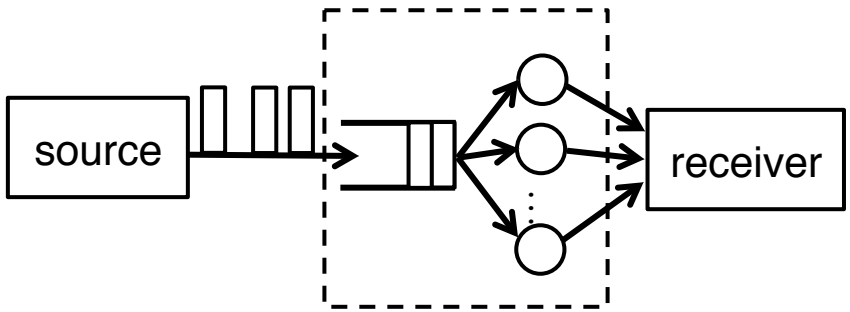
Age-Optimal Updates of Multiple Information Flows

Yin Sun
Dept. ECE, Auburn University

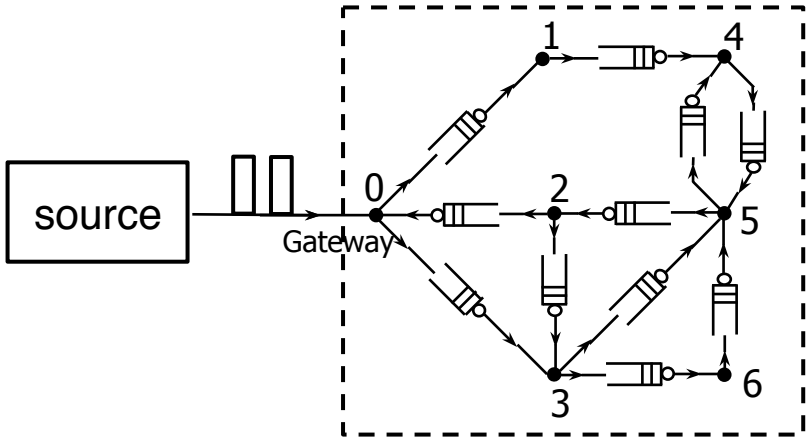
Joint work with Elif Uysal-Biyikoglu, Sastry Kompella

AUBURN
UNIVERSITY

Prior Art: Single-flow Networks



Single-hop multi-channel network



Multi-hop network

- Last-Generated, First-Served (LGFS) policy is (near) age-optimal in stochastic order [Bedewy-Sun-Shroff 'ISIT2016,2017]
  - **Pro:** These results are quite general, they hold for
    - Arbitrary generation times, arbitrary arrival times
    - Minimizing the age process $\{\Delta(t), t \geq 0\}$ in stochastic ordering
    - Minimizing any increasing functional $p(\{\Delta(t), t \geq 0\})$ of age process in stochastic ordering
    - Special case: LCFS-type policies for in-order arrivals

What is Age Penalty Functional?

- ***Age penalty functional*** $p(\{\Delta(t), t \geq 0\})$***:***
  Increasing mapping from the age process $\{\Delta(t), t \in [0, \infty)\}$ to a real number.
  - Larger age process ➔ higher penalty

- Previous age metrics as examples:
  - Time-avg. age: [Kaul-Yates-Gruteser'12, etc.]

$$p_1(\{\Delta(t), t \geq 0\}) = \frac{1}{T}\int_0^T \Delta(t)dt$$

  - Time-avg. age penalty function: [Sun-Uysal-Yates-Koksal-Shroff'16, etc.]

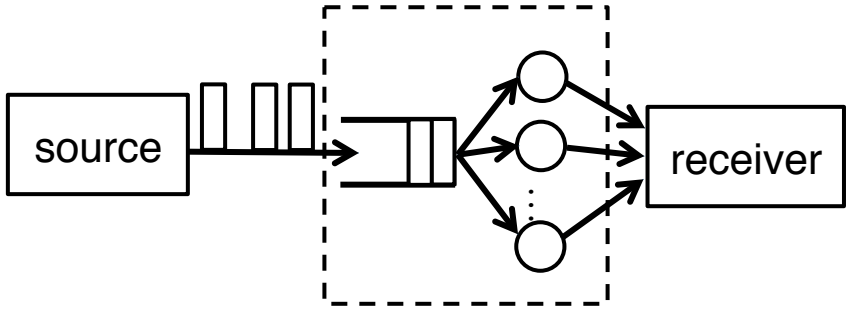
$$p_2(\{\Delta(t), t \geq 0\}) = \frac{1}{T}\int_0^T h(\Delta(t))dt$$

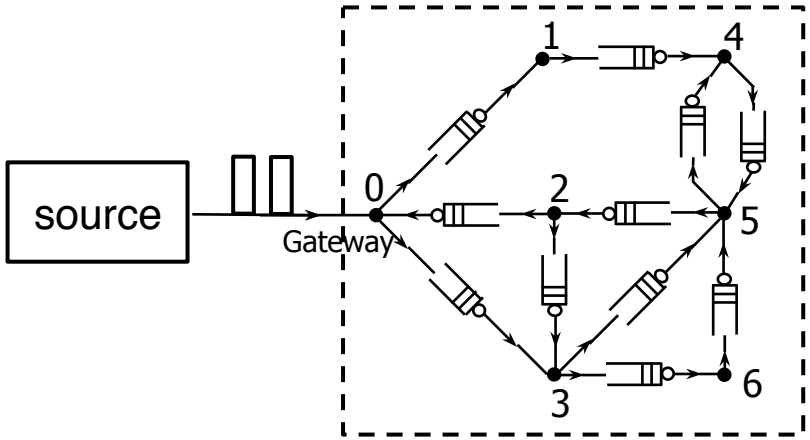
  - Age tail distribution

Indicator function

$$p_3(\{\Delta(t), t \geq 0\}) = \frac{1}{T}\int_0^T 1_{\{\Delta(t) \geq b\}}dt$$

Prior Art: Single-flow Networks (II)



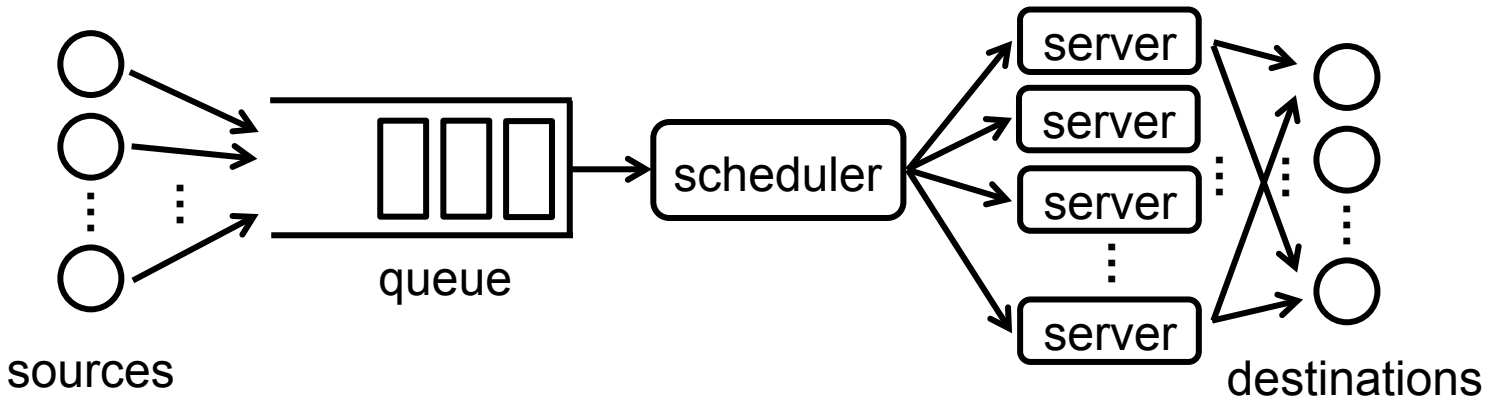


Single-flow multi-channel network

Single-flow multi-hop network

- Last-Generated, First-Served (LGFS) policy is (near) age-optimal in stochastic order [Bedewy-Sun-Shroff 'ISIT2016,2017]
  - **Pro:** These results are quite general, they hold for
    - Arbitrary generation times, arbitrary arrival times
    - Minimizing the age process $\{\Delta(t), t \geq 0\}$
    - Minimizing any increasing functional $p(\{\Delta(t), t \geq 0\})$ of age process
    - Special case: LCFS-type policies for in-order arrivals
  - Complement age analysis
  - How about multi-flow networks?

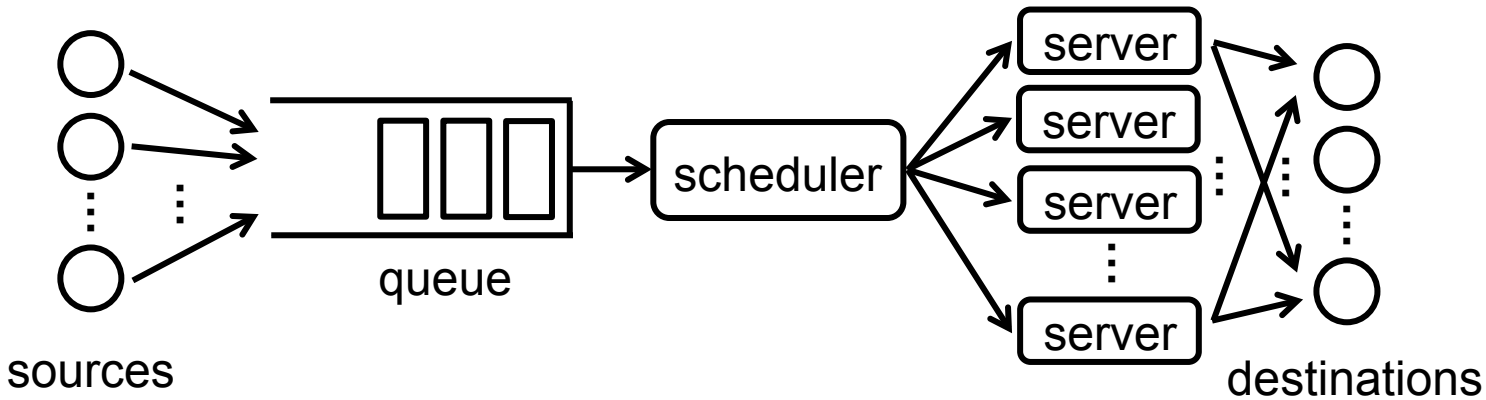
This work: Multi-flow, Multi-channel Networks



- In LGFS, the last generated packet is served the first
  - Ordering the packets from a single flow

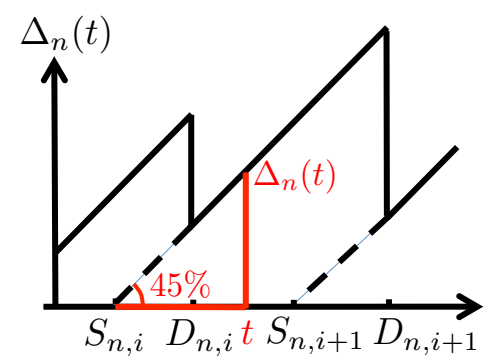
- **Difficulty:** How to compare the packets from multiple flows?

System Model: Notations

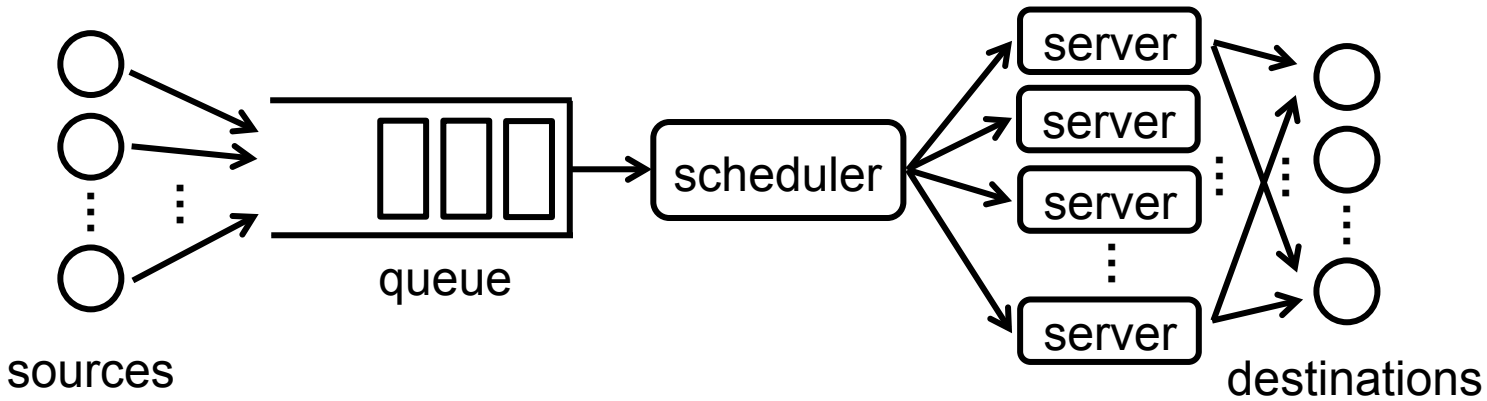


- Multiple sources-destination pairs, multiple servers
- Packet i of flow n is generated at time $S_{n,i}$, arrives at time $A_{n,i}$, and is delivered at time $D_{n,i}$
- **Age of Information:** The time difference between the current time t and the generation time of the freshest received sample

$$\Delta_n(t) = t - \max\{S_{n,i} : D_{n,i} \leq t\}$$



System Model: Scheduling Policies



- The scheduler decides which sample in the queue to send, and send on which server

- Preemptive policy:

  One sample can interrupt the transmission of another sample

- Causal policy:

  Scheduler makes decisions based on history and current information

Stochastic Ordering

- **Stochastic ordering of random variable**
  - means "**ordering of random variables in distribution**"

$$X \leq_{\mathrm{st}} Y \quad \Leftrightarrow \quad \Pr(X > t) \leq \Pr(Y > t), \ \forall\, t \in \mathbb{R}$$

$$\Leftrightarrow \quad \mathbb{E}[p(X)] \leq \mathbb{E}[p(Y)]$$

holds for all non-decreasing function p

- **Stochastic ordering of stochastic process (stronger)**
  - means "**ordering of stochastic processes in distribution**"

$$\{X(t), t \geq 0\} \leq_{\mathrm{st}} \{Y(t), t \geq 0\}$$

$$\Leftrightarrow \ \mathbb{E}[\phi(\{X(t), t \in [0, \infty)\})] \leq \mathbb{E}[\phi(\{Y(t), t \in [0, \infty)\})]$$

holds for all non-decreasing functional ϕ

Age Penalty Function of Multiple Sources

- Age penalty function: $p \circ \mathbf{\Delta} = p(\mathbf{\Delta}) = p(\Delta_1, \ldots, \Delta_N)$
  - Level of "dissatisfaction" for having stale data at multiple receivers
  - Examples:

$$p_{\text{avg}}(\mathbf{\Delta}) = \frac{1}{N}\sum_{n=1}^{N}\Delta_n \quad p_{\max}(\mathbf{\Delta}) = \max_{n=1,\ldots,N}\Delta_n \quad p_{l\text{-norm}}(\mathbf{\Delta}) = \left[\sum_{n=1}^{N}(\Delta_n)^l\right]^{\frac{1}{l}}, \; l \geq 1$$

- ***Symmetric***, ***non-decreasing***, and ***time-dependent*** age functions:
  - Define $\mathcal{P}_{\text{sym}} = \{p : [0,\infty)^N \to \mathbb{R} \text{ is symmetric and non-decreasing}\}$
  - We allow the age penalty function $p_t \in \mathcal{P}_{\text{sym}}$ to vary over time

$$p_t \circ \mathbf{\Delta}(t) = p_t(\mathbf{\Delta}(t)) = p_t(\Delta_1(t), \ldots, \Delta_{,N}(t))$$

  - Associated stochastic process under policy π: $\{p_t \circ \mathbf{\Delta}_\pi(t), t \geq 0\}$

- **Goal:** find scheduling policy to minimize $\{p_t \circ \mathbf{\Delta}_\pi(t), t \geq 0\}$ in distribution (stochastic ordering of stochastic process)

Scheduling Policy Design

- How to compare the packets from multiple flows?
  Maximum Age First (MAF) discipline
  - The flow with the maximum age is served the first
    [Li-Eryilmaz-Srikant'15, Kadota-Uysal-Singh-Modiano'16, Hsu-Modiano-Dua'17]

- **MAF + LGFS = Maximum Age First, Last Generated First Served (MAF-LGFS) policy:**
  - The last generated packet from the flow with the maximum age is served the first.

Single Server, Exponential Service Times

- **Assumption**: **(Synchronized Generations and Arrivals)** There exist two sequences $\{S_1, S_2, \ldots\}$ and$\{A_1, A_2, \ldots\}$ such that

$$S_{n,i} = S_i, \;\; A_{n,i} = A_i \text{ for all packet } i \text{ and flow } n$$

Theorem: If (i) packet generation and arrival times are synchronized across the flows, (ii) there is a single server, (iii) i.i.d. exponential service times, then for all $p_t \in \mathcal{P}_{\text{sym}}$ and all $\{S_1, S_2, \ldots, A_1, A_2, \ldots\}$, the preemptive MAF-LGFS policy minimizes $\{p_t \circ \mathbf{\Delta}_\pi(t), t \geq 0\}$ in distribution among all causal policies, i.e.,

$$\{p_t \circ \mathbf{\Delta}_{\text{prmp, MAF-LGFS}}(t), t \geq 0\} \leq_{\text{st}} \{p_t \circ \mathbf{\Delta}_\pi(t), t \geq 0\}, \;\; \pi \in \Pi$$

where Π is the set of causal policies.

The first to handle ***symmetric***, ***non-decreasing***, and ***time-dependent*** age penalty functions

Single-server v.s Multi-server Scheduling

- **Single-server systems:** Optimal scheduling is achievable
  - Deterministic scheduling:
    - Average delay: SRPT [Schrage'68,Smith'78]
  - Stochastic scheduling:
    - Weighted average delay: c-μ rule [Smith'56], Gittins index [Gittins'79], Klimov's model [Klimov'74]
- **Multi-server systems:** Optimal scheduling is difficult!!
  - Deterministic scheduling:
    - Average delay: NP hard for $\geq$ 2 servers [Leonardi, Raz'97]
  - Stochastic scheduling:
    - Average delay: "Notoriously hard" [Weiss'90'92'95], [Glazebrook et al'99]
    - Limiting regimes: Large system limits, heavy traffic limits, …

Our idea: Seek for near age optimality

Age Lower Bound

- Age of Information:

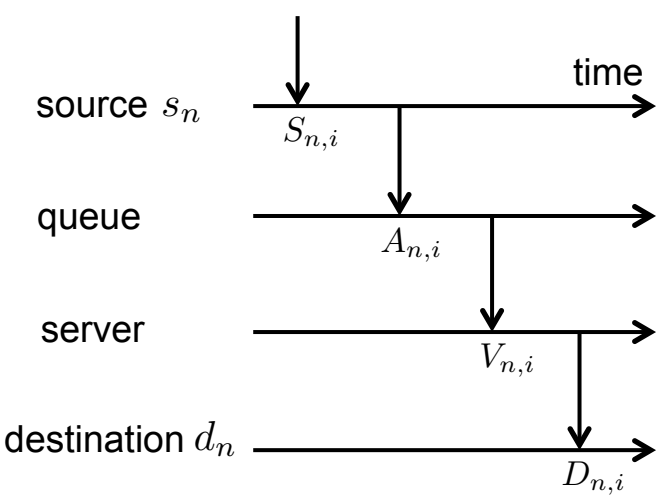
$$\Delta_n(t) = t - \max\{S_{n,i} : D_{n,i} \le t\}$$

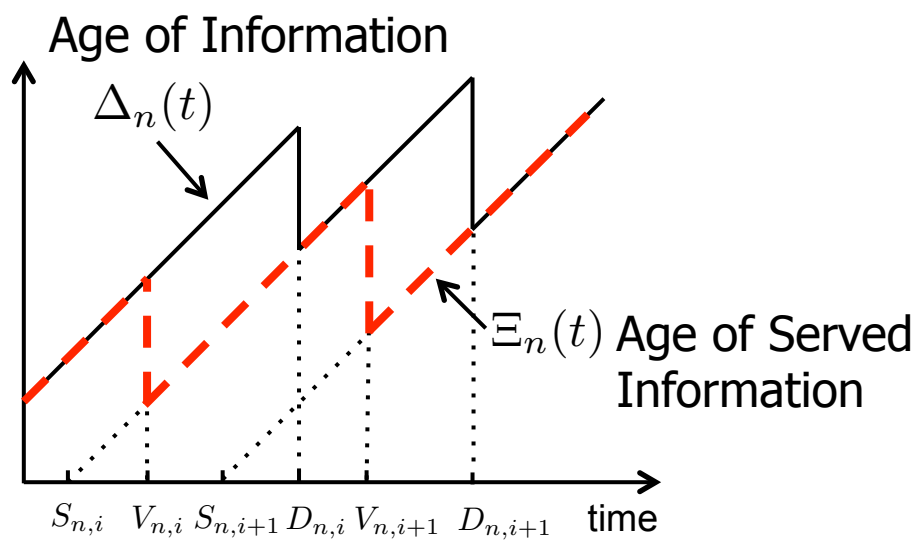
Deliver time

- **Age of Served Information: Age lower bound**

$$\Xi_n(t) = t - \max\{S_{n,i} : V_{n,i} \le t\}$$

Service starting time





Scheduling Policy Design

- How to compare the packets from multiple flows? **Maximum Age of Served Information First (MASIF) discipline**
  - The flow with the maximum Age of Served Information is served the first

- **MASIF + LGFS = Maximum Age of Served Information First, Last Generated First Served (MAISF-LGFS) policy:**
  - The last generated packet from the flow with the maximum Age of Served Information is served the first

Multiple Servers, NBU Service Times

- **New-Better-than-Used (NBU) distribution**:

$$\Pr[X > t] \geq \Pr[X > t + \tau | X > \tau], \ \forall\, t, \tau \geq 0.$$

P{New packet takes more than t sec to complete} $\geq$ P{Old packet takes more than another t sec to complete}

- E.g., exponential, geometric, gamma, negative binomial distributions

Theorem: If (i) packet generation and arrival times are synchronized across the flows, (ii) i.i.d. NBU service times, then for all $p_t \in \mathcal{P}_{\text{sym}}$ and all $\{S_1, S_2, \ldots, A_1, A_2, \ldots\}$, the non-preemptive MAISF-LGFS policy is near optimal for minimizing $\{p_t \circ \mathbf{\Delta}_\pi(t), t \geq 0\}$ in distribution among all non-preemptive causal policies, i.e.,

$$\{p_t \circ \mathbf{\Xi}_{\text{non-prmp, MAISF-LGFS}}(t), t \geq 0\} \leq_{\text{st}} \{p_t \circ \mathbf{\Delta}_\pi(t), t \geq 0\}, \ \pi \in \Pi_{np}$$

where Π_{np} is the set of non-preemptive causal policies.

MASIF-LGFS is near age-optimal in distribution

Small Sub-optimality Gap

- Example: average age $p_{\text{avg}}(\mathbf{\Delta}) = \frac{1}{N}\sum_{n=1}^{N}\Delta_n$, define

$$\bar{\Delta}_\pi = \lim\sup_{T\to\infty}\frac{1}{T}\mathbb{E}[\int_0^T p_{\text{avg}}\circ\mathbf{\Delta}_\pi(t)dt]$$

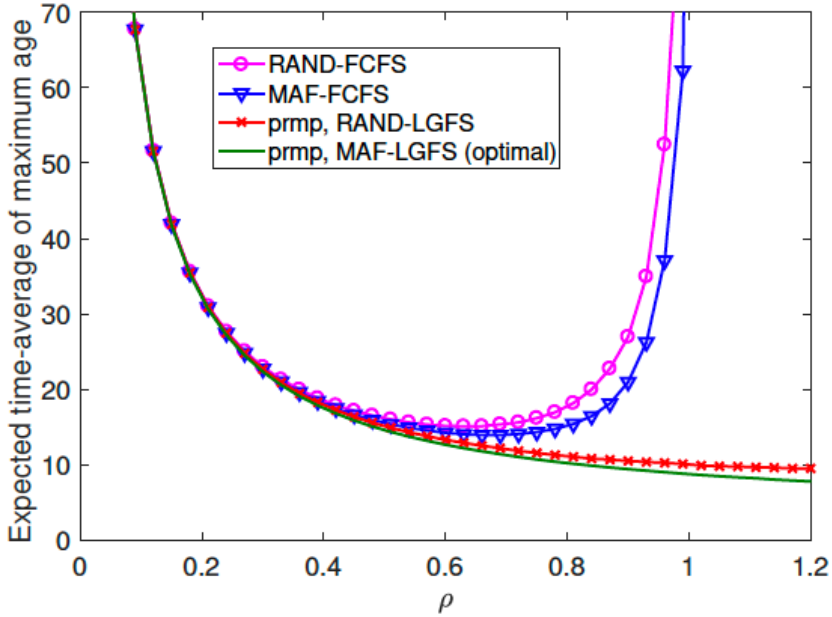
Corollary: Under the above conditions, it holds that

$$\min_{\pi\in\Pi_{np}}\bar{\Delta}_\pi \le \bar{\Delta}_{\text{non-prmp, MASIF-LGFS}} \le \min_{\pi\in\Pi_{np}}\bar{\Delta}_\pi + \mathbb{E}[X]$$

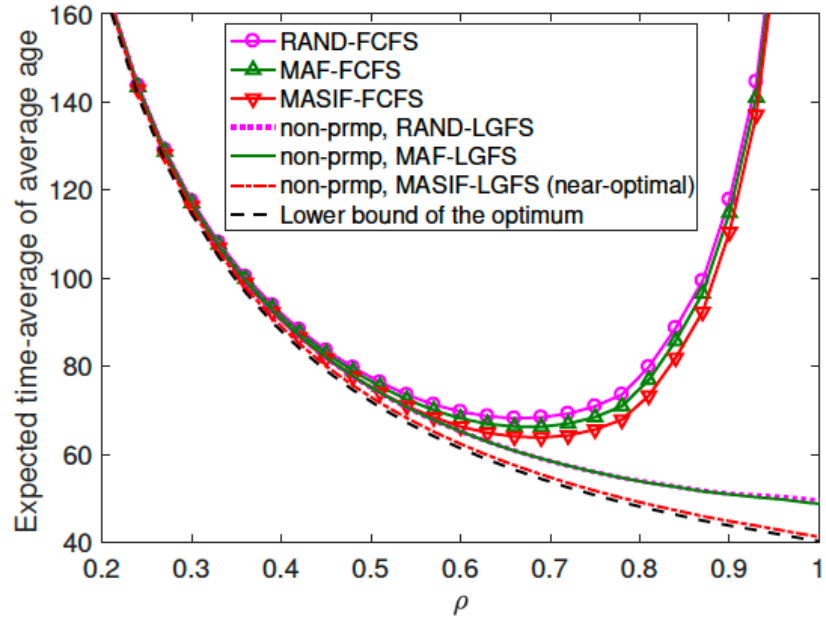
Mean service time

MASIF-LGFS is within a small constant gap from the optimum.
The gap is invariant of arrival process and number of servers

Numerical Results



Max age, 3 flows, 1 server, exp distr.



Avg. age, 50 flows, 3 servers, NBU distr.

Prmp, MAF-LGFS is age-optimal.

Its age is finite even for $\rho > 1$

Non-prmp, MASIF-LGFS is near age-optimal.

Remove stale packets cannot greatly reduce age.

Much better than MAF-LGFS

Relation with Maximum Age Matching

- ***Maximum Age First:***
  - Once a packet starts service, it will stay in the server for a while.
  - The flow still has the maximum age, and will be assigned to other servers
  - Waste server resources
- ***Maximum Age of Served Information First:***
  - Once a packet starts service, the flow's age of served information drops to a small value
  - Another flow with a large age of served information will be assigned to the next idle server
  - discrete-time version: Maximum age matching

Summary

- Multi-flow, single-server, exp service times:
  - MAF-LGFS is age-optimal
- Multi-flow, multi-server, NBU service times:
  - MASIF-LGFS is near age-optimal
  - ***Symmetric***, ***non-decreasing***, and ***time-dependent*** age penalty functions
- Future work:
  - Independent, non-i.d. service times
  - Replication over multi servers
  - Transmission error
- Limitation of sample-path methods:
  - Synchronized arrivals, symmetric age metrics


Thank
You!