IEEE PIMRC 2019, Tutorial 07

Age of Information: Providing Fresh Data to Real-time Applications

Part I

Yin Sun, Auburn University
Elif Uysal, Middle East Technical University

September 8, 2019

Tutorial Outline: Part I

(14:00-15:30 by Yin Sun, 15:30-15:45 coffee break)

- Motivation and Introduction:
  - Why is Fresh Data Important?
  - What is the Age of Information (a.k.a. Age or AoI)?
  - Is Age a Good Metric for Freshness?

- Information Freshness Optimization (3 Topics):
  - Sampling for Optimizing Nonlinear Age Functions
  - Sampling for Remote Signal Estimation
  - Age-Optimal Scheduling in Networks

Tutorial Outline: Part II

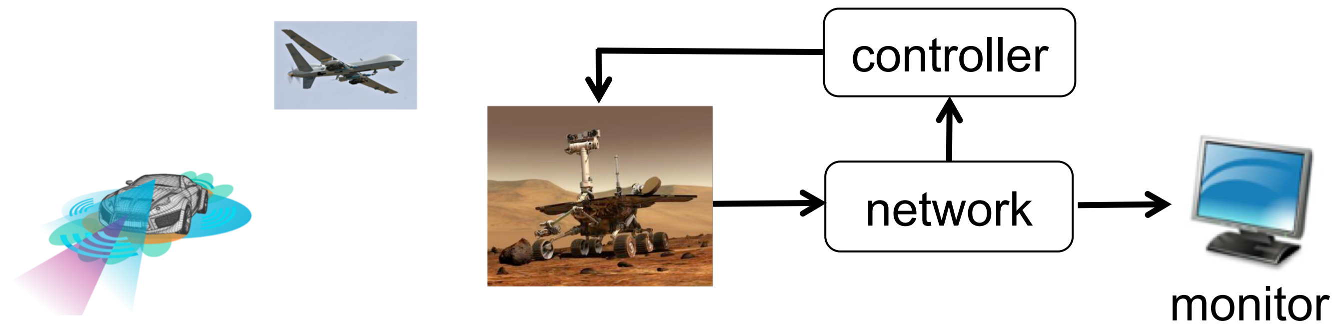
3 Topics: (15:45-17:00 by Elif Uysal)

- Age of Information and Energy
- Age of Information and Channel Coding
- Age of Information in the Internet

Motivation and Introduction

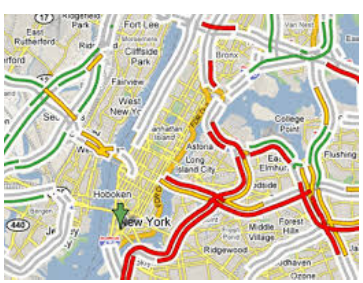
Real-time Services Need Fresh Data

- **Real-time Monitoring and Control**

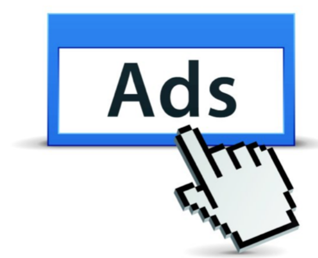


Sensor networks, Internet-of-Things, Cyber-Physical Systems, Robotics, Health, Security, etc.

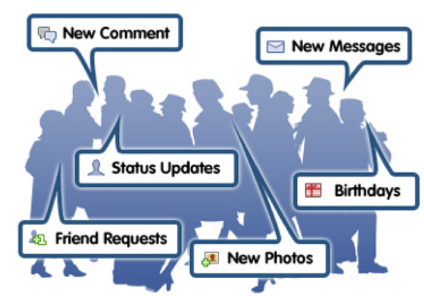
- **Real-time Data Analytics and Learning**



Crowdsourcing

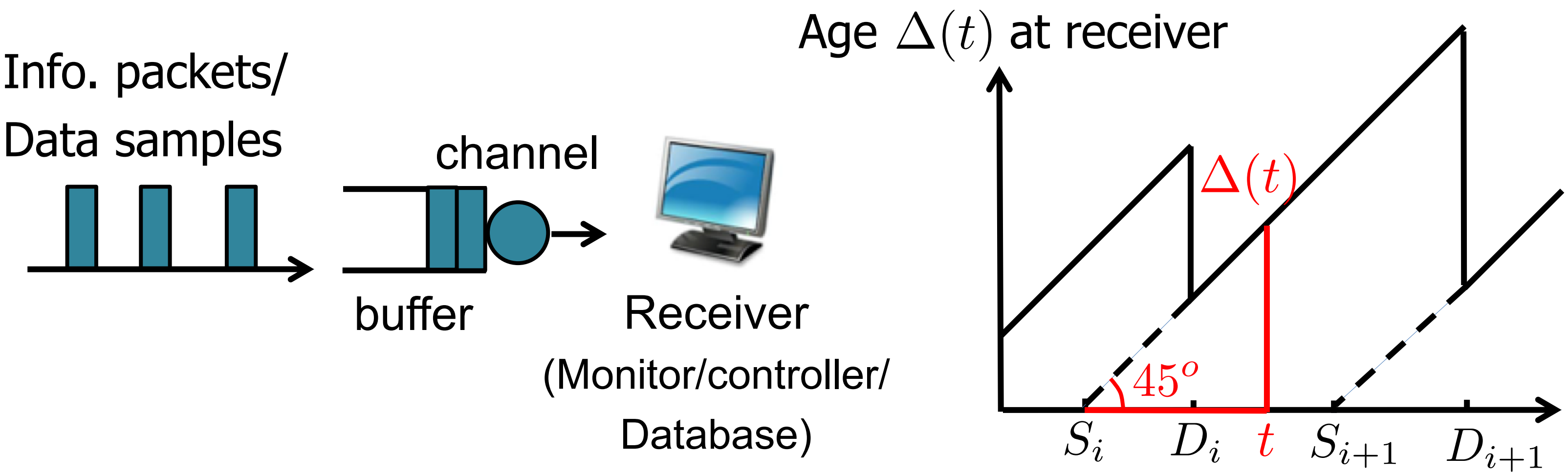


Online Learning (ads bidding)



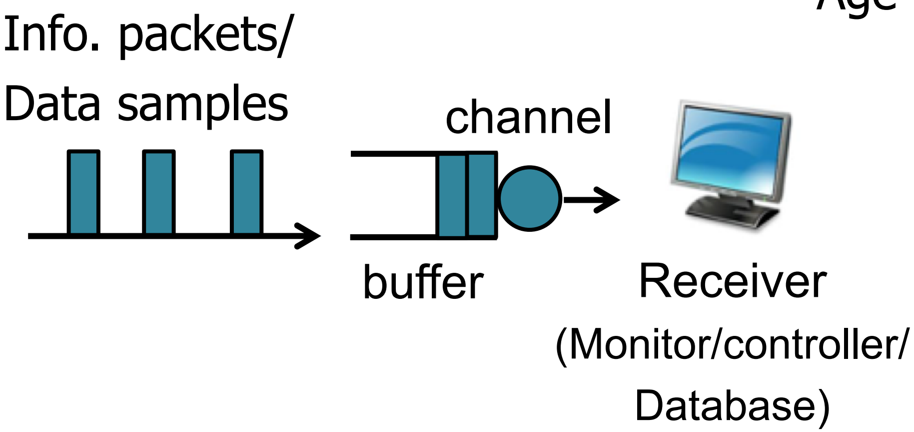
Social networks

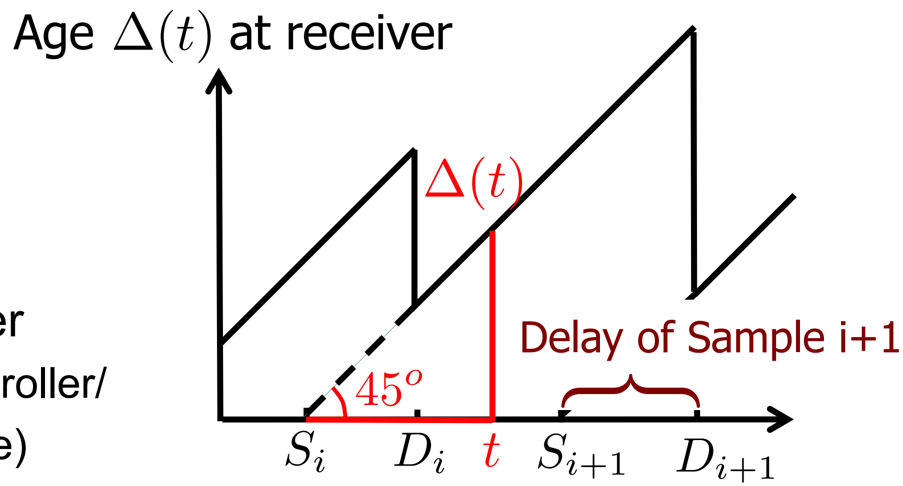
Data Freshness Metric: Age of Information



- In real-time applications, fresh data is more important than stale data
  - E.g., UAV/vehicle/robotic control, wild fire/tornado monitoring …

Age of Information: Definition



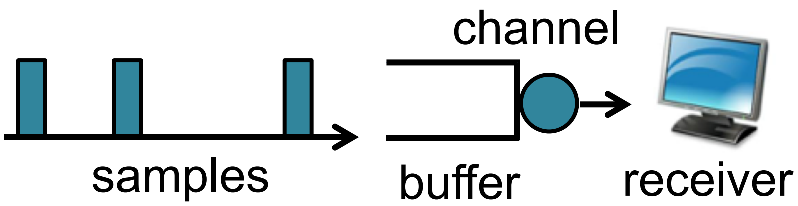


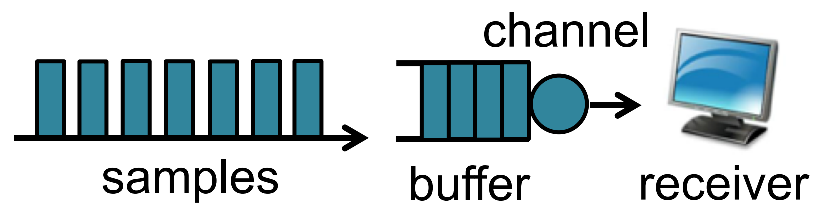
Definition: At time t, the **Age of Information** $\Delta(t)$ is time difference between the current time t and the generation time of the latest received data sample

- If sample i is generated at S_i and delivered at D_i

$$\Delta(t) = \underbrace{t - \max\{S_i : D_i \le t\}}_{\text{Time difference between data generation and usage}}$$

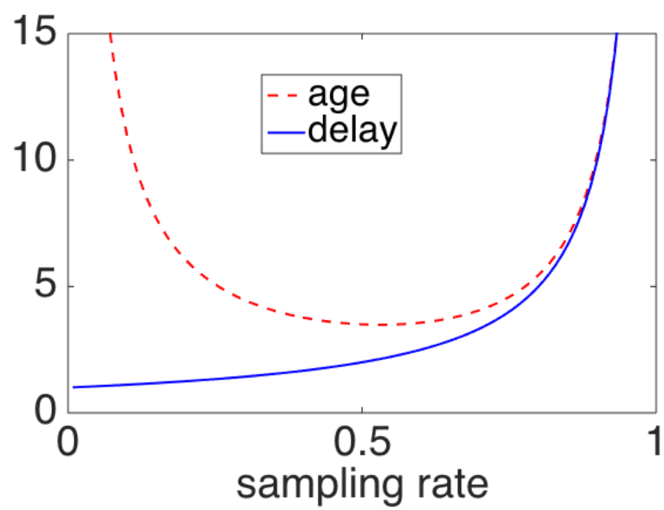
Difference between Age & Delay





In M/M/1 FIFO queues: [Kaul, Yates, Gruteser'12]

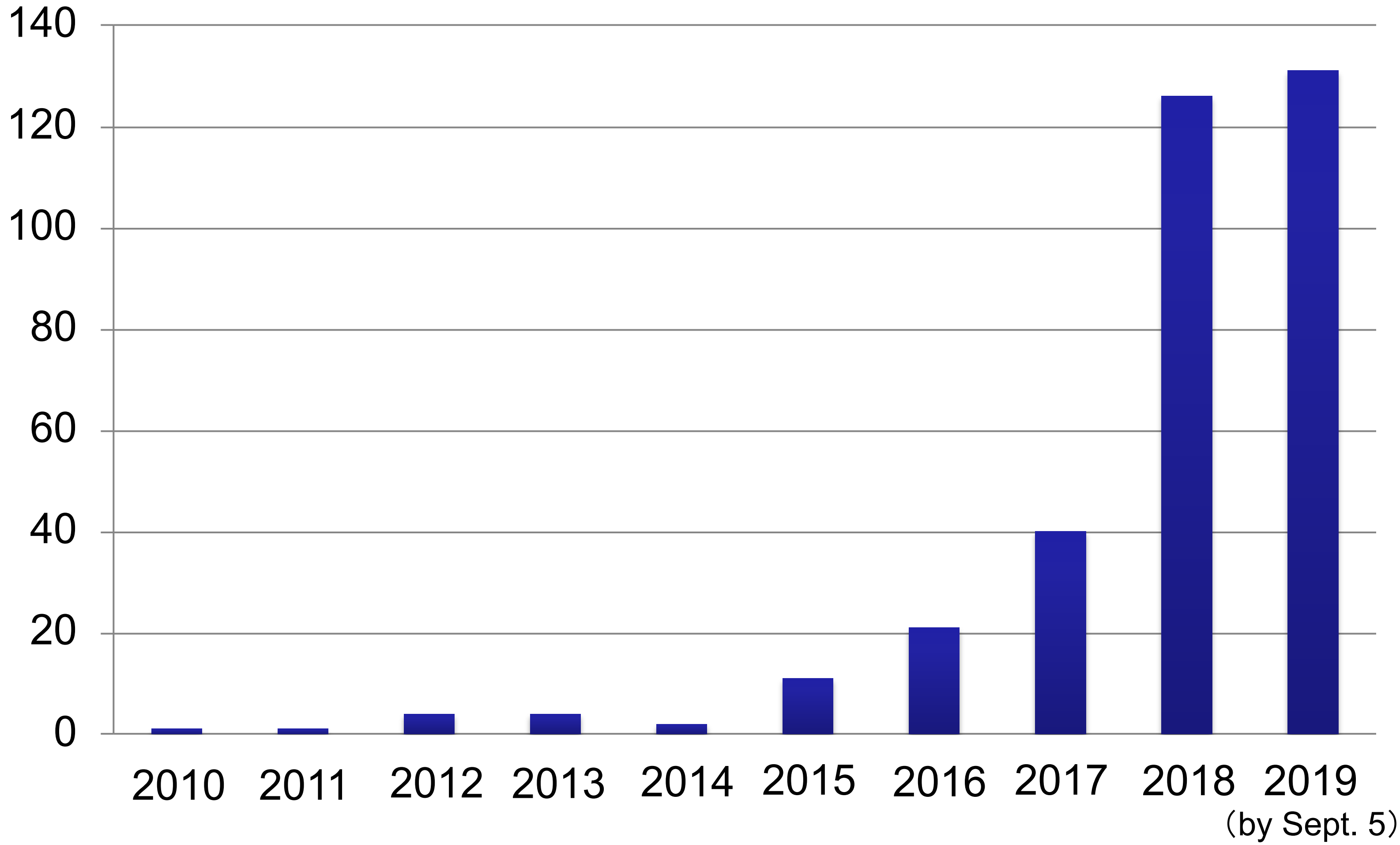
- Age first decreases, then increases with sampling rate
- Delay increases with sampling rate



A Brief History on Age of Information

- Defined in real-time databases [Song-Liu'90,…]
- Queueing analysis [Kaul-Gruteser-Rai-Kenney'11, Yates-Kaul'12, Kam-Kompella-Ephremides'13,14, …]
- Scheduling and age minimization [Altman et al.'10(19), Yates'15, He-Yuan-Ephremides'16, Sun et al.'16, Bedewy-Sun-Shroff'16&17,…]
- Source coding [Zhong-Yates'16, Zhong-Yates-Soljanin'17,…]
- Channel coding [Najm-Yates-Soljanin'17, Yates-Najm-Soljanin-Zhong'17,…]
- Energy harvesting [Bacinoglu-Ceran-Uysal'15, Wu-Yang-Wu'17, Bacinoglu-Uysal'17, Arafa-Ulukus'17,…]
- Freshness of Channel State Information [Costa-Valentin-Ephremides'15, Farazi-Klein-Brown'17,…]
- Caching [Yates-Ciblat-Yener-Wigger'17, Kam et al.'17]
- Game theory [Nguyen et al. 17, Xiao-Sun'18]
- Age of information and Information Theory [Baknina-Ozel-Yang-Ulukus-Yener'18]
- Age of information and learning [Sert-Sonmez-Baghaee-Uysal-Biyikoglu'18, Ceran-Gunduz-Gyorgy'18]
- Age of information and Control [Zhang-Wang'18, Soleymani-Baras-Johansson'18,…]
- Human in the loop [Bastopcu-Ulukus'18, Bin-Liu'19]

Number of papers on AoI



Sources: Google Scholar

AoI paper list: auburn.edu/~yzs0078/AoI.html

Is Age a Good Metric for Freshness?

- **Age:** Time difference between data generation and usage
- **Issue:**
  - Some information changes fast (e.g., car location)
    - Update frequently (sub-second), more bandwidth
  - Other information changes slowly (e.g., temperature)
    - Update infrequently (hourly), less bandwidth

Tony's Question

- Tony asked a question in the Open Problem Session at ITA 2015
  - …There is a correlation structure in the signal. Clearly, the actual age is not a good representation of freshness.
  - What is freshness, can we get an Information Theoretical measure of freshness?
  - There should be a concept of effective age…

Anthony Ephremides

A Possible Answer: Nonlinear Age Functions

- Penalty function $p(\Delta(t))$: p can be any non-decreasing function
  - Dissatisfaction for data staleness, eagerness for refreshing

- Utility function $u(\Delta(t))$: u can be any non-increasing function
  - Utility value of fresh data

- Examples:
  - Temporal Auto-Correlation Function $u(\Delta(t)) = |\mathbb{E}[X_t^* X_{t-\Delta(t)}]|$ [Kosta-Pappas-Ephremides-Angelakis'ISIT2017]
    - For example, auto-correlation function of CSI is a non-increasing age function for small age [Truong-Health'JCN2013]

Nonlinear Age Functions (cont'd)

- More Examples:
  - Estimation Error of Real-time Signal:

Estimate signal value based on causally received samples.

- If sampling times are independent of the signal, estimation error is an age penalty function. [Sun-Polyanskiy-Uysal'ISIT17]
- Wiener process: [Yates-Kaul'ISIT12, Sun-Polyanskiy-Uysal'ISIT17]

$$\mathbb{E}[(W_t - \hat{W}_t)^2] = \mathbb{E}[W^2_{\Delta(t)}] = \Delta(t)$$

- Ornstein-Uhlenbeck process: continuous-time AR (1) [Ornee-Sun'19]

$$\mathbb{E}\left[(X_t - \hat{X}_t)^2\right] = \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta\Delta(t)}\right) = p(\Delta(t))$$

- **Note:** If sampling times are determined based on causal knowledge of the signal, estimation error is not necessarily a function of age.

Nonlinear Age Functions (cont'd)

- More Examples:
  - Mutual Information and Conditional Entropy:
    - Let $\mathbf{W}_t$ be the set of samples received up to time t
    - Mutual information: $I(X_t; \mathbf{W}_t)$ amount of information contained in $\mathbf{W}_t$ [Sun-Cyr'SPAWC18]
    - Conditional Entropy: $H(X_t|\mathbf{W}_t)$ uncertainty after knowing $\mathbf{W}_t$ [Soleymani-Hirche-Baras'CDC, ECC, WODES2016]
    - If sampling times are independent of the signal and signal is a stationary Markov chain, mutual information is an age utility function, conditional entropy is an age penalty function [Sun-Cyr'JCN2019]
    - Otherwise, they are not necessarily functions of age

Nonlinear Age Functions (cont'd)

- More Examples:
  - Estimation Error in Feedback Control System:

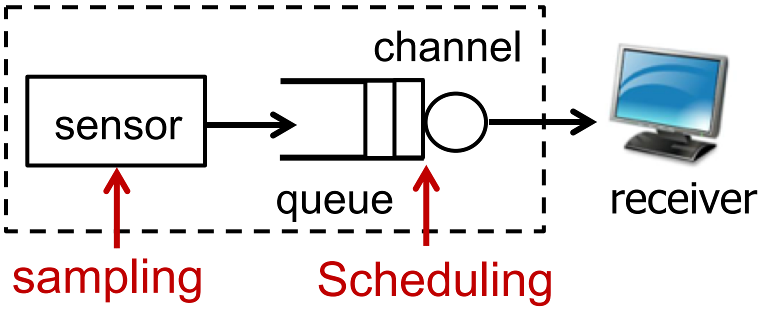
$$X_{t+1} = AX_t + BU_t + N_t$$

    - Samples of the system state X_t are sent to the controller.
    - The controller determines U_t based on causally received samples.
    - The estimation error $\mathbb{E}[(X_t - \hat{X}_t)^2]$ is independent of control.
    - If sampling times are independent of the signal, $\mathbb{E}[(X_t - \hat{X}_t)^2]$ is an age penalty function [Champati et. al'19, Klugel et. al'19]; otherwise, it is not necessarily a function of age.
- Nonlinear age functions date back to 2003. A short survey available in [Sun-Cyr'JCN2019]
- **Summary:** Nonlinear age functions are simple enough and general enough, but not always perfect

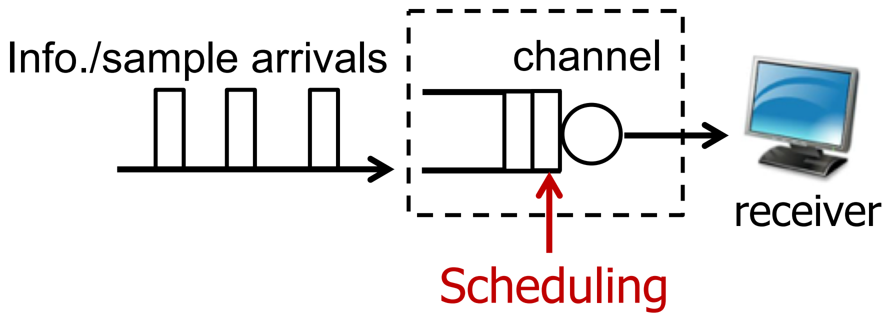
Sun and Cyr, "Sampling for Data Freshness Optimization: Non-linear Age Functions," JCN special issue on Age of Info., 2019.

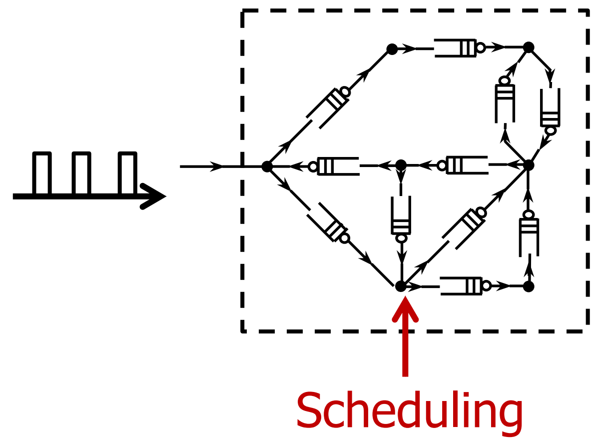
3 Topics on Information Freshness Optimization

- Optimal Sampling and Scheduling



  - From AoI to Sampling theory
- Optimal Scheduling Only





  - Given arrival process
  - optimal scheduling in networks with multiple channels, hops, and/or sources

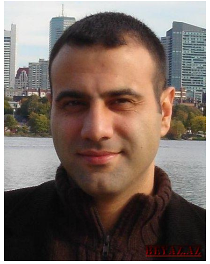
Topic I

Sampling for Optimizing Nonlinear Age Functions

Joint work with

Elif Uysal

Roy D. Yates

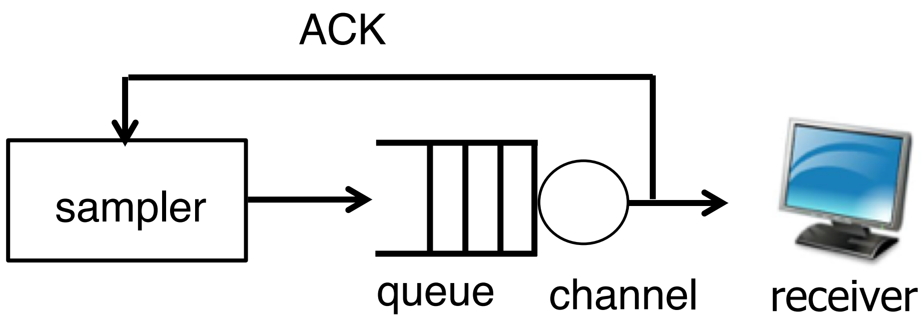


C. Emre Koksal

Ness B. Shroff.

Benjamin Cyr

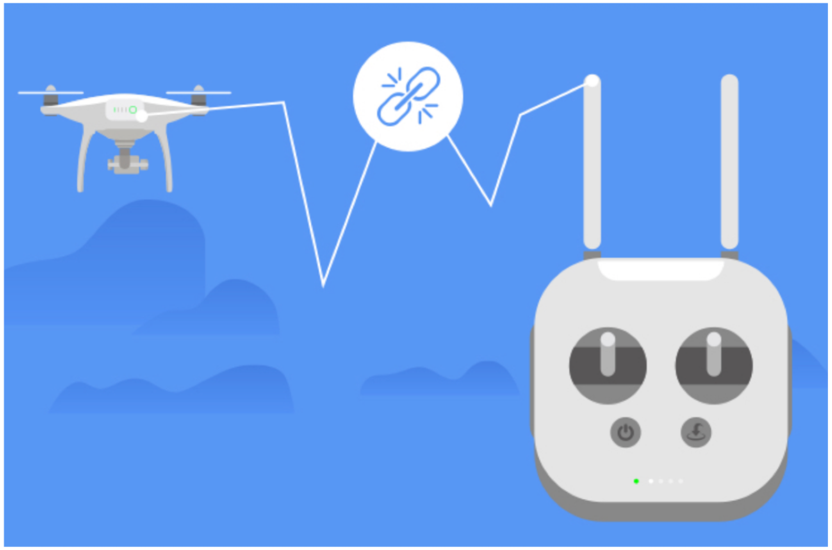
Model:
Fresh Samples through Queues



- Two terminals: **Sampler** sends samples to **Receiver**
- **Channel**: FIFO queue with *i.i.d.* transmission times Y_i
- **Feedback:** ACK
  - Zero delay ➔ sampler knows channel state (idle/busy)

- Sample i is generated at S_i, with channel transmission time Y_i
  - $Y_i \geq 0$ has a **general** distribution
  - Can be discrete/continuous random variable (finite moments)
  - Robustness of CPS under occasionally long commun. delay

Example:
Remote Drone Control Under Interference



Long communication delay
due to interference and jamming

Source: www.dji.com

Source: www.battelle.org

Problem: Continuous-time Sampling

$$\inf_{\pi \in \Pi} \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\int_0^T \overbrace{p(\Delta(t))}^{\text{age penalty function}} dt\right]$$

Expected time-average age penalty

$$\text{s.t.} \quad \liminf_{n \to \infty} \frac{1}{n} \mathbb{E}\left[\sum_{i=1}^{n} \underbrace{(S_{i+1} - S_i)}_{\text{inter-sample time}}\right] \geq \frac{1}{f_{\max}}$$

Avg. sampling-rate constraint

- $p(\Delta(t))$ is arbitrary non-decreasing function of the age $\Delta(t)$
- Age utility maximization handled by choosing $p(\Delta(t)) = -u(\Delta(t))$
- Sampling policy: $\pi = (S_1, S_2, \ldots)$ sequence of sampling times
- Space of causal policies: Π

  Sampling time S_i decided by the history information of the system

Question: Which sampling policy minimizes the age penalty?

Problem: Continuous-time Sampling

$$\inf_{\pi\in\Pi} \quad \limsup_{T\to\infty} \frac{1}{T}\,\mathbb{E}\left[\int_0^T \overbrace{p(\Delta(t))}^{\text{age penalty function}} dt\right]$$

Expected time-average age penalty

$$\text{s.t.} \quad \liminf_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} \underbrace{(S_{i+1} - S_i)}_{\text{inter-sample time}}\right] \geq \frac{1}{f_{\max}}$$

Avg. sampling-rate constraint

- Constrained continuous-time MDP with a continuous state space
  - Usually difficult to solve

Unconstrained Continuous-time Sampling

Theorem 1. For unconstrained problem, optimal sampling times are given by

$$S_{i+1}(\beta) = \inf\{\underbrace{t \geq D_i(\beta)}_{\text{After previous sample is delivered}} : \underbrace{\mathbb{E}[p(\Delta(t+Y_{i+1}))] \geq \beta}_{\text{Expected age penalty exceeds threshold}}\}$$

Deterministic threshold policy

where $D_i(\beta) = S_i(\beta) + Y_i$ and β is the root of

$$\beta = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} p(\Delta(t))dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]} = \bar{p}_{\text{opt}}$$

Bisection search (no "curse of dimensionality")

Threshold = optimal objective value

General penalty function, general service time distribution, with finite moments

$$\bar{p}_{\text{opt}} = \inf_{\pi \in \Pi} \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\int_0^T p(\Delta(t))dt\right]$$

Structure more insightful than [Sun-Uysal-Yates-Koksal-Shroff, TIT17]

Sun and Cyr, "Sampling for Data Freshness Optimization: Non-linear Age Functions," JCN special issue on Age of Info., 2019.

Bisection Search for β

- **Goal:** Solve $\beta = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} p(\Delta(t))dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}$

- **Method:** Bisection search

Algorithm 1 Bisection method for solving (17)

given l, u, tolerance $\epsilon > 0$.
repeat
  $\beta := (l+u)/2$.
  $o := \beta - \frac{\mathbb{E}[v(D_{i+1}(\beta) - S_i(\beta)) - v(Y_i)]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}$.
  if $o \geq 0$, $u := \beta$; **else**, $l := \beta$.
until $u - l \leq \epsilon$.
return β.

- where $v(t) = \int_0^t p(\Delta(s))ds$

$$\int_{D_i(\beta)}^{D_{i+1}(\beta)} p(\Delta_t)dt = v(D_{i+1}(\beta) - S_i(\beta)) - v(Y_i)$$

Constrained Continuous-time Sampling

Theorem 2. The optimal sampling times are given by **Theorem 1** if

$$\mathbb{E}[S_{i+1}(\beta) - S_i(\beta)] > \frac{1}{f_{\max}}$$

Otherwise, $S_{i+1}(\beta) = \begin{cases} T_{i,\min}(\beta) \text{ with probability } \lambda, \\ T_{i,\max}(\beta) \text{ with probability } 1-\lambda, \end{cases}$

Randomized threshold policy

where $T_{i,\min}(\beta) = \inf\{t \geq D_i(\beta) : \mathbb{E}[p(\Delta(t+Y_{i+1}))] \geq \beta\},$

$T_{i,\max}(\beta) = \inf\{t \geq D_i(\beta) : \mathbb{E}[p(\Delta(t+Y_{i+1}))] > \beta\},$

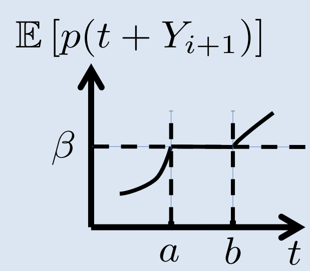
β is determined by solving

Bisection search. No curse of dimensionality

$$\mathbb{E}[T_{i,\min}(\beta) - S_i(\beta)] \leq \frac{1}{f_{\max}} \leq \mathbb{E}[T_{i,\max}(\beta) - S_i(\beta)],$$

Then, λ is given by

$$\lambda = \frac{\mathbb{E}[T_{i,\max}(\beta) - S_i(\beta)] - \frac{1}{f_{\max}}}{\mathbb{E}[T_{i,\max}(\beta) - T_{i,\min}(\beta)]}.$$



Sun and Cyr, "Sampling for Data Freshness Optimization: Non-linear Age Functions," JCN special issue on Age of Info., 2019.

Constrained Continuous-time Sampling

Corollary. Suppose $p(\cdot)$ is strictly increasing. Then, the optimal sampling times are given by **Theorem 1** if

$$\mathbb{E}[S_{i+1}(\beta) - S_i(\beta)] > \frac{1}{f_{\max}}$$

Otherwise, $S_{i+1}(\beta) = \inf\{t \geq D_i(\beta) : \mathbb{E}[p(\Delta(t + Y_{i+1}))] \geq \beta\}$

two thresholds are the same

where β is determined by solving

$$\mathbb{E}[S_{i+1}(\beta) - S_i(\beta)] = \frac{1}{f_{\max}}$$

Bisection search. No curse of dimensionality

Sun and Cyr, "Sampling for Data Freshness Optimization: Non-linear Age Functions," JCN special issue on Age of Info., 2019.

Unconstrained Discrete-time Sampling

Theorem 3. The optimal sampling times are given by

$$S_{i+1}(\beta)=\inf\{\underbrace{t\in\mathbb{N}}_{\text{discrete-time}}:\underbrace{t\geq D_i(\beta)}_{\text{After previous sample is delivered}},\underbrace{\mathbb{E}[p(\Delta(t+Y_{i+1}))]\geq\beta}_{\text{Expected age penalty exceeds threshold}}\}$$

Deterministic threshold policy

where $D_i(\beta)=S_i(\beta)+Y_i$ and β is the root of

$$\beta=\frac{\mathbb{E}\left[\sum_{t=D_i(\beta)}^{D_{i+1}(\beta)-1}p(\Delta(t))\right]}{\mathbb{E}[D_{i+1}(\beta)-D_i(\beta)]}=\bar{p}_{\text{opt}}$$

Threshold = optimal objective value

Bisection search

No curse of dimensionality

$$\bar{p}_{\text{opt}}=\inf_{\pi\in\Pi}\ \limsup_{T\to\infty}\frac{1}{T}\ \mathbb{E}\left[\sum_{t=1}^{T}p(\Delta(t))\right]$$

t are integers.

Scalable with clock rate

Sun and Cyr, "Sampling for Data Freshness Optimization: Non-linear Age Functions," JCN special issue on Age of Info., 2019.

Constrained Discrete-time Sampling

Theorem 4. The optimal sampling times are given by **Theorem 3** if

$$\mathbb{E}[S_{i+1}(\beta) - S_i(\beta)] > \frac{1}{f_{\max}}$$

Otherwise,
$$S_{i+1}(\beta) = \begin{cases} T_{i,\min}(\beta) \text{ with probability } \lambda, \\ T_{i,\max}(\beta) \text{ with probability } 1-\lambda, \end{cases}$$

Randomized threshold policy

where
$$T_{i,\min}(\beta) = \min\{t \in \mathbb{N} : t \geq D_i(\beta), \mathbb{E}\left[p(\Delta(t+Y_{i+1}))\right] \geq \beta\},$$
$$T_{i,\max}(\beta) = \min\{t \in \mathbb{N} : t \geq D_i(\beta), \mathbb{E}\left[p(\Delta(t+Y_{i+1}))\right] > \beta\},$$

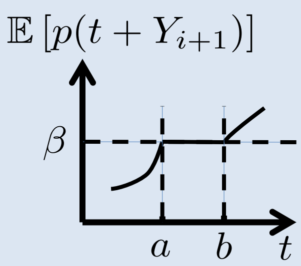
β is determined by solving

Bisection search. No curse of dimensionality

$$\mathbb{E}[T_{i,\min}(\beta) - S_i(\beta)] \leq \frac{1}{f_{\max}} \leq \mathbb{E}[T_{i,\max}(\beta) - S_i(\beta)],$$

Then, λ is given by

$$\lambda = \frac{\mathbb{E}[T_{i,\max}(\beta) - S_i(\beta)] - \frac{1}{f_{\max}}}{\mathbb{E}[T_{i,\max}(\beta) - T_{i,\min}(\beta)]}.$$



Sun and Cyr, "Sampling for Data Freshness Optimization: Non-linear Age Functions," JCN special issue on Age of Info., 2019.

Is the optimal sampling policy deterministic or randomized?

Continuous or discrete time	Constrained or not	Age penalty function	Deterministic policy or not
Continuous time	unconstrained	Non-decreasing	Deterministic
Continuous time	constrained	Non-decreasing	Randomized
Continuous time	constrained	Strict increasing	Deterministic
discrete time	unconstrained	Non-decreasing	Deterministic
discrete time	constrained	Non-decreasing	Randomized

- Why randomized?
  - If one policy cannot meet the sampling rate constraint with equality, a randomized mixture of two policies can.

Three Equivalent Threshold Policies

- **Threshold policy on age penalty:**

$$S_{i+1}(\beta) = \inf\{t \geq D_i(\beta) : \mathbb{E}[p(\Delta(t + Y_{i+1}))] \geq \beta\}$$

  - Define the threshold function

$$w(\beta) = \inf\{\Delta \geq 0 : \mathbb{E}[p(\Delta + Y_{i+1})] \geq \beta\}$$

- **Threshold policy on age:**

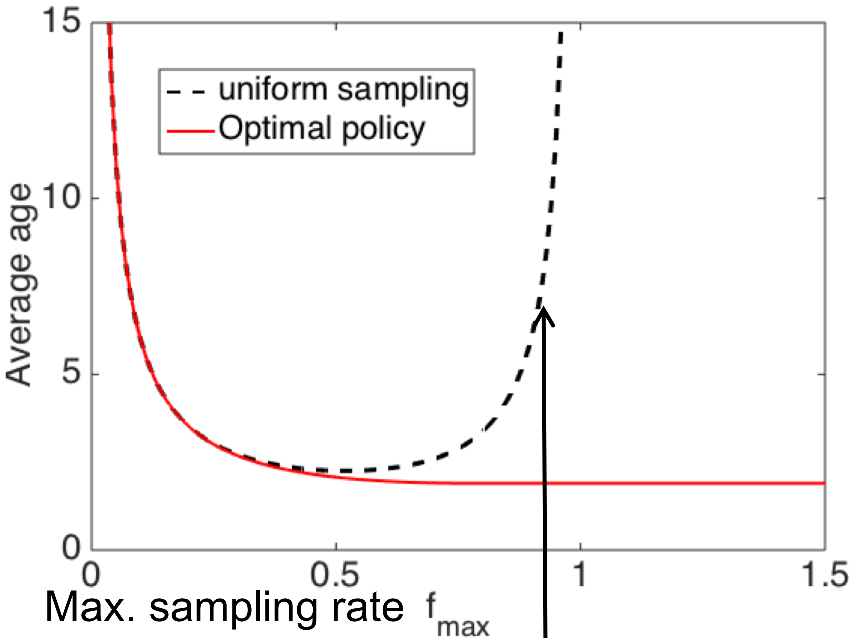
$$S_{i+1}(\beta) = \inf\{t \geq D_i(\beta) : \Delta(t) \geq w(\beta)\}$$

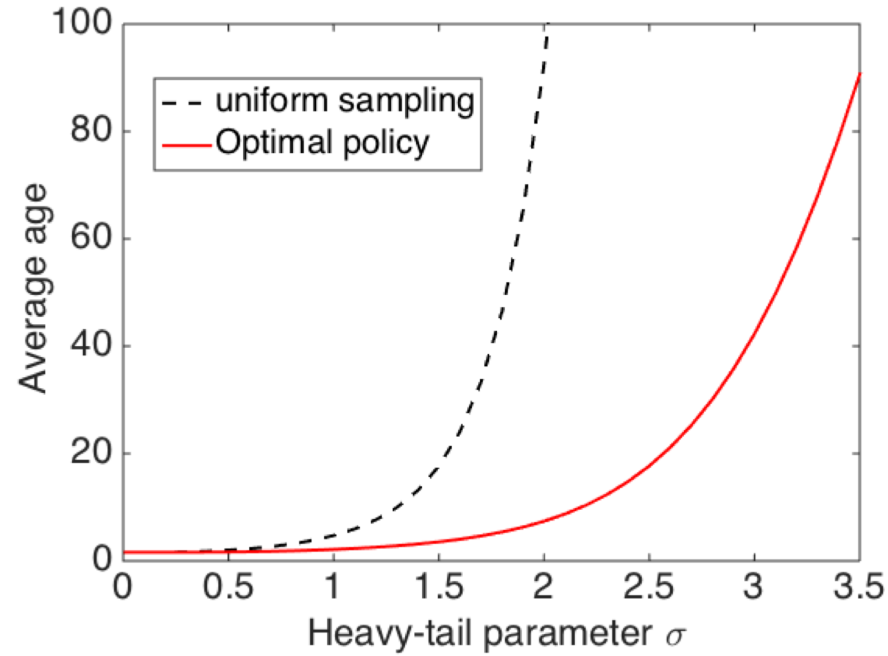
- **Water-filling solution for waiting time:**

$$Z_i(\beta) = \max\{w(\beta) - Y_i, 0\}$$

  - where $Z_i(\beta) = S_{i+1}(\beta) - D_i(\beta)$ is the waiting time after previous sample is delivered
  - All of them are insightful, appeared in different studies

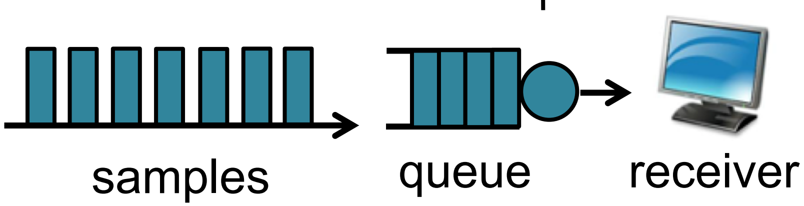
Uniform Sampling can be FAR from Optimal





i.i.d. exponential trans. times $\mu = 1$

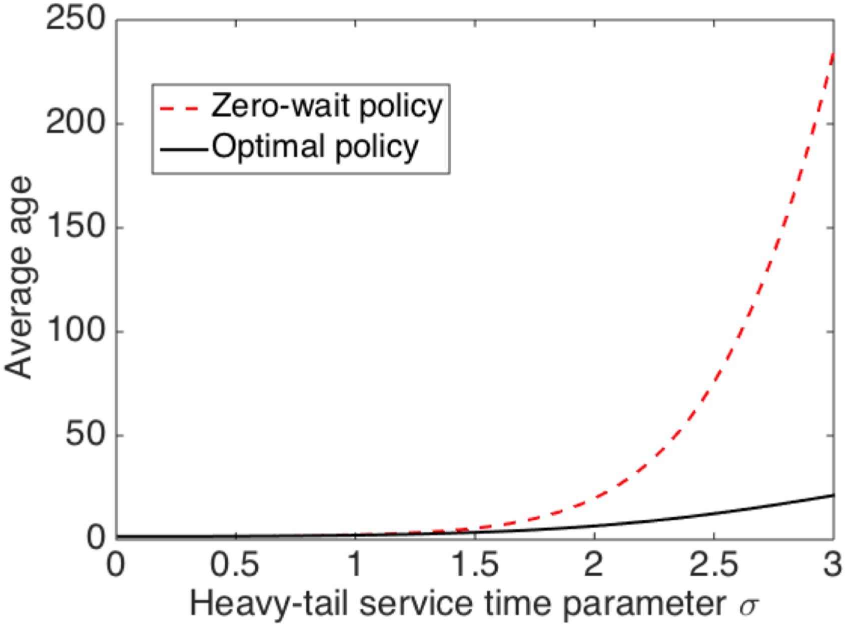
i.i.d. log-normal trans. times $\mu = 1$



Become stale while waiting

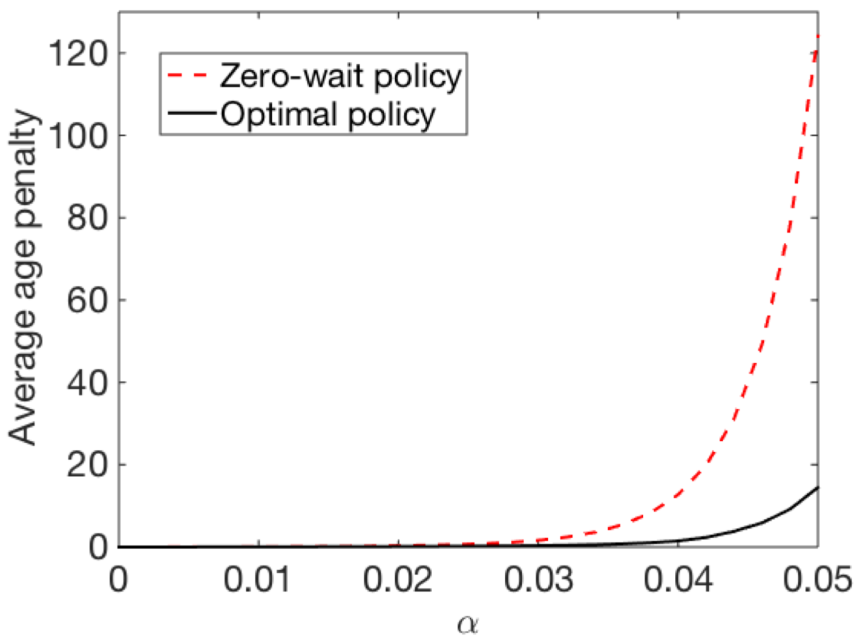
Uniform sampling is far from optimal if
(1) Sampling rate is high
(2) Transmission time is highly random

Zero-Wait can be FAR from Optimal



i.i.d. log-normal trans. times $\mu = 1$

$$S_{i+1} = S_i + Y_i$$



i.i.d. log-normal trans. times $\mu = 1$

$$g(\Delta) = e^{\alpha\Delta} - 1$$

Zero-wait is far from optimal if

(1) Transmission time is highly random

(2) Age penalty grows quickly with age (impatient user, real-time control)

Summary

- Developed optimal sampling policies for optimizing time-average nonlinear age functions

- The optimal policy is a deterministic or randomized threshold policy and the optimal threshold is characterized exactly
  - Both continuous-time & discrete-time sampling
  - Both with & without sampling rate constraint
  - General service time distributions
  - General monotonic age functions
    - Represents the need of real-time applications

- Uniform sampling and Zero-Wait sampling can be far from optimal

Topic II

Sampling for Remote Signal Estimation

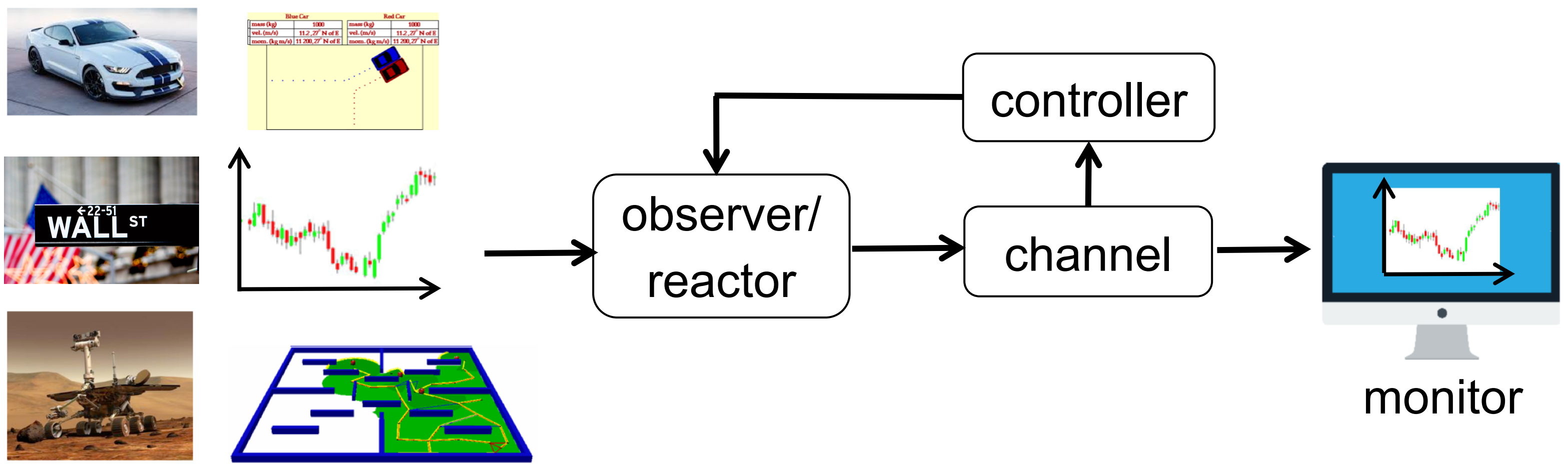
Joint work with

Elif Uysal

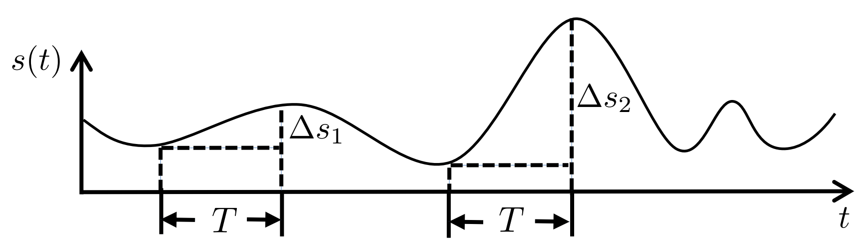
Yury Polyanskiy

Tasmeen Zaman Ornee

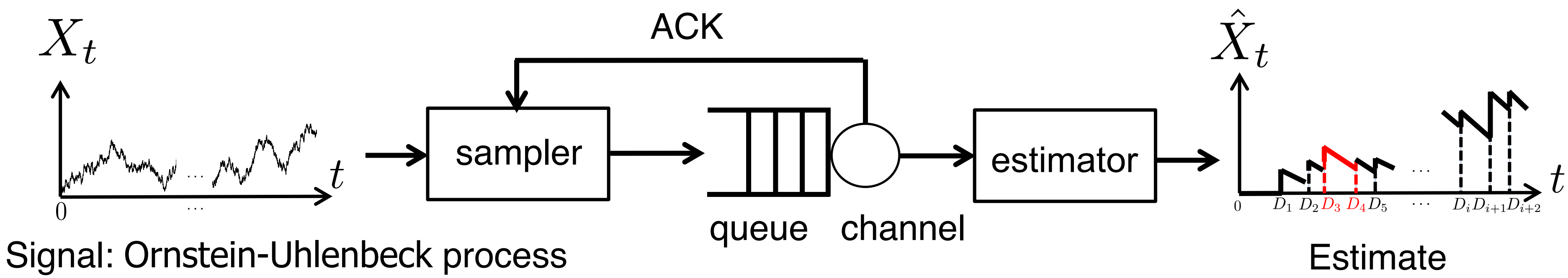
Motivation: Real-time Monitoring & Control



- Real-time information usually in the form of a signal
- Age-of-information: time-difference between source and destination
  - time-difference $\neq$ signal-difference



Model: Timely **Signal** Updates



- **Sampler**: Take samples of an Ornstein-Uhlenbeck (OU) process X_t
  - continuous-time analogue of Autoregressive(1) model
  - Defined by $dX_t = \theta(\mu - X_t)dt + \sigma dW_t$
    - $\theta > 0, \sigma > 0$ and W_t is Wiener process
- **MMSE Estimator**: Use causally received samples to estimate the real-time value of X_t
- **Channel**: FIFO queue with *i.i.d.* service times
  - ACK: Zero delay

Goal: Find optimal sampling policy that minimizes **estimation error** between signal X_t and estimate $\hat{X}_t$

Problem: Optimal Sampling

$$\min_{\pi\in\Pi} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T \overbrace{(X_t-\hat{X}_t)^2 dt}^{\text{Estimation error}}\right]$$

Time-avg. estimation error (MSE)

$$\text{s.t.} \liminf_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} \underbrace{(S_{i+1}-S_i)}_{\text{inter-sample time}}\right] \geq \frac{1}{f_{\max}}$$

Avg. sampling-rate constraint

- Sampling policy: $(S_1, S_2, ...)$
  - Causal policy space Π: Sampling time S_i decided by the history of signal and channel processes (in math, stopping time)
  - Continuous-time MDP with continuous state space and a constraint
    - Quite challenging to solve

Questions: Which sampling policy is optimal?

What is the relationship with the age of information?

Question 1: Which sampling policy is optimal?

Theorem. For unconstrained problem, a **threshold-based sampler** is optimal:

$$S_{i+1}(\beta) = \inf \Big\{ \underbrace{t \geq D_i(\beta)}_{\text{After previous sample is delivered}} : \underbrace{\big|X_t - \hat{X}_t\big| \geq v(\beta)}_{\text{Signal difference exceeds threshold}} \Big\},$$

Threshold function

After previous sample is delivered   Signal difference exceeds threshold

The optimal β is determined by solving:

$$\beta = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]} = \text{mse}_{\text{opt}}$$

Bisection search
(no "curse of dimensionality")

β = optimal objective value

$$G(x) = \frac{e^{x^2}}{x}\int_0^x e^{-t^2} dt = \frac{e^{x^2}}{x}\frac{\sqrt{\pi}}{2}\operatorname{erf}(x)$$

$$v(\beta) = \frac{\sigma}{\sqrt{\theta}} G^{-1}\left(\frac{\text{mse}_\infty - \text{mse}_{Y_i}}{\text{mse}_\infty - \beta}\right)$$

$$\text{mse}_{Y_i} = \frac{\sigma^2}{2\theta}\mathbb{E}[1 - e^{-2\theta Y_i}] \qquad \text{mse}_\infty = \frac{\sigma^2}{2\theta}$$

$$\text{mse}_{Y_i} \leq \text{mse}_{\text{opt}} \leq \beta \leq \text{mse}_\infty$$

$$\text{mse}_{\text{opt}} = \min_{\pi\in\Pi}\ \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T (X_t - \hat{X}_t)^2 dt\right]$$

Ornee and Sun, "Sampling for Remote Estimation through Queues: Age of Information and Beyond," IEEE WiOpt, 2019, Best Paper Award.

Evaluating Expectations

By Dynkin's formula and optional stopping theorem, we get

$$
\begin{aligned}
&\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] \\
=&\mathbb{E}[\max\{R_1(v(\beta)) - R_1(O_{Y_i})\}, 0] + \mathbb{E}[Y_i], \\
&\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right] \\
=&\mathbb{E}[\max\{R_2(v(\beta)) - R_2(O_{Y_i})\}, 0] \\
&+ \text{mse}_\infty[\mathbb{E}(Y_i) - \gamma] + \mathbb{E}\left[\max\{v^2(\beta), O_{Y_i}^2\}\right]\gamma,
\end{aligned}
$$

where
$$
\gamma = \frac{1}{2\theta}\mathbb{E}[1 - e^{-2\theta Y_i}],
$$

$$
R_1(v) = \frac{v^2}{\sigma^2}\,{}_2F_2\left(1, 1; \frac{3}{2}, 2; \frac{\theta}{\sigma^2}v^2\right),
$$

$$
R_2(v) = -\frac{v^2}{2\theta} + \frac{v^2}{2\theta}\,{}_2F_2\left(1, 1; \frac{3}{2}, 2; \frac{\theta}{\sigma^2}v^2\right),
$$

Wiener Process vs. Ornstein-Uhlenbeck Process

Theorem: (Wiener Process)

The MSE-optimal sampling policy is

$$S_{i+1}(\beta) = \inf\left\{t \geq D_i(\beta) : \underbrace{\left|X_t - \hat{X}_t\right|}_{\text{instantaneous estimation error}} \geq v(\beta)\right\}$$

optimal β is determined by solving

$$\beta = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]} = \text{mse}_{\text{opt}}$$

β = optimal objective value

threshold function

$$v(\beta) = \sqrt{3(\beta - \mathbb{E}[Y_i])}$$

Theorem: (Ornstein-Uhlenbeck Process)

The MSE-optimal sampling policy is

$$S_{i+1}(\beta) = \inf\left\{t \geq D_i(\beta) : \underbrace{\left|X_t - \hat{X}_t\right|}_{\text{instantaneous estimation error}} \geq v(\beta)\right\}$$

optimal β is determined by solving

$$\beta = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]} = \text{mse}_{\text{opt}}$$

β = optimal objective value

threshold function

$$v(\beta) = \frac{\sigma}{\sqrt{\theta}} G^{-1}\left(\frac{\text{mse}_\infty - \text{mse}_{Y_i}}{\text{mse}_\infty - \beta}\right)$$

The same solution structure, different threshold functions

Ornee and Sun, "Sampling for Remote Estimation through Queues: Age of Information and Beyond," IEEE WiOpt, 2019, Best Paper Award.

Yin Sun, Yury Polyanskiy, and Elif Uysal, "Sampling of the Wiener Process for Remote Estimation over a Channel with Random Delay", *IEEE Transactions on Information Theory*, 2019.

Question 2: Relation with Age of Information?

- Recall that sampling time S_i decided by the history of signal and channel processes

- If the sampling times are independent of signal $\{X_t, t \geq 0\}$, then

$$\mathbb{E}\left[(X_t - \hat{X}_t)^2\right] = \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta\Delta(t)}\right) = p(\Delta(t))$$

(MSE = Age penalty function)

Our problem reduces to an age penalty minimizaition problem.

$$\inf_{\pi \in \Pi} \quad \limsup_{T \to \infty} \frac{1}{T} \, \mathbb{E}\left[\int_0^T p(\Delta(t))dt\right]$$

$$\text{s.t.} \quad \liminf_{n \to \infty} \frac{1}{n} \mathbb{E}\left[\sum_{i=1}^{n} (S_{i+1} - S_i)\right] \geq \frac{1}{f_{\max}}$$

Age-Optimal vs. MSE-Optimal Sampling (Signal-ignorant vs Signal-aware Sampling)

Theorem: (Age-Optimal Sampling)

The age-optimal sampling policy is

$$S_{i+1}(\beta) = \inf \Big\{ t \geq D_i(\beta) : \underbrace{\mathbb{E}[(X_{t+Y_{i+1}} - \hat{X}_{t+Y_{i+1}})^2]}_{\text{expected estimation error}} \geq \beta \Big\}$$

optimal β is determined by solving

$$\beta = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]} = \text{mse}_{\text{age-opt}}$$

β = optimal objective value

Theorem: (MSE-Optimal Sampling)

The MSE-optimal sampling policy is

$$S_{i+1}(\beta) = \inf \Big\{ t \geq D_i(\beta) : \underbrace{|X_t - \hat{X}_t|}_{\text{instantaneous estimation error}} \geq v(\beta) \Big\},$$

Threshold function

optimal β is determined by solving

$$\beta = \frac{\mathbb{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} (X_t - \hat{X}_t)^2 dt\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]} = \text{mse}_{\text{opt}}$$

β = optimal objective value

Solutions to constrained problems available in the papers.

Sun and Cyr, "Sampling for Data Freshness Optimization: Non-linear Age Functions," JCN special issue on Age of Info., 2019.

Ornee and Sun, "Sampling for Remote Estimation through Queues: Age of Information and Beyond," IEEE WiOpt, 2019, Best Paper Award.

Summary

- Optimal sampler follows a threshold policy:
  - For signal-aware sampling, the threshold is on instantaneous estimation error, and it is a function of optimal objective value.
  - For signal-ignorant sampling, the threshold is on expected estimation error, and it is equal to the optimal objective value.
  - It is conjectured that this structure holds for a number of strong Markov signals.
  - Tool: Free-boundary method for solving optimal stopping problems of diffusion processes.

- More general than the Queueing model
  - Results should hold for other channel models (e.g., erasure channel)

Topic III

Age-Optimal Scheduling in Networks

Joint work with

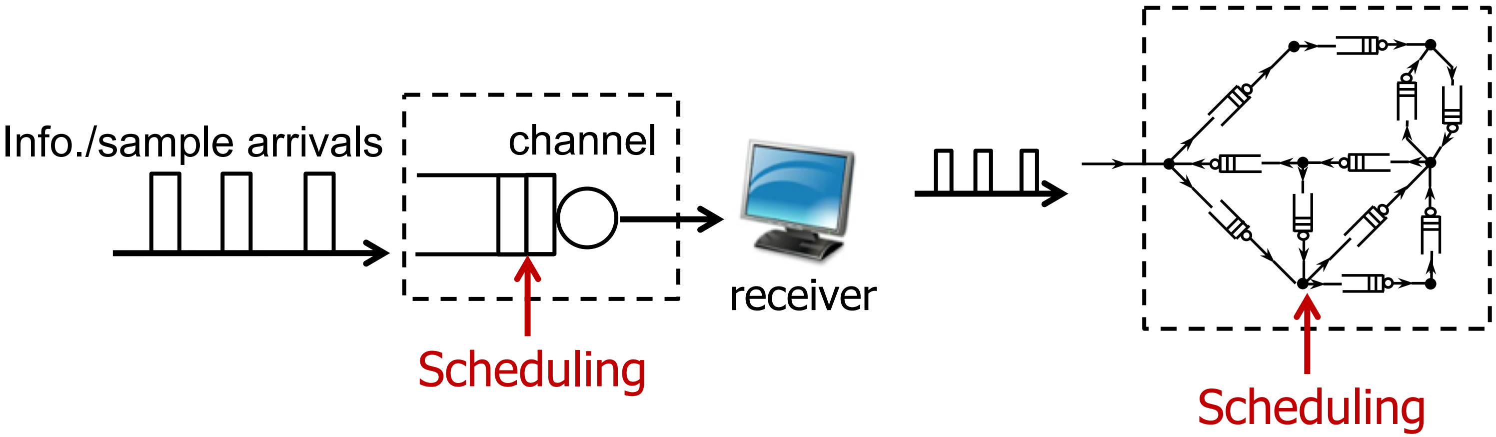
Ahmed M. Bedewy

Ness B. Shroff

Elif Uysal

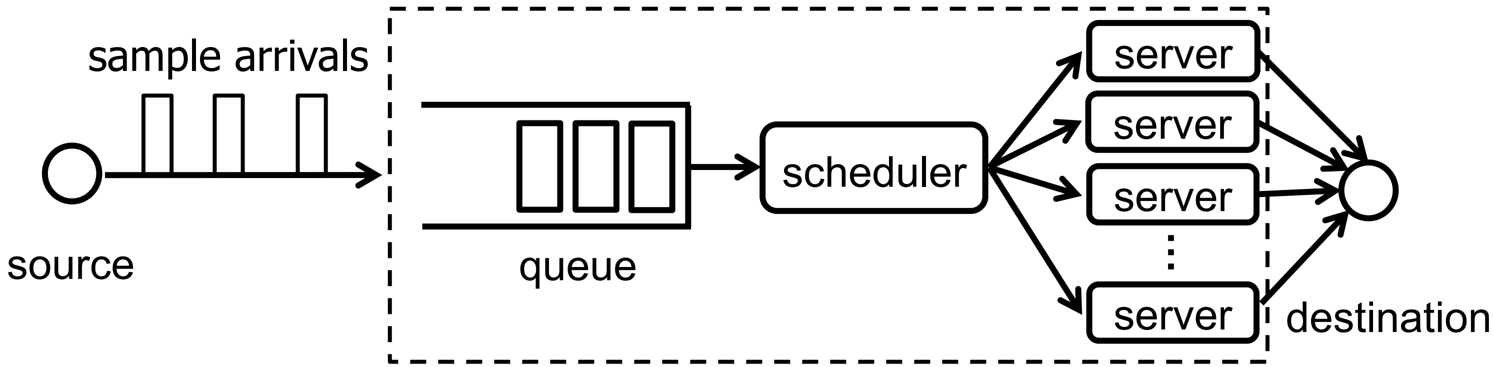
Sastry Kompella

Motivation:



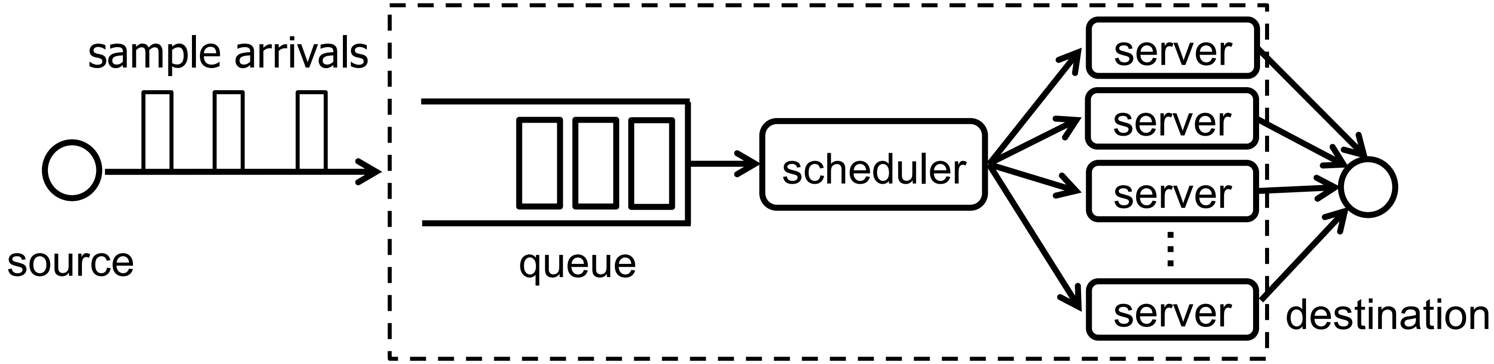
- In news, email, and social updates, the arrival process of data samples cannot be controlled
- Can we improve data freshness by designing the service order of the samples?

Single-Source, Multi-Channel Networks



- Single source, **arbitrarily given arrival** process:
  - Sample i is generated at S_i, arrives at A_i. Hence, $S_i \leq A_i$
  - Arbitrarily given sample generation & arrival times
    - Out-of-order arrivals are possible, e.g., $A_i < A_j, S_i > S_j$
- Multiple **channels,** *i.i.d.* transmission times
- Scheduling Policy:
  - Given sample generation and arrival times, **scheduler** decides which samples to send over time

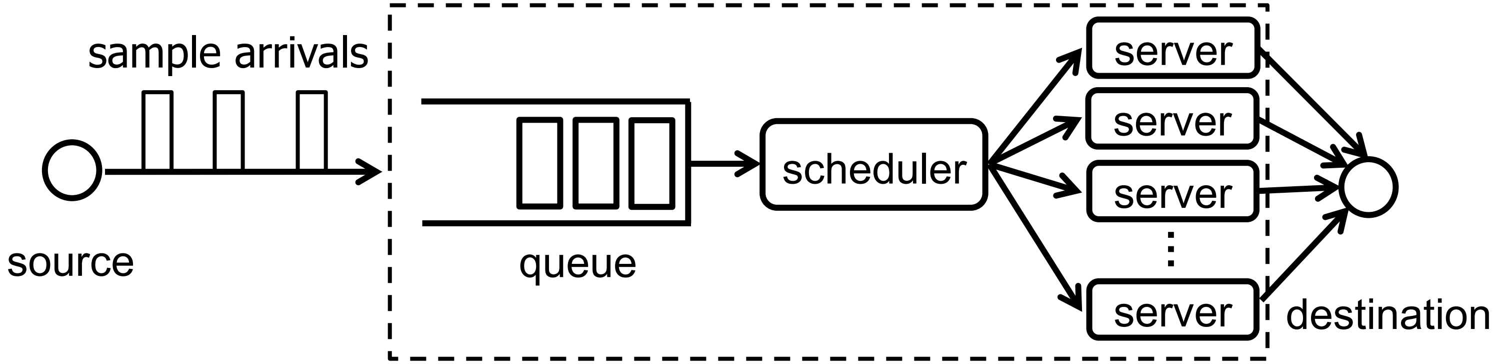
Scheduling Policy Space



- Preemptive Policy:
  - One sample can interrupt the transmission of another sample
- Causal Policy:
  - Scheduler makes decisions based on history and current information
  - Non-causal policies are not practical

Goal: Find the optimal policy that minimizes the age among all casual policies?

Last Generated First Served (LGFS) Policy



Last Generated, First Served (LGFS) policy:

- The last generated sample is sent first, with ties broken arbitrarily

Last Come, First Served (LCFS) policy:

- The last arrived sample is sent first, with ties broken arbitrarily.

- If packets arrive in the same order of their generation times, i.e., $(A_i - A_j)(S_i - S_j) \geq 0$, LGFS becomes LCFS

Metrics: Age Penalty Functional

- ***Age penalty functional*** $p(\{\Delta(t), t \geq 0\})$
  - Any non-decreasing functional p of age process $\{\Delta(t), t \in [0, \infty)\}$, i.e.,

$$\text{If } \Delta_1(t) \leq \Delta_2(t), t \geq 0, \text{ then } p(\{\Delta_1(t), t \geq 0\}) \leq p(\{\Delta_2(t), t \geq 0\})$$

- Examples of age penalty functionals:
  - Avg. age: [Kaul, Yates, Gruteser'12, etc.]

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

  - 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

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

Very general age metrics

Preemptive LGFS Minimizes Age Process!

Theorem: If the transmission times are i.i.d. exponential, then for all non-decreasing functional p and all sample generation/arrival times $\mathcal{I} = \{n, (S_i, A_i)_{i=1}^n\}$, it holds that

$$\mathbb{E}[p(\{\Delta_{\text{prmp-LGFS}}(t), t \geq 0\})|\mathcal{I}] = \min_{\pi \in \Pi} \mathbb{E}[p(\{\Delta_\pi(t), t \geq 0\})|\mathcal{I}], \ \forall\, \mathcal{I}$$

where Π is the set of all causal policies.

Corollary: If, in addition, packets arrive in the same order of their generation times, then preemptive LCFS is age-optimal, i.e.,

$$\mathbb{E}[p(\{\Delta_{\text{prmp-LCFS}}(t), t \geq 0\})|\mathcal{I}] = \min_{\pi \in \Pi} \mathbb{E}[p(\{\Delta_\pi(t), t \geq 0\})|\mathcal{I}], \ \forall\, \mathcal{I}$$

Preemptive LGFS and preemptive LCFS are age-optimal in a quite strong sense, if service times are i.i.d. exponential.

Bedewy, Sun, Shroff, "Optimizing Data Freshness, Throughput, and Delay in Multi-Server Information-Update Systems", *IEEE ISIT*, 2016.

Bedewy, Sun, Shroff, "Minimizing the Age of Information through Queues", *IEEE Trans. on Inf. Theory*, 2019.

How about Non-Exponential Distributions?

- **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., constant time, exponential, geometric, gamma distributions

Theorem: If the transmission times are i.i.d. NBU, then for all non-decreasing functional p and all sample generation/arrival times $\mathcal{I} = \{n, (S_i, A_i)_{i=1}^n\}$, it holds that

$$\min_{\pi \in \Pi_{np}} \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\int_0^T \Delta_\pi(t) dt \middle| \mathcal{I} \right] \leq \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\int_0^T \Delta_{\text{non-prmp-LGFS}}(t) dt \middle| \mathcal{I} \right]$$

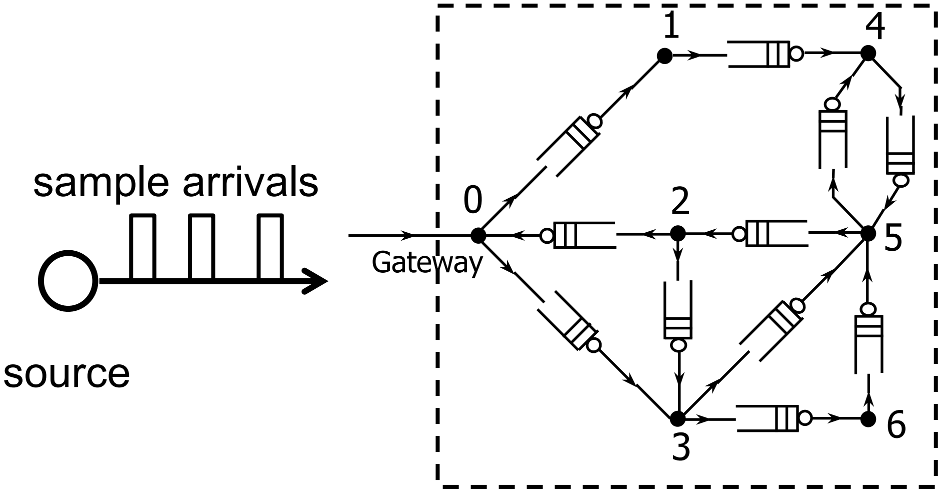
$$\leq \min_{\pi \in \Pi_{np}} \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\int_0^T \Delta_\pi(t) dt \middle| \mathcal{I} \right] + \mathbb{E}[X],$$

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

Mean service time of a packet

Non-preemptive LGFS is within a small additive gap from age-optimum.

Single-Source, Multi-Hop Networks



Age vector of all nodes:

$$\mathbf{\Delta}(t) = (\Delta_1(t), \ldots, \Delta_N(t))$$

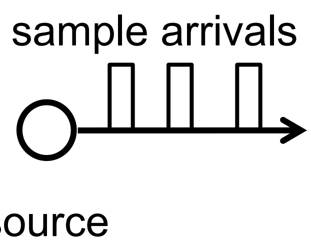
Theorem:

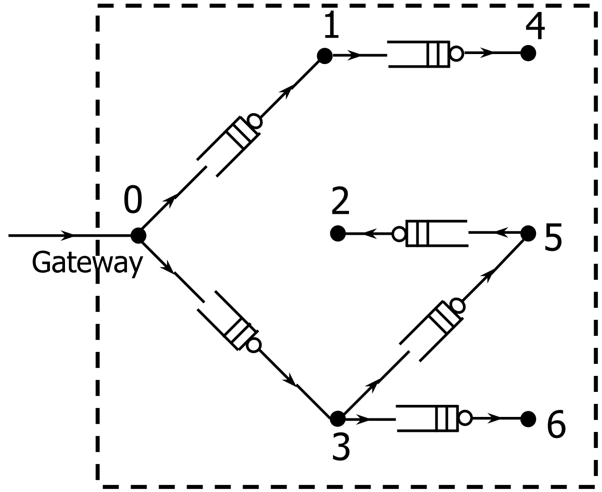
- For independent exponential service times, arbitrary topology
  - Preemptive LGFS minimizes all non-decreasing functional of the age processes at all nodes:

$$\mathbb{E}[p(\{\mathbf{\Delta}_{\text{prmp-LGFS}}(t), t \geq 0\})|\mathcal{I}] = \min_{\pi \in \Pi} \mathbb{E}[p(\{\mathbf{\Delta}_{\pi}(t), t \geq 0\})|\mathcal{I}], \ \forall\ \mathcal{I}$$

Bedewy, Sun, Shroff, "Age-Optimal Information Updates in Multihop Networks", *IEEE ISIT*, 2017.
Bedewy, Sun, Shroff, "Age of Information in Multihop Networks", *IEEE/ACM Transactions on Networking*, 2019.

Single-Source, Multi-Hop Networks





tree topology

Age vector of all nodes:

$$\mathbf{\Delta}(t) = (\Delta_1(t), \ldots, \Delta_N(t))$$

Theorem:

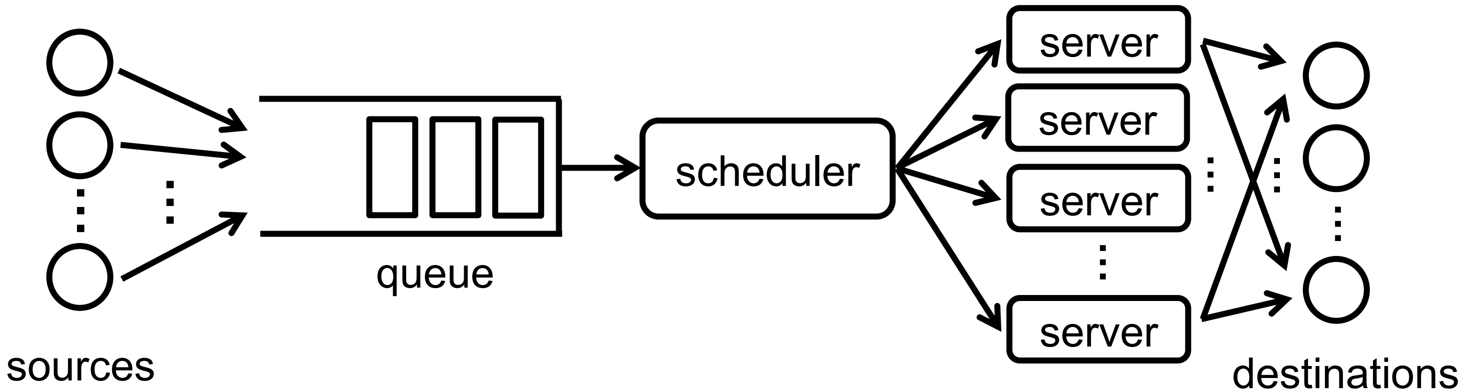
- For independent exponential service times, arbitrary topology
  - Preemptive LGFS minimizes all non-decreasing functional of the age processes at all nodes:

$$\mathbb{E}[p(\{\mathbf{\Delta}_{\text{prmp-LGFS}}(t), t \geq 0\})|\mathcal{I}] = \min_{\pi \in \Pi} \mathbb{E}[p(\{\mathbf{\Delta}_{\pi}(t), t \geq 0\})|\mathcal{I}], \ \forall \ \mathcal{I}$$

- For independent NBU service times, tree topology
  - Non-preemptive LGFS is near age-optimal.

Bedewy, Sun, Shroff, "Age-Optimal Information Updates in Multihop Networks", *IEEE ISIT*, 2017.
Bedewy, Sun, Shroff, "Age of Information in Multihop Networks", *IEEE/ACM Transactions on Networking*, 2019.

Multi-Flow, Multi-Channel Networks



- Multiple flows, each flow with a source and a destination.
- Let $\boldsymbol{\Delta}(t) = (\Delta_1(t), \ldots, \Delta_N(t))$ be the age vector of N flows.
- **Maximum Age First, Last Generated First Served (MAF-LGFS) policy**: The last generated packet from the flow with the maximum age is sent first, with ties broken arbitrarily.

Theorem: Under certain conditions (synchronized arrivals, one server, Exp distribution), preemptive MAF-LGFS minimizes any *symmetric* and *non-decreasing* penalty function of the age vector over time.

- Near age-optimal results for NBU distributions are in

Yin Sun, Elif Uysal-Biyikoglu, and Sastry Kompella, Age-Optimal Updates of Multiple Information Flows, *IEEE INFOCOM Age of Information Workshop (AoI Workshop)*, 2018.

Summary

- For i.i.d. Exponential service times, preemptive LGFS is age-optimal.
- For i.i.d. New-Better-than-Used (NBU) service times, non-preemptive LGFS is near age-optimal.
  - These results hold for
    - Minimizing the age of information process
    - Minimizing non-decreasing functionals of age process
    - Arbitrarily given packet generation times and arrival times (limitations exist for multi-flow case)
    - Several queueing network models
  - In-order arrivals ➔ LGFS becomes LCFS
- Tools:
  - Sample-path method
  - Stochastic ordering
- Some Ads:
  - 3rd Age of Information Workshop

The 3rd Age of Information Workshop

- April 27th 2020 at Beijing, with IEEE INFOCOM 2020
- General Chairs:
  - Ness B. Shroff (The Ohio State University)
  - Zhisheng Niu (Tsinghua University)
- TPC Chairs:
  - Sennur Ulukus (University of Maryland)
  - Sheng Zhou (Tsinghua University)

References

- Yin Sun, Yury Polyanskiy, and Elif Uysal, Sampling of the Wiener Process for Remote Estimation over a Channel with Random Delay, *IEEE Transactions on Information Theory*, 2019.
- Ahmed M. Bedewy, Yin Sun, and Ness B. Shroff, The Age of Information in Multihop Networks, *IEEE/ACM Transactions on Networking*, 2019.
- Ahmed M. Bedewy, Yin Sun, and Ness B. Shroff, Minimizing the Age of Information through Queues, *IEEE Transactions on Information Theory,* 2019.
- Yin Sun, Elif Uysal, Roy D. Yates, C. Emre Koksal, and Ness B. Shroff, Update or Wait: How to Keep Your Data Fresh, *IEEE Transactions on Information Theory,* vol. 63, no. 11, Nov. 2017.
- Yin Sun and Benjinmin Cyr, "Sampling for Data Freshness Optimization: Non-linear Age Functions," *Journal of Communications and Networks,* 2019.
- Yin Sun, Elif Uysal, and Sastry Kompella, Age-Optimal Updates of Multiple Information Flows, *INFOCOM Age of Information Workshop (AoI Workshop)*, 2018.
- Tasmeen Zaman Ornee and Yin Sun, "Sampling for Remote Estimation through Queues: Age of Information and Beyond," *IEEE/IFIP WiOpt,* 2019, Best Paper Award.