Calculating Call Blocking, Preemption Probabilities, and Bandwidth Utilization for Satellite Communication Systems.

Aero RIPS 2011
Leah Rosenbaum, Mohit Agrawal, Leah Birch, and Yacoub Kureh

Academic Mentor
Nam Lee

Aerospace Mentors
James Hant, Eric Campbell, Krista O'Neill, Brian Wood, and James Gidney

Final Presentation
19 August 2011

Outline

1. Introduction
2. Single-Type Model
3. Priority Model
4. Bandwidth Model
5. Static vs. Dynamic Resource Allocation Schemes
6. Conclusions and Future Work

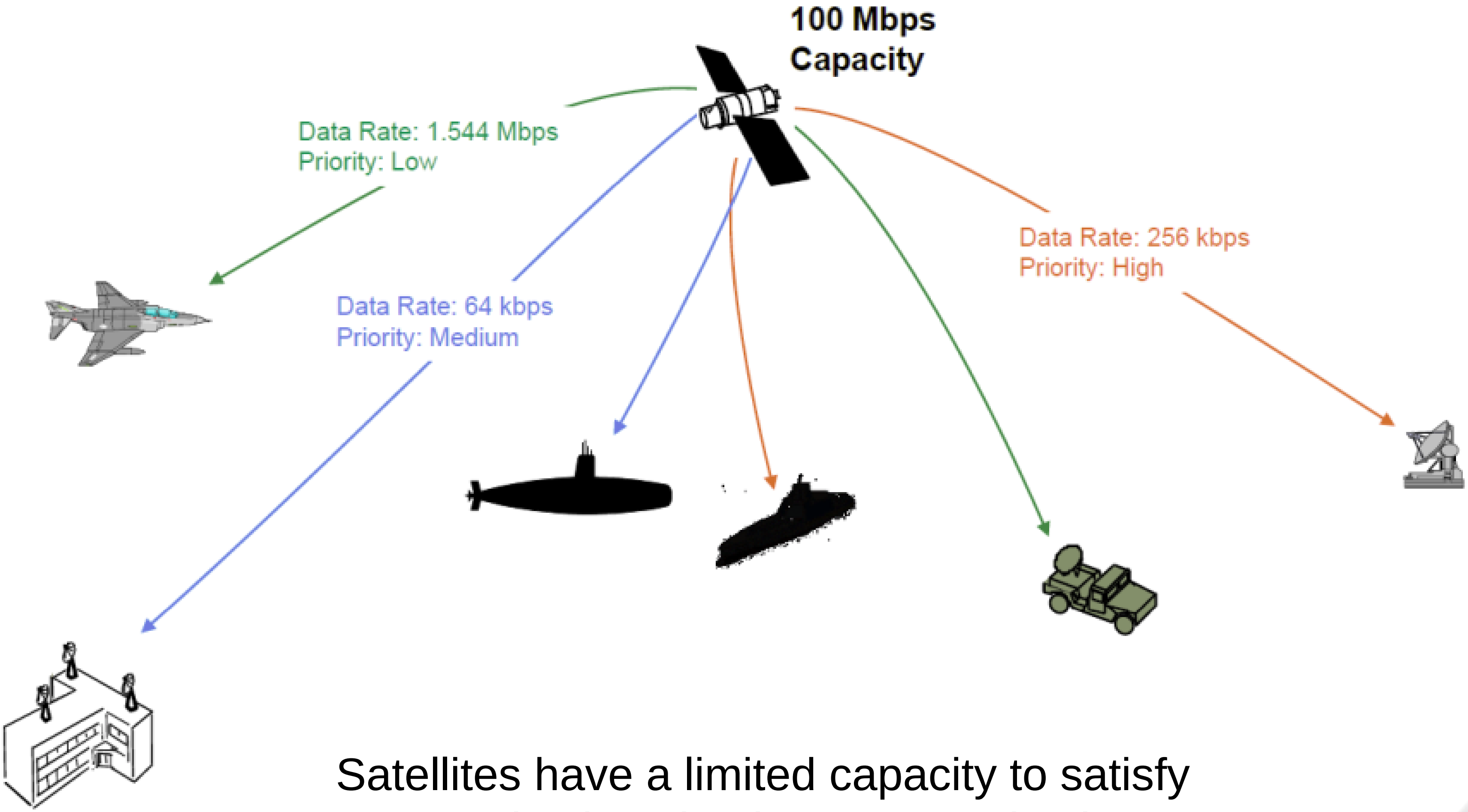
1. Introduction

About the Aerospace Corporation

- Aero is a federal R&D center.
- Aero manages the development, launch, and maintenance of satellites.

We were there! →

Satellite Communication System



Satellites have a limited capacity to satisfy communication circuits. Communication requests have varying priorities and data rates.

Static Allocation Is Inefficient

Circuit #	Priority	Data Rate	Time										Duty
1	High	5 Mbps	X	X			X	X	X	X			0.6
2	High	10 Mbps			X	X							0.2
3	High	20 Mbps					X	X	X	X			0.4
4	High	30 Mbps				X							0.1
5	High	5 Mbps	X	X		X		X				X	0.5
6	High	15 Mbps			X			X	X	X			0.4
7	High	4 Mbps					X				X		0.2
8	High	15 Mbps	X	X	X								0.3
9	Medium	10 Mbps						X	X				0.2
10	Medium	15 Mbps		X									0.1
11	Medium	50 Mbps	X	X	X		X				X		0.5
12	Medium	3 Mbps							X				0.1
13	Medium	10 Mbps			X	X		X		X			0.4
14	Low	15 Mbps		X							X		0.2
15	Low	8 Mbps	X				X	X					0.3
16	Low	8 Mbps		X			X		X				0.3
17	Low	8 Mbps								X	X		0.2
18	Low	8 Mbps	X	X	X							X	0.4
Server Utilization			75	90	100	55	79	65	53	50	54	5	

X Circuit Resources Requested

Circuit Resources Allocated

Circuit Resources Not Allocated

Based on figure from June 28, 2011 Aerospace presentation.

Approach

- Generate theoretical and simulation models to measure <u>user satisfaction</u> and <u>resource utilization</u> for **single-type traffic**.
- Expand the theoretical and simulation model to measure user satisfaction and resource utilization for competing traffic classes with **different priorities or bandwidths**.
- Use the new model to compare user satisfaction and resource utilization for **dynamic and static allocation** schemes.

2. Single-Type Model

The Model

- The times when jobs enter the system and leave the system are random variables.
- A queueing model is not fully appropriate.
- Best modeled by a stochastic processing network with no queues.
- Stochastic processing networks can be
  - *simulated by discrete-event simulation, and*
  - *modeled theoretically using Markov chains.*

Definitions and Parameters

M/M/m/0

Arrival distribution | Service distribution | # servers | Queue capacity

- $\lambda =$ Arrival rate (birth rate)
- $\mu =$ Service rate (death rate)

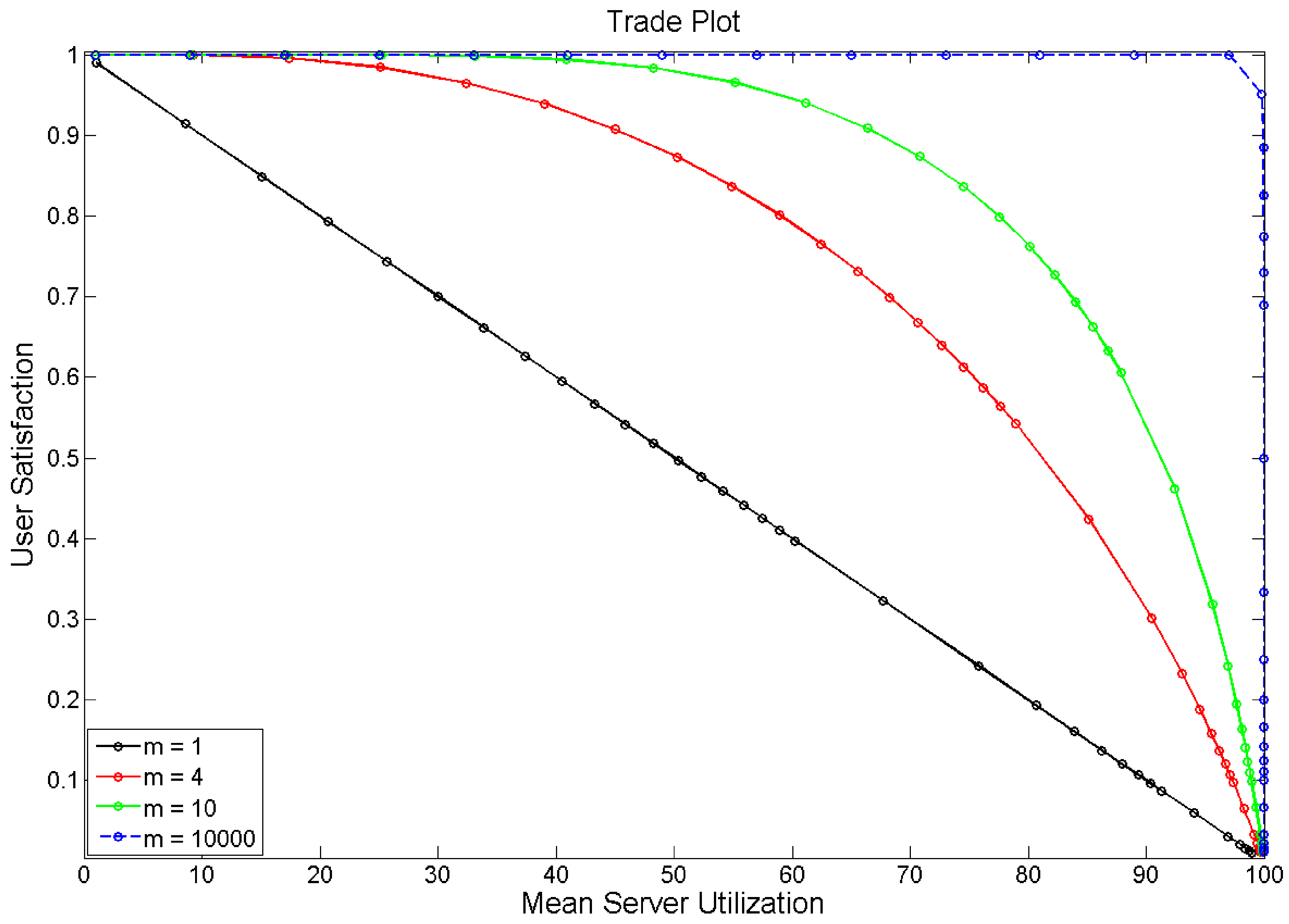
Parameters for exponential distribution

- $\rho = \lambda / m\mu =$ Traffic intensity

System Properties

For each system, defined by the parameters (ρ, m), we:

- Enumerate the state space.
- π_i = Steady state probability of the system being in state i
- User satisfaction: $S_{\rho,\mathrm{m}} = 1 - \sum_{i \in A} \pi_i$, where $A = \{states\ in\ which\ a\ job\ is\ blocked\}$.
- Resource utilization: $\mathrm{Ut}_{\rho,\mathrm{m}} = \sum_i \beta(i)\pi_i$, where $\beta(i) = \{bandwidth\ of\ state\ i\}$.

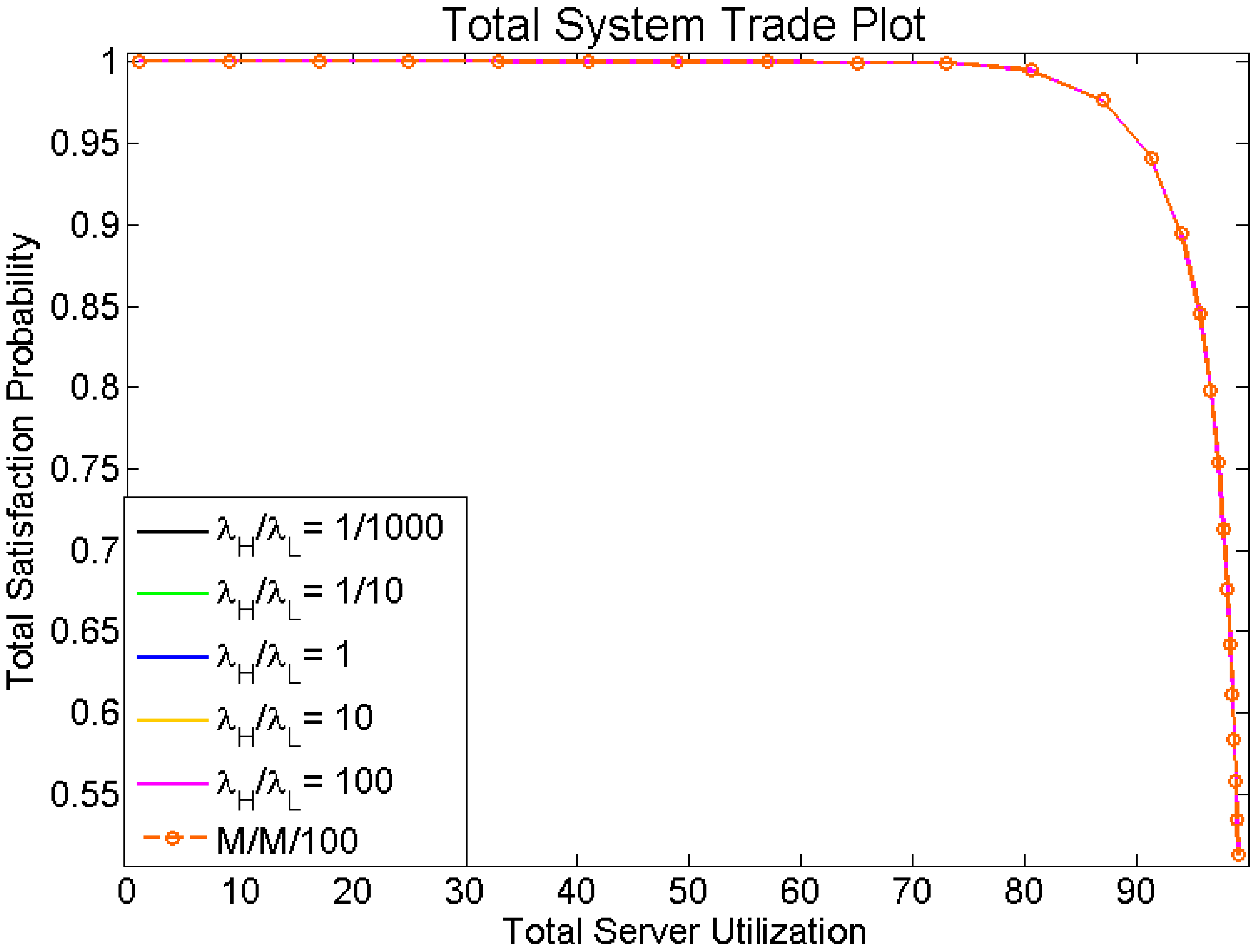


Higher bandwidth resources result in a better tradeoff.

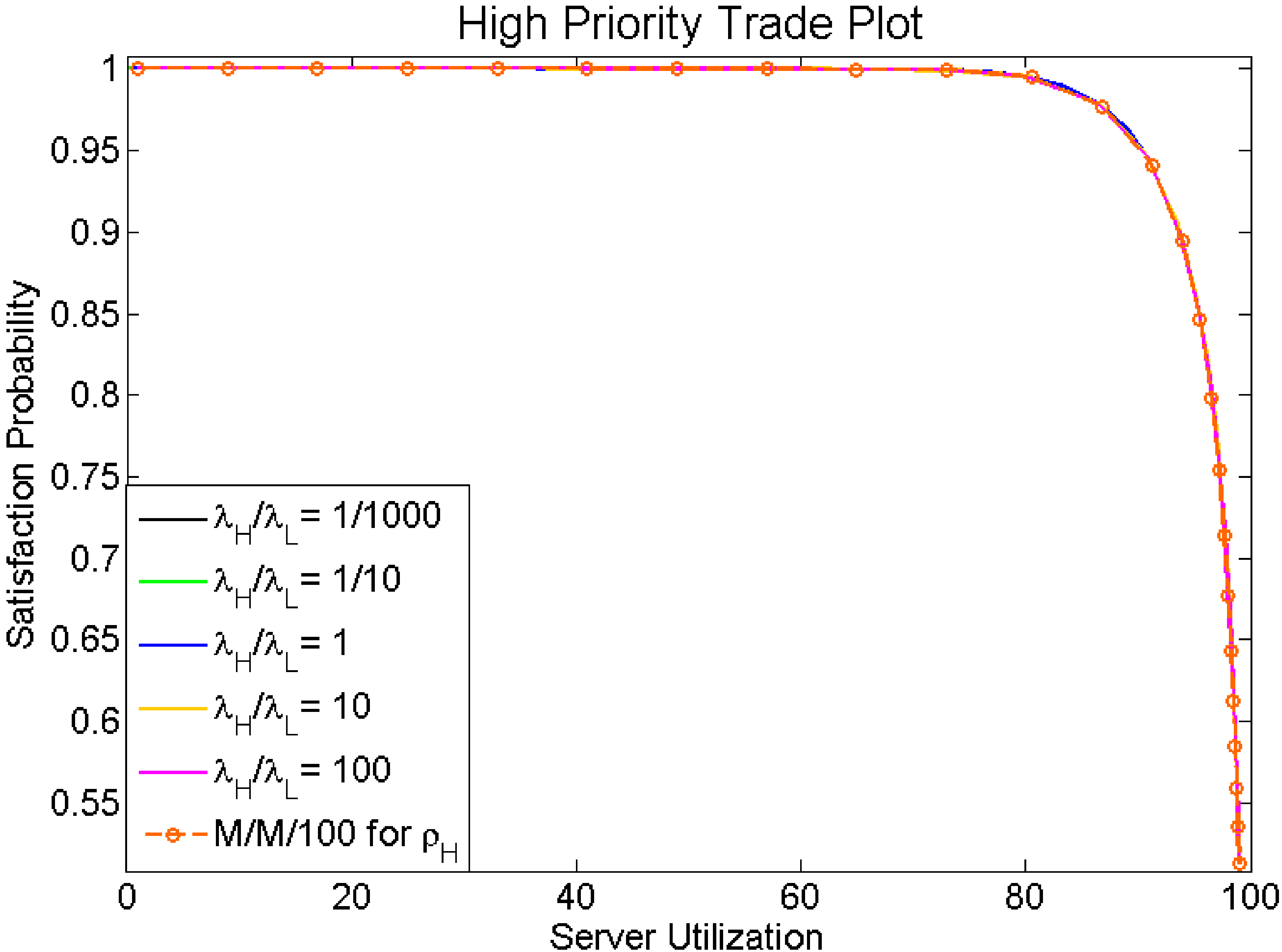
3. Priority Model

Background

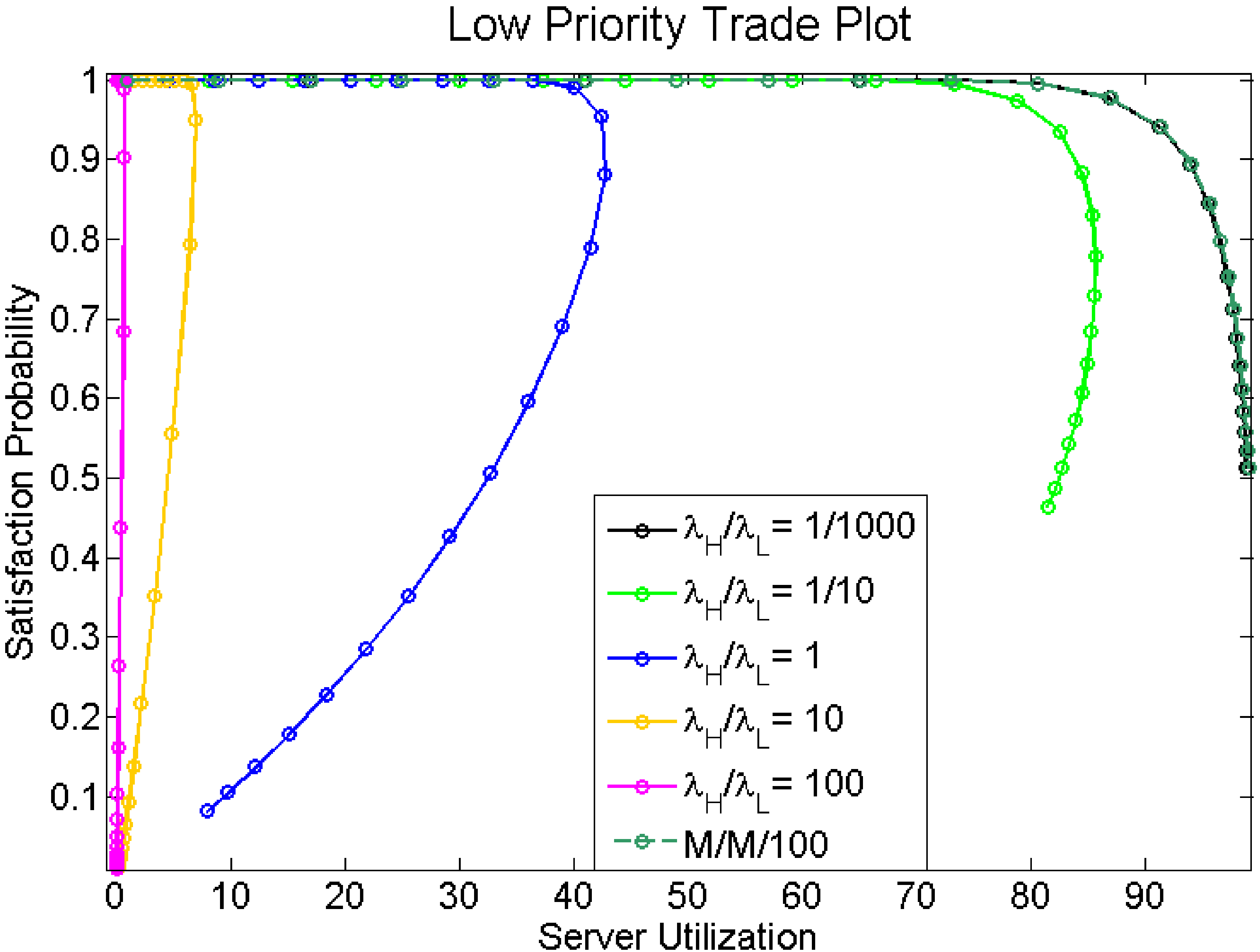
- Total Traffic Intensity: ρ_{total}
  - $\rho_{total} = \rho_H + \rho_L = \frac{\lambda_H + \lambda_L}{m\,\mu}$
    - $\rho_H = \frac{\lambda_H}{m\,\mu}$
    - $\rho_L = \frac{\lambda_L}{m\,\mu}$
- Our Model
  - $\mu = 1$, $m = 100$
  - *Ratios of arrivals,* $\frac{\lambda_H}{\lambda_L}$*, varied from 1/1000 to 100*
  - *Results compared to M/M/100 model*



Prioritization does not cause the system on a whole to deviate from the non-prioritized M/M/100.



High priority jobs act like the non-prioritized M/M/100 in satisfaction and server utilization.



Higher ratios of λ_H to λ_L cause low priority jobs to be preempted more often and deviate from the non-prioritized M/M/100.

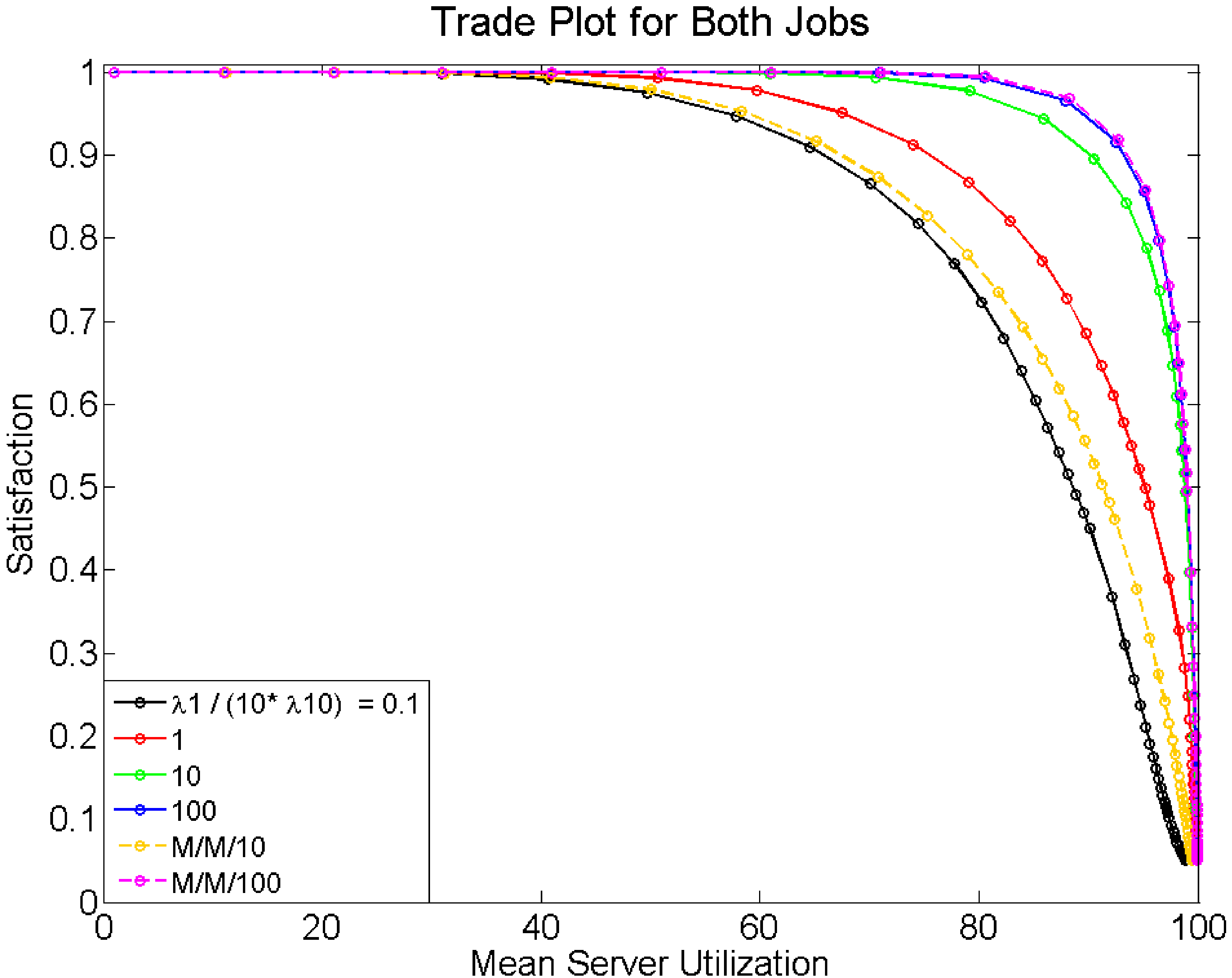
4. Bandwidth Model

Background

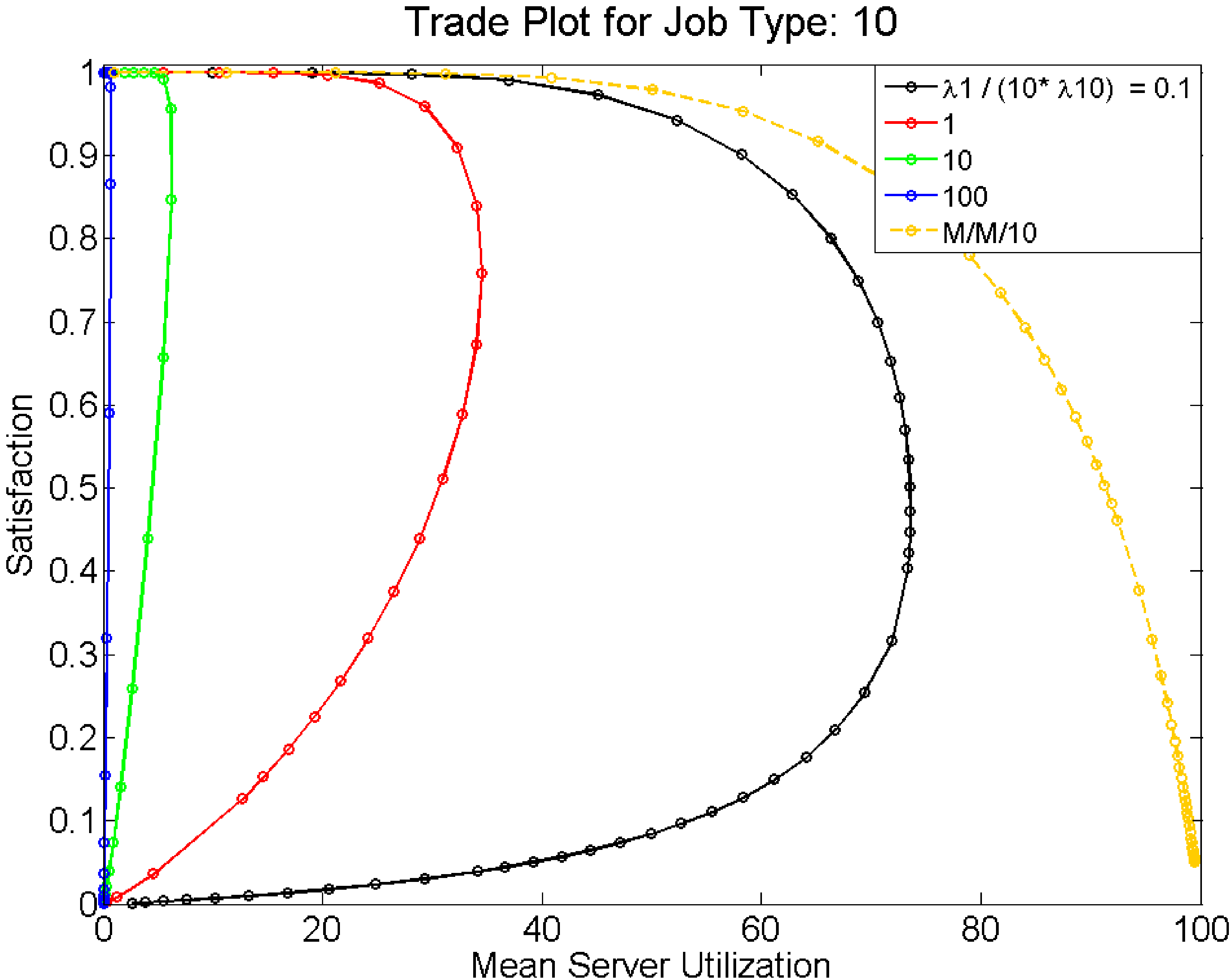
- Total Traffic Intensity, ρ_{total}
  - $\rho_{total} = \frac{\sum \lambda_i \, B_i}{m \, \mu}$
    - λ_i = Arrival rate of jobs of class i
    - B_i = Bandwidth requested by jobs of class i
- Our Model
  - *$\mu = 1$, $m = 100$*
  - *Two competing job classes. Large bandwidth jobs require 10 servers, and small bandwidth jobs require 1 server.*

Background

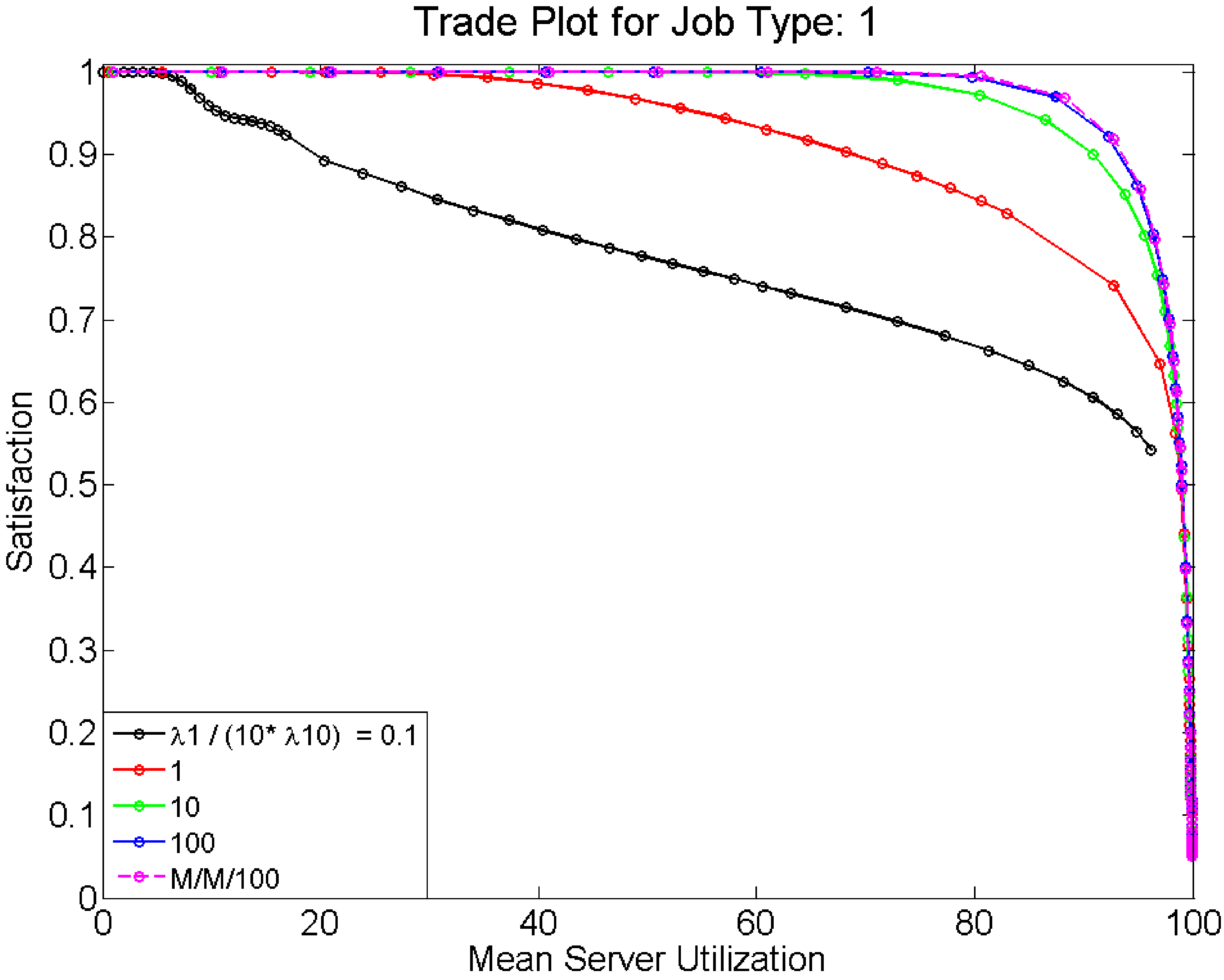
- Ratio of arrivals is weighted by bandwidth: $\frac{\lambda_1}{10\ \lambda_{10}}$
- Ratio was varied from 1/10 to 100.
- User satisfaction and resource utilization were measured.
- Results were compared to classical models.
  - *All traffic is type large bandwidth: M/M/10*
  - *All traffic is type small bandwidth: M/M/100*



Certain ratios are less effective than a pure M/M/10 system.



Across all ratios, big bandwidth jobs have decreased satisfaction as ρ increases.

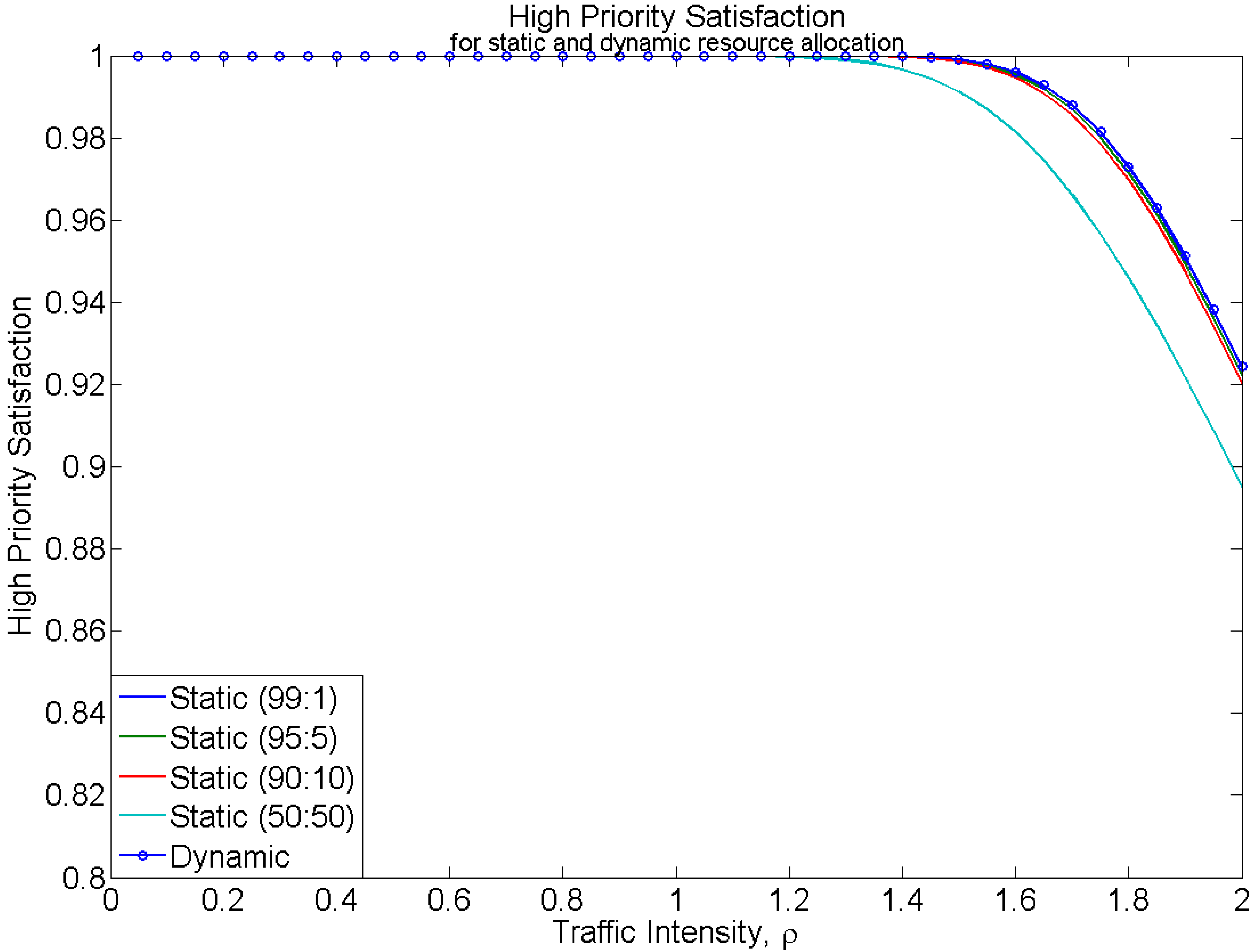


Small bandwidth jobs tend to have high satisfaction and are better at filling up the system.

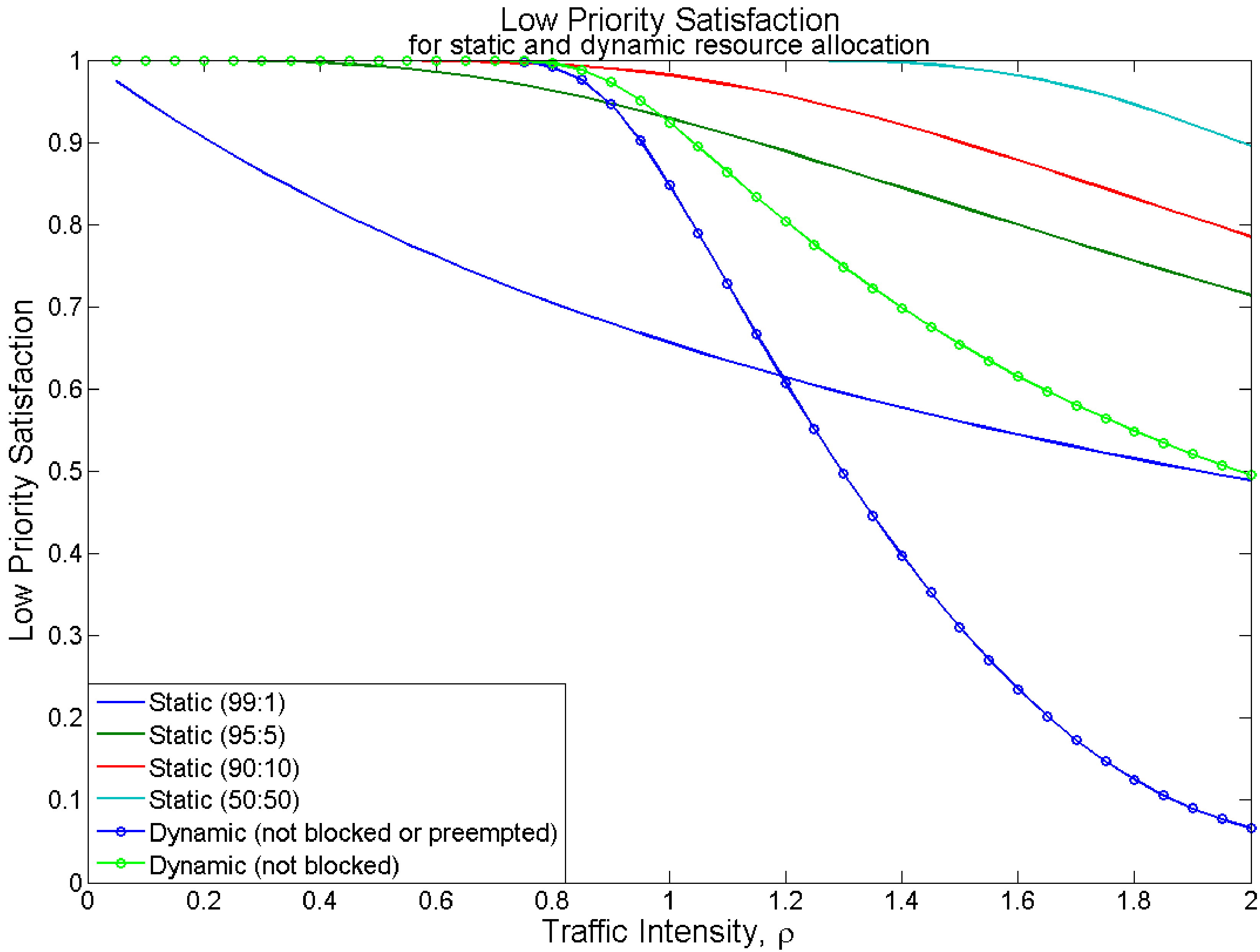
5. Static vs. Dynamic Allocation Schemes

Competing Priority Classes

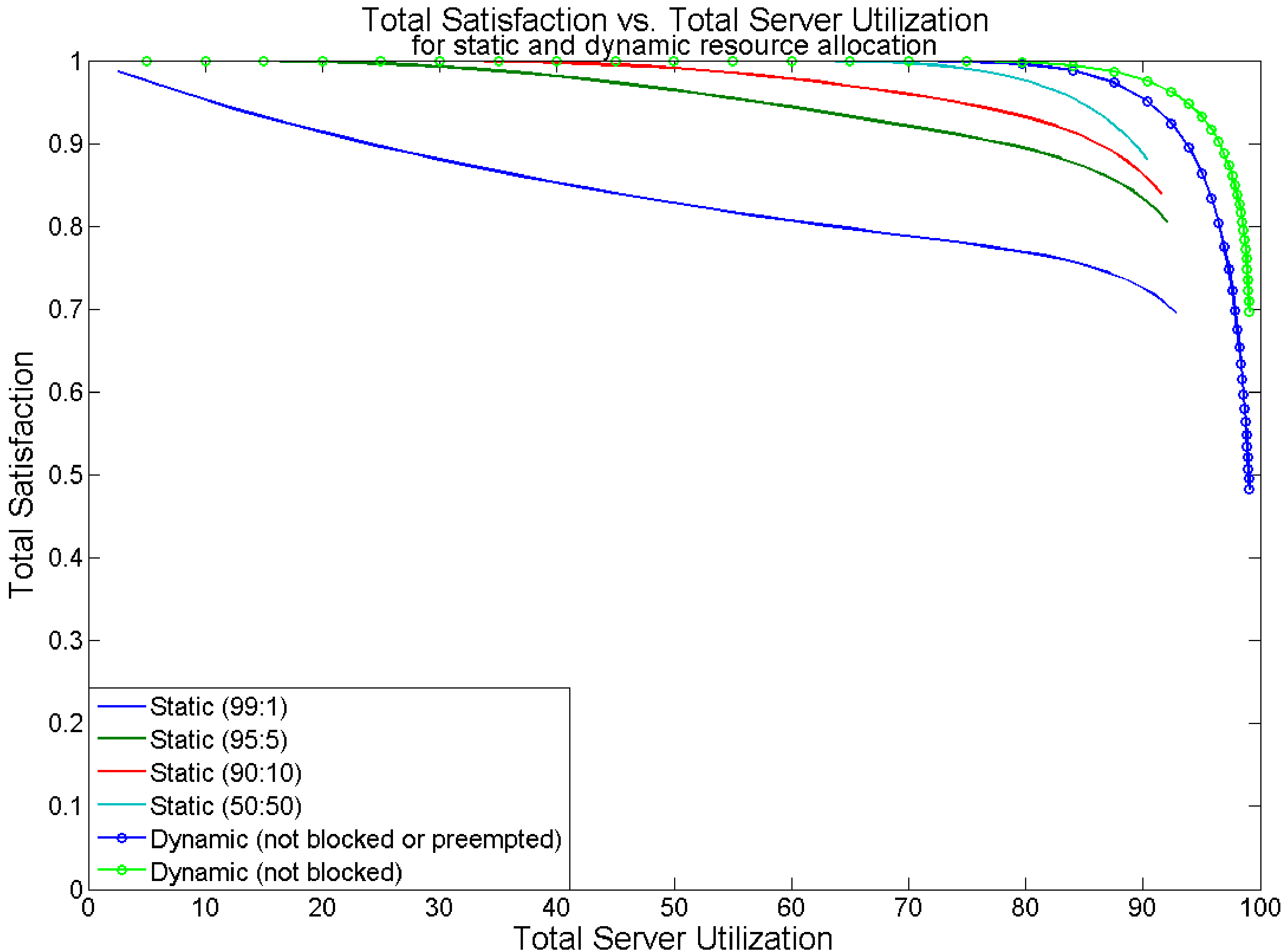
- Server capacity $m = 100$
- Fix $\lambda_H = \lambda_L$
- Static Allocation Approach
  - *Each priority class pre-allocated a number of servers (high:low)*
    - 99:1, 95:5, 90:10, 50:50
  - *M/M/m with m = to those number of servers*
- Dynamic Allocation Approach
  - *Dynamic allocation model for competing priority classes*
- Measures
  - *User satisfaction*
  - *Resource utilization*



Dynamic allocation provides the best high priority satisfaction.



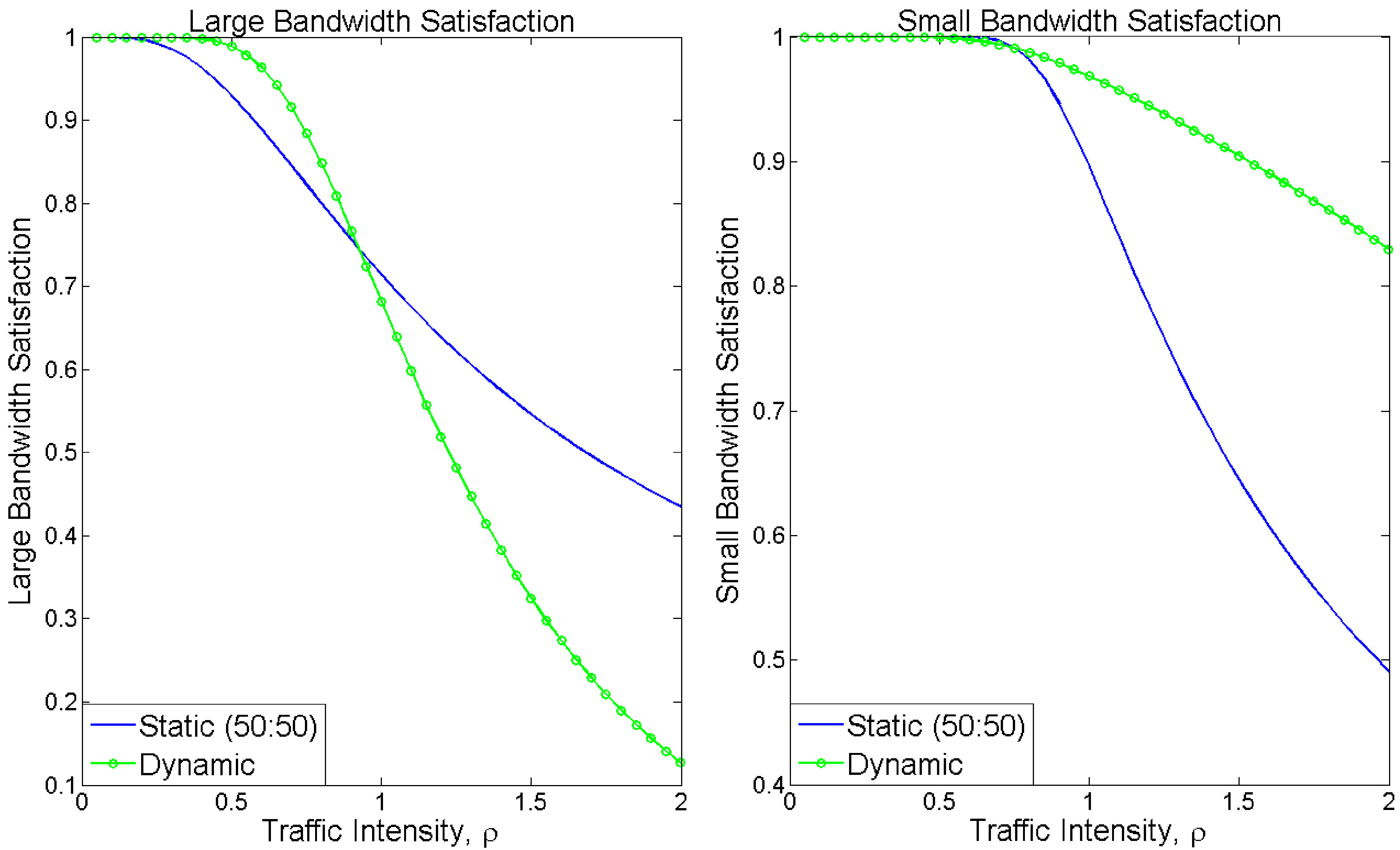
Though better at lower traffic intensities, dynamic allocation allows high priority requests to outcompete low priority, leading to low satisfaction for low priority.



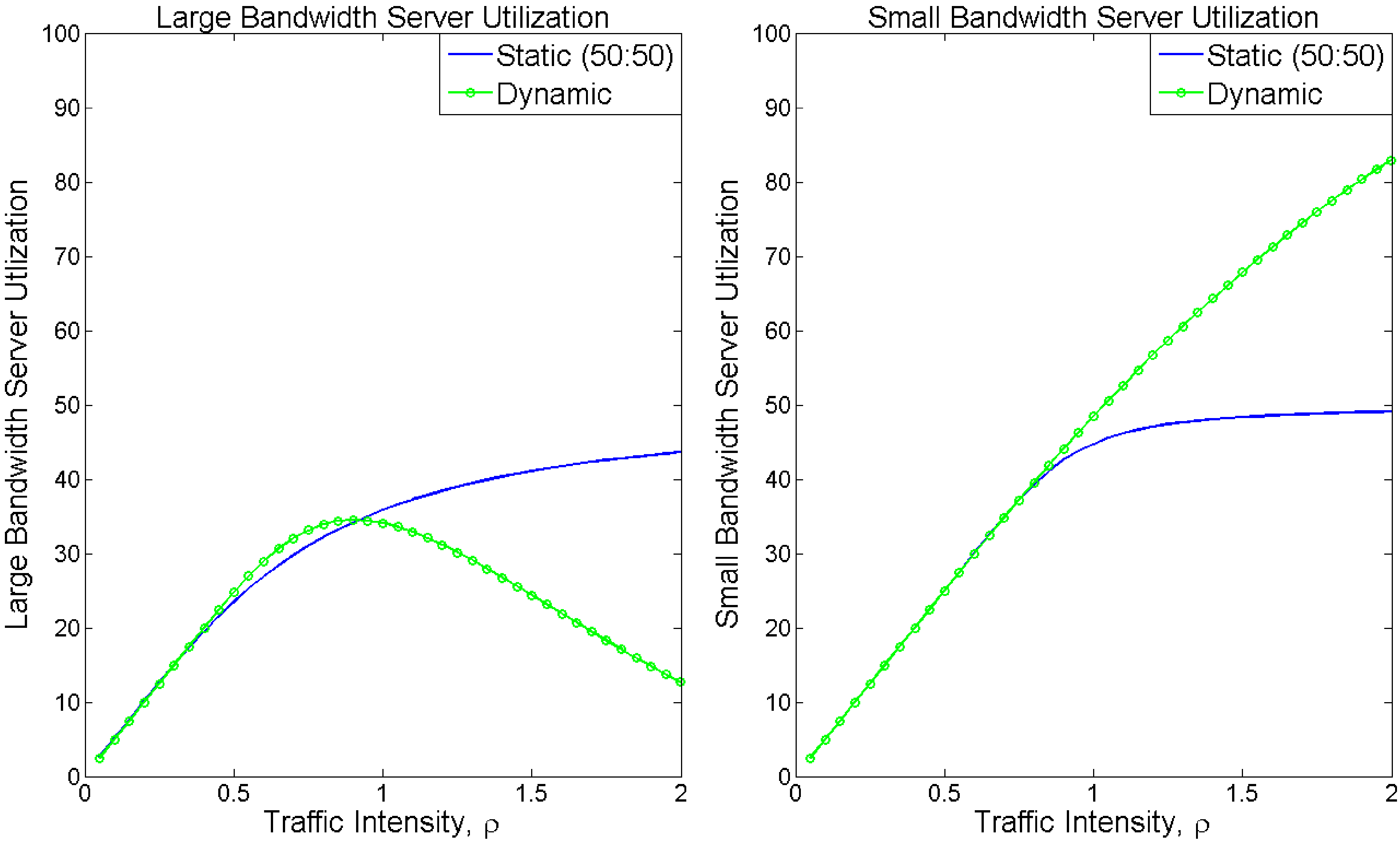
For a given utilization value, dynamic allocation provides better satisfaction and creates a more efficient system.

Competing Bandwidth Classes

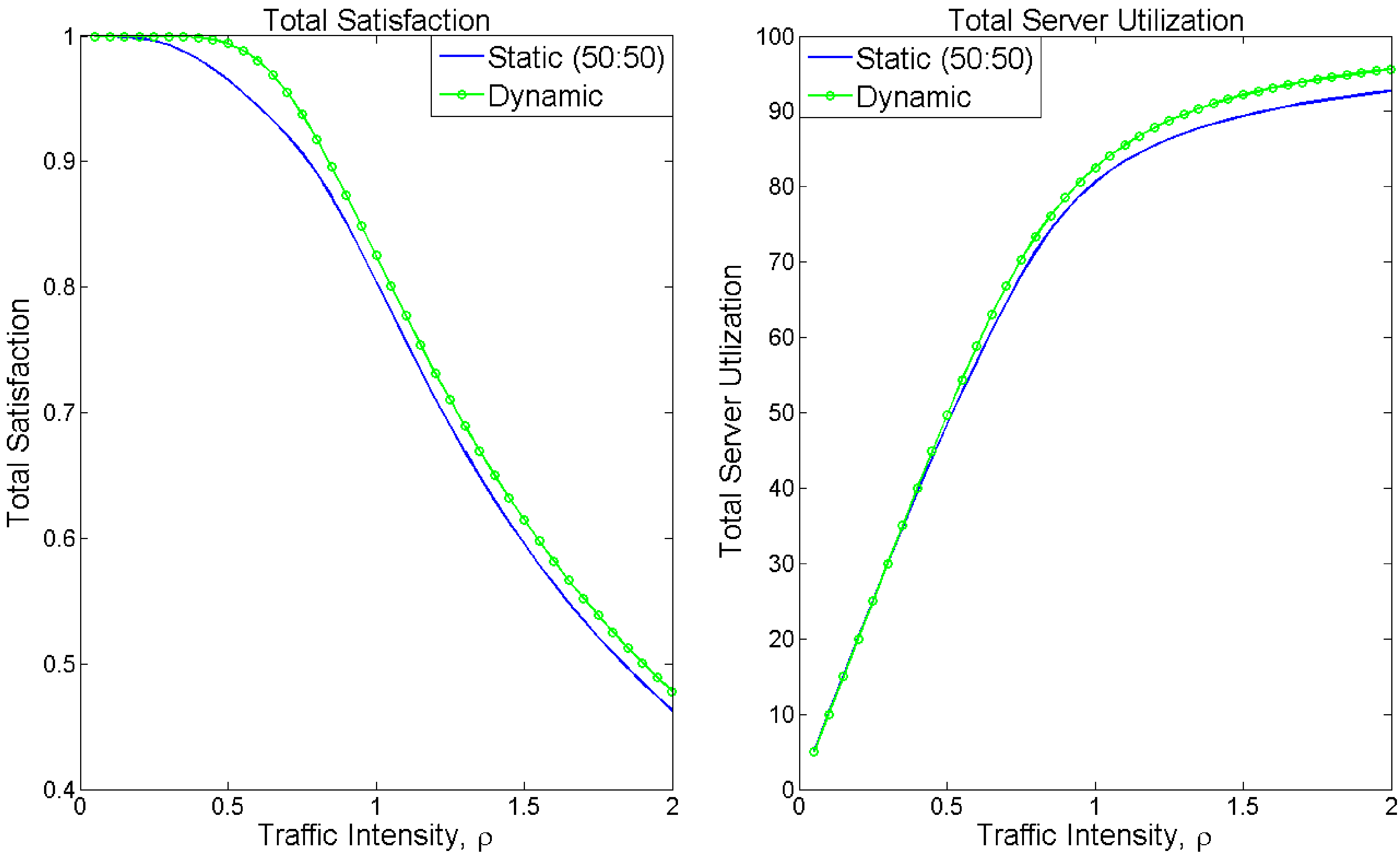
- Server capacity $m = 100$
- Competing bandwidth classes: 1 and 10
- Fix λ_1 = $10\ \lambda_{10}$
- Static Allocation Approach
  - *Each bandwidth class allotted ½ of server capacity*
  - *M/M/50 for each class*
- Dynamic Allocation Approach
  - *Dynamic allocation model for competing bandwidth classes*
- Measures
  - *User satisfaction*
  - *Resource utilization*



While large bandwidth sees decreased satisfaction for high traffic intensities, small bandwidth jobs have better satisfaction under dynamic allocation.



Under dynamic allocation, small bandwidth jobs outcompete large bandwidth jobs and seize more resources.



Overall, dynamic allocation schemes provides higher satisfaction and more efficient use of resources.

6. Conclusions and Future Work

Conclusions

- Single-type Traffic
  - *Higher bandwidth resources result in a better tradeoff.*
- Priority
  - *High priority users enter the system as if low priority users do not exist.*
  - *Low priority satisfaction is dependent on the arrival rate of high priority users.*
- Bandwidth
  - *Small bandwidth jobs always experience great user satisfaction, while big jobs may experience low satisfaction for high traffic intensity systems.*
- Comparison of Dynamic and Static Allocation Methods
  - *Though results mixed by class, overall more efficient systems and better service under dynamic allocation than under static*

Future Work

- Allow for more than two competing priority and bandwidth classes
- Consider reentry procedures for preempted jobs
- Consider varying arrival and service distributions
- Develop approximation algorithms to reduce computational cost

Acknowledgements

- We would like to thank:
  - *Nam Lee*
  - *Industry mentors from Aerospace*
  - *Amber Puha*