia traffic streams
  - The superposition of multiple independent M/Pareto processes is an M/Pareto process with a combined Poisson rate, λ

LRD traffic models

- M/Pareto
  - With an appropriate choice of λ the M/Pareto process provides an accurate prediction of the queuing performance.
- Some of drawbacks
  - There is no systematic way of calculating the appropriate value of λ
  - Difficult to estimate the Hurst parameter, H, from a finite data set

Traffic model	Applications	Mathematical complexity	Computing complexity	Advantages	Disadvantages
Poisson	• Voice • Large number of independent traffic streams	Low	Low	• Oldest and commonly used model • Superposition of Poisson process is a new Poisson • Memory-less process	• Fails to capture autocorrelation • Optimistic estimation of queuing performance for burst traffic
Markov	N/A	High	High	• Capable of capturing correlation of traffic (i.e. nonzero autocorrelations)	• Inflexible • Complexity overshadows accuracy
MMPP	• A single traffic source with variable rates	Low	Low	• Simple and flexible • Possible to capture some degree of correlation of traffic	• Inadequate autocorrelation • Unsuitable for LRD traffic
Fluid	• ATM traffic • Bursty traffic	Medium	Low	• Simple • Fast simulation • Suitable to model bursty traffic with ON/OFF patterns	• Unsuitable for variable rate traffic
Fractional ARIMA	• Voice • Bursty data and multimedia traffic	Low	Medium-high	• Flexible • Suitable for self-similar traffic with SRD and LRD	• High computing complexity
TES	• Broadband traffic streams • Nonstationary traffic	Medium	Low	• Fast simulation • Suitable to capture both marginal and autocorrelation function of the traffic	• Requires high programming complexity

Self-similar and LRD traffic models: Traditional, current, and future traffic models

Traffic model	Applications	Mathematical complexity	Computing complexity	Advantages	Disadvantages
Gaussian	• Aggregated network traffic	Low	Low	• Simple • Good representation of network traffic as more traffic is aggregated together	• Overly optimistic estimation of network performance if the aggregation level is low
FBm (continuous-time)	• Real-audio • Real-video • Aggregated network traffic	Low	Medium-high	• Flexible • No need to select a sampling interval • Simplest Gaussian model to capture today's network traffic	• Unsuitable for small timescales simulation • Optimistic estimation of queuing performance
Fractional Gaussian noise (Discrete-time)	• Burst data & multimedia application traffic	Medium	Medium	• Simple and flexible • Possible to capture some degree of correlation of traffic	• Unsuitable for self-similar traffic with both SRD and LRD
Hyper-Erlang	• User mobility • Self-similar traffic	Low	Low	• Simple and general • Provides a good user mobility model in wireless and mobile networks	• Unsuitable in traffic management context
M/Pareto	• Broadband traffic streams (Ethernet, IP)	Low	Low-medium	• Simple • Suitable for current network traffic where traffic is not Gaussian enough • Good estimation of queuing performance	• Inadequate marginal distribution or autocorrelation function • No simple formula to determine the appropriate value for λ or H
M/G/∞	• Aggregated network traffic	Medium	Medium	• Introduce multifractal behavior at small/medium timescales • Good estimation of queuing performance	

Self-similar and LRD traffic models: Traditional, current, and future traffic models

Traffic models for applications

E-mail traffic

- ON – Weibull distribution
  - The message is downloaded from the mail server to the mobile terminal during the ON period
  - The length of the ON period depends on the message size and the instantaneous throughput available to the user

Traffic models for applications

E-mail traffic

- OFF – Pareto distribution
  - probability that users will finish reading an email in X time

Traffic models for applications

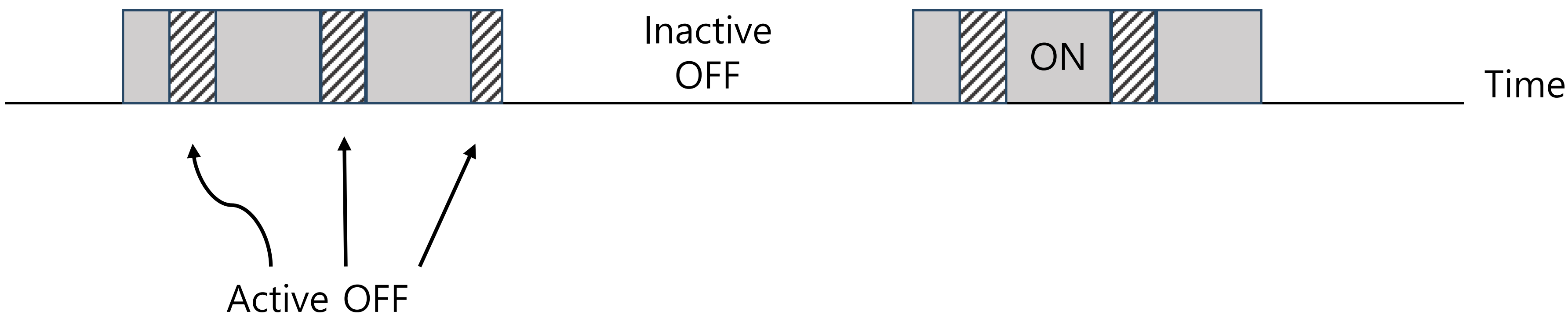
WWW traffic

- ON – Pareto distribution
  - The file is transferred on the downlink and the ON period depends on the file size and the available downlink bandwidth

Traffic models for applications

WWW traffic

- Active OFF time – Weibull distribution
  - The time needed to processes transmitted files (format, display a document component)
- Inactive OFF time – Pareto distribution
  - User reading time



Active and inactive OFF patterns in WWW traffic

Traffic models for applications

- Web file size
  - Web file system prefers documents in the 256-512 byte range
  - Web file systems are currently more biased toward small files than UNIX systems

Traffic models for applications

- Web file size
  - Text (smaller than 1000 bytes)
  - image(1000-30000 bytes)
  - audio(30000-3000000 bytes)
  - video(300000 bytes)

Traffic models for applications

FTP traffic

- The behavior of the FTP sessions is similar to e-mail but with larger file sizes and longer ON periods
- ON – Pareto
- OFF – Weibull
  - Depends on the user-induced delay such as user think time and typing speed

Self-similar traffic models

- FARIMA
  - Used in voice and bursty data & multimedia traffic
  - Self-similar traffic with both SRD and LRD
  - Ethernet traffic modeling, LAN, cooperate network

Self-similar traffic models

- TES
  - Used in Broadband traffic streams and nonstationary traffic
  - Self-similar traffic with both SRD and LRD
  - LAN, cooperate network traffic modeling

Self-similar traffic models

- FGN
  - Used in Burst data & multimedia application traffic
  - Self-similar traffic with LRD only
  - WAN

Self-similar traffic models

- Although it is hard to determine the sufficient aggregation level where short-range dependence(SRD) effects can be ignored, if the traffic is aggregated enough, SRD would be averaged out. We only need to consider the LRD properties.

Self-similar traffic models

- M/Pareto
  - Used in Burst data & multimedia application traffic
  - LRD
  - Multimedia traffic, broadband traffic in general

Self-similar traffic models

- M/G/∞
  - Used in aggregated network traffic
  - Multifractal LRD traffic
  - WAN