

# *Data Staging Effects in Wide Area Task Farming Applications*

**Wael Elwasif**

Computer Science and Math. Division  
Oak Ridge National Laboratory.

**James Plank, Rich Wolski**

Department of Computer Science  
Univ. Of Tennessee, Knoxville.

# The Question

---

*How does data pre-staging affect task throughput in wide area task farming applications?*

# Contents

---

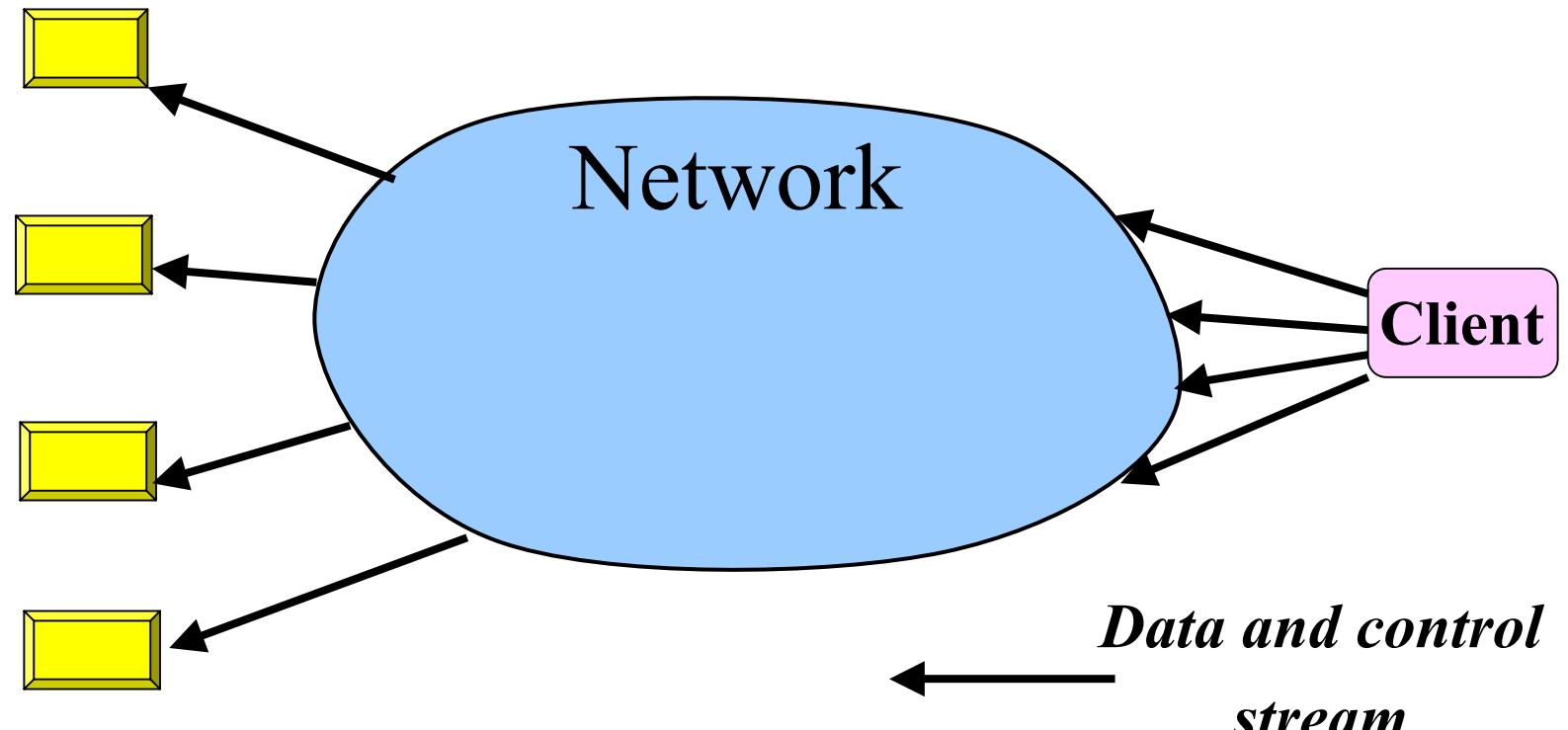
- The problem.
- Task farming mathematical model.
- Experimental results.
- Simulation results.
- Conclusion.

# Data Pre-staging

---

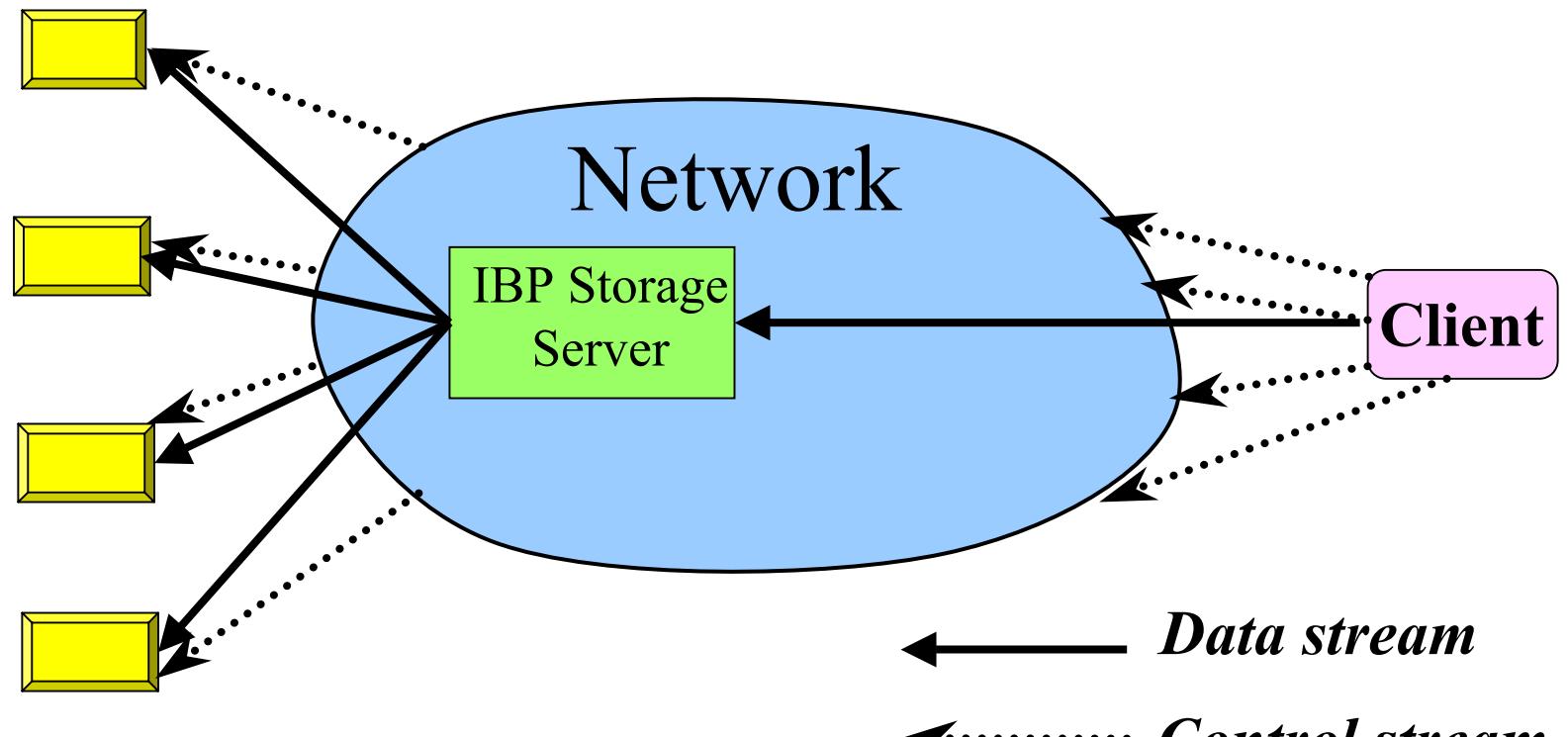
- Problem: **wide area task farming.**
- Characteristics:
  - Embarrassingly parallel application.
  - Large common input data set.
  - Small control data set.
  - Small output data set.
- Examples:
  - Time series analysis.
  - Physical systems design/simulations.
- Execution model: client-agent-server (NetSolve).

# No-staging Model



Computational Servers

# Staging Model



# IBP: Storage in the Network

---

- Internet Backplane Protocol
- User level client-server remote storage access architecture developed at UTK.
- Files accessible at the byte level.
- Users contribute local storage to a common pool for general access.
- Users manage how their local storage can be used (e.g. data lifetime).

# Anatomy of a Task



- $T_f$  task forking time.
  - $T_t$  input data transfer time.
  - $T_c$  task computational time.
- } **No staging**

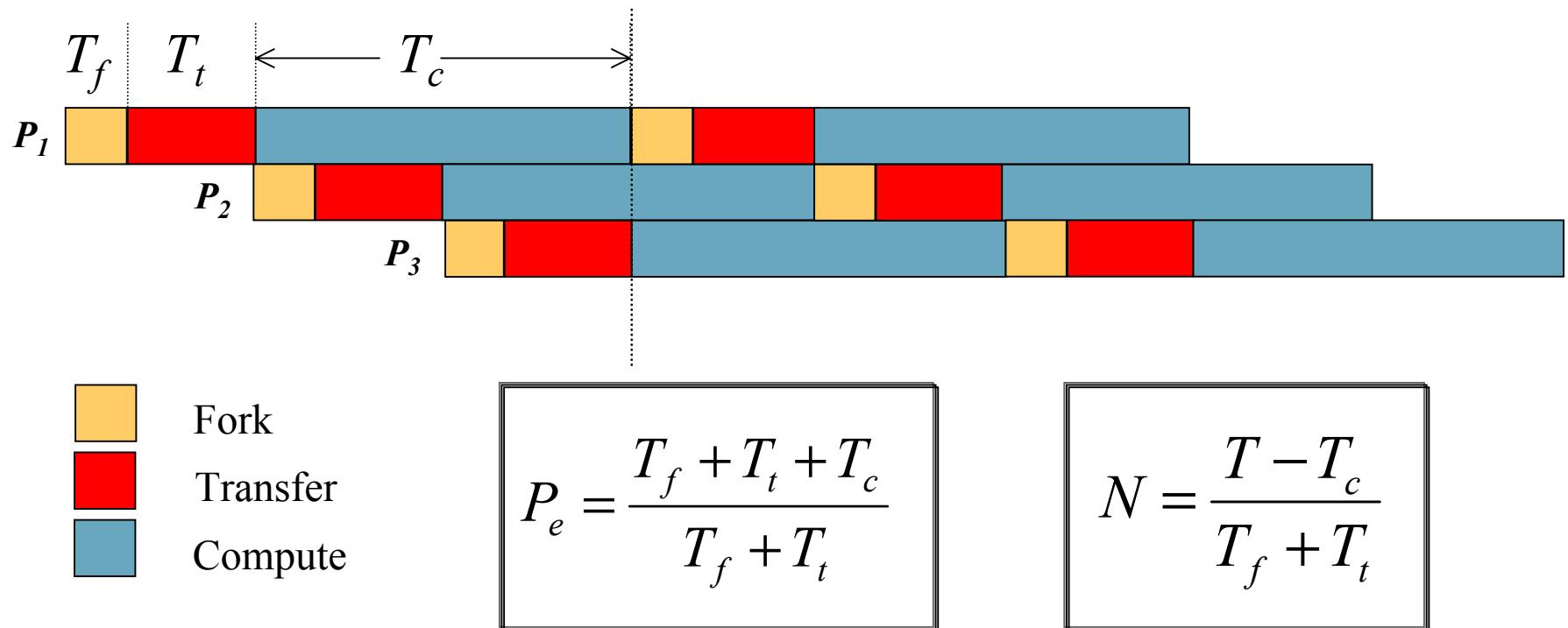
For staging, use  $T_f'$ ,  $T_t'$ ,  $T_c'$

# Application Parameters

---

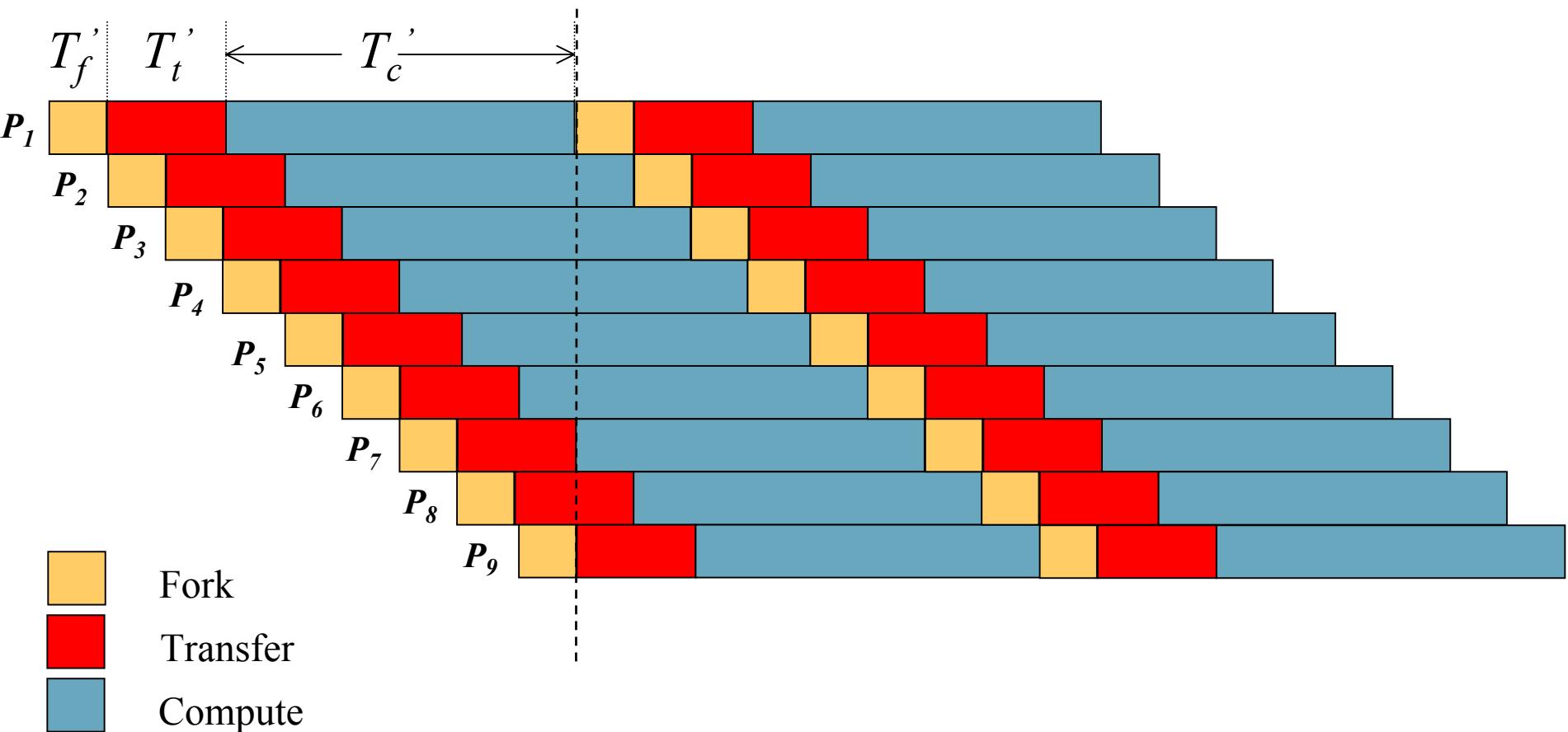
- $T$  total run time.
- $P$  number of available servers.
- $N$  number of completed tasks in  $T$  s.  
(No staging).
- $N'$  number of completed tasks in  $T$  s.  
(Using staging).

# No-staging Model

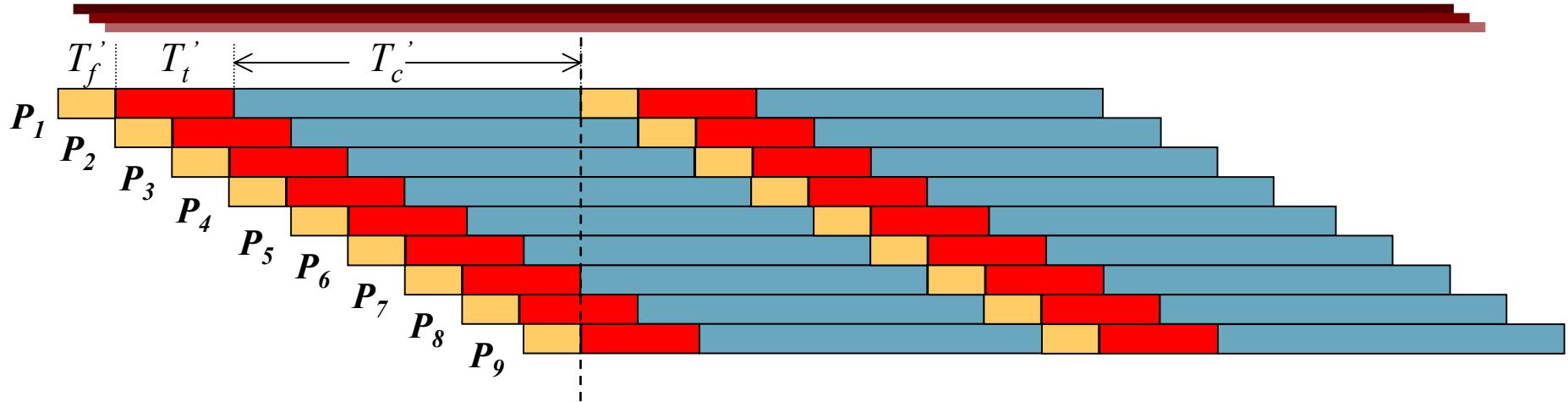


$P_e$  “effective” number of servers

# Staging: How Does It Work?



# Staging Model



$$P_e' = \frac{T_f' + T_t' + T_c'}{T_f}$$

$$N' = \frac{T - T_t' - T_c'}{T_f}$$

# Throughput Ratio (1)

- $P_e < P$

$$\frac{N'}{N} = \frac{T - T_c - T_t}{T_f} \frac{T_f + T_t}{T - T_c}$$
$$\approx 1 + \frac{T_t}{T_f}$$

$$N' = \frac{T - T_t - T_c}{T_f}$$

$$N = \frac{T - T_c}{T_f + T_t}$$

# Throughput Ratio (2)

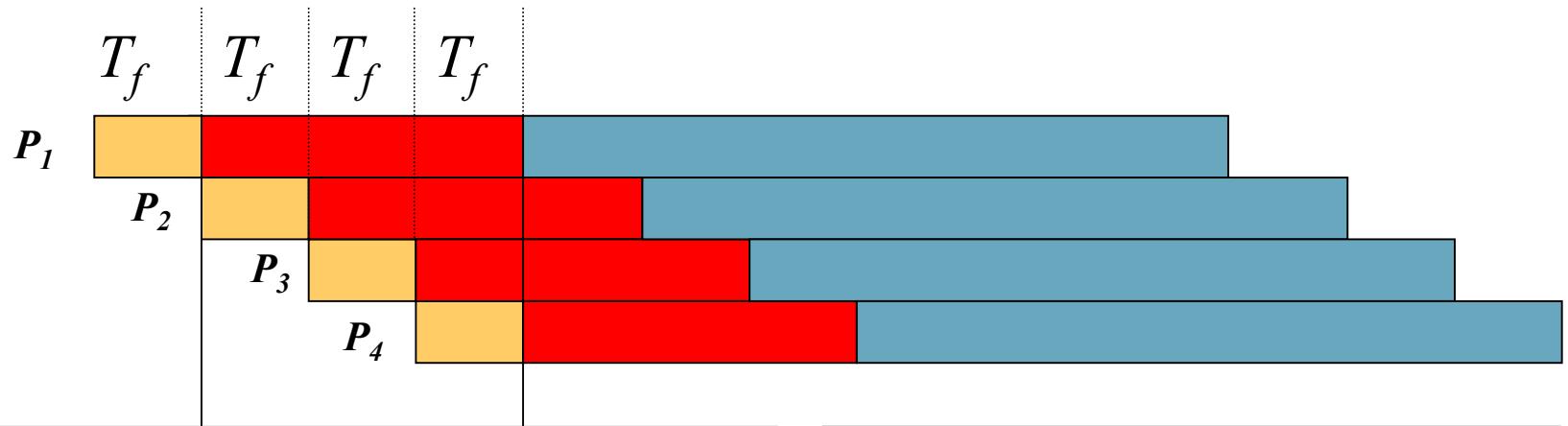
- $P_e > P$

$$\frac{N'}{N} \approx \frac{T_f + T_t + T_c}{T_f + T_t' + T_c}$$

$$N' = P \frac{T}{T_f + T_t + T_c} - (P-1)$$

$$N = P \frac{T}{T_f + T_t + T_c} - (P-1)$$

# Bandwidth Sharing Effects



$$S = T_f BW + T_f \frac{BW}{2} + T_f \frac{BW}{3}$$

$$= T_f BW \sum_{i=1}^n \frac{1}{i}$$

$S$  Input data size

$BW$  Staging-to-computing servers BW.

# Bandwidth Sharing (Cont.)

$$S \approx T_f BW[\ln(n) + \gamma]$$

$\gamma \Rightarrow$  Euler-Mascheroni constant  $\approx 0.5772156649$

$$n \approx e^{\frac{s}{T_f BW} - \gamma}$$

**Effective bandwidth  $BW_e \leq BW/n$**

# Experimental Setup

---

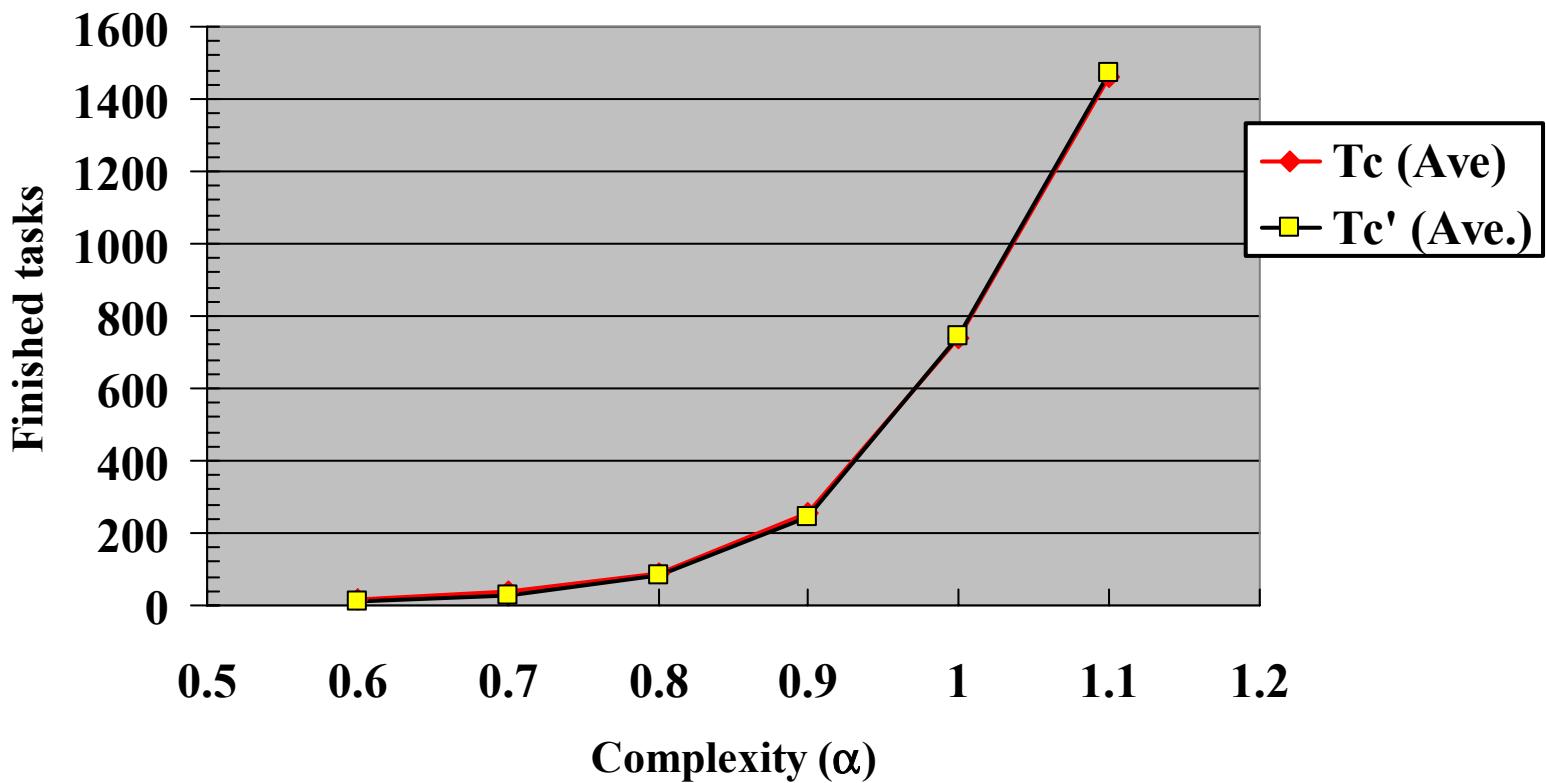
- NetSolve servers + Agent on 41 UTK-CS hosts.
- Client running at Princeton.
- “Work stealing“ mode (**nice** apps.)
- Synthetic, tunable algorithm.
  - Complexity =  $c n^\alpha$ , where n = Input size.
- IBP on DSI storage server at UTK (100 Mb/S.)
- Input data size : 832 KB.
- Number of computational servers: 35 and 6.

# Experiment Observations

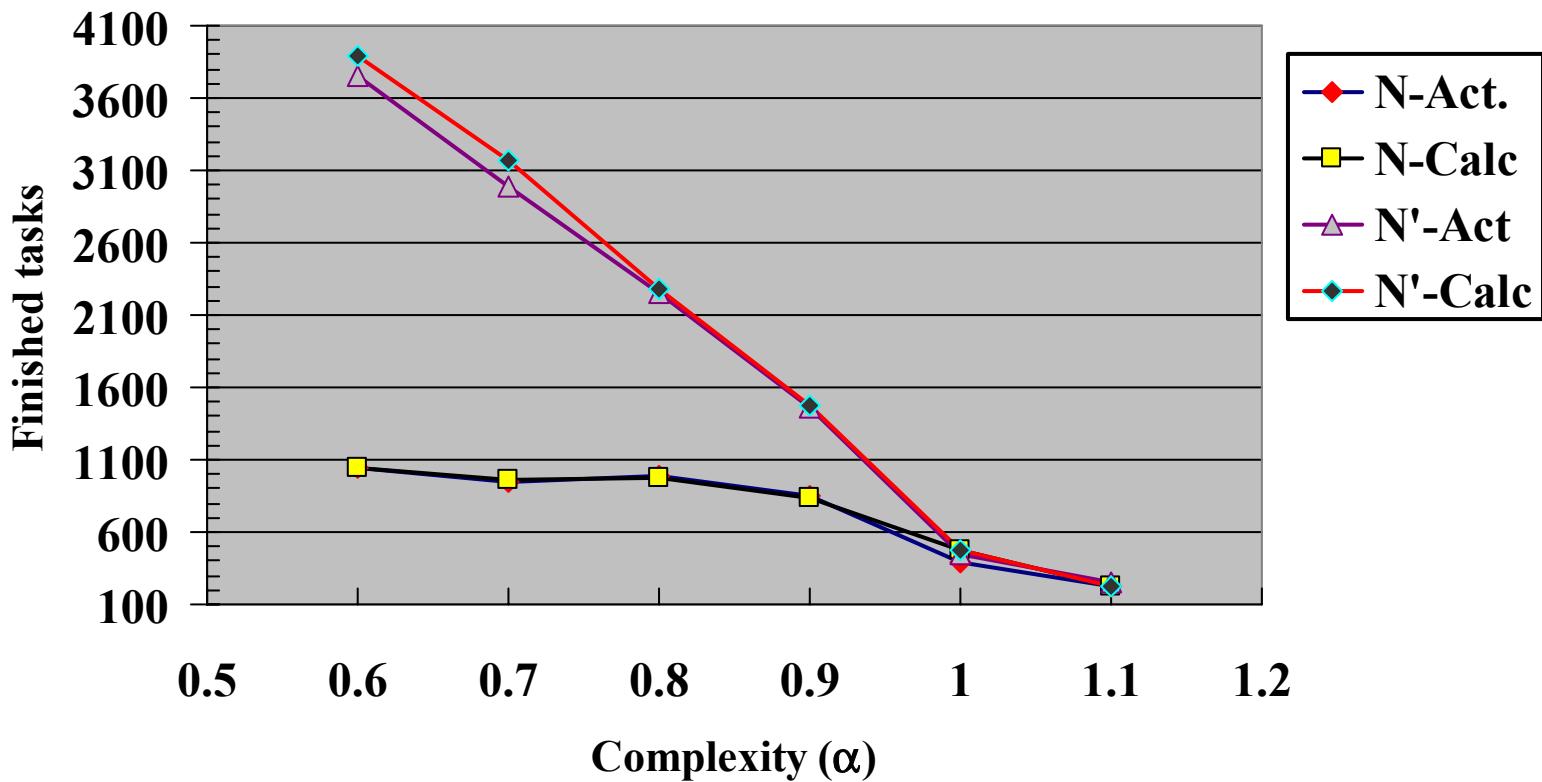
---

- Average  $T_f$  : 2.6 - 7.1 sec.
- Average  $T_t$  : 6.0 – 8.5 sec.
- Average  $T_t'$  : < 1 sec.

# Computational Time

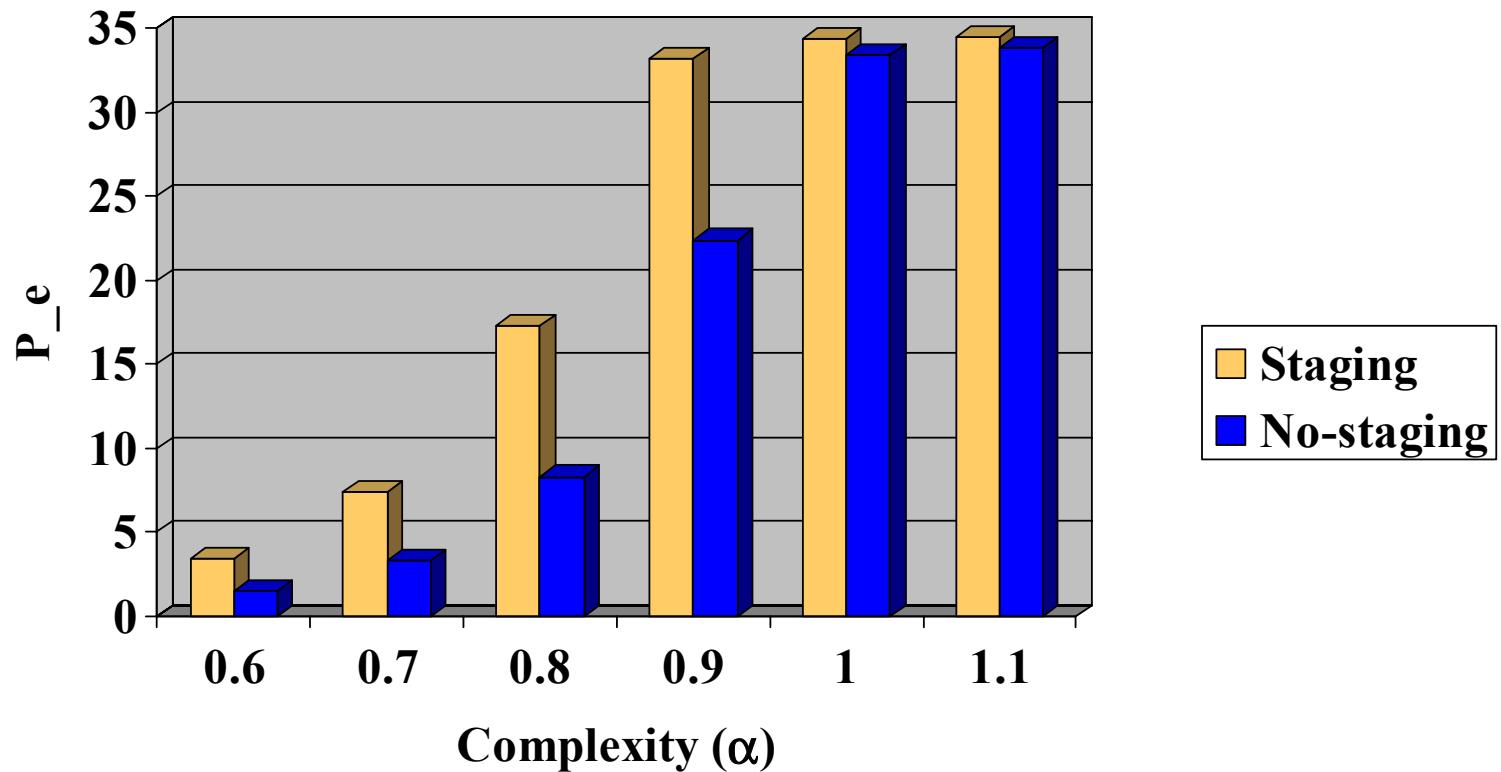


# Results: 35 Servers

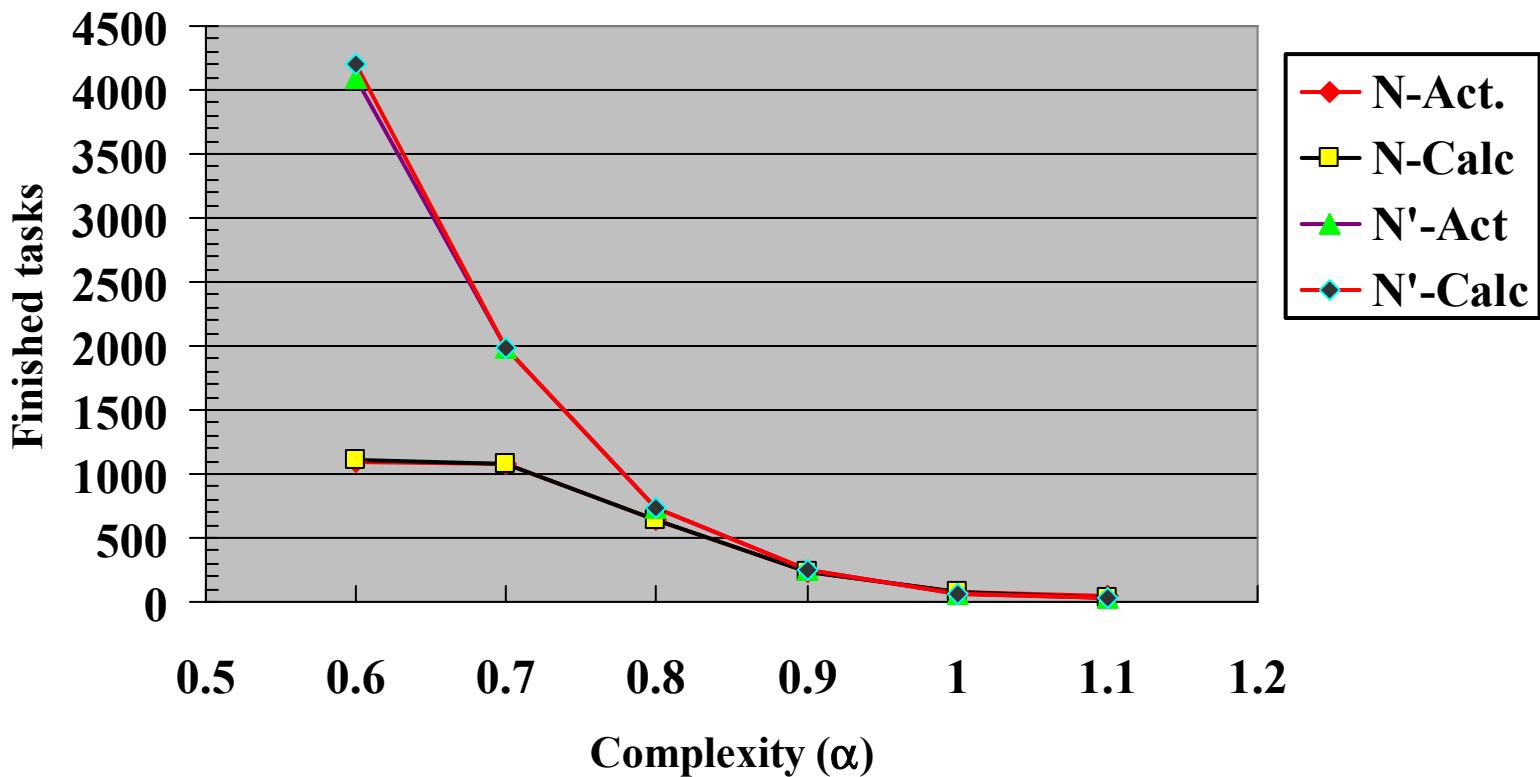


# Results: 35 Servers (Cont.)

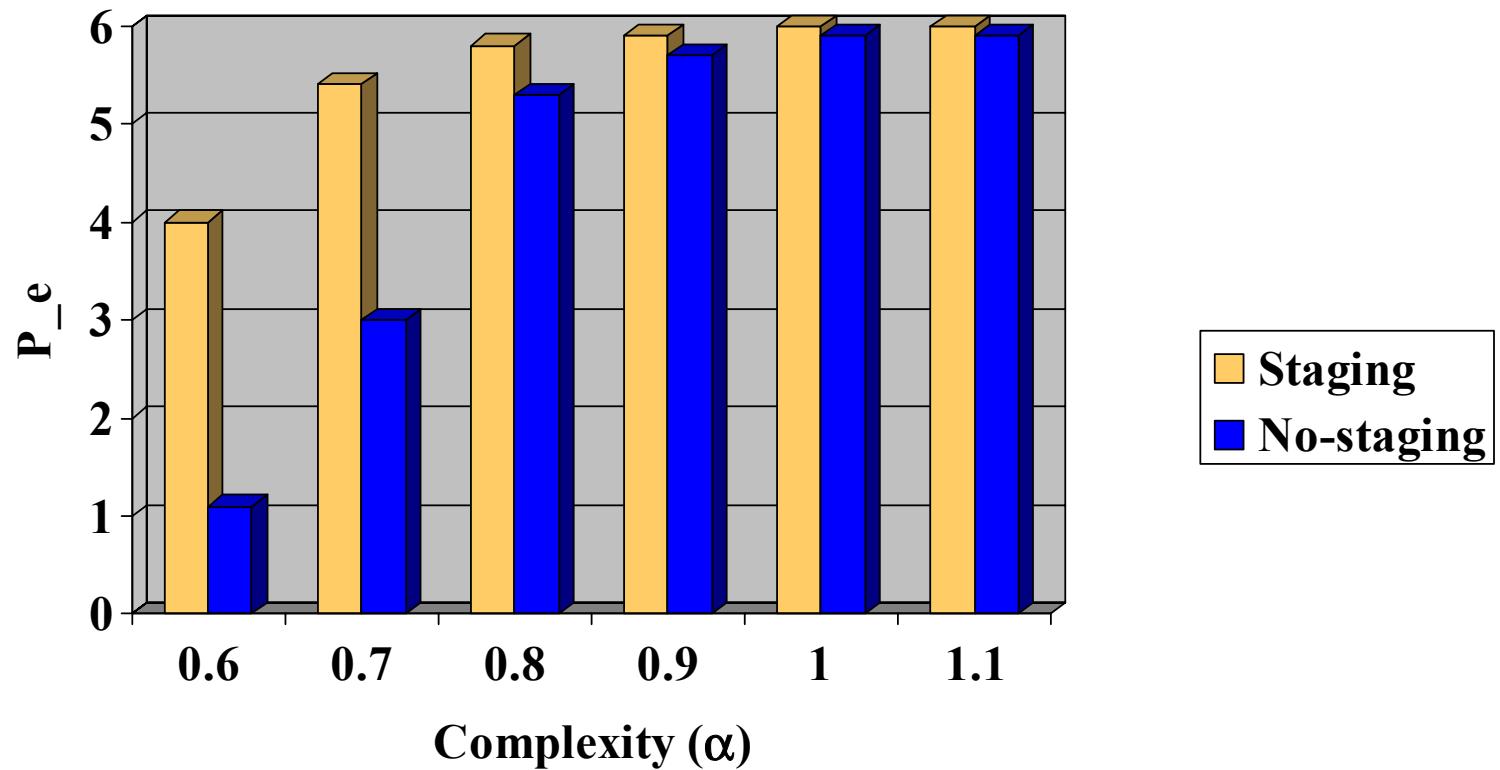
---



# Results: 6-servers



# Results: 6-servers (Cont.)

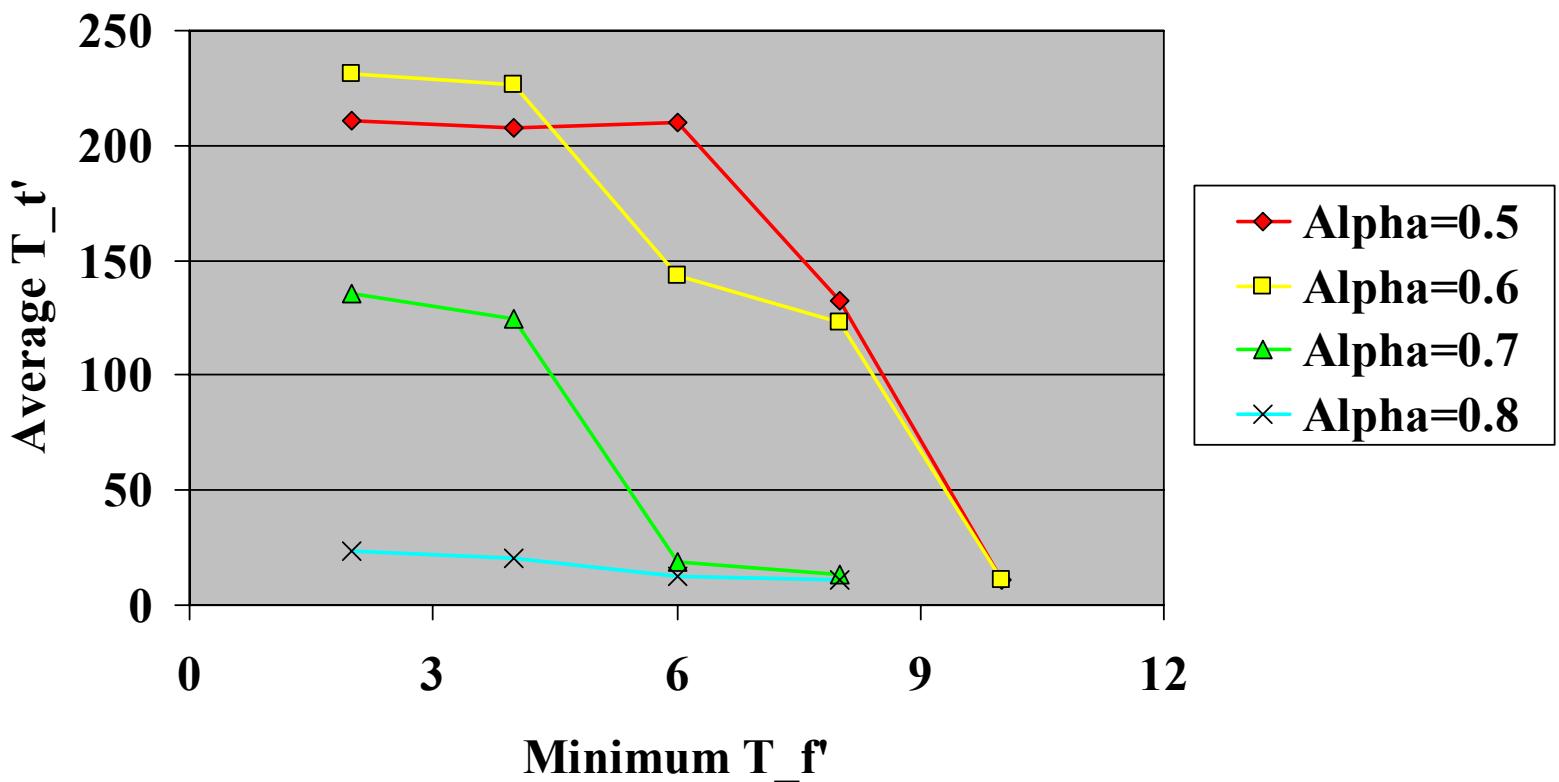


# Experimental Setup (2)

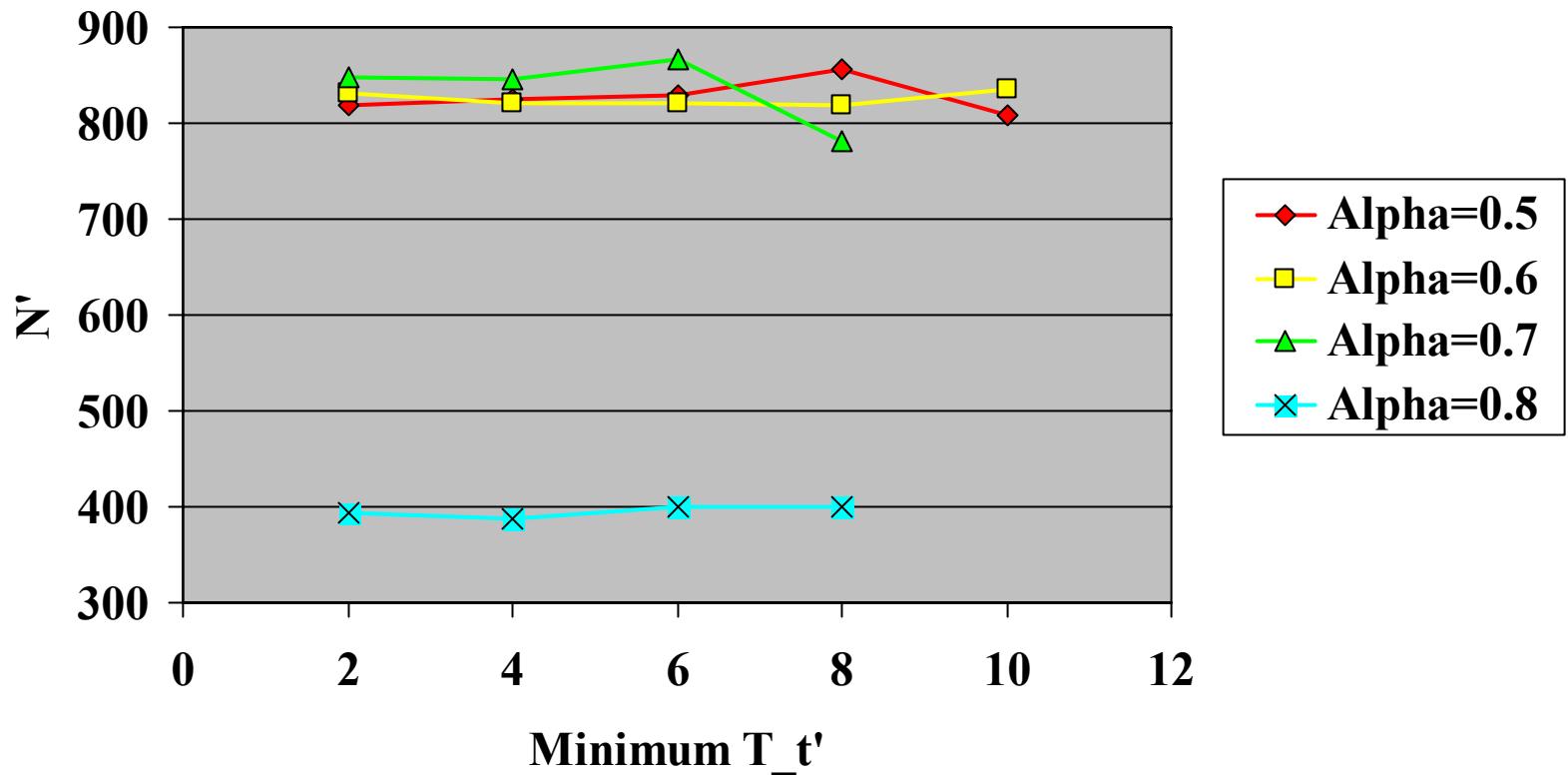
---

- To study effect of bandwidth sharing on staging performance.
- Data size = 9MB.
- 28 computational servers + agent at UTK.
- Client at Princeton.
- Staging server: SUN 10, 10 Mb/s. NIC.
- Vary  $T_f$  through induced delays at client.

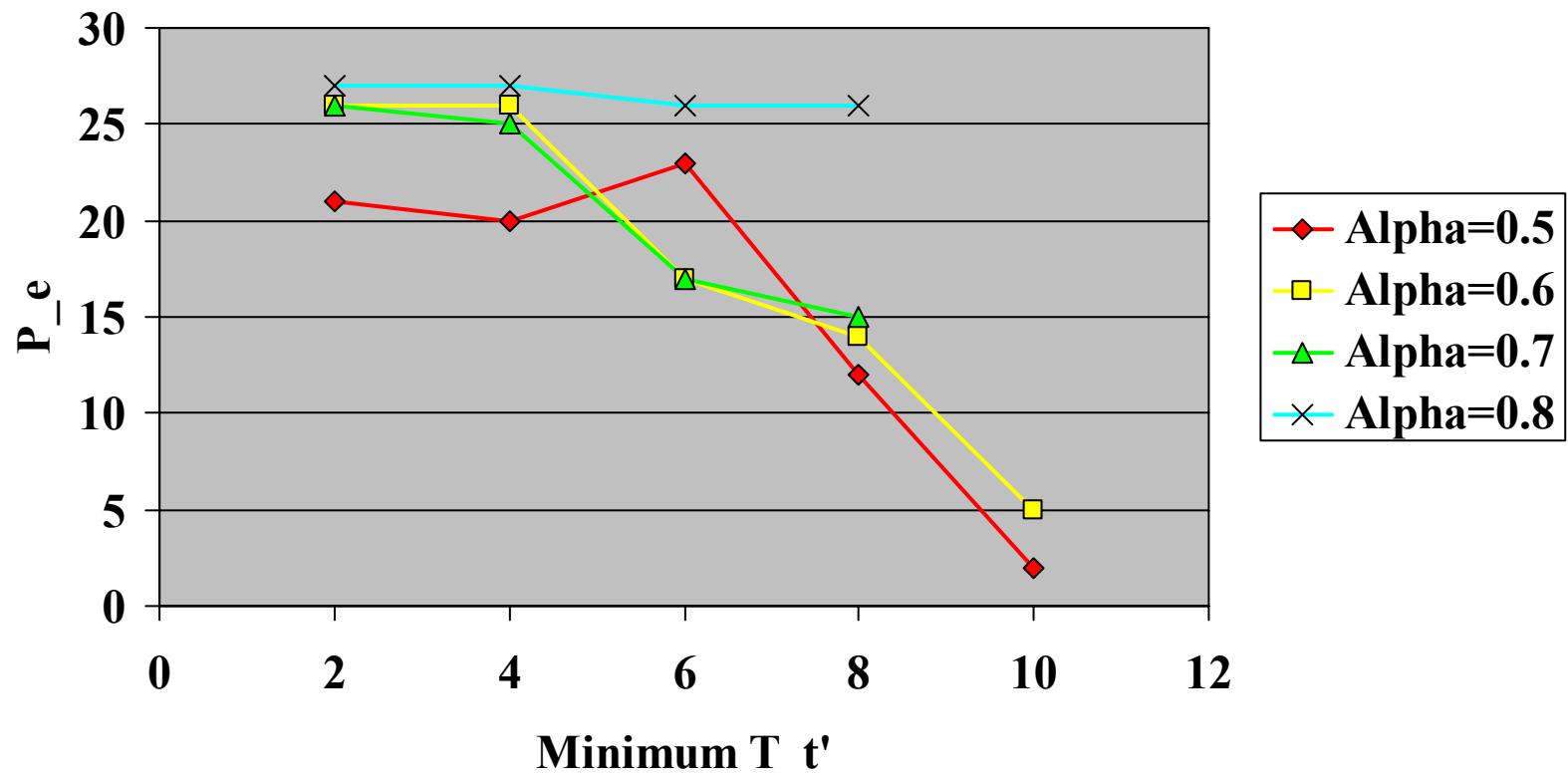
# Data Transfer Time



# Task Throughput



# Effective Number of Servers

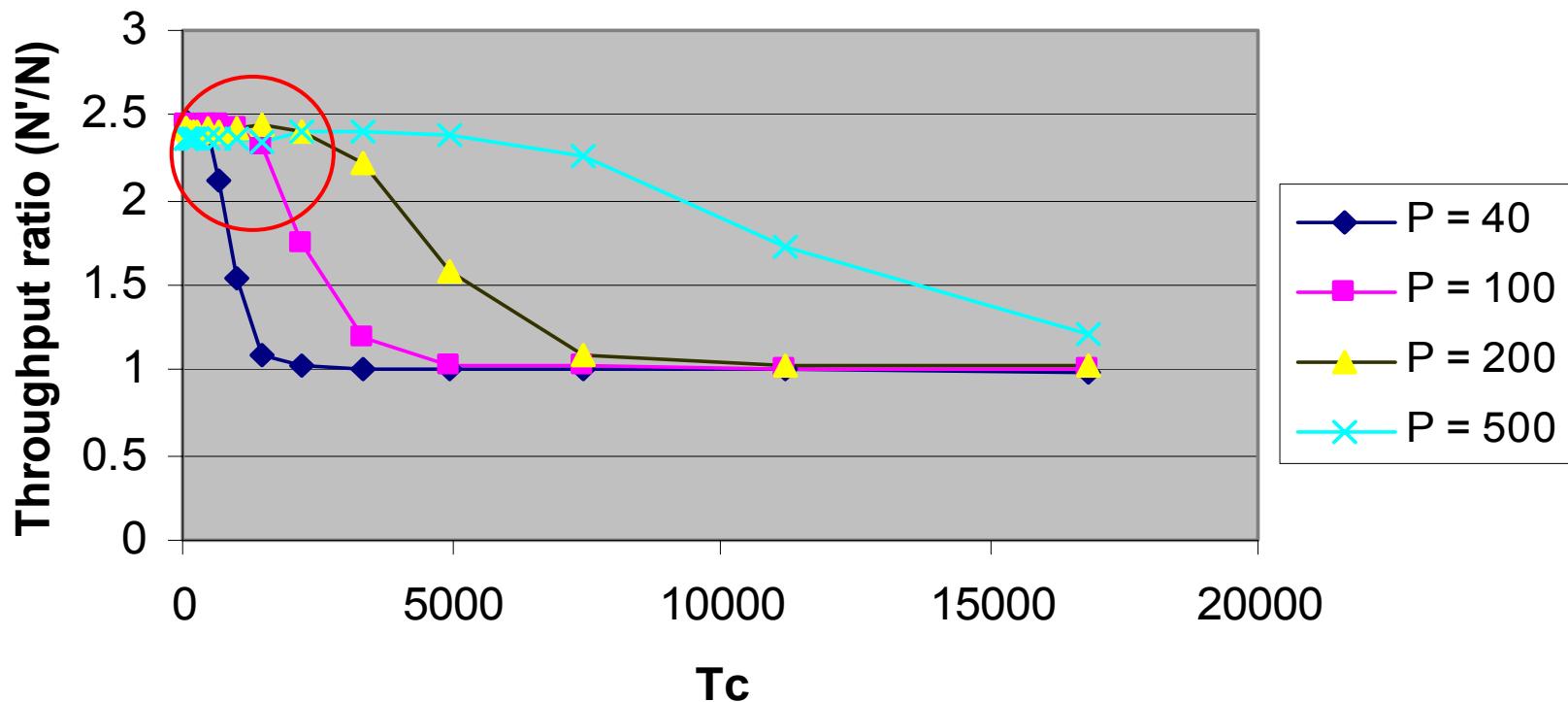


# Task Farming Simulator

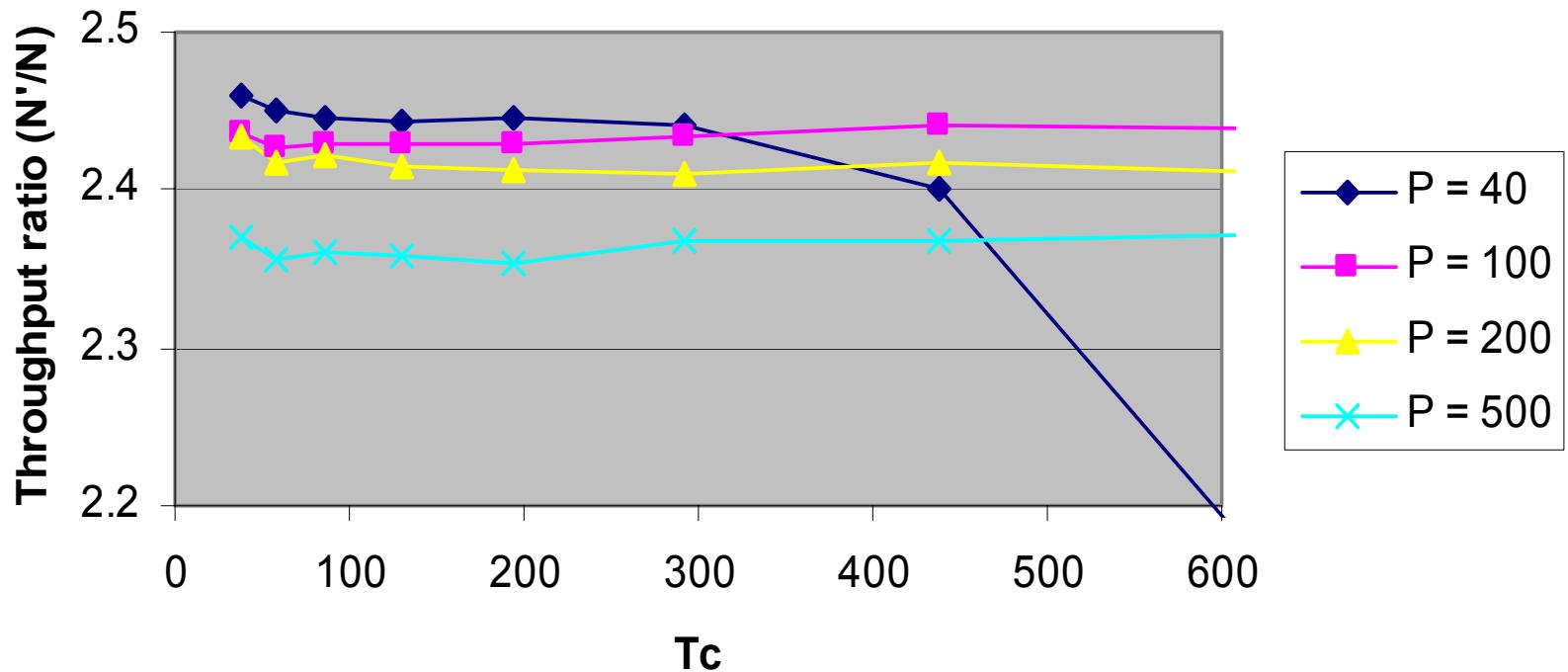
---

- Based on the two previous models.
- Accepts probability distributions for  $T_c$ ,  $T_f$ , and remote and local bandwidth.
- Results obtained using average of 10 runs:
  - $T_f$  uniform in [3,7] sec.
  - Local BW uniform, mean 1.0 MB/sec.
  - Remote BW uniform, mean 0.5 MB/sec.

# Simulator Results: 16 MB



# Simulator Results: 16 MB (Cont.)



# Summary

---

- Mathematical model for staging in task farming applications.
- Task throughput improvement due mainly to increased server utilization.
- Bandwidth sharing effects depends on the ratio  $S/(T_f BW)$ .
- *Future work:* multiple staging servers.

# Questions??

---