Are utility, price and satisfaction based resource allocation models suitable for large-scale distributed systems?

Xin Bai, Ladislau Bölöni, Dan C. Marinescu
School of Electrical Engineering and Computer Science
University of Central Florida

Howard Jay Siegel
Department of Electrical and Computer Engineering
and Department of Computer Science
Colorado State University

Rose A. Daley and I-Jeng Wang
Applied Physics Laboratory
Johns Hopkins University

Introduction

- Resource management in large scale distributed systems is challenging
  - Scale
  - Selfishness of providers and consumers
  - Different objectives of different providers
  - Societal objectives needs to be accomplished as well
- Direct negotiation between consumers and providers is difficult if n > 100.
- Agents which mediate between consumers and providers:
  - Brokers, matchmakers, market makers etc.
  - We express our valuation of the resources in monetary terms (usually virtual money)
  - But how do we calculate the valuation?

Middle agents

- Matchmakers:
  - Find providers who can potentially satisfy the resource requirements
  - Negotiation is one-by-one
- Broker
  - Mediates the negotiation
  - Potentially simultaneous negotiations with multiple providers
- Market maker
  - Improves the resource allocation by acquiring and reselling resources
  - It is the only player interested in efficiency – it derives its profits from efficiency improvements

Price, utility, satisfaction

- Price function p(r)
  - Determined by the provider
- Utility u(r)
  - The utility of a set of resources r for a user
  - Application dependent
  - Always a monotonic function: more resources are always more useful.
  - Generally, there is a level after which more resources give less and less utility.
- Consumer satisfaction s(r)
  - Takes into account the price paid for the resource.
  - It is not monotonic, it will exhibit one or more local maximums.

In this paper:

- We explore the implications of a resource allocation model which relies on price, utility and satisfaction functions.
- Some potential objectives:
  - Achieve maximum societal benefits
  - Achieve best benefits for the given provider or consumer
    - E.g. complex negotiation, without disclosing your true valuation…
- Potential things to consider:
  - Variety of function shapes
  - Every provider and consumer can come with their own function
  - Providers can play favorites: use different pricing models for different customers.
  - Co-allocation: make “packaged deals”
- This paper assumes that all participants use the same functions
- Our main goal is societal objectives.

Utility function

- Resource vector $r_{ij} = (r_{ij}^1\ r_{ij}^2 \ldots r_{ij}^l)$
- Utility function for consumer i, for resource type k is a sigmoid

$$u_{ij}^k = u(r_{ij}^k) = \frac{(r_{ij}^k/\omega_i^k)^{\zeta_i^k}}{1 + (r_{ij}^k/\omega_i^k)^{\zeta_i^k}}$$

- Overall utility can be:
  - The product of individual utilities
  - The weighted sum of individual utilities

Price and satisfaction

- We consider a linear pricing function

$$p_{ij} = \sum_{k=1}^{l} p_{ij}^k \times t_{ij}$$

- We consider a satisfaction function

$$s_{ij}^k(u_{ij}^k, p_{ij}^k) = 1 - e^{-\kappa_i^k {u_{ij}^k}^{\mu_i^k} (p_{ij}^k/\phi_i^k)^{-\epsilon_i^k}}, \qquad \kappa_i^k, \phi_i^k, \mu_i^k, \epsilon_i^k > 0$$

- ϵ_i^k and μ_i^k control the sensitivity of the satisfaction function to utility and price
- ϕ_i^k and κ_i^k are normalization constants

Composition of the satisfaction function

- Different ways to compose the satisfaction
  - A product over the set of resources – if one of the resources does not provide any satisfaction, the result will be zero as well.
  - A weighted sum over the set of resources – ability to set more or less resources.
  - More complex functions?

Consumer – broker – provider model

- In this model, the amount of resources to be allocated is determined according to a target utility τ, i.e., the broker allocates an amount of resources such that the utility of each type of resource to the consumer reaches this τ value.
- The broker also has "societal goals" and attempts to maximize the average utility and revenue, as opposed to providers and consumers that have individualistic goals.
- To reconcile the requirements of a consumer and the candidate providers, a broker chooses a subset of providers such that the satisfaction is above a threshold and all providers in the subset have equal chances to be chosen by the consumer. We call the size of this subset satisficing size σ.

Measured parameters

- The average hourly revenue for providers. The revenue is the sum of revenues for all of its resource types. This average is over the set of all providers connected to broker *B*.
- The consumer admission ratio. This ratio is the number of admitted consumers over the number of all consumers connected to *B*. A consumer is admitted into the system when there is a provider able to allocate some of the resources requested by the consumer, otherwise the consumer is dropped.
- The average consumer overall utility. This average is over the set of all admitted consumers connected to broker *B*.
- The average consumer overall satisfaction. This average is over the set of all admitted consumers connected to broker *B*.

Simulation results

- 100 clusters
- 1 broker
- number of nodes in the clusters: is a random variable normally distributed with the mean of 50 and the standard deviation of 30.
- resource vector: the CPU rate, the amount of main memory, and the disk capacity
  - For example, the resource vector for a node with one 2 GHz CPU, 1 GB of memory, and a 40 GB disk is (2*GHz;* 1*GB;* 40*GB*).
- Initially, there is no consumer in the system. Consumers arrive with an inter-arrival time exponentially distributed with the mean of 2 seconds.
- Demand / capacity ratio η

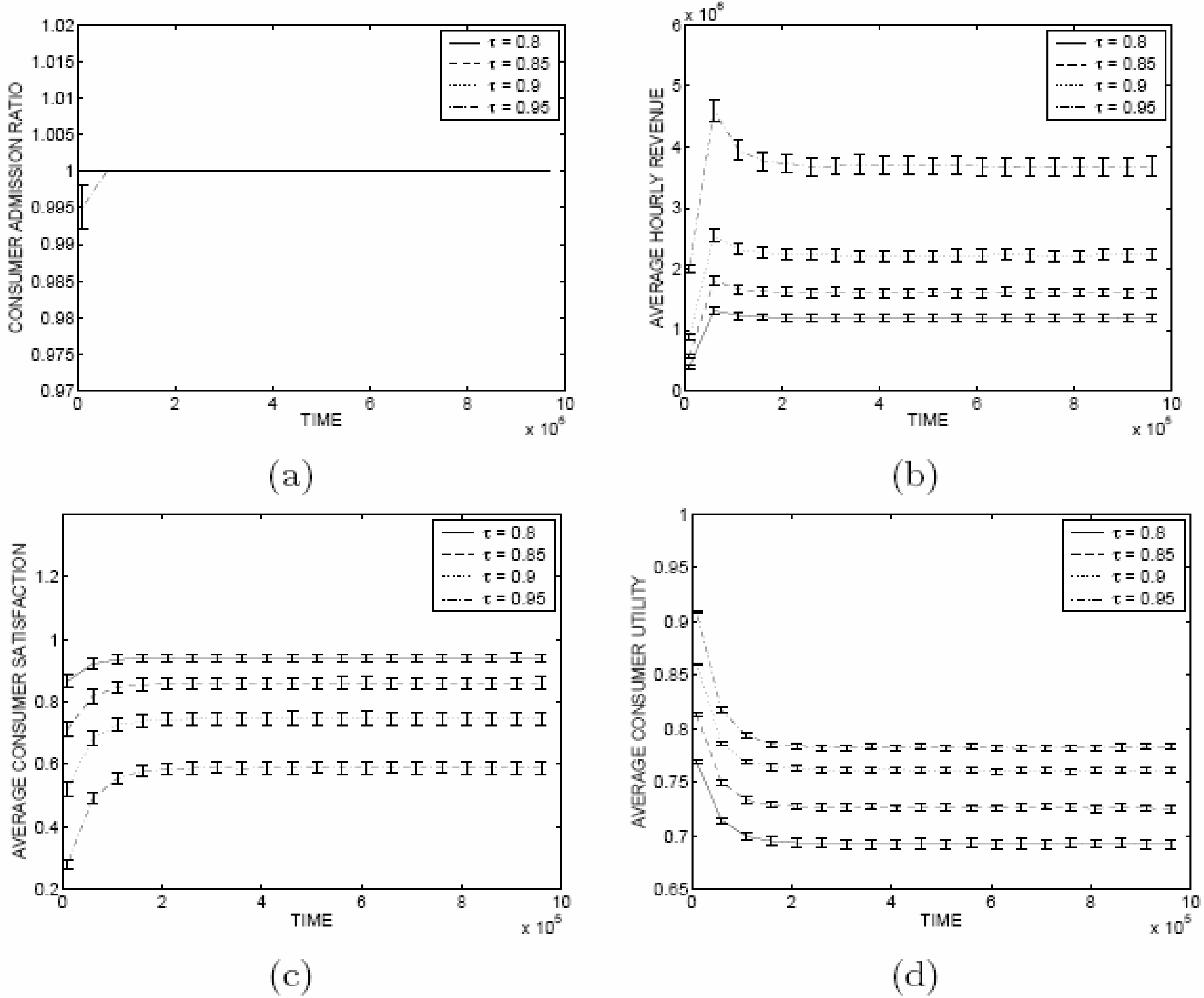

Figure 1. Consumer admission ratio (a), average hourly revenue (b), average consumer satisfaction (c), and average consumer utility (d) vs. time (in seconds) for $\sigma = 1$, $\eta = 1.0$, and $\tau = 0.8$, 0.85, 0.9, and 0.95.

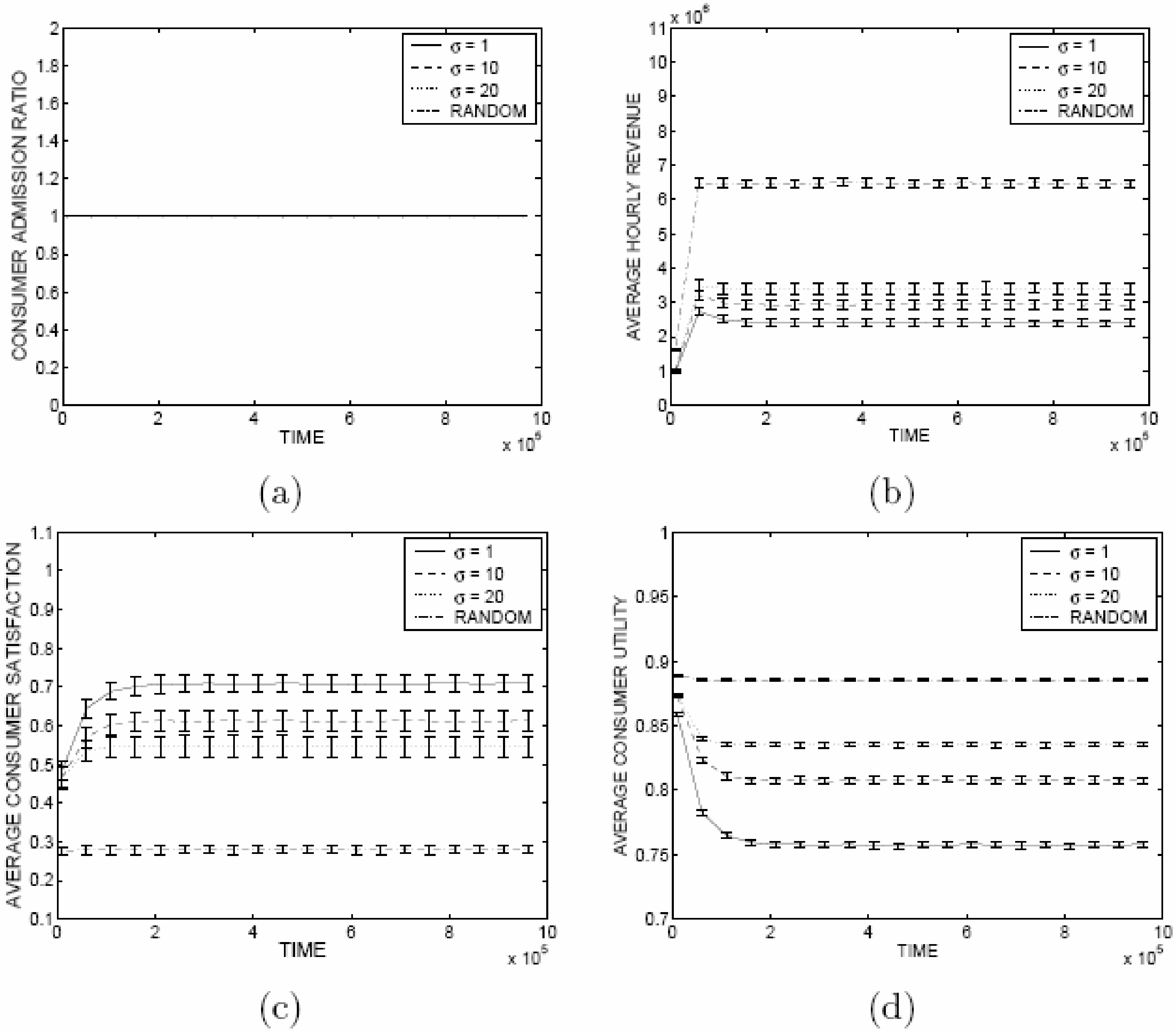

Figure 2. Consumer admission ratio (a), average hourly revenue (b), average consumer satisfaction (c), and average consumer utility (d) vs. time (in seconds) for $\tau = 0.9$, $\eta = 0.5$, and $\sigma = 1$, 10, and 20. For the random strategy, $\sigma = |\mathcal{R}| = 50$.

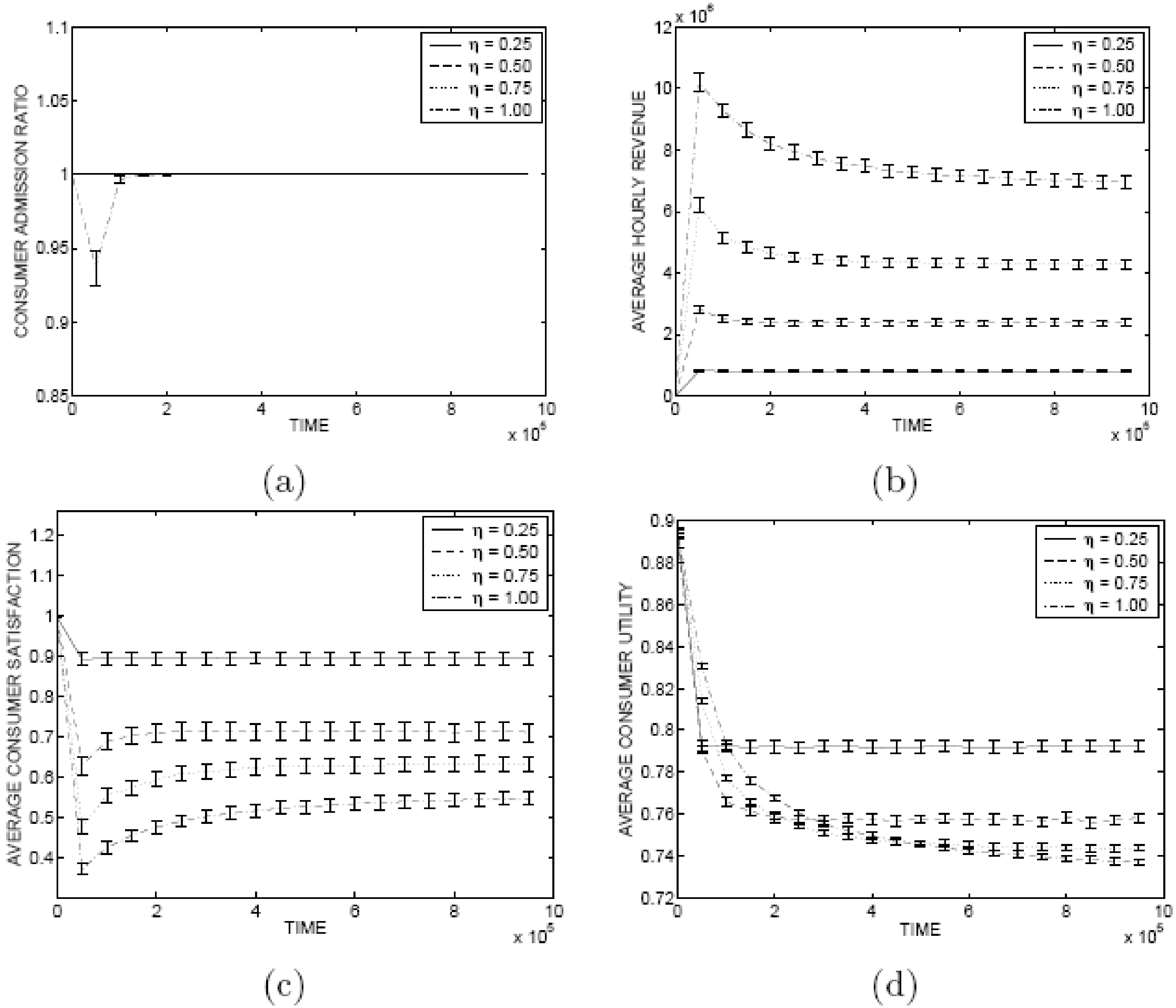

Figure 3. Consumer admission ratio (a), average hourly revenue (b), average consumer satisfaction (c), and average consumer utility (d) vs. time (in seconds) for $\tau = 0.9$ and $\sigma = 1$.

How we utilize the results in practice?

- The function of a broker is to monitor the system and set σ and τ for optimal performance.
- For example, if the broker perceives that the average consumer utility is too low, it has two choices: increase σ or increase τ. At the same time, the system experiences an increase of average hourly revenue and a decrease of average consumer satisfaction.
- We note that while the utility is always increasing with the amount of allocated resources, the satisfaction also takes into account the price paid and exhibits an optimum at a certain level of resources. Increasing the resources beyond the optimum will still increase the utility but yield lower satisfaction, because the additional utility was paid an unjustifably high price.

Conclusions

- The complexity of the utility, price, and satisfaction based models precludes analytical studies and in this paper we report on a simulation study.
- The goal of our simulation study is to validate our choice of utility, price, and satisfaction functions, to study the effect of the many parameters that characterize our model, and to get some feeling regarding the transient and the steady-state behavior of our models. We are primarily interested in qualitative rather than quantitative results, and we are interested in trends rather than actual numbers.
- It is too early to compare our model with other economic models proposed for resource allocation in distributed systems, but we are confident that a model that formalizes the selfish goals of consumers and providers, as well as societal goals, has a significant potential.

Acknowledgements

- This research was supported in part by National Science Foundation grants MCB9527131, DBI0296035, ACI0296035, and EIA0296179, the Colorado State University George T. Abell Endowment, and the DARPA Information Exploitation Office under contract No. NBCHC030137.

Thank You

- Questions?