Hydrodynamics limits for Randomized Load Balancing

Joint work with Kavita Ramanan

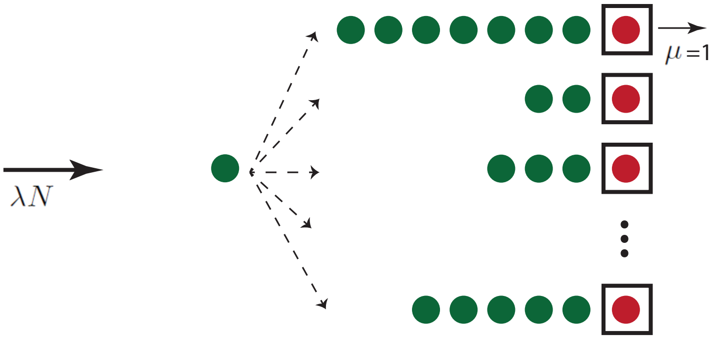
Brown University

November 2014

Model of Interest

Network with

- N servers
- an infinite capacity queue for each server
- a common arrival process
- FCFS service discipline within each queue (no processor sharing)

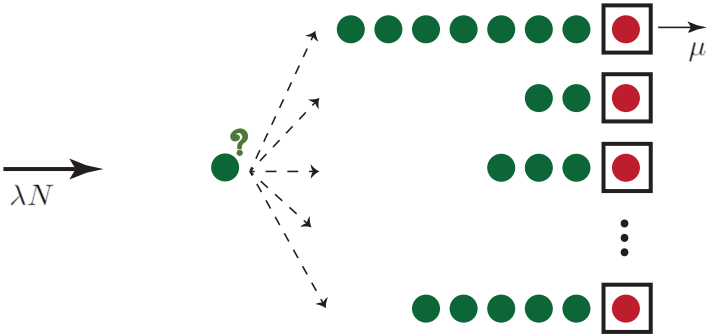


Model of Interest

Load Balancing Algorithm:

- How to assign incoming jobs to servers?
- Aim to achieve good performance with low computational cost

Goal: Analysis and comparison of different load balancing algorithms



Model of Interest

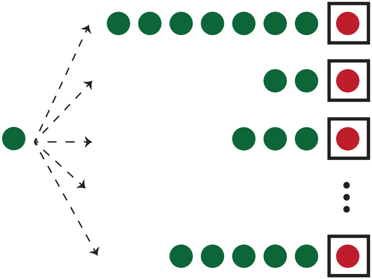
Appears in:

- Supermarkets
- Hash tables
- Distributed memory machines
- Path selection in networks
- Web Servers
- etc.

Routing Algorithm: Supermarket Model

Each arriving job

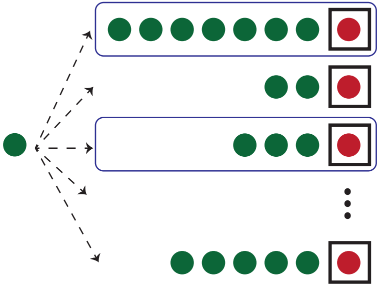
- chooses d queues out of N, uniformly at random,
- joins the shortest queue among the chosen d.
- ties broken uniformly at random.



Routing Algorithm: Supermarket Model

Each arriving job

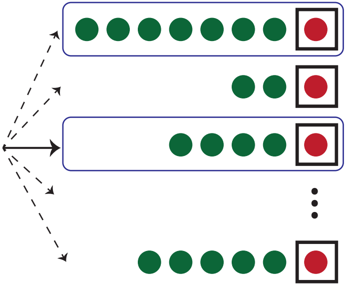
- chooses d queues out of N, uniformly at random,
- joins the shortest queue among the chosen d.
- ties broken uniformly at random.



Routing Algorithm: Supermarket Model

Each arriving job

- chooses d queues out of N, uniformly at random,
- joins the shortest queue among the chosen d.
- ties broken uniformly at random.



Prior Work - Exponential Service Distribution

Supermarket model for exponential service time

Fluid limit and steady state queue length decay rate is obtained

- case $d = 2$, [Vvedenskaya-Dobrushin-Karpelevich '96]
- case $d \geq 2$, [Mitzenmacher '01]

Prior Work - Exponential Service Distribution

Supermarket model for exponential service time

Fluid limit and steady state queue length decay rate is obtained

- case $d = 2$, [Vvedenskaya-Dobrushin-Karpelevich '96]
- case $d \geq 2$, [Mitzenmacher '01]

General Approach

Using Markovian state descriptor $\{S_\ell^N(t); \ell \geq 1, t \geq 0\}$

- $S_\ell^N(t)$: fraction of stations with at least ℓ jobs
- Convergence as $N \to \infty$ proved using an extension of Kurtz's theorem
- The limit process is a solution to a sequence of ODEs
- Steady state queue length distribution is obtained by the fixed point of the ODE sequence

Prior Work - Exponential Service Distribution

Summary of Results

- Joint the Shortest Queue (JSQ)
  - Performance: $P(X^N(\infty) > \ell) \to 0$ for $\ell \geq 1$
  - Computational Cost: N comparison per routing (**not feasible**)

Prior Work - Exponential Service Distribution

Summary of Results

- Joint the Shortest Queue (JSQ)
  - Performance: $P(X^N(\infty) > \ell) \to 0$ for $\ell \geq 1$
  - Computational Cost: N comparison per routing (**not feasible**)

- $d = 1$ (random routing, decoupled $M/M/1$ queues):
  - Performance: $P(X^N(\infty) > \ell) \to c\lambda^\ell$
  - Computational cost: one random flip per routing

Prior Work - Exponential Service Distribution

Summary of Results

- Joint the Shortest Queue (JSQ)
  - Performance: $P(X^N(\infty) > \ell) \to 0$ for $\ell \geq 1$
  - Computational Cost: N comparison per routing (**not feasible**)
- $d \geq 2$ (supermarket model):
  - Performance: $P(X^N(\infty) > \ell) \to \lambda^{(d^\ell - 1)/(d-1)}$
  - Computational Cost: d random flips and $d-1$ comparison per routing
- $d = 1$ (random routing, decoupled $M/M/1$ queues):
  - Performance: $P(X^N(\infty) > \ell) \to c\lambda^\ell$
  - Computational cost: one random flip per routing

Prior Work - Exponential Service Distribution

Summary of Results

- Joint the Shortest Queue (JSQ)
  - Performance: $P(X^N(\infty) > \ell) \to 0$ for $\ell \geq 1$
  - Computational Cost: N comparison per routing (**not feasible**)
- $d \geq 2$ (supermarket model):
  - Performance: $P(X^N(\infty) > \ell) \to \lambda^{(d^\ell - 1)/(d-1)}$
  - Computational Cost: d random flips and $d-1$ comparison per routing
- $d = 1$ (random routing, decoupled $M/M/1$ queues):
  - Performance: $P(X^N(\infty) > \ell) \to c\lambda^\ell$
  - Computational cost: one random flip per routing

Power of two Choices: double-exponential decay for $d \geq 2$

Prior Work -General Service Distribution

Our Focus: General service time distribution

- almost nothing was known 5 years ago
- Mathematical Challenge:
    - $\{S_\ell^N\}$ is no longer Markovian
    - need to keep track of more information
    - No finite dimensional common state space for Markovian Representations

Prior Work -General Service Distribution

Recent Progress

- Stability of pre-limit systems [Foss-Chernova'98]
- Tightness of stationary distributions sequence [Bramson'10]
- Stationary queue length decay [Bramson-Lu-Prabhakar'13]

Prior Work -General Service Distribution

Recent Progress

- Stability of pre-limit systems [Foss-Chernova'98]
- Tightness of stationary distributions sequence [Bramson'10]
- Stationary queue length decay [Bramson-Lu-Prabhakar'13]

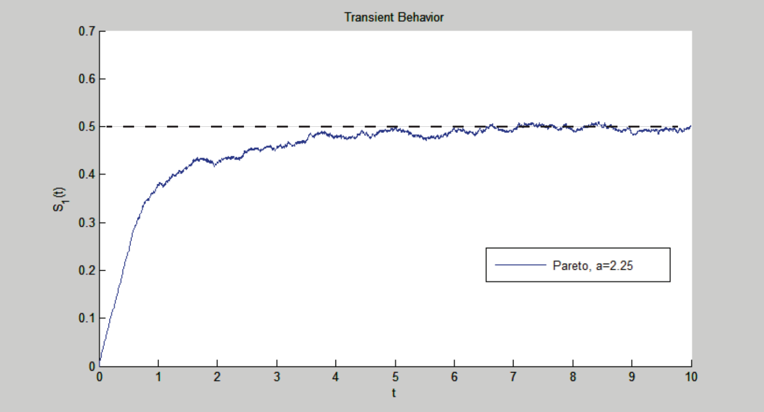
Approach Taken in [Bramson-Lu-Prabhakar'13]:
Cavity Method

- only proved for service distribution with decreasing hazard rate
- assumes Poisson arrival (uses Poisson splitting)
- only applicable for steady-state distribution
- Pro: can also be easily applied to processor sharing

Transient Behavior - Simulation

Simulation results for *fraction of busy servery**

- Poisson arrival with $\lambda = 0.5$
- 1000 servers
- empty initial condition

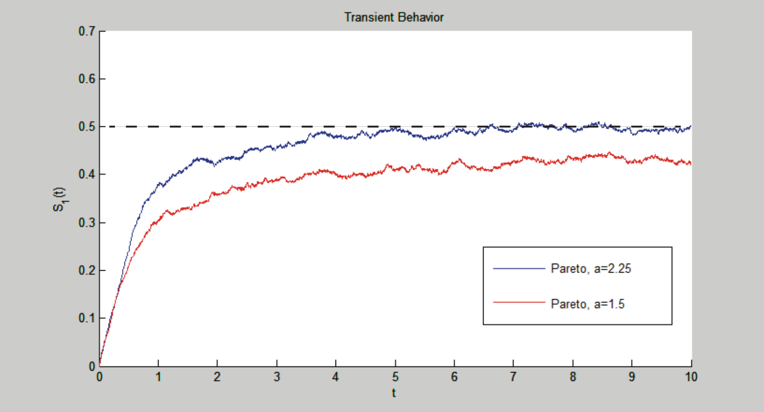


*Simulation results by Xingjie Li, Brown University

Transient Behavior - Simulation

Simulation results for *fraction of busy servery**

- Poisson arrival with $\lambda = 0.5$
- 1000 servers
- empty initial condition

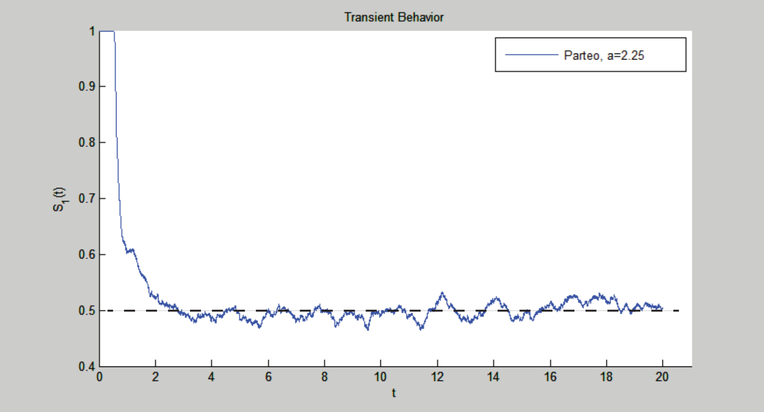


*Simulation results by Xingjie Li, Brown University

Transient Behavior - Simulation

Simulation results for *fraction of busy servery*†

- Poisson arrival with $\lambda = 0.5$
- 1000 servers
- initially one job in each server

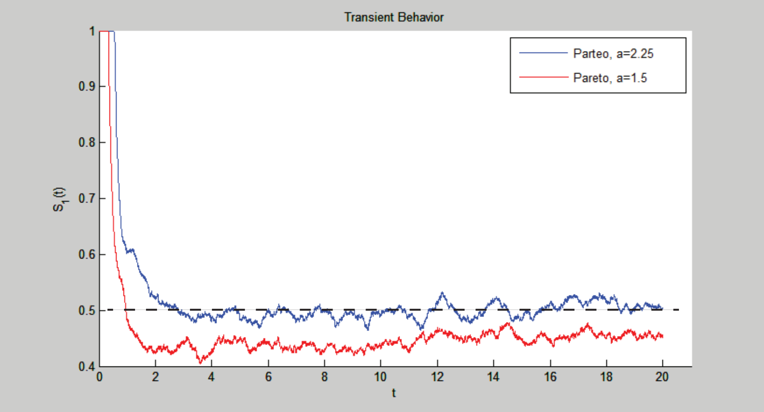


†Simulation results by Xingjie Li, Brown University

Transient Behavior - Simulation

Simulation results for *fraction of busy servery*†

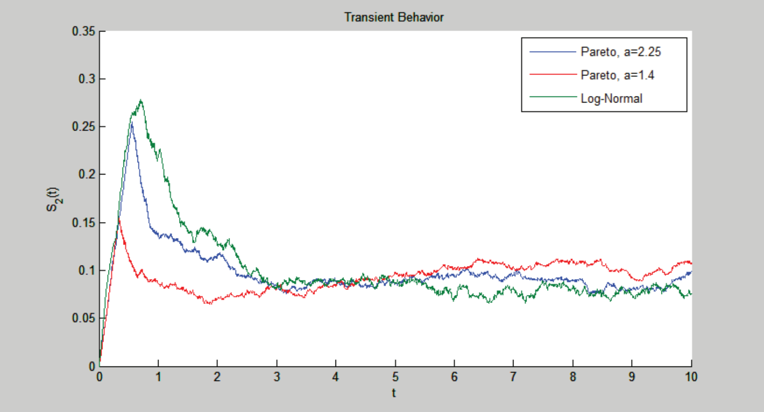
- Poisson arrival with $\lambda = 0.5$
- 1000 servers
- initially one job in each server



†Simulation results by Xingjie Li, Brown University

Simulation results for *fraction of queues with queue length at least* 2[‡]

- Poisson arrival with $\lambda = 0.5$
- 1000 servers
- initially one job in each server



[‡]Simulation results by Xingjie Li, Brown University

Our Goal

Observations:

- No result on the time scale to reach equilibrium
- Transient behavior is also important
- No result on distributions without decreasing hazard rate

Our Goal

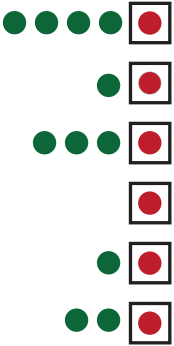
Observations:

- No result on the time scale to reach equilibrium
- Transient behavior is also important
- No result on distributions without decreasing hazard rate

Our Goal:

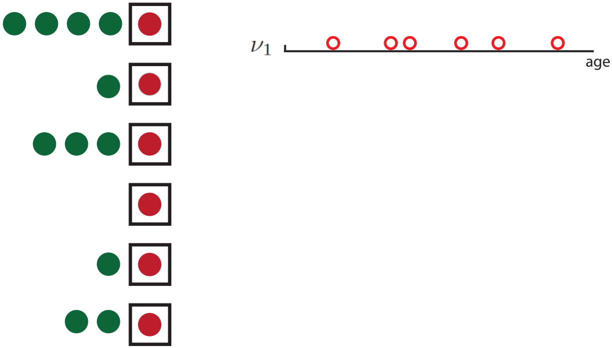
Introduce a new approach: Interacting Measure-valued Processes

ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .



Interacting Measure-Valued Processes Representation

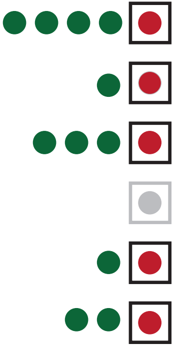
ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .

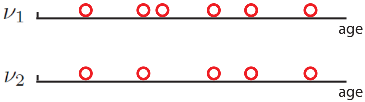


at least one jobs

Interacting Measure-Valued Processes Representation

ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .

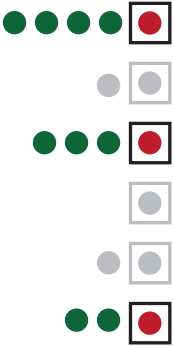


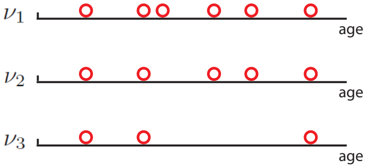


at least two jobs

Interacting Measure-Valued Processes Representation

ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .

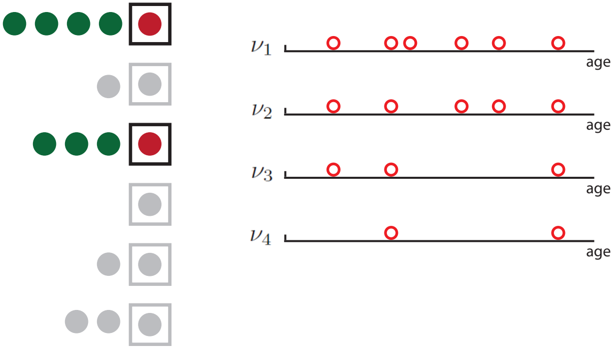




at least three jobs

Interacting Measure-Valued Processes Representation

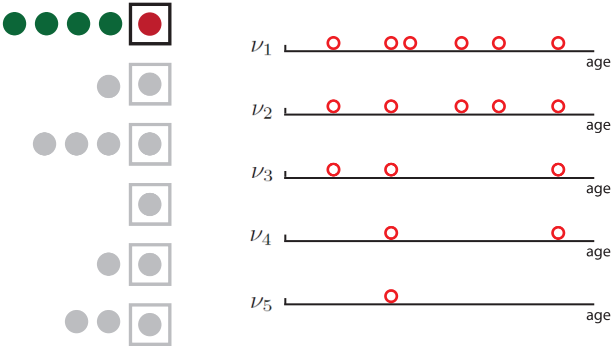
ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .



at least four jobs

Interacting Measure-Valued Processes Representation

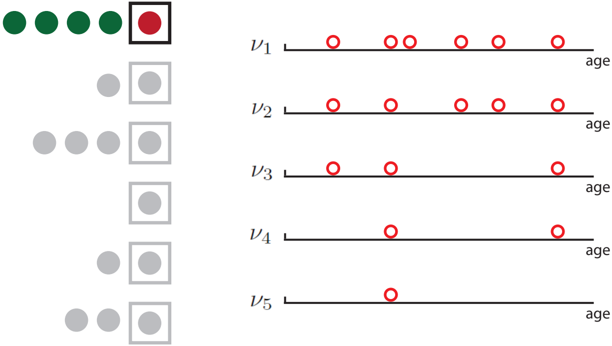
ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .



at least five jobs

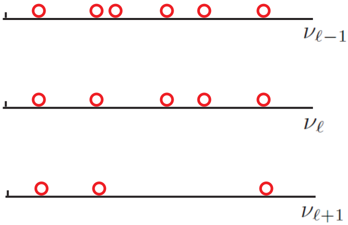
Interacting Measure-Valued Processes Representation

ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .

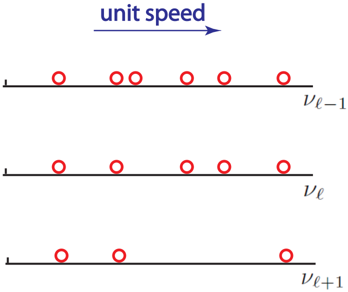


Analogous to [Kaspi-Ramanan'11]

I. when no arrival/departure is happening, the masses move to the right with unit speed.



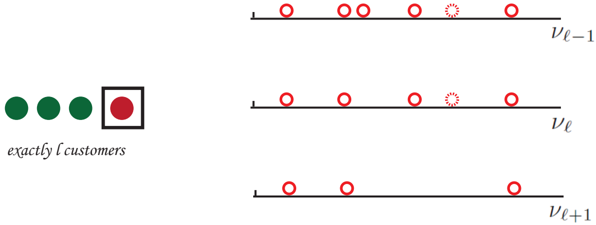
I. when no arrival/departure is happening, the masses move to the right with unit speed.



Dynamics of Measure-Valued Processes

II. Upon departure from a queue with ℓ jobs,

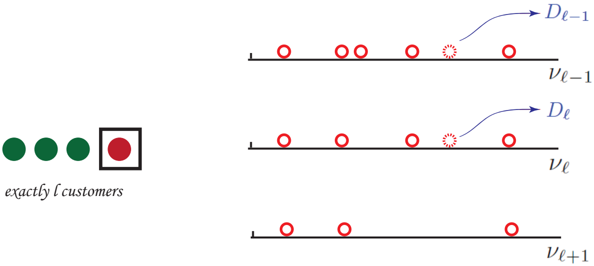
- the corresponding mass departs from all $\nu_j, j \leq \ell$
- a new mass at zero is added to all $\nu_j, j \leq \ell - 1$



Dynamics of Measure-Valued Processes

II. Upon departure from a queue with ℓ jobs,

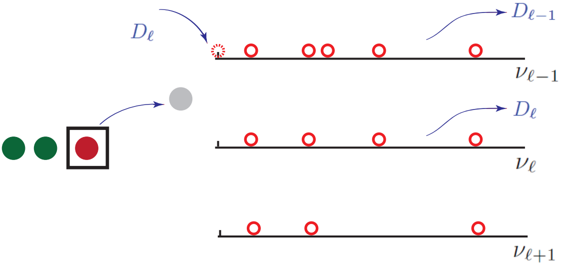
- the corresponding mass departs from all $\nu_j, j \leq \ell$
- a new mass at zero is added to all $\nu_j, j \leq \ell - 1$



- D_ℓ: cumulative departure process from servers with at least ℓ jobs before departure.

II. Upon departure from a queue with ℓ jobs,

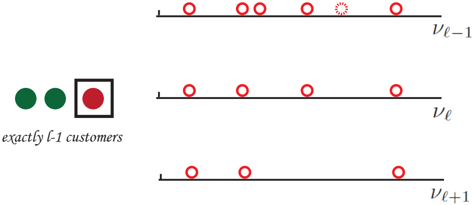
- the corresponding mass departs from all $\nu_j, j \leq \ell$
- a new mass at zero is added to all $\nu_j, j \leq \ell - 1$



- D_ℓ: cumulative departure process from servers with at least ℓ jobs before departure.

III. Upon arrival a queue with $\ell - 1$ jobs right before arrival,

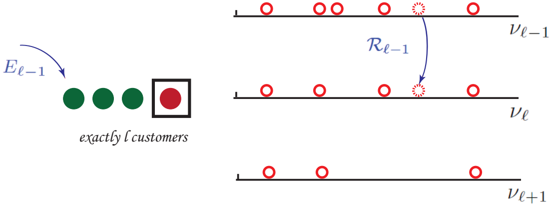
- if $\ell = 1$, a mass at zero joins ν_1
- if $\ell \geq 2$, the mass corresponding to the age of job in that particular server is added to ν_ℓ



Dynamics of Measure-Valued Processes

III. Upon arrival a queue with $\ell - 1$ jobs right before arrival,

- if $\ell = 1$, a mass at zero joins ν_1
- if $\ell \geq 2$, the mass corresponding to the age of job in that particular server is added to ν_ℓ



- $\mathcal{R}_\ell$: *routing measure process*

Routing Probabilities Super-Market Model

Upon arrival of j^{th} job,

- server i has ℓ job: $X^i = \ell$.
- ζ_j is the index of the server to which job j is routed

what is the probability $\{\zeta_j = i | X^i = \ell\}$?

Routing Probabilities Super-Market Model

Upon arrival of j^{th} job,

- server i has ℓ job: $X^i = \ell$.
- ζ_j is the index of the server to which job j is routed

$$\text{what is the probability } \{\zeta_j = i | X^i = \ell\}?$$

1. $\mathbb{P}\{\text{server } i \text{ has queue length} \geq \ell\} = \mathbb{P}\{\text{all picks have queue length} \geq \ell\} = S_\ell^d$.

$$S_\ell = \frac{1}{N}\langle 1, \nu_\ell \rangle : \text{portion of servers with at least} \leq \ell$$

Routing Probabilities Super-Market Model

Upon arrival of j^{th} job,

- server i has ℓ job: $X^i = \ell$.
- ζ_j is the index of the server to which job j is routed

$$\text{what is the probability } \{\zeta_j = i | X^i = \ell\}?$$

1. $\mathbb{P}\{\text{server } i \text{ has queue length} \geq \ell\} = \mathbb{P}\{\text{all picks have queue length} \geq \ell\} = S_\ell^d$.

$$S_\ell = \frac{1}{N}\langle 1, \nu_\ell \rangle : \text{portion of servers with at least} \leq \ell$$

2. $\mathbb{P}\{\text{server } \zeta_j \text{ has exactly } \ell \text{ job}\} = S_\ell^d - S_{\ell+1}^d$.

Routing Probabilities Super-Market Model

Upon arrival of j^{th} job,

- server i has ℓ job: $X^i = \ell$.
- ζ_j is the index of the server to which job j is routed

$$\text{what is the probability } \{\zeta_j = i | X^i = \ell\}?$$

1. $\mathbb{P}\{\text{server } i \text{ has queue length} \geq \ell\} = \mathbb{P}\{\text{all picks have queue length} \geq \ell\} = S_\ell^d.$

$$S_\ell = \frac{1}{N}\langle 1, \nu_\ell\rangle : \text{portion of servers with at least} \leq \ell$$

2. $\mathbb{P}\{\text{server } \zeta_j \text{ has exactly } \ell \text{ job}\} = S_\ell^d - S_{\ell+1}^d.$

3. $\mathbb{P}\{\zeta_j = i \big| X^i = \ell\} = \frac{1}{N} \frac{S_\ell^2 - S_{\ell+1}^2}{S_\ell - S_{\ell+1}}$

Definition A process $\bar{\nu} = \{\bar{\nu}_\ell\}_{\ell \geq 0}$ solves the *age equations* if for all $f \in \mathbb{C}_b^1[0, \infty)$,

$$\langle f, \nu_\ell(t) \rangle = \langle f, \nu_\ell(0) \rangle$$

initial jobs

Definition A process $\bar{\nu} = \{\bar{\nu}_\ell\}_{\ell \geq 0}$ solves the *age equations* if for all $f \in \mathbb{C}_b^1[0,\infty)$,

$$\langle f, \nu_\ell(t) \rangle = \langle f, \nu_\ell(0) \rangle + \int_0^t \langle f', \nu_\ell(s) \rangle ds$$

linear growth of ages

Definition A process $\bar{\nu} = \{\bar{\nu}_\ell\}_{\ell \geq 0}$ solves the *age equations* if for all $f \in \mathbb{C}_b^1[0, \infty)$,

$$\langle f, \nu_\ell(t)\rangle = \langle f, \nu_\ell(0)\rangle + \int_0^t \langle f', \nu_\ell(s)\rangle ds + f(0) D_{\ell+1}(t)$$

→ service entry

Main Results

Definition A process $\bar{\nu} = \{\bar{\nu}_\ell\}_{\ell \geq 0}$ solves the *age equations* if for all $f \in \mathbb{C}_b^1[0,\infty)$,

$$\langle f, \nu_\ell(t)\rangle = \langle f, \nu_\ell(0)\rangle + \int_0^t \langle f', \nu_\ell(s)\rangle ds + f(0) D_{\ell+1}(t) - \int_0^t \langle h\, f, \nu_\ell(s)\rangle ds$$

departure

Definition A process $\bar{\nu} = \{\bar{\nu}_\ell\}_{\ell \geq 0}$ solves the *age equations* if for all $f \in \mathbb{C}^1_b[0,\infty)$,

$$\langle f, \nu_\ell(t)\rangle = \langle f, \nu_\ell(0)\rangle + \int_0^t \langle f', \nu_\ell(s)\rangle ds + f(0) D_{\ell+1}(t) - \int_0^t \langle h\, f, \nu_\ell(s)\rangle ds + \int_0^t \langle f, \eta_\ell(s)\rangle ds.$$

Routing process

Main Results

Definition A process $\bar{\nu} = \{\bar{\nu}_\ell\}_{\ell \geq 0}$ solves the *age equations* if for all $f \in \mathbb{C}_b^1[0,\infty)$,

$$\langle f, \nu_\ell(t)\rangle = \langle f, \nu_\ell(0)\rangle + \int_0^t \langle f', \nu_\ell(s)\rangle ds + f(0) D_{\ell+1}(t) - \int_0^t \langle h\, f, \nu_\ell(s)\rangle ds + \int_0^t \langle f, \eta_\ell(s)\rangle ds.$$

$$\langle 1, \nu_\ell(t)\rangle - \langle 1, \nu_\ell(0)\rangle = D_{\ell+1}(t) + \int_0^t \langle 1, \eta_\ell(s)\rangle ds - D_\ell(t),$$

→ mass balance

Main Results

Definition A process $\bar{\nu} = \{\bar{\nu}_\ell\}_{\ell \geq 0}$ solves the *age equations* if for all $f \in \mathbb{C}_b^1[0, \infty)$,

$$\langle f, \nu_\ell(t)\rangle = \langle f, \nu_\ell(0)\rangle + \int_0^t \langle f', \nu_\ell(s)\rangle ds + f(0) D_{\ell+1}(t) - \int_0^t \langle h\, f, \nu_\ell(s)\rangle ds + \int_0^t \langle f, \eta_\ell(s)\rangle ds.$$

$$\langle \mathbf{1}, \nu_\ell(t)\rangle - \langle \mathbf{1}, \nu_\ell(0)\rangle = D_{\ell+1}(t) + \int_0^t \langle \mathbf{1}, \eta_\ell(s)\rangle ds - D_\ell(t),$$

$$D_\ell(t) = \int_0^t \langle h, \nu_\ell(s)\rangle ds,$$

departure rate

Main Results

Definition A process $\bar{\nu} = \{\bar{\nu}_\ell\}_{\ell \geq 0}$ solves the *age equations* if for all $f \in \mathbb{C}_b^1[0,\infty)$,

$$\langle f, \nu_\ell(t)\rangle = \langle f, \nu_\ell(0)\rangle + \int_0^t \langle f', \nu_\ell(s)\rangle ds + f(0) D_{\ell+1}(t) - \int_0^t \langle h\, f, \nu_\ell(s)\rangle ds + \int_0^t \langle f, \eta_\ell(s)\rangle ds.$$

$$\langle \mathbf{1}, \nu_\ell(t)\rangle - \langle \mathbf{1}, \nu_\ell(0)\rangle = D_{\ell+1}(t) + \int_0^t \langle \mathbf{1}, \eta_\ell(s)\rangle ds - D_\ell(t),$$

routing measure

$$D_\ell(t) = \int_0^t \langle h, \nu_\ell(s)\rangle ds,$$

$$\eta_\ell(t) = \begin{cases} \lambda(1 - \langle \mathbf{1}, \nu_1(t)\rangle^2)\delta_0 & \text{if } \ell = 1, \\ \lambda\langle \mathbf{1}, \nu_{\ell-1}(t) + \nu_\ell(t)\rangle(\nu_{\ell-1}(t) - \nu_\ell(t)) & \text{if } \ell \geq 2. \end{cases}$$

Main Results

Definition A process $\bar{\nu} = \{\bar{\nu}_\ell\}_{\ell \geq 0}$ solves the *age equations* if for all $f \in \mathbb{C}_b^1[0,\infty)$,

$$\langle f, \nu_\ell(t)\rangle = \langle f, \nu_\ell(0)\rangle + \int_0^t \langle f', \nu_\ell(s)\rangle ds + f(0) D_{\ell+1}(t) - \int_0^t \langle h\, f, \nu_\ell(s)\rangle ds + \int_0^t \langle f, \eta_\ell(s)\rangle ds.$$

$$\langle \mathbf{1}, \nu_\ell(t)\rangle - \langle \mathbf{1}, \nu_\ell(0)\rangle = D_{\ell+1}(t) + \int_0^t \langle \mathbf{1}, \eta_\ell(s)\rangle ds - D_\ell(t),$$

routing measure

$$D_\ell(t) = \int_0^t \langle h, \nu_\ell(s)\rangle ds,$$

routing probabilities

$$\eta_\ell(t) = \begin{cases} \lambda\left(1 - \langle \mathbf{1}, \nu_1(t)\rangle^2\right)\delta_0 & \text{if } \ell = 1, \\ \lambda\langle \mathbf{1}, \nu_{\ell-1}(t) + \nu_\ell(t)\rangle(\nu_{\ell-1}(t) - \nu_\ell(t)) & \text{if } \ell \geq 2. \end{cases}$$

Main Result

Theorem

Let $\{\nu^{(N)}(t) = (\nu_\ell^{(N)}(t))_\ell; t \geq 0\}$ *be the measure-valued representation for the* N*-server system with initial condition* $\nu^{(N)}(0)$*. If for some* $\nu_\ell(0)$

1. *arrival process* $E^{(N)}$ *is a renewal process with rate* λ^N*, and* $\lambda^N/N \to \lambda$*,*
2. *service distribution* G *has mean* 1 *and density* g*,*
3. *for every* $\ell \geq 1$*,* $\nu_\ell^{(N)}(0)/N \to \nu_\ell(0)$*,*

then

$$\frac{1}{N}\nu_\ell^{(N)} \to \nu_\ell,$$

where ν *is the unique solution to the age equation corresponding to* $\nu(0)$*.*

Main Result

Theorem

Let $\{\nu^{(N)}(t) = (\nu_\ell^{(N)}(t))_\ell; t \geq 0\}$ *be the measure-valued representation for the* N*-server system with initial condition* $\nu^{(N)}(0)$*. If for some* $\nu_\ell(0)$

1. *arrival process* $E^{(N)}$ *is a renewal process with rate* λ^N*, and* $\lambda^N/N \to \lambda$*,*
2. *service distribution* G *has mean* 1 *and density* g*,*
3. *for every* $\ell \geq 1$*,* $\nu_\ell^{(N)}(0)/N \to \nu_\ell(0)$*,*

then

$$\frac{1}{N}\nu_\ell^{(N)} \to \nu_\ell,$$

where ν *is the unique solution to the age equation corresponding to* $\nu(0)$*.*

Proof sketch.

- show the tightness of the sequence $\{\frac{1}{N}\nu^{(N)}\}$.

Main Result

Theorem

Let $\{\nu^{(N)}(t) = (\nu_\ell^{(N)}(t))_\ell; t \geq 0\}$ be the measure-valued representation for the N-server system with initial condition $\nu^{(N)}(0)$. If for some $\nu_\ell(0)$

1. *arrival process $E^{(N)}$ is a renewal process with rate λ^N, and $\lambda^N/N \to \lambda$,*
2. *service distribution G has mean 1 and density g,*
3. *for every $\ell \geq 1$, $\nu_\ell^{(N)}(0)/N \to \nu_\ell(0)$,*

then

$$\frac{1}{N}\nu_\ell^{(N)} \to \nu_\ell,$$

where ν is the unique solution to the age equation corresponding to $\nu(0)$.

Proof sketch.

- show the tightness of the sequence $\{\frac{1}{N}\nu^{(N)}\}$.
- show that every sub-sequential limit solves the age equation.

Main Result

Theorem

Let $\{\nu^{(N)}(t) = (\nu_\ell^{(N)}(t))_\ell; t \geq 0\}$ be the measure-valued representation for the N-server system with initial condition $\nu^{(N)}(0)$. If for some $\nu_\ell(0)$

1. *arrival process $E^{(N)}$ is a renewal process with rate λ^N, and $\lambda^N/N \to \lambda$,*
2. *service distribution G has mean 1 and density g,*
3. *for every $\ell \geq 1$, $\nu_\ell^{(N)}(0)/N \to \nu_\ell(0)$,*

then

$$\frac{1}{N}\nu_\ell^{(N)} \to \nu_\ell,$$

where ν is the unique solution to the age equation corresponding to $\nu(0)$.

Proof sketch.

- show the tightness of the sequence $\{\frac{1}{N}\nu^{(N)}\}$.
- show that every sub-sequential limit solves the age equation.
- use the uniqueness theorem for a unique characterization of sub-sequential limits.

Hydrodynamics Equations

We can partially solve the age equation: for every $f \in \mathbb{C}_b[0,\infty)$

$$\begin{aligned}\langle f, \nu_\ell(t)\rangle =& \langle f(\cdot + t)\frac{\bar{G}(\cdot + t)}{\bar{G}(\cdot)}, \nu_\ell(0)\rangle + \int_{[0,t]} f(t-s)\bar{G}(t-s)dD_{\ell+1}(s) \\ &+ \int_0^t \langle f(\cdot + t - s)\frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s)\rangle ds \qquad (1)\end{aligned}$$

Hydrodynamics Equations

We can partially solve the age equation: for every $f \in \mathbb{C}_b[0,\infty)$

$$\langle f, \nu_\ell(t)\rangle = \langle f(\cdot + t)\frac{\bar{G}(\cdot + t)}{\bar{G}(\cdot)}, \nu_\ell(0)\rangle + \int_{[0,t]} f(t-s)\bar{G}(t-s)dD_{\ell+1}(s)$$

$$+ \int_0^t \langle f(\cdot + t - s)\frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s)\rangle ds \qquad (1)$$

and

$$\langle \mathbf{1}, \nu_\ell(t)\rangle - \langle \mathbf{1}, \nu_\ell(0)\rangle = D_\ell(t) + \int_0^t \langle \mathbf{1}, \eta_\ell(s)ds - D_\ell(t), \qquad (2)$$

with

$$D_\ell(t) = \int_0^t \langle h, \nu_\ell(s)\rangle ds \qquad (3)$$

$$\eta_\ell(t) = \begin{cases} \lambda(1 - \langle \mathbf{1}, \nu_1(t)\rangle^2)\delta_0 & \text{if } \ell = 1, \\ \lambda\langle \mathbf{1}, \nu_{\ell-1}(t) + \nu_\ell(t)\rangle(\nu_{\ell-1}(t) - \nu_\ell(t)) & \text{if } \ell \geq 2. \end{cases} \qquad (4)$$

Hydrodynamics Equations

We can partially solve the age equation: for every $f \in \mathbb{C}_b[0,\infty)$

$$\begin{aligned}\langle f, \nu_\ell(t)\rangle =& \langle f(\cdot + t)\frac{\bar{G}(\cdot + t)}{\bar{G}(\cdot)}, \nu_\ell(0)\rangle + \int_{[0,t]} f(t-s)\bar{G}(t-s)dD_{\ell+1}(s) \\ &+ \int_0^t \langle f(\cdot + t - s)\frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s)\rangle ds \end{aligned} \quad (1)$$

and

$$\langle \mathbf{1}, \nu_\ell(t)\rangle - \langle \mathbf{1}, \nu_\ell(0)\rangle = D_\ell(t) + \int_0^t \langle \mathbf{1}, \eta_\ell(s) ds - D_\ell(t), \quad (2)$$

with

$$D_\ell(t) = \int_0^t \langle h, \nu_\ell(s)\rangle ds \quad (3)$$

$$\eta_\ell(t) = \begin{cases} \lambda(1 - \langle \mathbf{1}, \nu_1(t)\rangle^2)\delta_0 & \text{if } \ell = 1, \\ \lambda\langle \mathbf{1}, \nu_{\ell-1}(t) + \nu_\ell(t)\rangle(\nu_{\ell-1}(t) - \nu_\ell(t)) & \text{if } \ell \geq 2. \end{cases} \quad (4)$$

Equations (1)-(4) are called Hydrodynamics Equations.

A PDE representation

If one is only interested in $S_\ell(t) = \langle \mathbf{1}, \nu_\ell(t) \rangle$,

$$\langle \mathbf{1}, \nu_\ell(t) \rangle = \langle \frac{\bar{G}(\cdot + t)}{\bar{G}(\cdot)}, \nu_\ell(0) \rangle + \int_{[0,t]} \bar{G}(t-s) dD_{\ell+1}(s) + \int_0^t \langle \frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s) \rangle ds \tag{5}$$

A PDE representation

If one is only interested in $S_\ell(t) = \langle \mathbf{1}, \nu_\ell(t) \rangle$,

$$\langle \mathbf{1}, \nu_\ell(t) \rangle = \langle \frac{\bar{G}(\cdot + t)}{\bar{G}(\cdot)}, \nu_\ell(0) \rangle + \int_{[0,t]} \bar{G}(t-s) dD_{\ell+1}(s) + \int_0^t \langle \frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s) \rangle ds \tag{5}$$

define

$$f^r(x) = \frac{1 - G(x+r)}{1 - G(x)}, \qquad \xi_\ell(t,r) = \langle f^r, \nu_\ell(t) \rangle.$$

A PDE representation

If one is only interested in $S_\ell(t) = \langle \mathbf{1}, \nu_\ell(t) \rangle$,

$$\langle \mathbf{1}, \nu_\ell(t) \rangle = \langle \frac{\bar{G}(\cdot + t)}{\bar{G}(\cdot)}, \nu_\ell(0) \rangle + \int_{[0,t]} \bar{G}(t-s) dD_{\ell+1}(s) + \int_0^t \langle \frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s) \rangle ds \qquad (5)$$

define

$$f^r(x) = \frac{1 - G(x+r)}{1 - G(x)}, \qquad \xi_\ell(t, r) = \langle f^r, \nu_\ell(t) \rangle.$$

Then, we have $D_\ell(t) = -\int_0^t \partial_r \xi_\ell(s, 0) ds$, and the PDE

A PDE representation

If one is only interested in $S_\ell(t) = \langle \mathbf{1}, \nu_\ell(t) \rangle$,

$$\langle \mathbf{1}, \nu_\ell(t) \rangle = \langle \frac{\bar{G}(\cdot + t)}{\bar{G}(\cdot)}, \nu_\ell(0) \rangle + \int_{[0,t]} \bar{G}(t-s) dD_{\ell+1}(s) + \int_0^t \langle \frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s) \rangle ds \quad (5)$$

define

$$f^r(x) = \frac{1 - G(x+r)}{1 - G(x)}, \qquad \xi_\ell(t,r) = \langle f^r, \nu_\ell(t) \rangle.$$

Then, we have $D_\ell(t) = -\int_0^t \partial_r \xi_\ell(s,0) ds$, and the PDE

$$\xi_\ell(t,r) = \xi_\ell(0, t+r) - \int_0^t \bar{G}(t+r-u) \xi'_{\ell+1}(u,0) du, + \lambda \int_0^t F(\xi_{\ell-1}(u,r), \xi_\ell(u,r)) du,$$

A PDE representation

If one is only interested in $S_\ell(t) = \langle \mathbf{1}, \nu_\ell(t) \rangle$,

$$\langle \mathbf{1}, \nu_\ell(t) \rangle = \langle \frac{\bar{G}(\cdot + t)}{\bar{G}(\cdot)}, \nu_\ell(0) \rangle + \int_{[0,t]} \bar{G}(t-s) dD_{\ell+1}(s) + \int_0^t \langle \frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s) \rangle ds \tag{5}$$

define

$$f^r(x) = \frac{1 - G(x+r)}{1 - G(x)}, \qquad \xi_\ell(t,r) = \langle f^r, \nu_\ell(t) \rangle.$$

Then, we have $D_\ell(t) = -\int_0^t \partial_r \xi_\ell(s,0) ds$, and the PDE

$$\xi_\ell(t,r) = \xi_\ell(0, t+r) - \int_0^t \bar{G}(t+r-u) \xi'_{\ell+1}(u,0) du, + \lambda \int_0^t F(\xi_{\ell-1}(u,r), \xi_\ell(u,r)) du,$$

with boundary condition

$$\xi_\ell(t,0) - \xi_\ell(0,0) = \int_0^t \lambda(u) \left(\xi_{\ell-1}(u,0)^2 - \xi_\ell(u,0)^2 \right) - \left(\xi'_{\ell-1}(u,0) - \xi'_\ell(u,0)^2 \right) du,$$

Conclusion

We introduced a framework to analysis the load balancing algorithm, featuring

- Hydrodynamics limit which captures transient behavior
- Applicable for general service distributions
- Applicable for more general time varying arrival processes

Conclusion

We introduced a framework to analysis the load balancing algorithm, featuring

- Hydrodynamics limit which captures transient behavior
- Applicable for general service distributions
- Applicable for more general time varying arrival processes

For Exponential service distribution:

- limit process is characterized by a solution of a sequence of ODEs

Conclusion

We introduced a framework to analysis the load balancing algorithm, featuring

- Hydrodynamics limit which captures transient behavior
- Applicable for general service distributions
- Applicable for more general time varying arrival processes

For Exponential service distribution:

- limit process is characterized by a solution of a sequence of ODEs

For General service distribution:

- limit process is characterized by a solution of a sequence of PDEs

Conclusion

We introduced a framework to analysis the load balancing algorithm, featuring

- Hydrodynamics limit which captures transient behavior
- Applicable for general service distributions
- Applicable for more general time varying arrival processes

For Exponential service distribution:

- limit process is characterized by a solution of a sequence of ODEs

For General service distribution:

- limit process is characterized by a solution of a sequence of PDEs

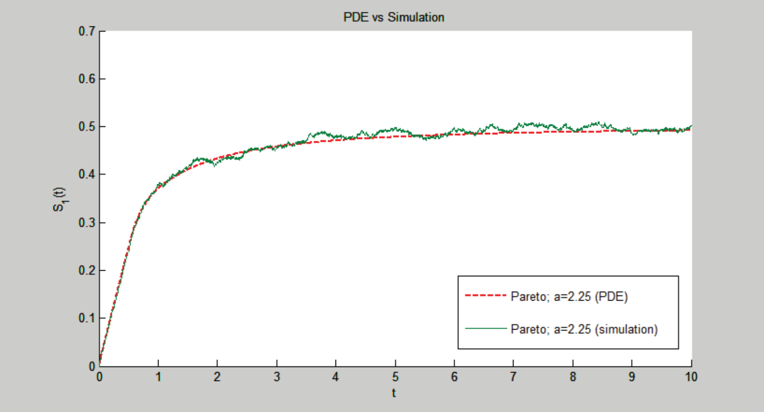
Equilibrium distributions are characterized by the fixed point of the PDEs.

Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw

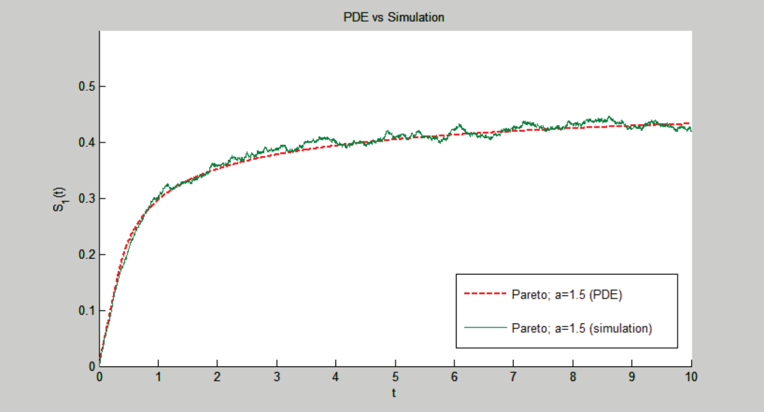
Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw



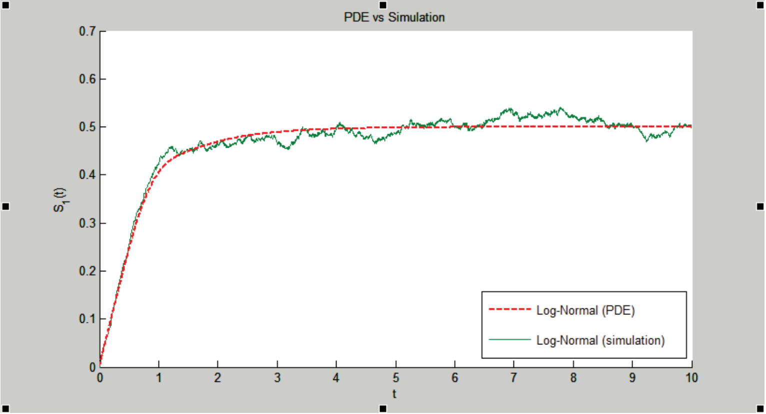
Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw



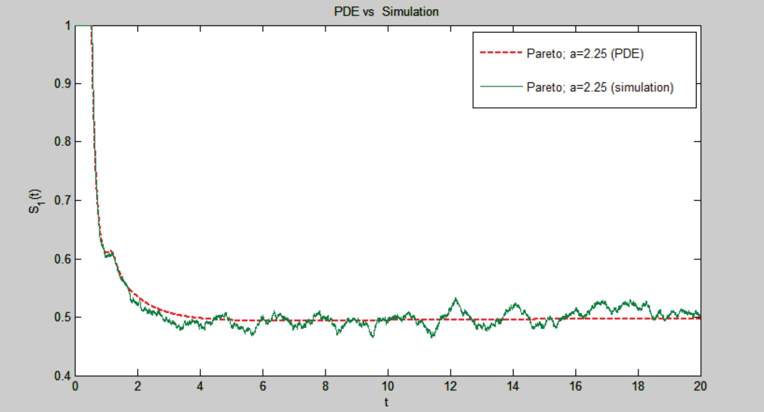
Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw



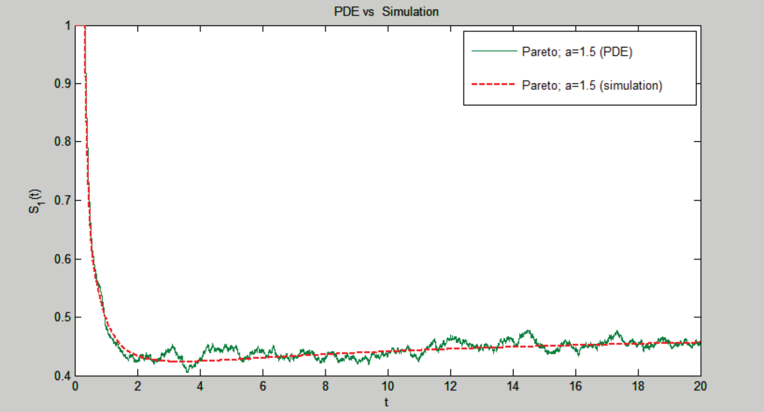
Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw



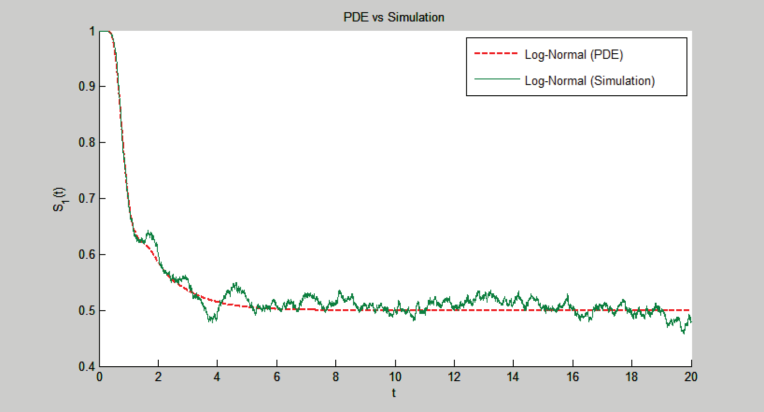
Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw



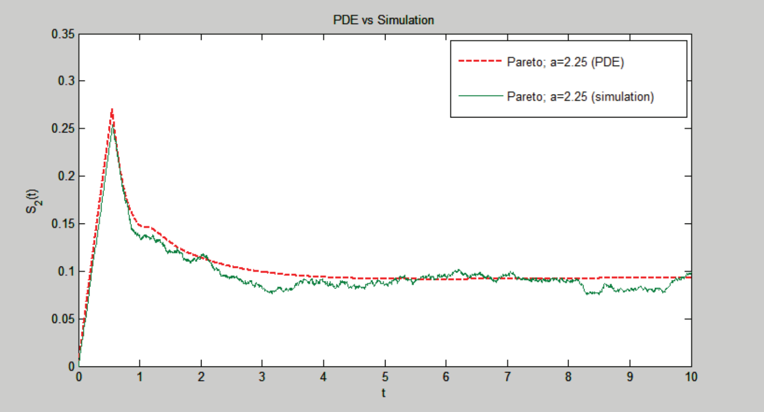
Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw



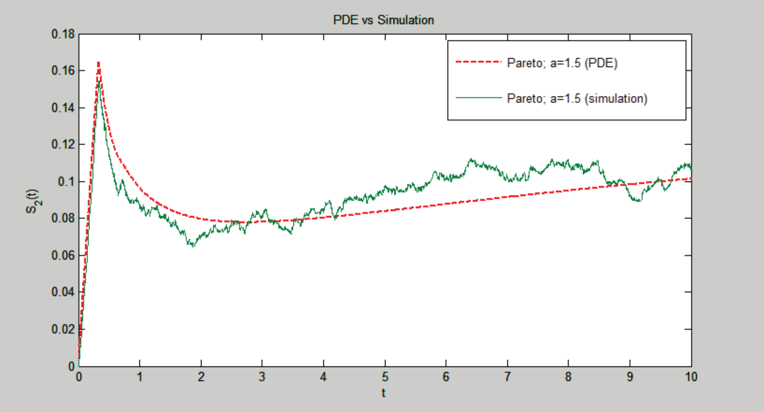
Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw



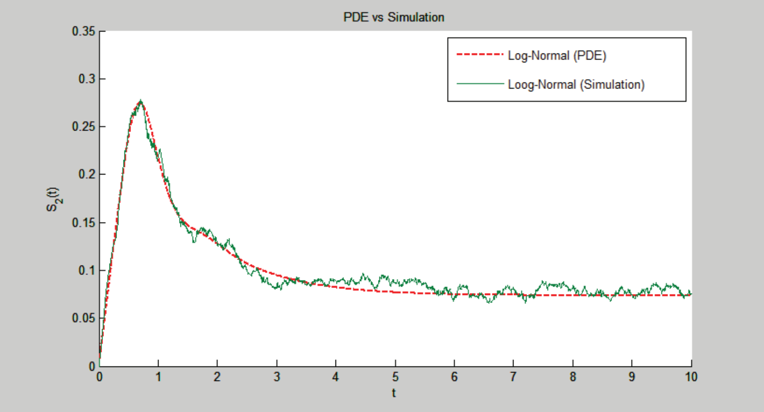
Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw



Simulation Result

We can numerically solve the PDE and compare to the simulation result we previously saw



Discussion and Ongoing Work

- The PDE provides more efficient alternative to simulations in order to address network optimization and design questions

Discussion and Ongoing Work

- The PDE provides more efficient alternative to simulations in order to address network optimization and design questions

Our Interacting measure-valued processes framework is general

- Applicable for different load balancing algorithms
- Applied to the analysis of the Serve the Longest Queue (SLQ) service discipline [Ramanan, Ganguly, Robert]

Discussion and Ongoing Work

- The PDE provides more efficient alternative to simulations in order to address network optimization and design questions

Our Interacting measure-valued processes framework is general

- Applicable for different load balancing algorithms
- Applied to the analysis of the Serve the Longest Queue (SLQ) service discipline [Ramanan, Ganguly, Robert]

Ongoing Work

- Rate of Convergence Result
- More on Numerical solution for the PDEs
- Gaining insight to specific time-varying scenarios
- Fixed point analysis for the PDE to derive stationary distribution