ODE and PDE Methods for Analysis of Large-Scale Load-Balancing Networks

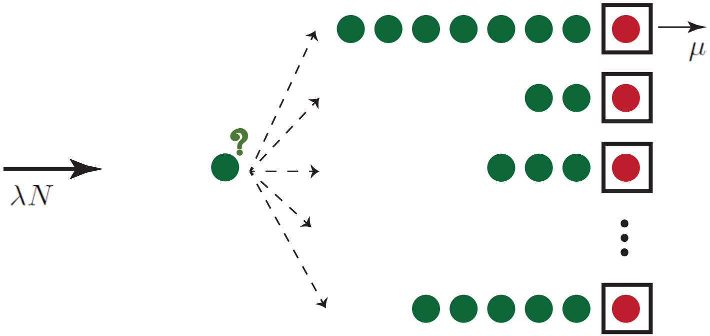
Reza Aghajani

UCSD

Joint work with Kavita Ramanan

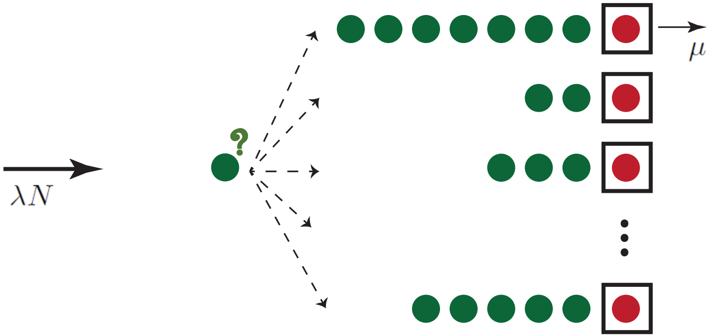
Aug 2016

Load-Balancing Network



Large-Scale Load-Balancing Networks

Load-Balancing Network



Load Balancing Algorithm:

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

Appear in:

- supermarket

Large Scale Load-Balancing Networks

Appear in:

- supermarket
- server farms

Appear in:

- supermarket
- server farms
- distributed memory machines

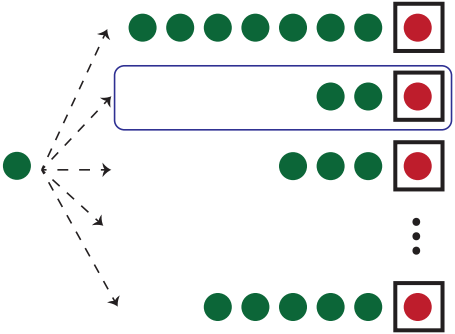
Large Scale Load-Balancing Networks

Appear in:

- supermarket
- server farms
- distributed memory machines
- hash tables

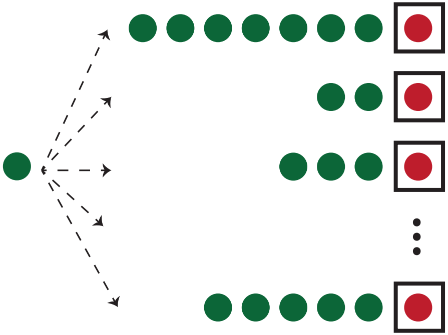
Common Load Balancing Algorithms

- ~~Joins the Shortest Queue~~ not feasible for large N



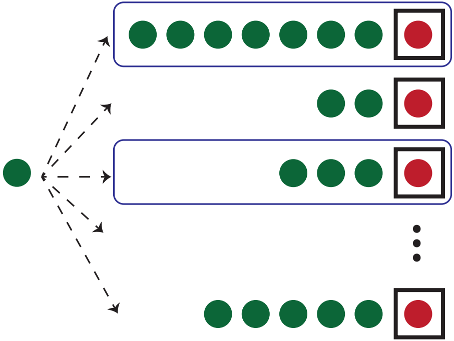
Common Load Balancing Algorithms

- ~~Joins the Shortest Queue~~ not feasible for large N
- $SQ(d)$ algorithm:



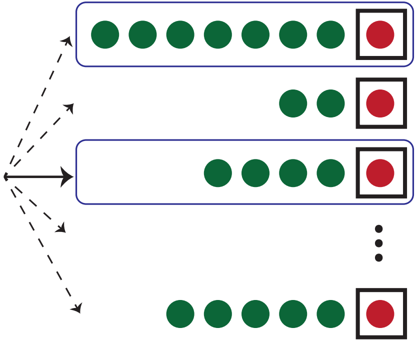
Common Load Balancing Algorithms

- ~~Joins the Shortest Queue~~ not feasible for large N
- $SQ(d)$ algorithm:
  - chooses d queues out of N, uniformly at random



Common Load Balancing Algorithms

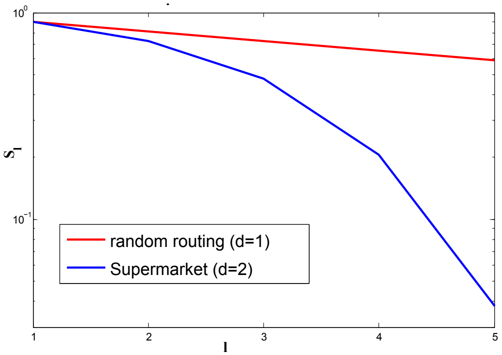
- ~~Joins the Shortest Queue~~ not feasible for large N
- $SQ(d)$ algorithm:
  - chooses d queues out of N, uniformly at random
  - joins the shortest queue among the chosen d



Supermarket model for exponential service time

Steady-State Queue Length Probabilities:

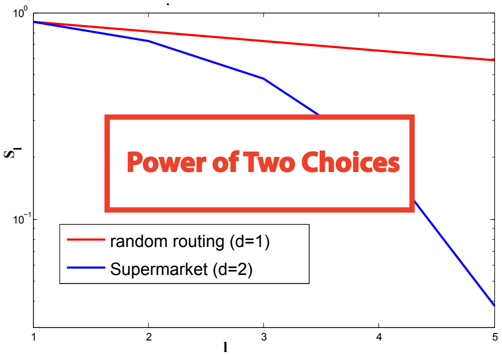
$$S_\ell = \mathbb{P}_{ss}\{\text{a typical queue length} \geq \ell\}$$



Supermarket model for exponential service time

Steady-State Queue Length Probabilities:

$$S_\ell = \mathbb{P}_{ss}\{\text{a typical queue length} \geq \ell\}$$



Typically,

stochastic networks are too complex

Large-Scale Stochastic Networks

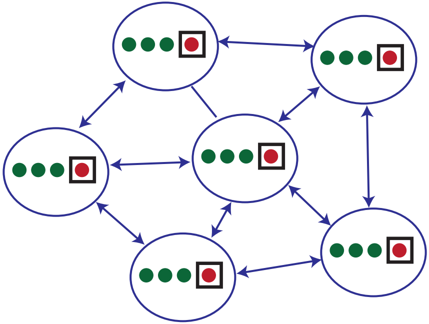
Typically,

stochastic networks are too complex

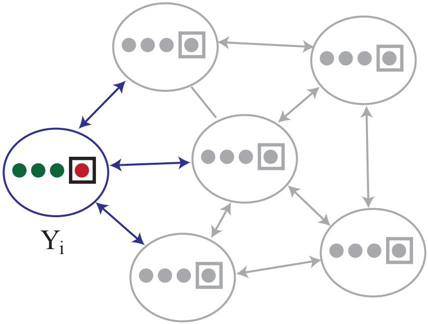
- not amenable to exact analysis
- should look for approximate solutions
- natural approximation for large-scale networks:

$$\text{number of servers } (N) \;\rightarrow\; \infty$$

Approach 1: Mean-Field Method (cavity method)

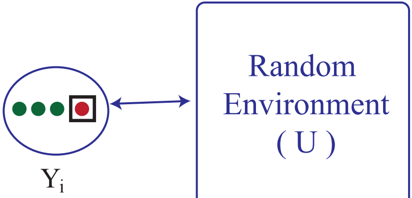


Approach 1: Mean-Field Method (cavity method)



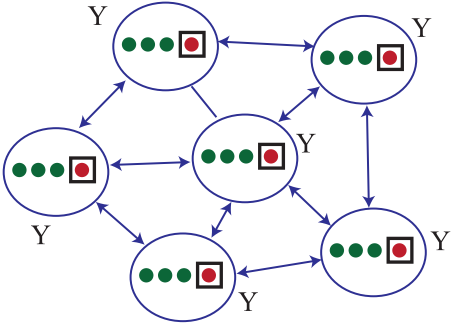
- local representation Y_i

Approach 1: Mean-Field Method (cavity method)



- local representation Y_i
- given an environment U, compute $Y_i = F(U)$

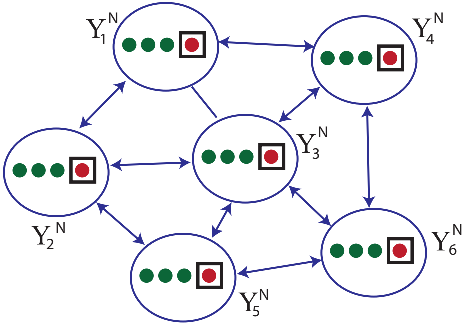
Approach 1: Mean-Field Method (cavity method)



- local representation Y_i
- given an environment U, compute $Y_i = F(U)$
- prove asymptotic independence as $N \to \infty$ (propagation of chaos)
- solve the distributional fixed-point equation

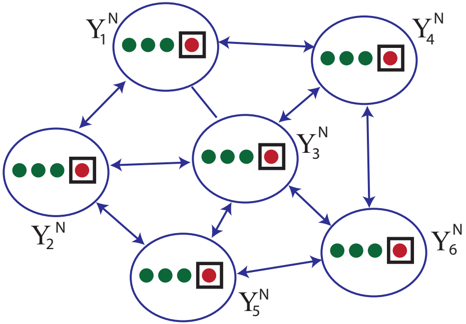
Asymptotic Analysis

Approach 2: ODE Method



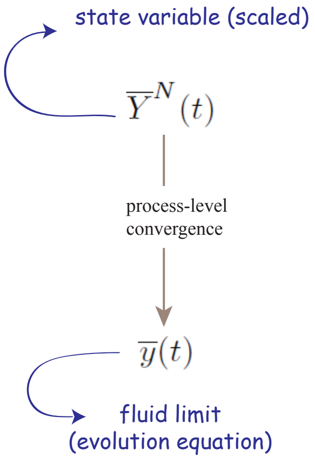
- Markovian (global) representation $Y^{(N)} = F(Y_1^{(N)}, \ldots, Y_N^{(N)})$

Approach 2: ODE Method

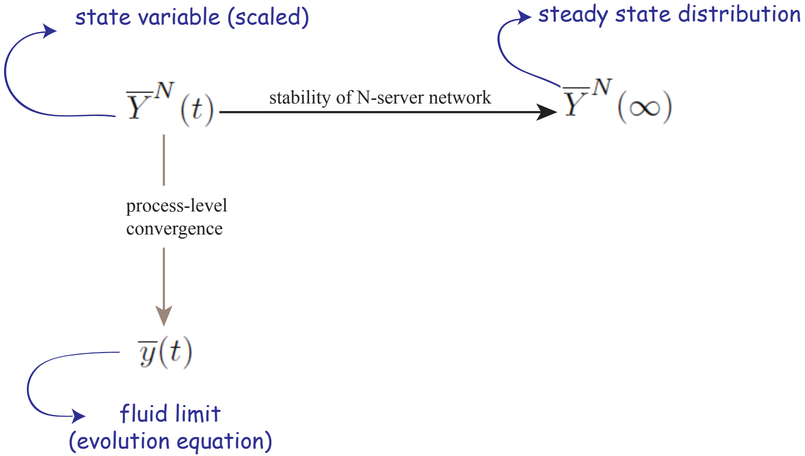


- Markovian (global) representation $Y^{(N)} = F(Y_1^{(N)}, ..., Y_N^{(N)})$
- establish limit theorems for $Y^{(N)}$

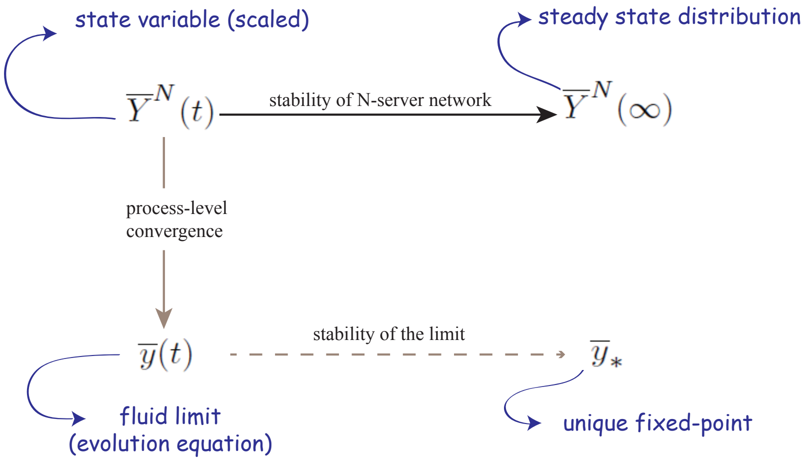
Hydrodynamic Approximation



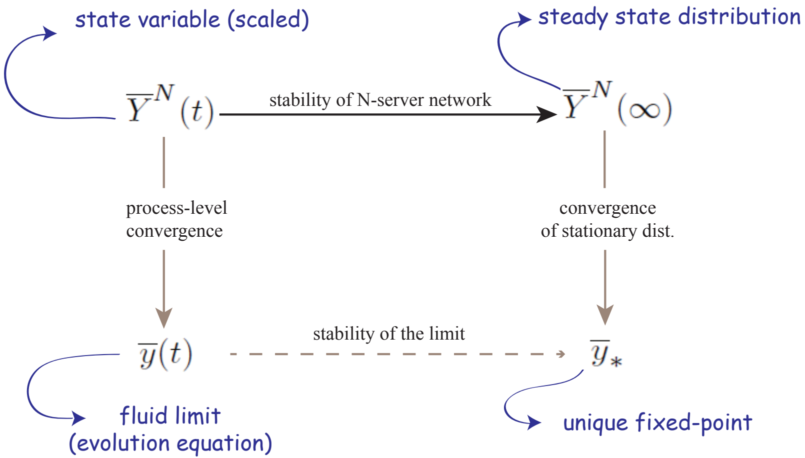
Hydrodynamic Approximation



Hydrodynamic Approximation



Hydrodynamic Approximation



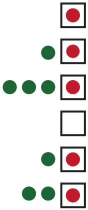
$$Y^{(N)}(\infty) \approx N\overline{y}_* + o(N)$$

Analysis of $SQ(d)$ algorithm:

To compute the transition probabilities:

- routing probabilities are to be computed
- to compute these probabilities, one needs the empirical distribution of queue lengths $S^N = (S_1^N, S_2^N, \ldots)$

$$S_\ell^N = \#\text{ of queues with length of at least } \ell.$$



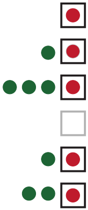
Analysis of $SQ(d)$ algorithm:

To compute the transition probabilities:

- routing probabilities are to be computed
- to compute these probabilities, one needs the empirical distribution of queue lengths $S^N = (S_1^N, S_2^N, \ldots)$

$$S_\ell^N = \#\text{ of queues with length of at least } \ell.$$

- $S_1^N = 5$



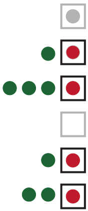
Analysis of $SQ(d)$ algorithm:

To compute the transition probabilities:

- routing probabilities are to be computed
- to compute these probabilities, one needs the empirical distribution of queue lengths $S^N = (S_1^N, S_2^N, \ldots)$

$$S_\ell^N = \# \text{ of queues with length of at least } \ell.$$

- $S_1^N = 5$
- $S_2^N = 4$



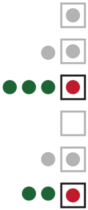
Analysis of $SQ(d)$ algorithm:

To compute the transition probabilities:

- routing probabilities are to be computed
- to compute these probabilities, one needs the empirical distribution of queue lengths $S^N = (S_1^N, S_2^N, \ldots)$

$$S_\ell^N = \#\text{ of queues with length of at least } \ell.$$

- $S_1^N = 5$
- $S_2^N = 4$
- $S_3^N = 2$



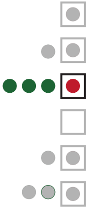
Analysis of $SQ(d)$ algorithm:

To compute the transition probabilities:

- routing probabilities are to be computed
- to compute these probabilities, one needs the empirical distribution of queue lengths $S^N = (S_1^N, S_2^N, \ldots)$

$$S_\ell^N = \# \text{ of queues with length of at least } \ell.$$

- $S_1^N = 5$
- $S_2^N = 4$
- $S_3^N = 2$
- $S_4^N = 1$



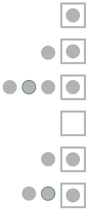
Analysis of $SQ(d)$ algorithm:

To compute the transition probabilities:

- routing probabilities are to be computed
- to compute these probabilities, one needs the empirical distribution of queue lengths $S^N = (S_1^N, S_2^N, \ldots)$

$$S_\ell^N = \# \text{ of queues with length of at least } \ell.$$

- $S_1^N = 5$
- $S_2^N = 4$
- $S_3^N = 2$
- $S_4^N = 1$
- $S_5^N = 0$



Exponential Service Time

When service time distribution is exponential [Vvedenskaya et. al. 96]:

- The empirical queue length $\{S_\ell^N(t); \ell \geq 1, t \geq 0\}$ is Markovian
- Convergence as $N \to \infty$ proved using 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

Exponential Service Time

When service time distribution is exponential [Vvedenskaya et. al. 96]:

- The empirical queue length $\{S_\ell^N(t); \ell \geq 1, t \geq 0\}$ is Markovian
- Convergence as $N \to \infty$ proved using 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

The following results are obtained:

- if $d = 1$: $P(X^N(\infty) > \ell) \to c\lambda^\ell$.
- if $d \geq 2$: $P(X^N(\infty) > \ell) \to \lambda^{(d^\ell - 1)/(d-1)}$

Exponential Service Time

When service time distribution is exponential [Vvedenskaya et. al. 96]:

- The empirical queue length $\{S_\ell^N(t); \ell \geq 1, t \geq 0\}$ is Markovian
- Convergence as $N \to \infty$ proved using 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

The following results are obtained:

- if $d = 1$: $P(X^N(\infty) > \ell) \to c\lambda^\ell$.
- if $d \geq 2$: $P(X^N(\infty) > \ell) \to \lambda^{(d^\ell - 1)/(d-1)}$

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

How about General Service Time Distributions?

Goal: To extend the result for general service distribution

How about General Service Time Distributions?

Goal: To extend the result for general service distribution

- Almost nothing was known 5 years ago
- Mathematical Challenge:
  - $\{S_\ell^N\}$ is no longer Markovian
  - need to keep track of more information: how long each job has been in service (ages)
  - No finite dimensional common state space for Markovian Representations
- Partial results by [Bramson-Lu-Prabhakar '13] using the cavity method

How about General Service Time Distributions?

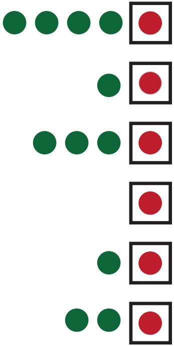
Goal: To extend the result for general service distribution

- Almost nothing was known 5 years ago
- Mathematical Challenge:
  - $\{S_\ell^N\}$ is no longer Markovian
  - need to keep track of more information: how long each job has been in service (ages)
  - No finite dimensional common state space for Markovian Representations
- Partial results by [Bramson-Lu-Prabhakar '13] using the cavity method

Our Approach:

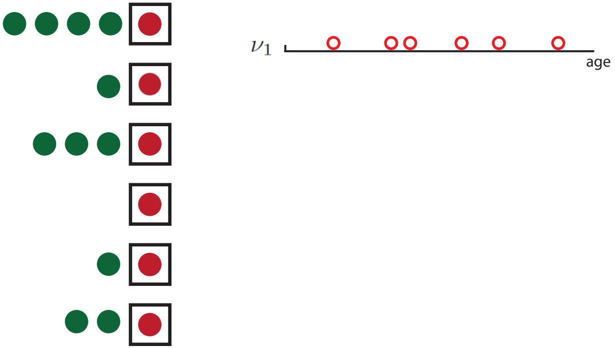
New representation: Interacting Measure-valued Processes

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



Interacting Measure-Valued Processes Representation

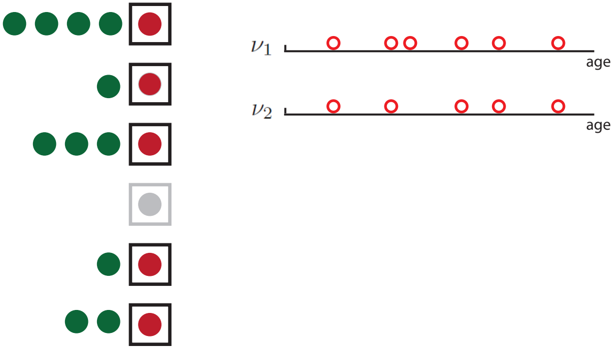
ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .



at least one jobs

Interacting Measure-Valued Processes Representation

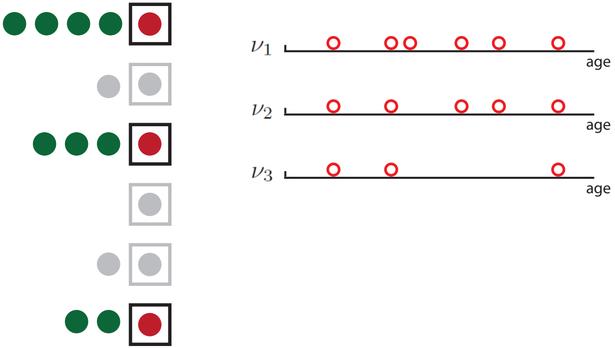
ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .



at least two jobs

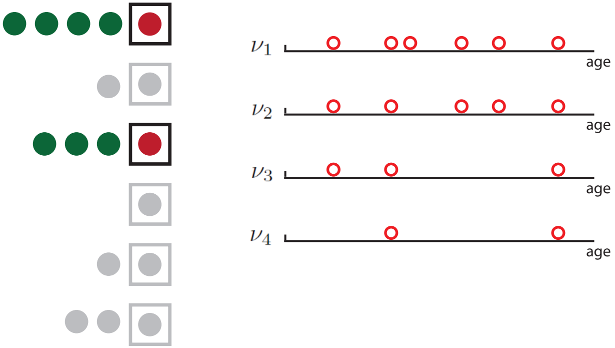
Interacting Measure-Valued Processes Representation

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



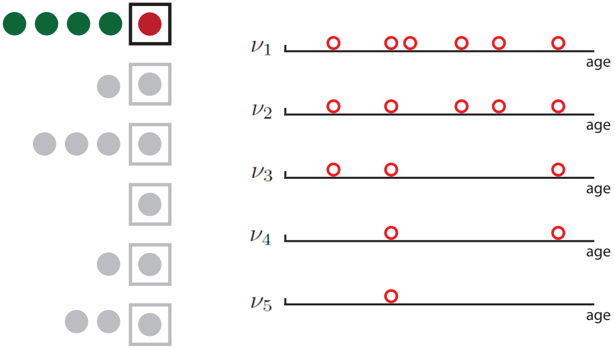
at least three jobs

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



at least four jobs

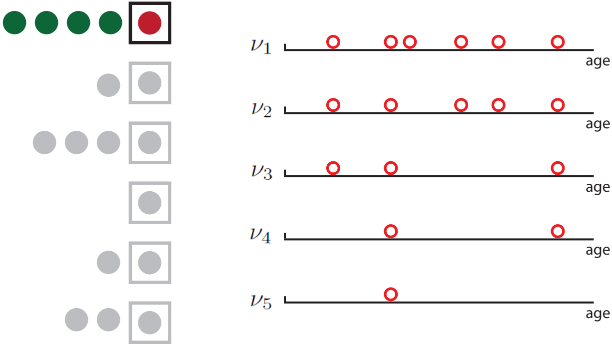
ν_ℓ: unit mass at the ages of jobs in servers with queues of length at least ℓ .



at least five jobs

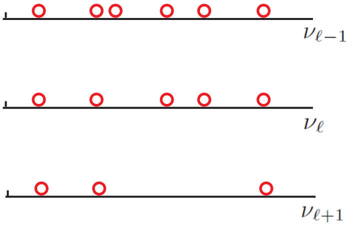
Interacting Measure-Valued Processes Representation

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

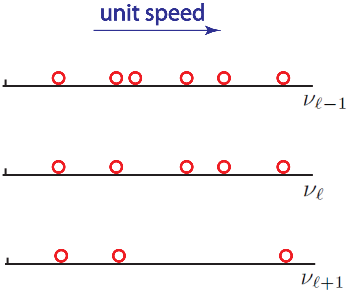


inspired by [Kaspi-Ramanan'11]

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



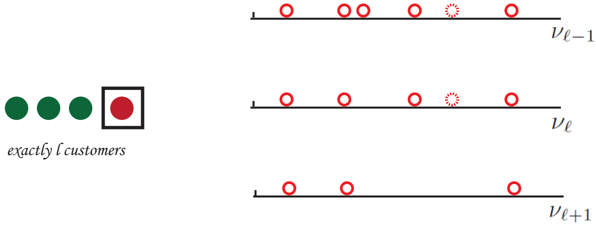
I. when no arrival/departure is happening, the masses move to the right with unit speed.



Dynamics of Measure-Valued Processes

II. Upon departure from a queue with ℓ jobs,

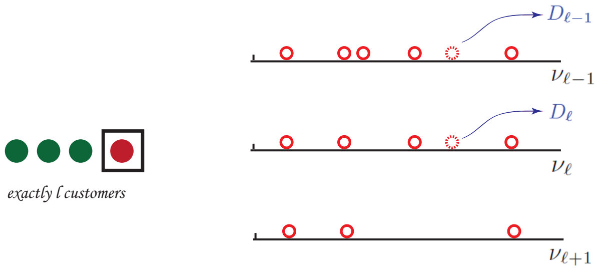
- the corresponding mass departs from all $\nu_j, j \leq \ell$
- a new mass at zero is added to all $\nu_j, j \leq \ell - 1$



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 $\nu_j, j \leq \ell - 1$

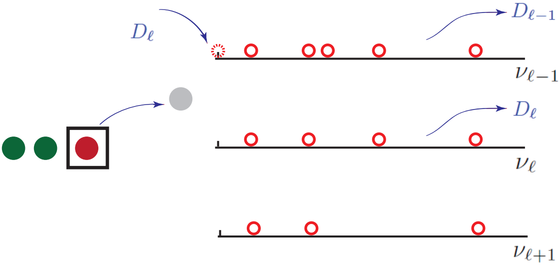


- D_ℓ: cumulative departure process from servers with at least ℓ jobs before departure.

Dynamics of Measure-Valued Processes

II. Upon departure from a queue with ℓ jobs,

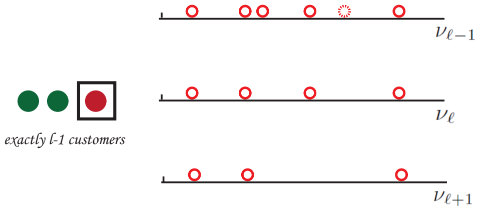
- the corresponding mass departs from all $\nu_j, j \leq \ell$
- a new mass at zero is added to all $\nu_j, j \leq \ell - 1$



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

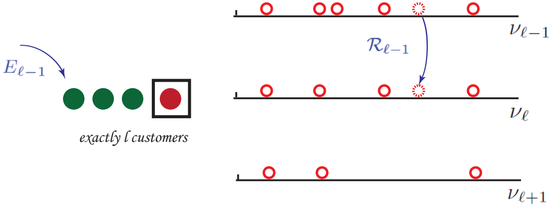
III. Upon arrival a queue with $\ell - 1$ jobs right before arrival,

- 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



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*

Hydrodynamics Equations

The following equations describe fluid limit of $\nu^{(N)}$:

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

for every $f \in \mathbb{C}_b[0,\infty)$, and

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

with

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

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

Hydrodynamics Equations

The following equations describe fluid limit of $\nu^{(N)}$:

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

for every $f \in \mathbb{C}_b[0,\infty)$, and

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

with

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

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

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

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

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

where ν *is the unique solution to the hydrodynamics equations 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

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

where ν is the unique solution to the hydrodynamics equations 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

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

where ν *is the unique solution to the hydrodynamics equations 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

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

where ν *is the unique solution to the hydrodynamics equations 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.
- show that the hydrodynamics equations have a unique solutions

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 $\{ f^r(x) = \frac{1-G(x+r)}{1-G(x)}; r \geq 0 \}$ and

$$Z_\ell(t,r) = \langle f^r, \nu_\ell(t) \rangle.$$

A PDE representation

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

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

define $\{f^r(x) = \frac{1-G(x+r)}{1-G(x)}; r \geq 0\}$ and

$$Z_\ell(t,r) = \langle f^r, \nu_\ell(t)\rangle.$$

Then, we have $D_\ell(t) = -\int_0^t \partial_r Z_\ell(s,0) ds$.

Hydrodynamic PDEs:

$Z = (Z_\ell)$ satisfies the following countable set of PDEs

$$\partial_t Z_\ell(t,r) - \partial_r Z_\ell(t,r) = -\overline{G}(r)\partial_{\ell+1} Z(t,0) + \lambda(t)(Z_{\ell-1}(t,0) + Z_\ell(t,0)) \times (Z_{\ell-1}(t,r) - Z_\ell(t,r))$$

for $\ell \geq 1$, with initial conditions $(Z_\ell(0,\cdot); \ell \geq 1)$.

- countable number of interacting equations
- non-linear
- non-standard: boundary condition appears as the external force

Hydrodynamic PDEs:

$Z = (Z_\ell)$ satisfies the following countable set of PDEs

$$\partial_t Z_\ell(t,r) - \partial_r Z_\ell(t,r) = -\overline{G}(r)\partial_{\ell+1} Z(t,0) + \lambda(t)(Z_{\ell-1}(t,0) + Z_\ell(t,0)) \times (Z_{\ell-1}(t,r) - Z_\ell(t,r))$$

for $\ell \geq 1$, with initial conditions $(Z_\ell(0,\cdot); \ell \geq 1)$.

- countable number of interacting equations
- non-linear
- non-standard: boundary condition appears as the external force

Theorem

If h is bounded, then hydrodynamic PDEs have a unique solution in a suitable subspace of $\mathbb{C}_b^1[0,\infty)^{\mathbb{N}}$.

challenge: infinite set of inter-dependent PDEs.

Main Results

The solution to the Hydrodynamic PDEs can be used to approximate the queue length probabilities as well as other quantities such as the virtual waiting time.

Theorem

Under Assumptions of Theorem 1,

$$\lim_{N\to\infty} \mathbb{P}\left\{X^{(N),1}(t) \geq \ell, X^{(N),2}(t) \geq k\right\} = Z_\ell(t,0)Z_k(t,0),$$

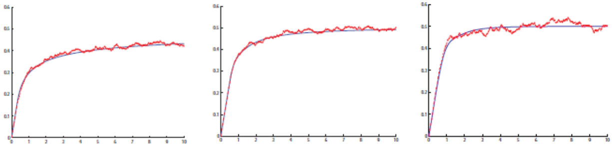
where $\{Z_\ell; \ell \geq 1\}$ is the unique solution to the hydrodynamic PDEs. Moreover,

$$\lim_{N\to\infty} \mathbb{E}\left[W^{(N)}(t)\right] = \sum_{\ell\geq 2} Z_\ell(t,0)^2 + \sum_{\ell\geq 1} \left[Z_\ell(t,0) + Z_{\ell+1}(t,0)\right] \times \int_0^\infty \left[Z_\ell(t,r) - Z_{\ell+1}(t,r)\right] dr.$$

Simulation Result

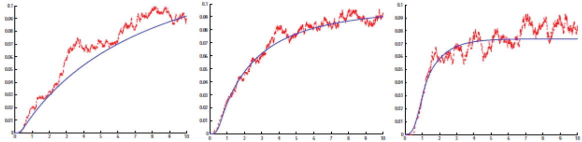
We can numerically solve the PDE, and obtain:

- **fraction of busy servers:**



(a) Pareto $b = 1.50$ (b) Pareto $b = 2.25$ (c) Log-Normal $\sigma = 0.33$

- **fraction of servers with queue length at least 2**



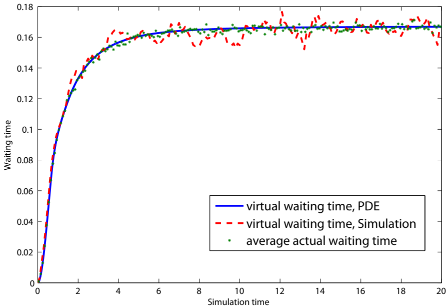
(d) Pareto $b = 1.50$ (e) Pareto $b = 2.25$ (f) Log-Normal $\sigma = 0.33$

Plots for network of 500 servers.

Simulation Results

We can numerically solve the PDE, and obtain:

- **Virtual waiting time**

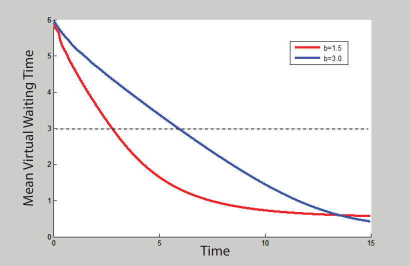


- The actual mean waiting time is also well-approximated by the same quantity obtained from PDEs.

Implication of Results

Example: Backlog Recovery

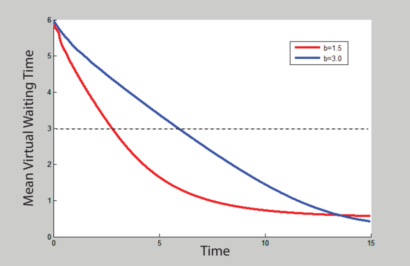
- Intermittently, jobs experience long waiting times due to a backlog
- How long would it take for the network to get rid of the backlog?
  - Relaxation Time: the time when virtual waiting time drops to half .



Implication of Results

Example: Backlog Recovery

- Intermittently, jobs experience long waiting times due to a backlog
- How long would it take for the network to get rid of the backlog?
  - Relaxation Time: the time when virtual waiting time drops to half .



Observation: For the infinite-variance (heavy tail) service distribution, the network gets rid of the backlog faster!

Implication of Results

- Comparison with equilibrium result for Pareto service distribution:
  - **Bramson-Lu-Prabakar '13:** when considering tail probabilities in equilibrium, finite variance is favorable.
  - **Our Observation:** when considering the mean virtual waiting time in network recovering form a backlog, infinite variance is favorable.

Implication of Results

- Comparison with equilibrium result for Pareto service distribution:
  - **Bramson-Lu-Prabakar '13:** when considering tail probabilities in equilibrium, finite variance is favorable.
  - **Our Observation:** when considering the mean virtual waiting time in network recovering form a backlog, infinite variance is favorable.

- Using the PDE, we observed an nonintuitive behavior of the load-balancing network

- The PDE provides more efficient alternative to simulations in order to address network optimization and design questions. Generating these kind of graphs with simulation would take much longer

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.