Self-Similarity On Network Systems With Finite Resources

Hsu-Hui Wang

Advisor: Prof. Po-Ning Chen

Department of Communications Engineering
National Chiao Tung University, Taiwan 30056, R.O.C.

2003.July.04

OUTLINE

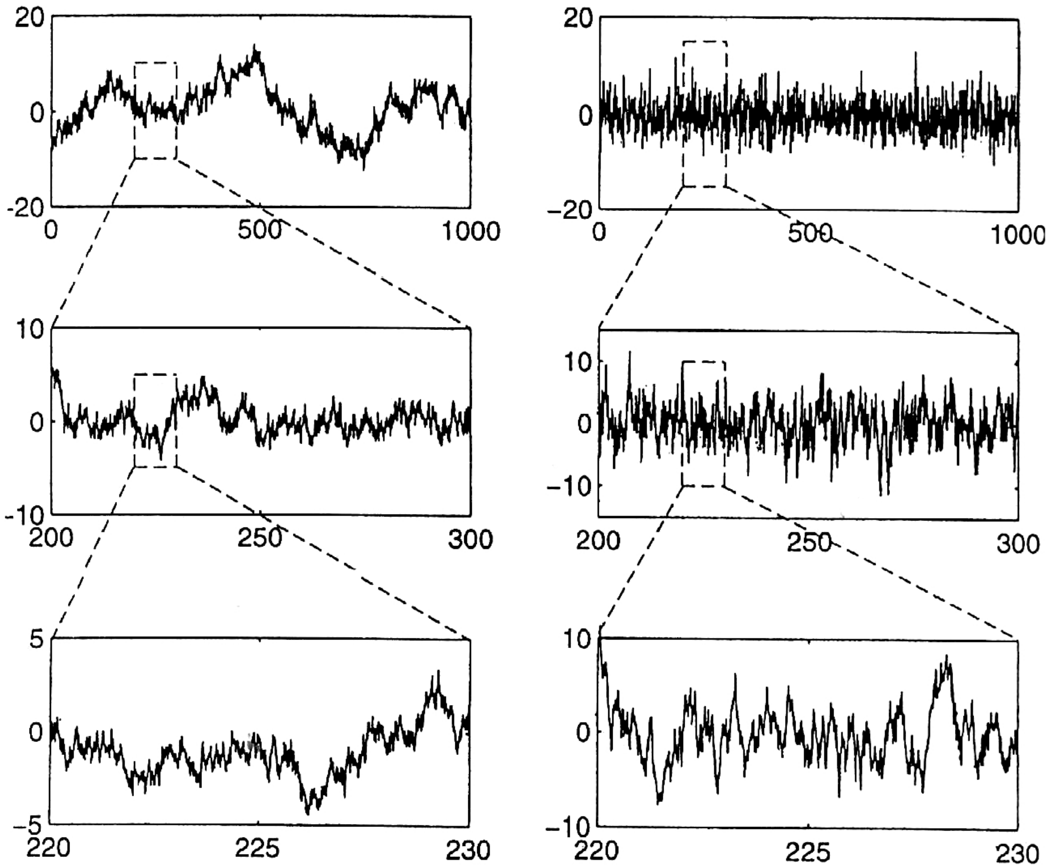
- Introduction
- Preliminary Background on Self-Similar Process and Its Determination
- Investigation of System Model with Finite Resources
- Simulation Results and Analysis
- Conclusion and Future Work

Introduction

- It is long conjectured that heavy-tail distributed data builds self-similarity. This conjecture has been recently proved under the assumption that infinite number of ON/OFF sources is aggregated.
- A nature question that follows is

"What if the number of sources aggregated is finite".

Background On Self-Similar Process And Its Determination



(a) Self-similar process (b) Non-self-similar process

The comparison of a self-similar and a non-self-similar processes in different time scale.

Continuous-time self-similar process

- $a^{-H}X(at)$ has the same distribution as $X(t)$ for any a.
- Statistical properties:

$$\begin{aligned} E[X(t)] &= \frac{E[X(at)]}{a^H} \\ \mathrm{Var}[X(t)] &= \frac{\mathrm{Var}[X(at)]}{a^{2H}} \\ R_X(t,s) &= \frac{R_X(at,as)}{a^{2H}}. \end{aligned}$$

Discrete-time self-similar process

$m^{-H}X^{(m)}$ has the same distribution as X, where $X^{(m)} = \dfrac{X_1 + \cdots + X_m}{m}$.

- Statistical properties: With $H = 1 - \beta/2$,
  - Exactly second-order self-similar

$$\mathrm{Var}\left[X^{(m)}\right] = \frac{\mathrm{Var}[X]}{m^{\beta}}.$$

  - Asymptotically second-order self-similar

$$\mathrm{Var}\left[X^{(m)}\right] \sim \frac{\mathrm{Var}[X]}{m^{\beta}}.$$

Heavy-tailed distributions

- Heavy-tail

$$1 - F(x) \;=\; \Pr[X > x] \;\geq\; a(x),$$

for some $a(x) \sim x^{-\alpha}$ as $x \to \infty$.

Pareto

- Complementary cumulative distribution function (CCDF)

$$\Pr[X > x] = \begin{cases} 1, & \text{for } x < k; \\ \left(\frac{k}{x}\right)^{\alpha}, & \text{for } x \geq k. \end{cases}$$

- 1st and 2nd moments

$$E[X] = \begin{cases} \frac{\alpha k}{\alpha - 1}, & \text{for } \alpha > 1; \\ \infty, & \text{otherwise,} \end{cases}$$

and

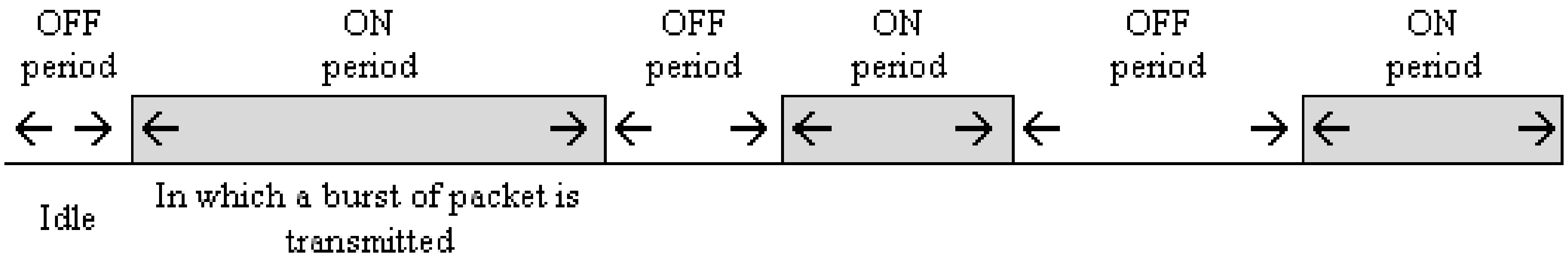
$$E[X^2] = \begin{cases} \frac{\alpha k^2}{\alpha - 2}, & \text{for } \alpha > 2 \\ \infty, & \text{otherwise.} \end{cases}$$

Fractional Brownian motion

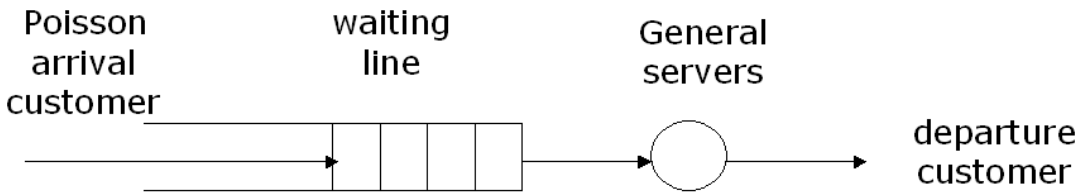
- Fractional Brownian motion is a self-similar Gaussian process with stationary increments.
- The distribution of $\{T^{-H}B_H(Tt), t \geq 0\}$ does not depend on T.
- For any $t \geq 0$ and $\Delta t > 0$, $B_H(t + \Delta t) - B_H(t)$ is normally distributed with mean 0 and variance $(\Delta t)^{2H}$

$$\Pr\left[B_H(t+\Delta t) - B_H(t) \leq x\right] = \frac{1}{\sqrt{2\pi}(\Delta t)^H}\int_{-\infty}^{x} e^{-y^2/[2(\Delta t)^{2H}]}dy.$$

ON/OFF source



M/G/1



Determination of heavy-tail distributions and self-similarity

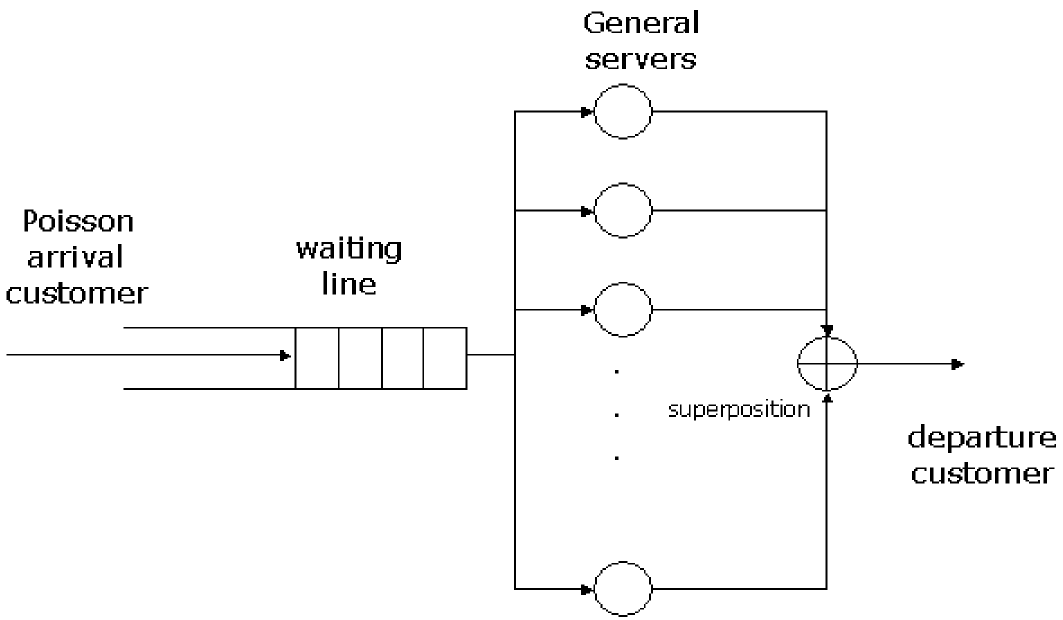
- Heave-tail determination by CCDF
- Variance-time plot

$$\log\left(\mathrm{Var}\left[X^{(m)}\right]\right) \;=\; \log(\mathrm{Var}[X]) - \beta\log(m).$$

Investigation of System Model with Finite Resources

Model of parallel servers

- This model is equivalent to M/G/L model.



System approximation technique

- **(Willinger, Taqqu, Sherman and Wilson 1997)** Define

$$W_M^*(Tt) = \int_0^{Tt} \left(\sum_{m=1}^{M} W^{(m)}(u) \right) du,$$

where $W^{(m)}(t)$ is the mth ON/OFF source with

$$[W^{(m)}(t) = 1] \equiv [\text{Source ON}] \quad \text{and} \quad [W^{(m)}(t) = 0] \equiv [\text{Source OFF}].$$

- For $M, T \to \infty$, the distribution of $W_M^*(Tt)$ becomes a function of $\{\sigma_{lim} B_H(t), t \geq 0\}$, where σ_{lim} is a finite positive constant and $\{B_H(t), t \geq 0\}$ is a fractional Brownian motion.

- The authors concluded that

$$H = (3 - \min\{\alpha_{\text{ON}}, \alpha_{\text{OFF}}\})/2,$$

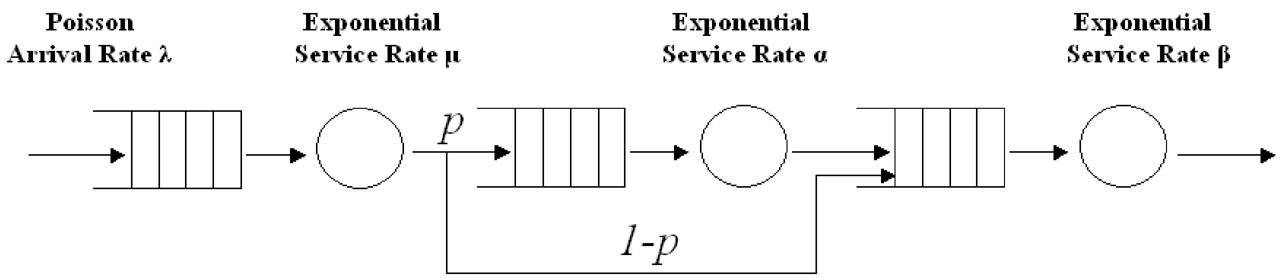
where the CCDFs of ON period and OFF period decay respectively as the orders of $x^{-\alpha_{ON}}$ and $x^{-\alpha_{\text{OFF}}}$ to zero.

Simulation concern

- The only source of dependence among departures of different servers is the shared queue.
- Our focus at case of small L
  — relation among α and H and L
- Our focus at case of large but finite L
  — the rate of convergence

Model of tandem

Self-similarity of inter-departures at the last server



Block diagram for simple network of tandem queues.

- Pick (2) and (3) from the 4 possible combinations of source inter-arrival and service time
  - (1) exponential interarrival versus exponential server;
  - (2) exponential interarrival versus Pareto server;
  - (3) Pareto interarrival versus exponential server;
  - (4) Pareto interarrival versus Pareto server.
- **(Anantharam and Verdu 1996)**
  Its analysis is another information-theoretical challenge since the outputs depend on inputs in a nonlinear fashion.

Entropy rate variation of a tandem network system

- **Definition** The entropy rate for a source $\boldsymbol{X} = \{X_1, X_2, \ldots\}$ is given by

$$\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$$

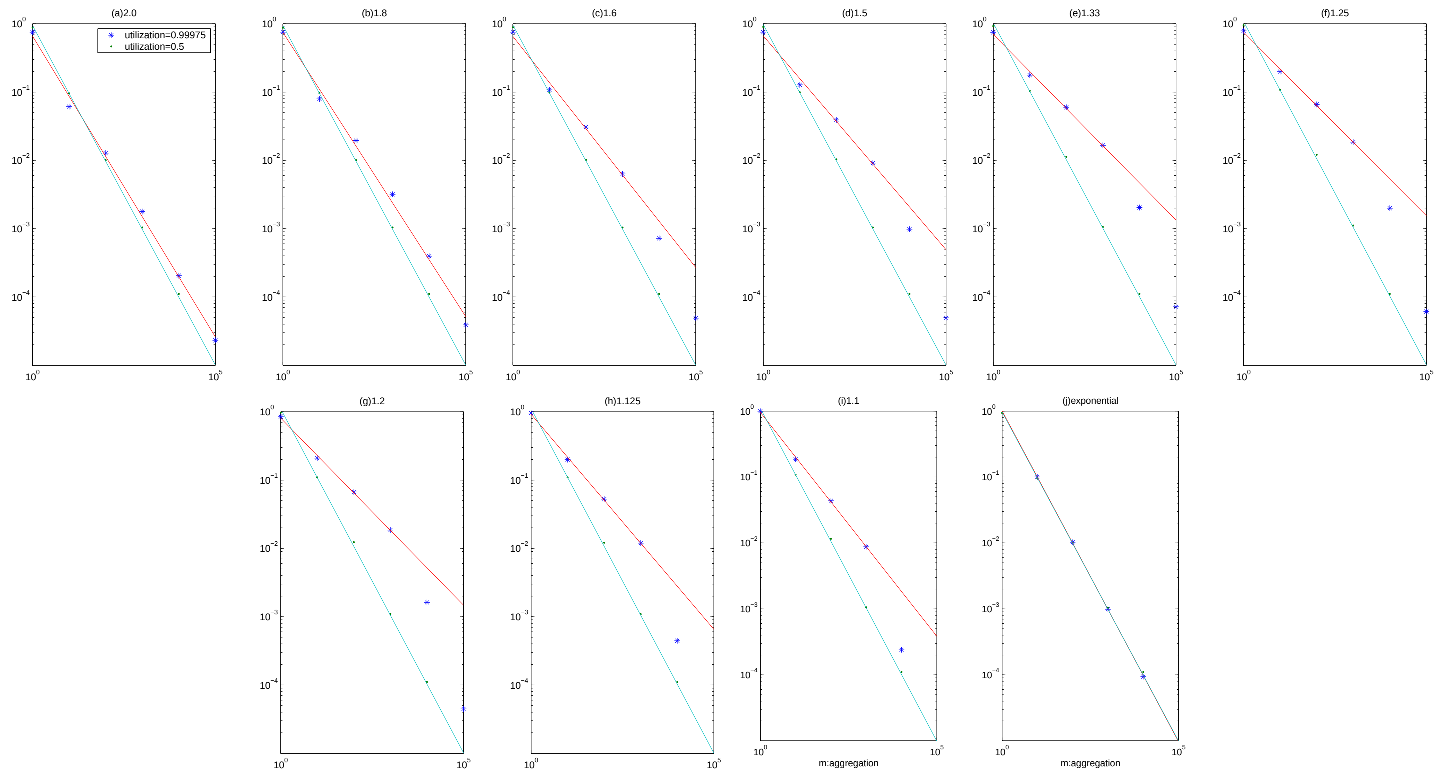
  provided the limit exists.

- The entropy rate for all stationary sources exist.
- The Lempel-Ziv coding algorithm is a universal coder for all stationary sources.
- It was shown that its compressing rate ultimately approaches the entropy rate of any stationary source.

Simulation Results and Analysis

System With Constant Service Mean Time

$\lambda = 0.01$ is Poisson arrival rate, $(1/\mu) = k\alpha/(\alpha - 1) = 200$ and 399.9 are means of Pareto service time.



The variance-time plots for 4 parallel servers ($L = 4$).

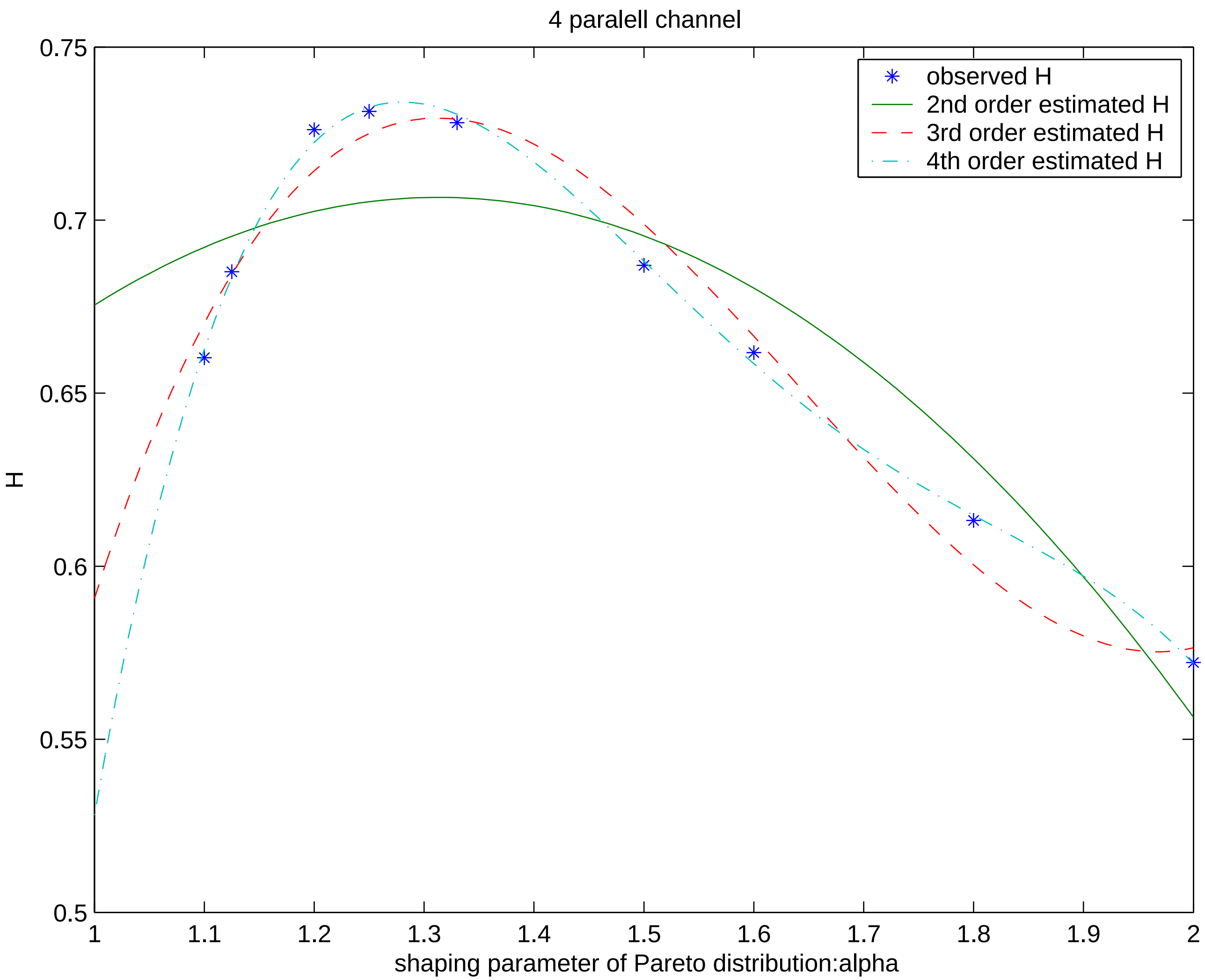
- For different utilizations, inconsistent slopes happen on $\rho \approx 1$ rather than $\rho = \lambda/(L\mu) = 0.5$

Estimated Hurst parameter for 4 parallel Pareto servers

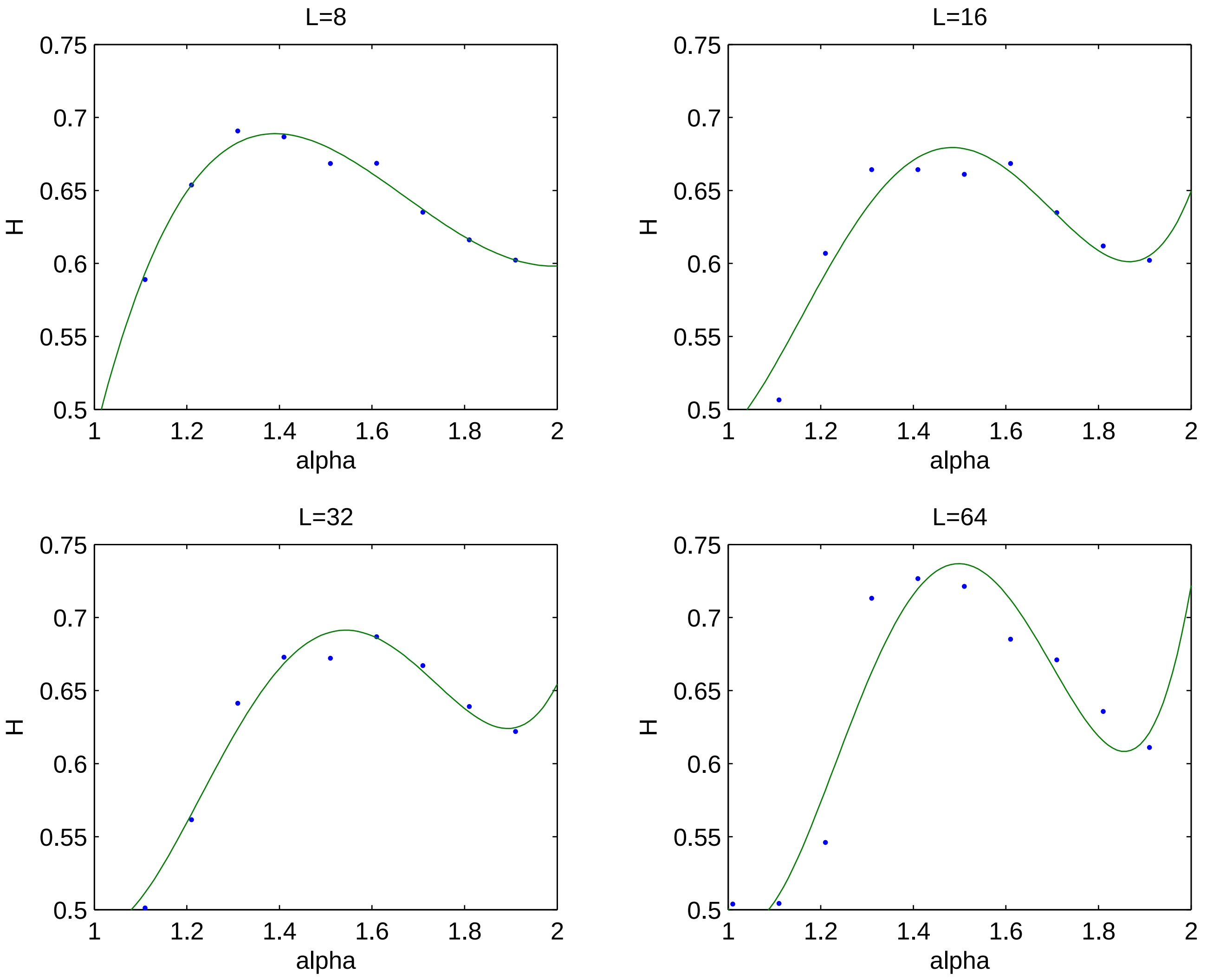
α	$\hat{H}$ (utilization=0.5)	$\hat{H}$ (utilization≈1.0)	$H = (3 - \alpha)/2$
2.0	0.5003	0.5722	0.5
1.8	0.5003	0.6133	0.6
1.6	0.5021	0.6617	0.7
1.5	0.5013	0.6869	0.75
1.33	0.4991	0.7282	0.835
1.25	0.4974	0.7314	0.875
1.2	0.4962	0.7262	0.9
1.125	0.4952	0.6851	0.9375
1.1	0.4953	0.66025	0.95
exp	0.5010	0.4981	N.A.

- Utilization and Self-Similarity:

$\rho \approx 1$ introduces self-similarity, while $\rho = 0.5$ implies no self-similarity.



Estimated H, as a function of α*, for 4 parallel servers with* $\rho = \lambda/(L\mu) = 0.99975$.



Observed Hurst parameter versus α under $\rho = \lambda/(L\mu) = 0.975$.
Fourth order polynomial approximation is also provided.

List of observed Hurst parameter and Pareto shaping parameter in previous figure.

α	$L=8$	$L=16$	$L=32$	$L=64$	$H=(3-\alpha)/2$
1.01	0.4955	0.498	0.493	0.5038	0.995
1.11	0.589	0.5065	0.50115	0.50425	0.945
1.21	0.6538	0.60685	0.56165	0.54605	0.895
1.31	0.69075	0.66425	0.64135	0.7132	0.845
1.41	0.6867	0.6642	0.67285	0.72665	0.795
1.51	0.6684	0.661	0.6721	0.7213	0.745
1.61	0.66855	0.6685	0.6868	0.68515	0.695
1.71	0.63515	0.63475	0.6671	0.67105	0.645
1.81	0.6162	0.61195	0.63905	0.6357	0.595
1.91	0.60225	0.60215	0.62205	0.6111	0.545

- Impact of Server Number on Self-Similarity:

Only when α is moderately large can the departure self-similar behavior be possibly close to departure behavior under infinite servers, $H=(3-\alpha)/2$

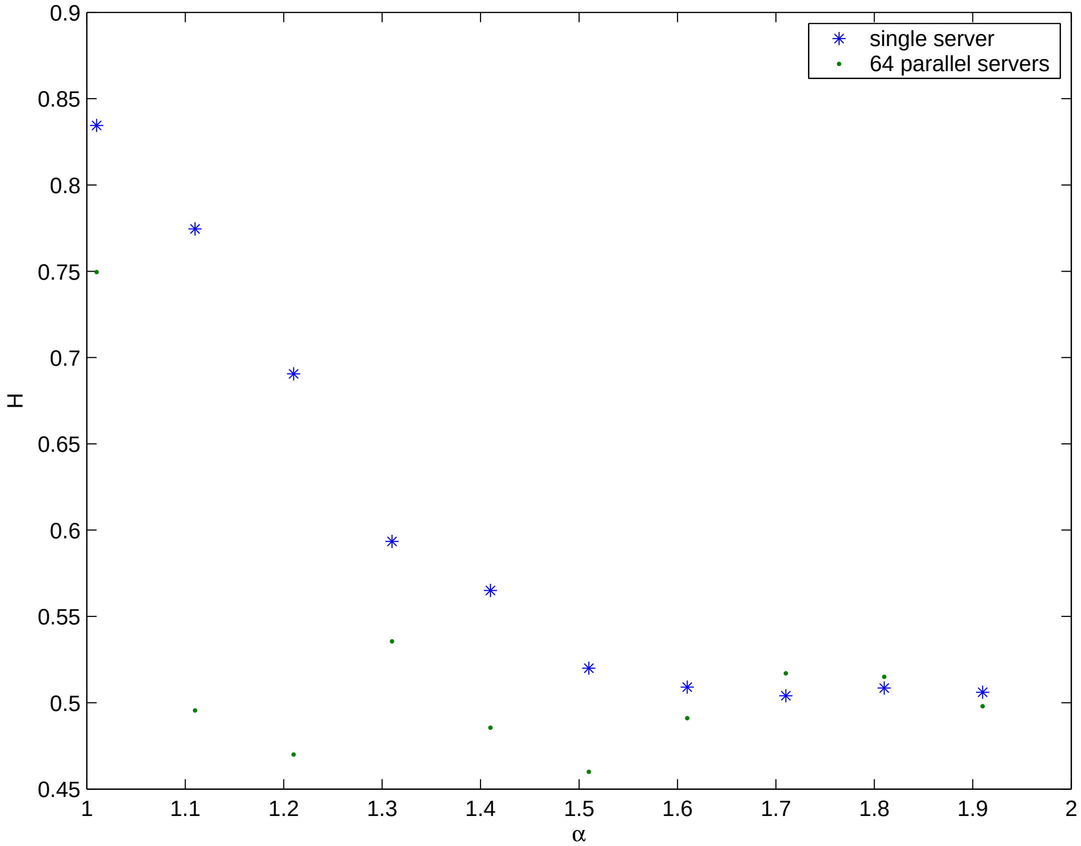
Why so? Interpretation of Non-Self-Similarity at Small α:

- When α is close to 1, k must be very small (in order to make the mean $k\alpha/(\alpha-1)$ constant), and deviates largely from the mean service time.
- This indicates that there will be a certain number of "very short" packets.
- Notably, if packet A enters the queue after packet B is served, the departure behaviors of packet A and packet B are statistically independent because the queue is the only place to introduce dependence in the system.
- Now, as a certain number of packets requires much less service time at α close to 1, the long-term dependence of the departure process should disappear at small average window m.

This explains why in our simulation, the degree of self-similarity decreases, rather than increases as expected by the analytical result under infinite number of servers, when α reduces to 1.

System With Varying Service Mean Time

- Based on the previous explanation, one way to obtain the theoretically anticipated trend at small α is to fix the parameter k in the Pareto distribution and the ratio of λ/L, and relax the restriction of constant utilization ρ.

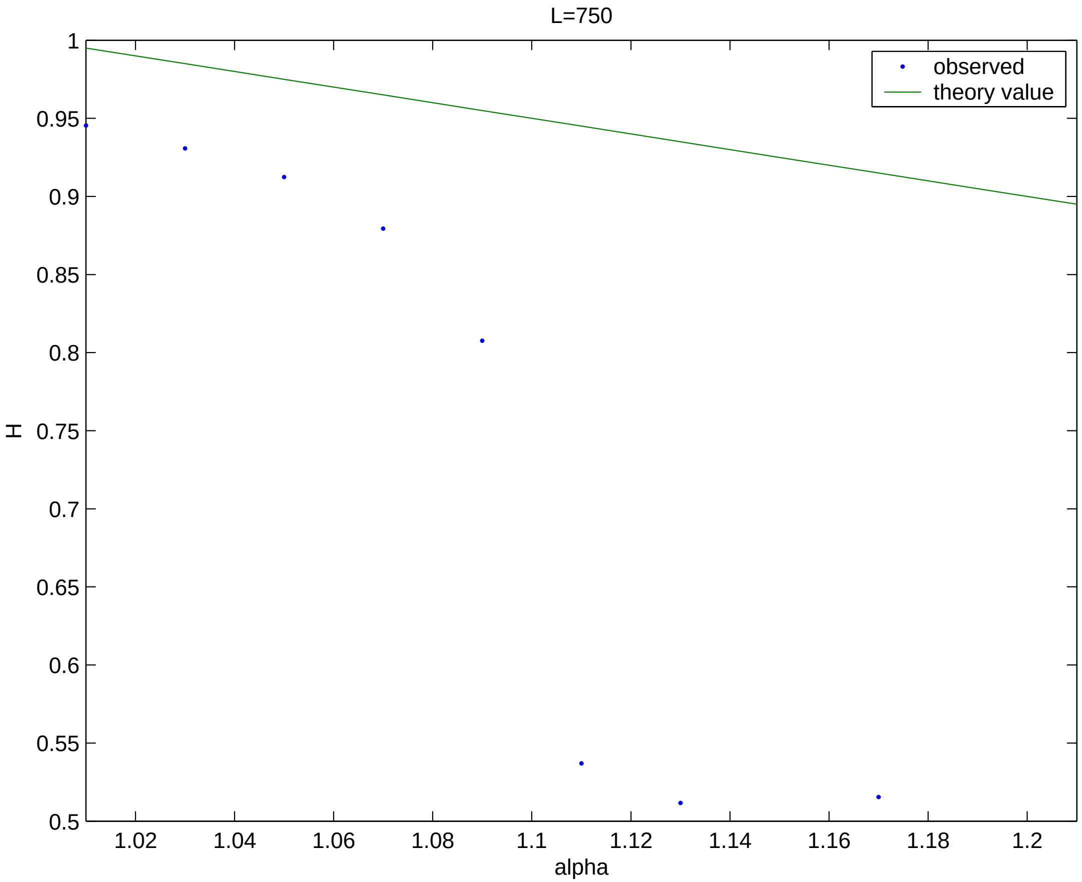


The H-versus-α plot for **single** *server and parallel* **64** *servers with fixed* $k = 40,\ \lambda/L = 0.0025.$

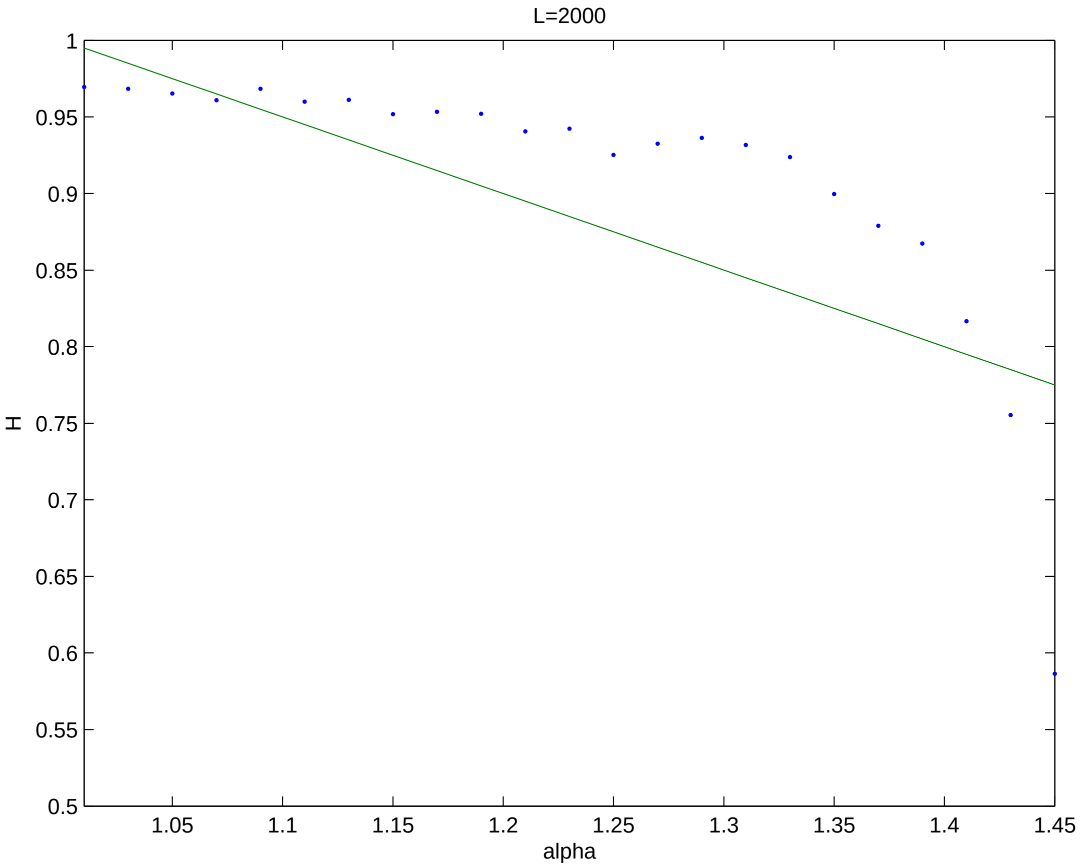
- The system is more self-similar at small α, but the system behaves non-self-similar while L modest large.

Large L

- By further increasing L (beyond 64) as shown in the next two slides, we observe that the relation between the observed $\hat{H}$ and α is getting closer to the theoretical $H = (3 - \alpha)/2$.



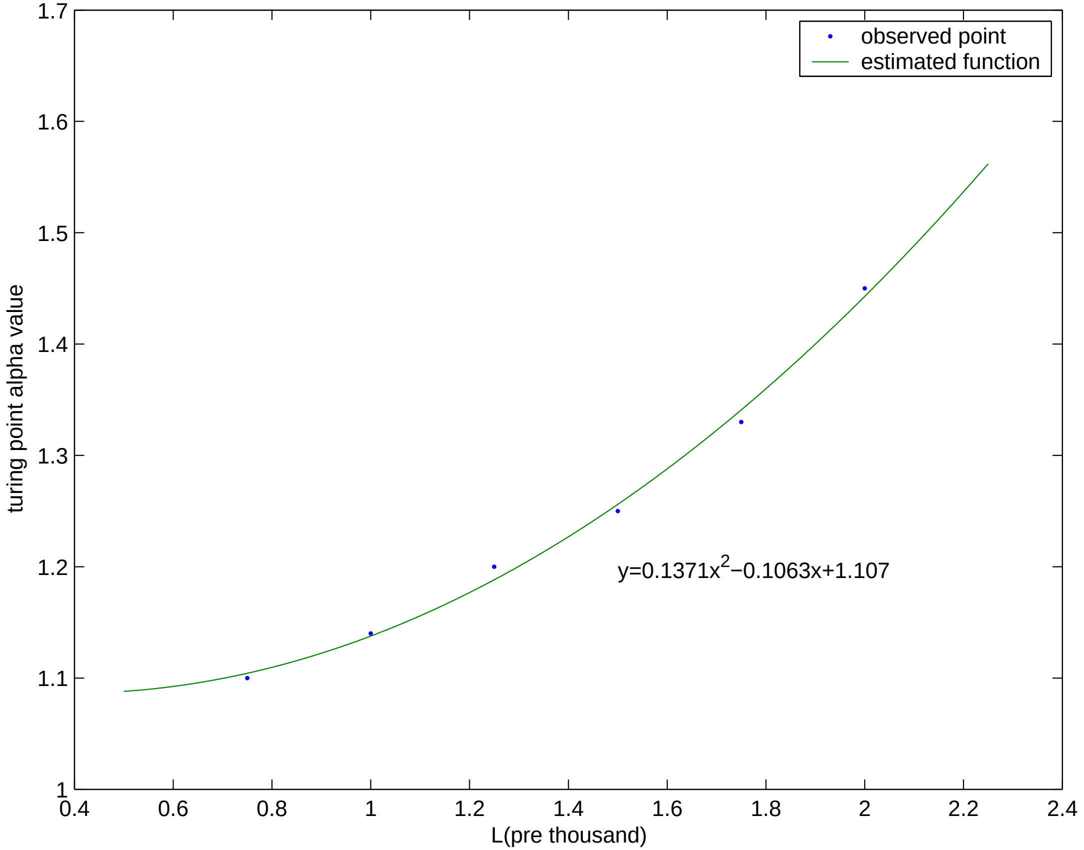
The H-versus-α plot for $L = 750$, and the curve of $H = (3 - \alpha)/2$. In these simulations, $k = 100$ and $\lambda/L = 0.0001$.



The H-versus-α plot for $L = 2000$, and the curve of $H = (3 - \alpha)/2$. In these simulations, $k = 100$ and $\lambda/L = 0.0001$.

- The observed points can be divided into two groups.

The first forms a line with a slope characterizing as selfsimilarity, while the second group belongs to "much-less-self-similar."

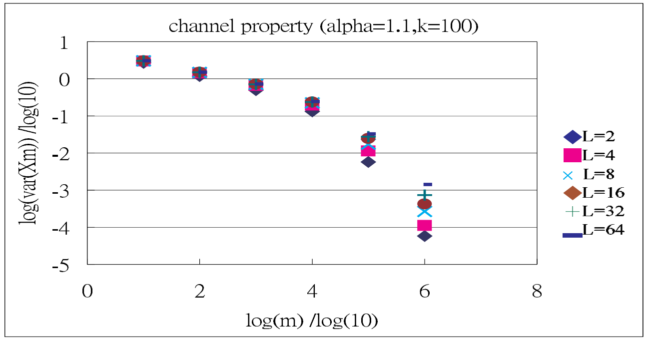


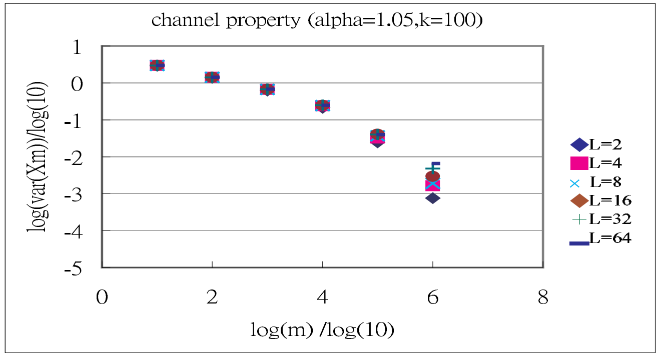
The turning points of α after which the degree of self-similarity drops.

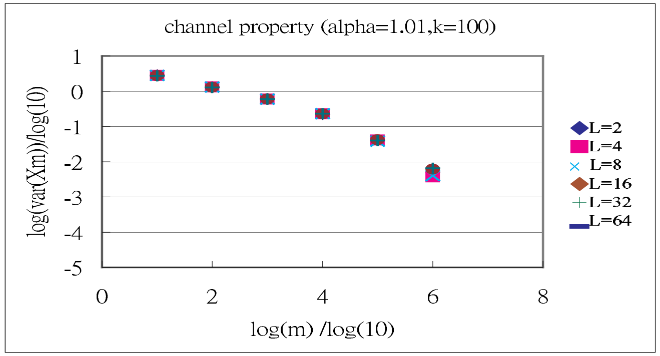
Tandem Server Systems

Variance-Time Analysis

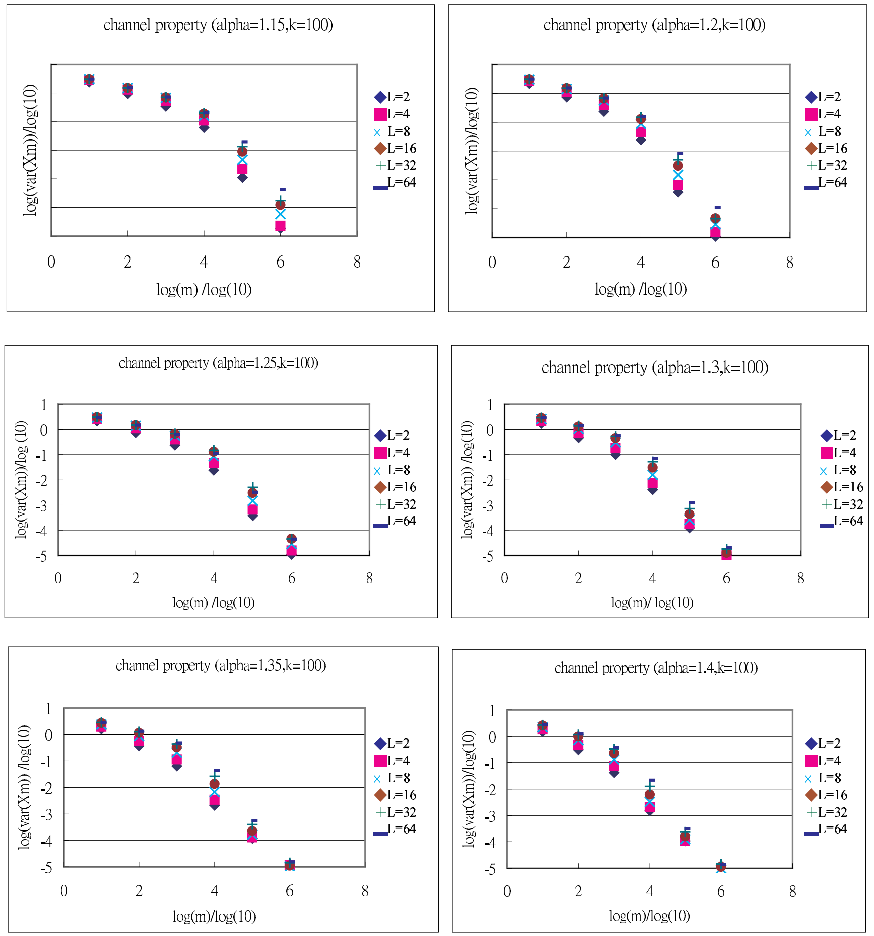
- The degree of dependence gradually decreases with the observation range.
- From the figure in the next slide, the variance-time analysis of the departures of the tandem server system does not appear to be straight line, nor to be a combination of two lines of different slopes as what obtained in parallel server system.
- For α small,
  - the variance of m-average departure process scatters more at $m = 10^6$ than at $m = 10$ and $m = 10^2$.
- For α large,
  - two kinds of behaviors for different regions of m can be observed: for some m, the variances of the m-average departure process diverge for different L, while for some other m, they actually converges.







(Continue on the next page.)



The variance-time plots of the departures for tandem L-server system. The originated Poisson arrival has mean $\lambda = 0.001$. The Pareto parameters, α and k, used for each Pareto server are indicated in each subfigure.

Analysis

- Let the queue lengths of tandem server system at time instant n be respectively denoted by Q_n^1, Q_n^2, ..., and Q_n^L.
- Denote by j_1, j_2, $\ldots$ the queues that has length zero, namely, $Q_n^{j_1} = Q_n^{j_2} = \cdots = 0$.
- The the longest length of consecutive non-zero queues is:

$$Q_{\max} = \max\left\{\sum_{j=0}^{j_1-1} Q_n^i, \sum_{j=j_1}^{j_2-1} Q_n^i, \sum_{j=j_2}^{j_3-1} Q_n^i, \ldots\right\}.$$

Relation between H, L and α

- For α which make utilization $\rho > 1$, by finding the best-fit line to the figure on the previous slide, we obtain a figure (as depicted in the next slide) which suggests that

$$H \approx f_1(\alpha)e^{-a(\alpha)\cdot L} + f_2(\alpha), \tag{1}$$

  where $f_1(\cdot)$, $f_2(\cdot)$ and $a(\cdot) > 0$ are some positive functions of α.

- A subsequent figure shows that H can be approximated by a linear function of α

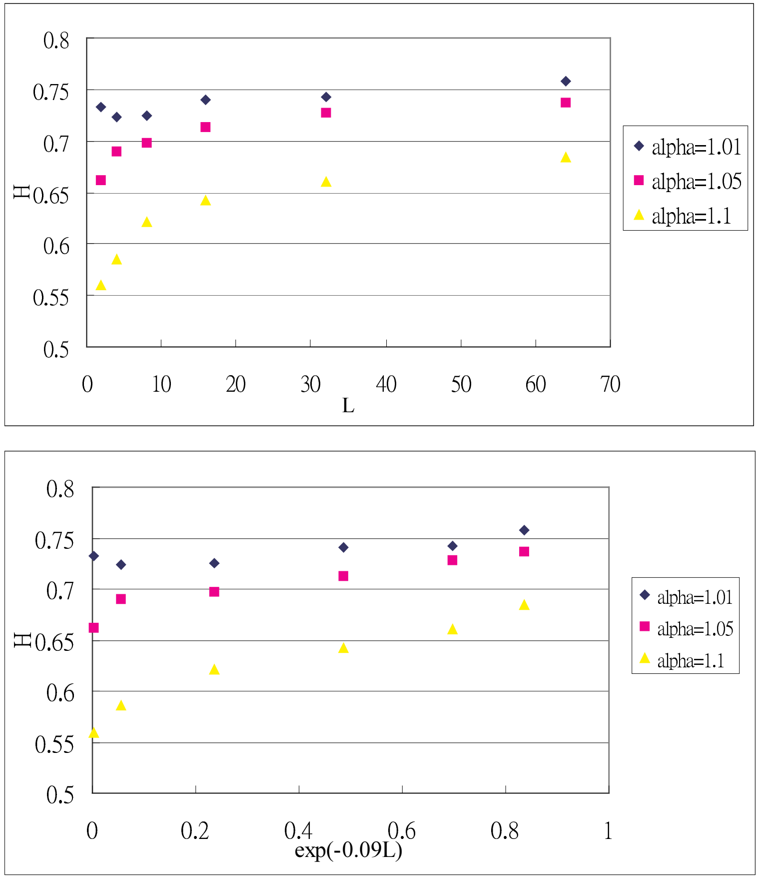
$$H \approx c - g(L)(\alpha - 1), \tag{2}$$

  where $g(\cdot)$ is some function of L, and c is a constant.

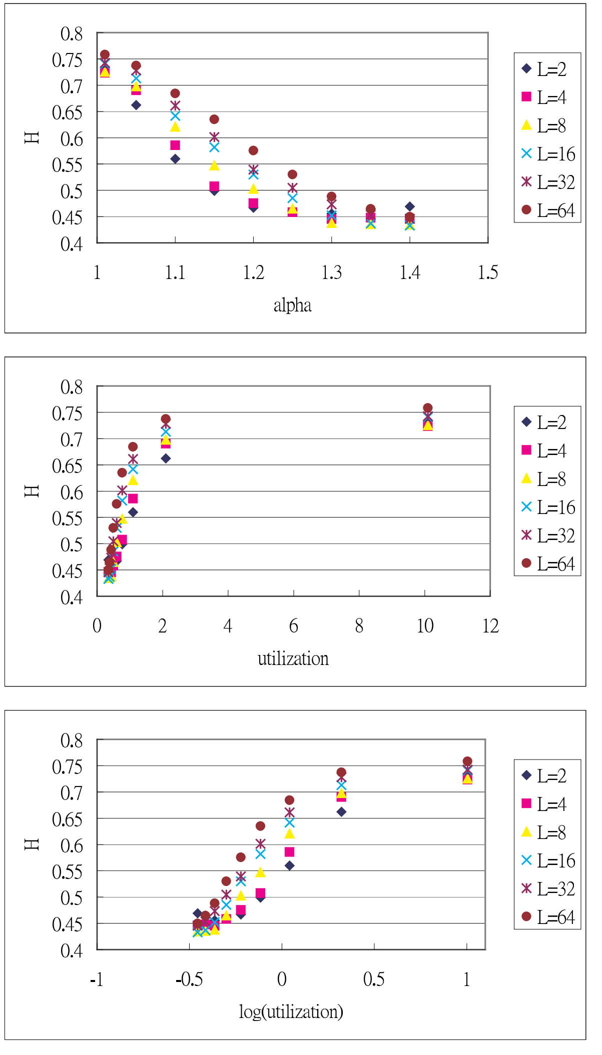
- Combine Eq. (1) and (2)

$$H \approx c - (\alpha - 1)(c_1 - e^{-c_2 L}),$$

- As $L \to \infty$, H approaches $c - c_1(\alpha - 1)$.



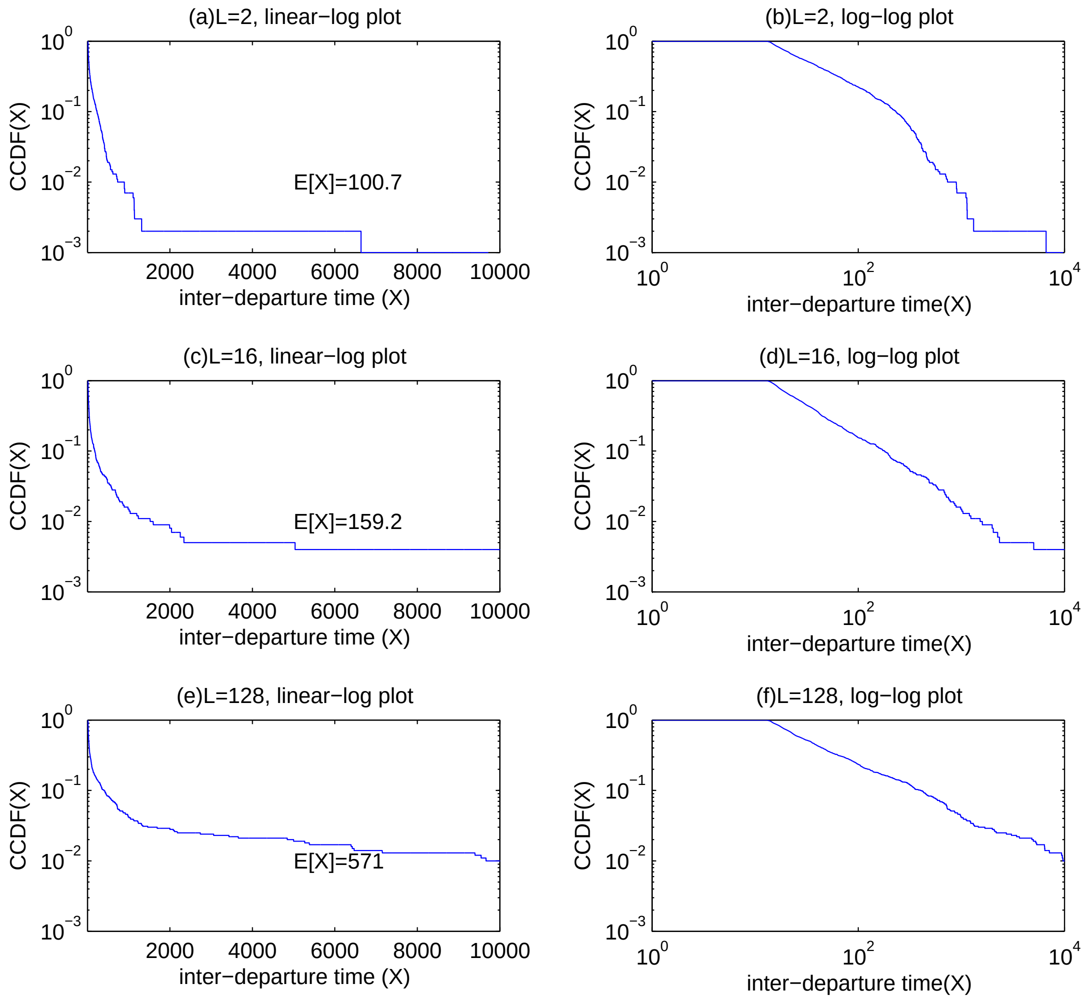
The observed Hurst parameter $\hat{H}$ *obtained by finding the best-fit lines. The upper subfigure gives a linear scale for* L*, and the lower subfigure presents with a logarithmic scale for* L*.*



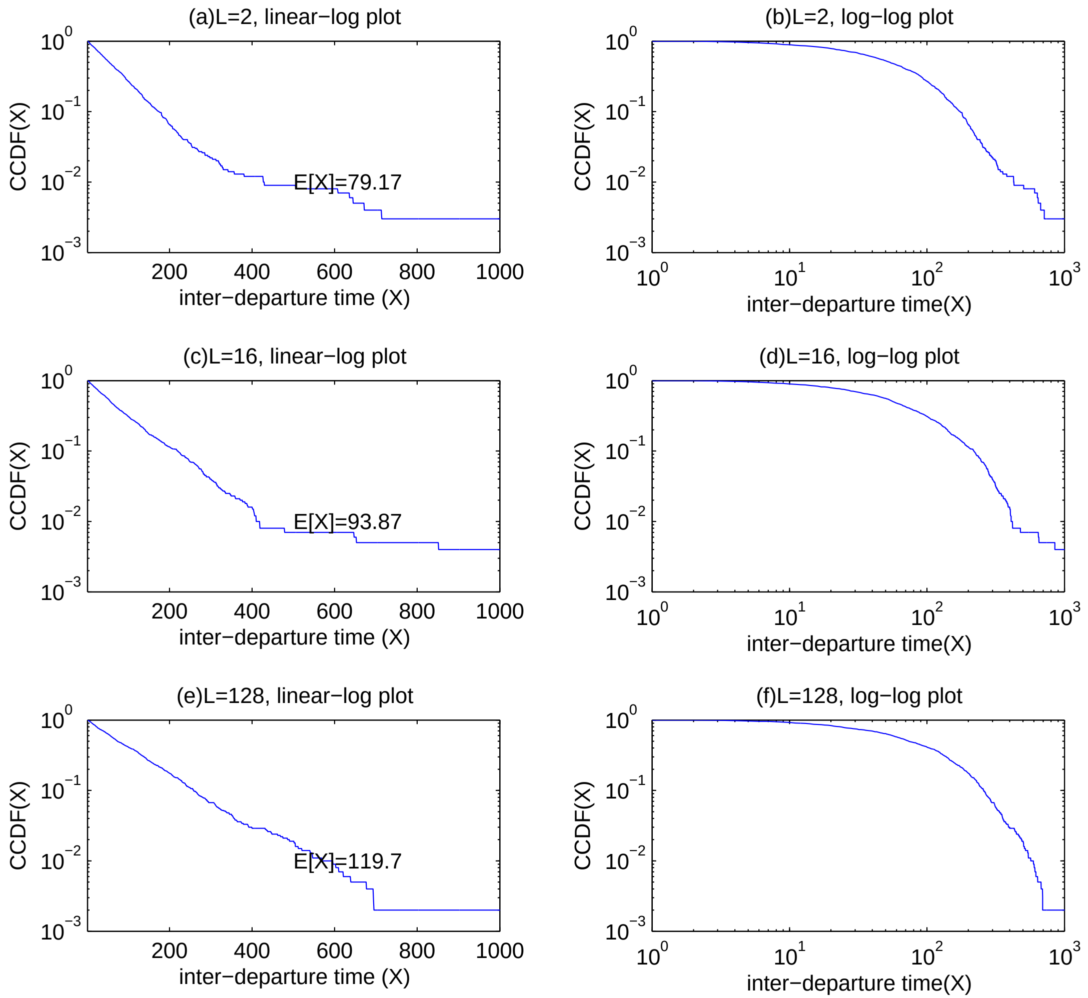
The Hurst parameter versus α and utilization.

CCDF Analysis Under L-Tandem Server System

- Poisson Arrival and Pareto Server
  - A larger interdeparture time can be observed for large L.
  - The slope of the CCDFs in log-log scale, which should equal the negation of Pareto shaping parameter α, is getting smaller (namely, decaying faster) for large L.
- Pareto Inter-Arrival and Tandem Exponential Servers
  - The interdeparture becomes more like a non-heavy-tailed exponential distribution.



The CCDF of inter-departure time obtained by passing Poisson arrivals through tandem Pareto servers. (Poisson arrival: $\lambda = 0.01$*, Pareto:* $\alpha = 1.25$ *and* $k = 12$*).*



The CCDF of inter-departure time obtained by passing Pareto inter-arrival process through tandem exponential servers. (exponential: $\lambda = 0.0167$*, Pareto:* $\alpha = 1.25$ *and* $k = 20$*).*

Entropy Rate Of The Interdeparture Process

- The results match what has been concluded in [**Anantharam and Verdu in 1996**], where the entropy rate is maximized for Poisson departure processes.

The sizes of Lempel-Ziv coded and uncoded files of inter-departure time with different number of tandem servers. The interarrival is exponential distributed with $\lambda = 0.01$*, and the Pareto service rate has distribution parameters* $\alpha = 1.25$ *and* $k = 12$*.*

L	2	32	256
original file size	400,000 Bytes	400,000 Bytes	400,000 Bytes
coded file size	101,775 Bytes	97,643 Bytes	94,236 Bytes
compression rate	0.254438	0.244108	0.23559

The sizes of Lempel-Ziv coded and uncoded files of inter-departure time with different number of tandem servers. The interarrival is Pareto distributed with $\alpha = 1.25$ *and* $k = 20$, *and the exponential service rate has distribution parameter* $\lambda = 0.0167$.

L	2	32	256
original file size	400,000 Bytes	400,000 Bytes	400,000 Bytes
coded file size	130,552 Bytes	137,445 Bytes	146,173 Bytes
compression rate	0.32638	0.343613	0.362933

Conclusion and Future Work

Conclusion

- In this thesis, finite resources, such as the number of servers L, is introduced, as contrary to the usual theoretical assumption of $L = \infty$.

Along this research line, we examine the M/G/L system

- Our simulations show that when the system utilization is fixed, adjusting the shaping parameter α (used in the Pareto server) does not make the inter-departure traffic as self-similar as $H = (3 - \alpha)/2$ suggests.
  - We realize from this simulation that the reduction of k neutralizes the anticipated degree of self-similarity at small α.
  - This leads to the subsequent simulations in which k is kept fixed.
- By fixing k rather than the mean in Pareto server
  - Increasing of α gives a monotone decreasing of self-similar parameter H.
  - A further increasing of L, which is the number of parallel Pareto servers, grows the degree of departure self-similarity.

For the tandem server system

- We perform the CCDF analysis on the interdeparture.
  - We observed an inconsistence on the CCDF behavior for different combinations of arrival and service time distributions.
  - Only the Pareto servers present apparent heavy probability tail.
- Even with similar CCDFs for both Pareto and exponential interarrivals passing to exponential servers, they have different (Lempel-Ziv bound of) entropy rates.
  - Our result suggests that Pareto servers may generate an interdeparture process consisting of the most redundancy among those cases we considered.

Thank you for giving advice

future Work

- Different queueing principles
- Provide more theoretical basis