Age-of-Information: Optimizing the Freshness of Real-Time Data

Yin Sun

Electrical and Computer Engineering
the Ohio State University

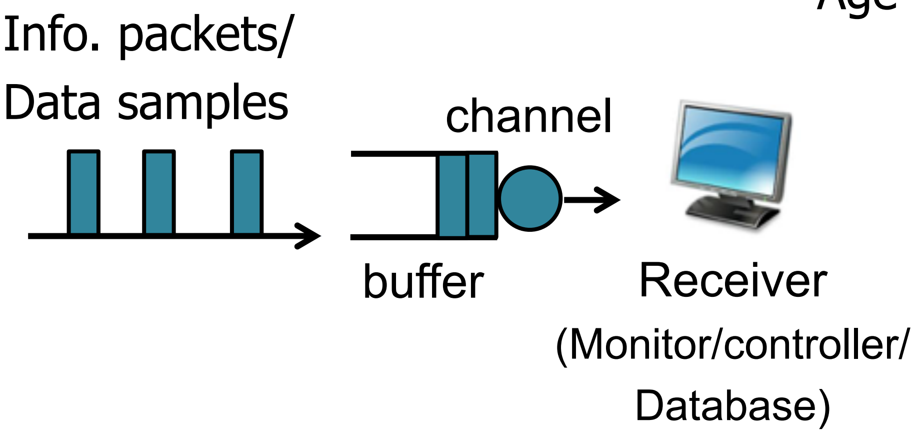
Workshop at MobiHoc TPC meeting, USC
March 24, 2017

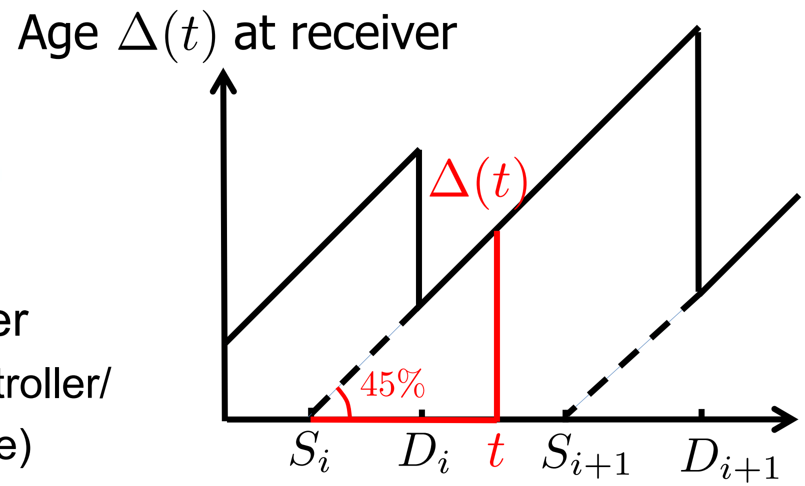
Outline

- Introduction and Motivation:
  Why is fresh data so important?
- How to optimize data freshness?
  - Model, Metrics, Challenges, Key Results, and Methodology
- Summary

Introduction and Motivation

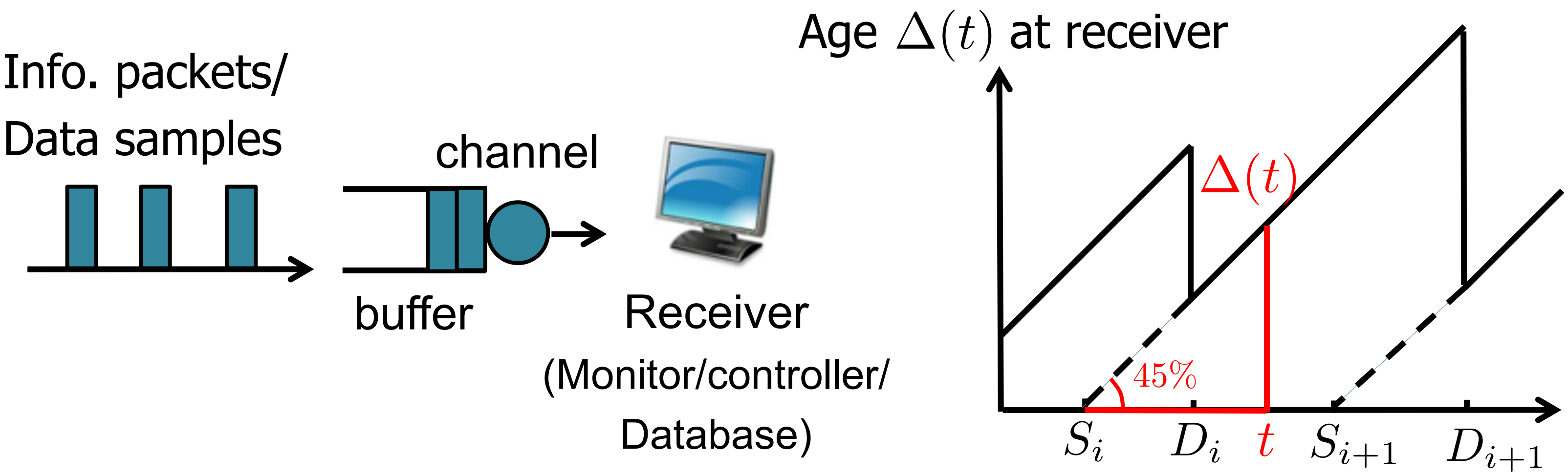
Data Freshness Metric: Age-of-Information





- In real-time applications, fresh data is more important than stale data
  - E.g., airplane/vehicle control, disaster monitoring,…

Data Freshness Metric: Age-of-Information



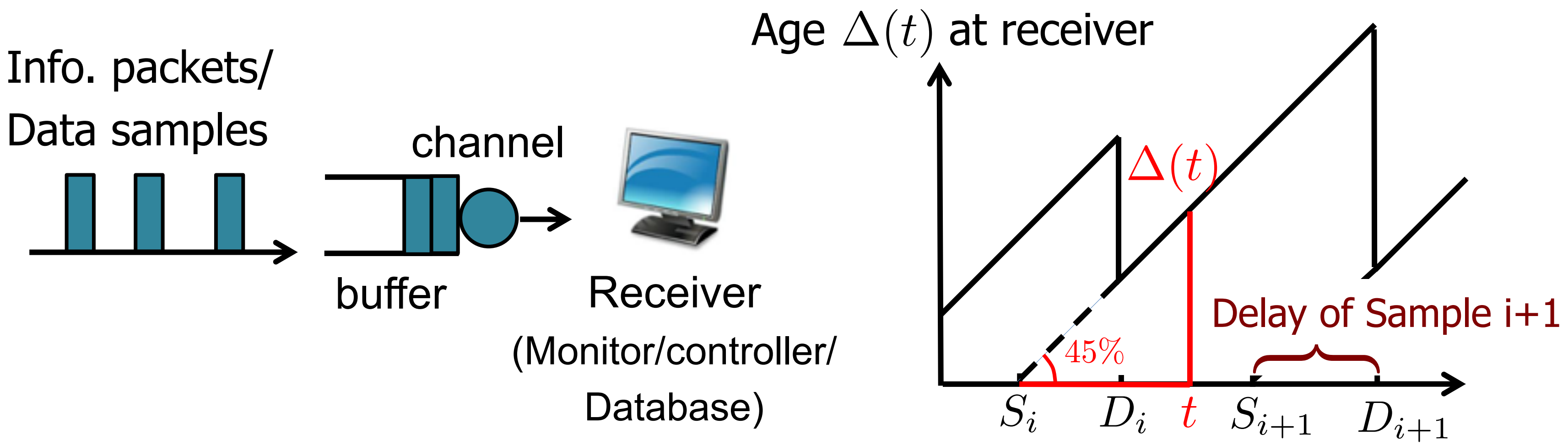
Definition: At time t, the **Age-of-Information** $\Delta(t)$ is the "age" of the "youngest" data sample delivered to the receiver before time t

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

$$\Delta(t) = t - \max\{S_i : D_i \leq t\}$$

- Age grows linearly, and drops upon new sample delivered

Data Freshness Metric: Age-of-Information



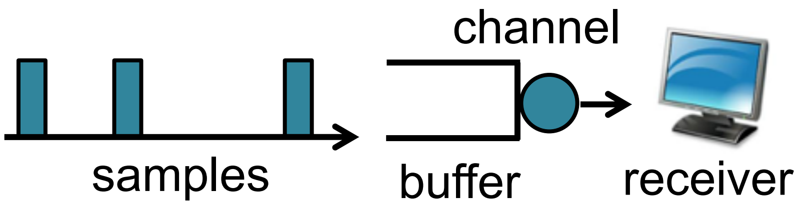
Definition: At time t, the **Age-of-Information** $\Delta(t)$ is the "age" of the "youngest" data sample delivered to the receiver before time t

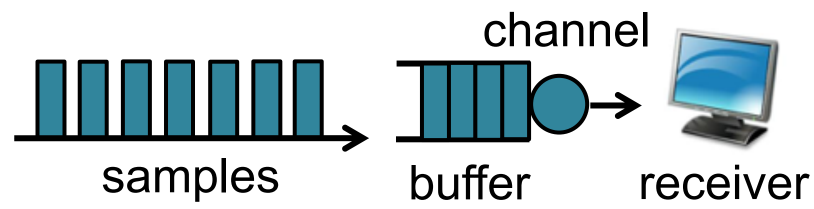
- If sample i is generated at S_i and delivered at D_i

$$\Delta(t) = t - \max\{S_i : D_i \leq t\}$$

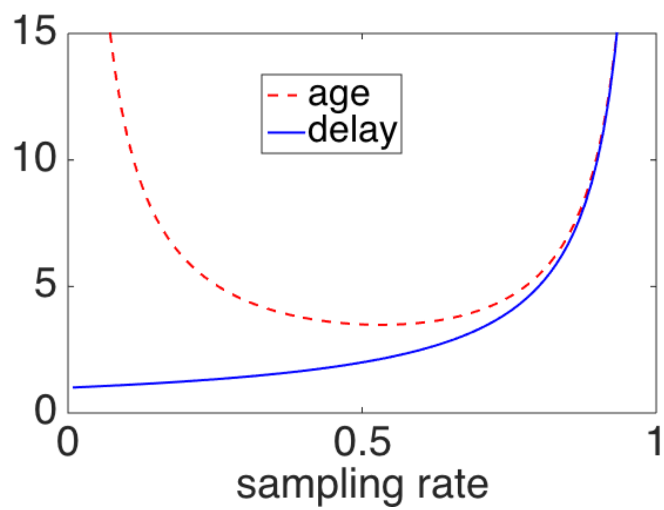
- Age grows linearly, and drops upon new sample delivered

Difference between Age & Delay

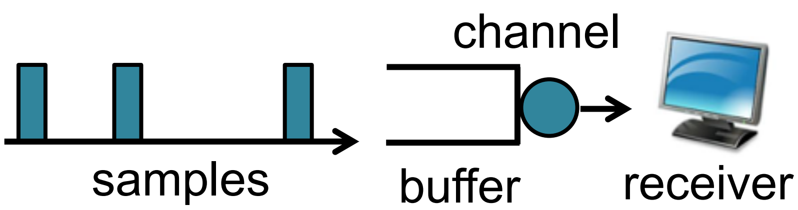


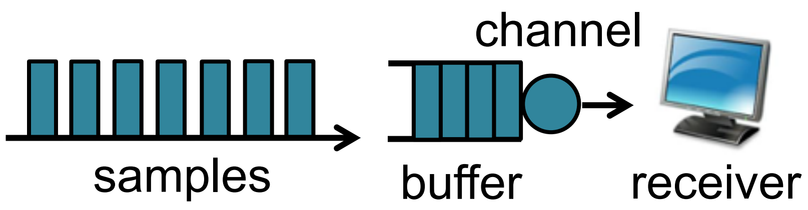


- Low sampling rate
  - Empty buffer ➔ low delay
  - Infrequent updates ➔ high age
- High sampling rate
  - Full buffer ➔ high delay
  - Become stale while waiting ➔ high age



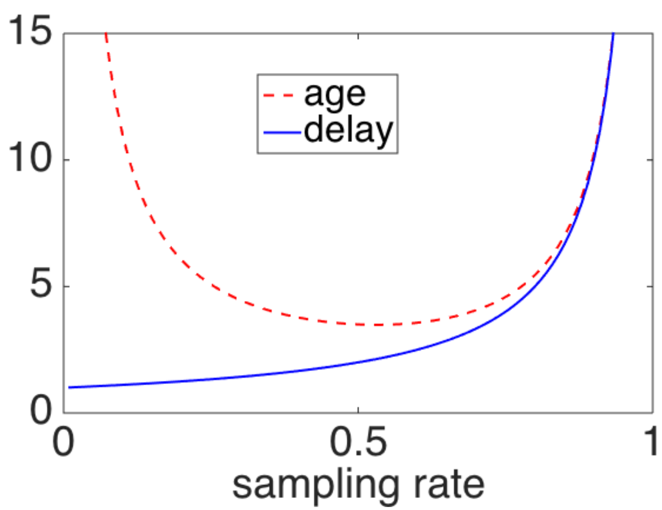
Difference between Age & Delay





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

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



Prior Research on Age-of-Information

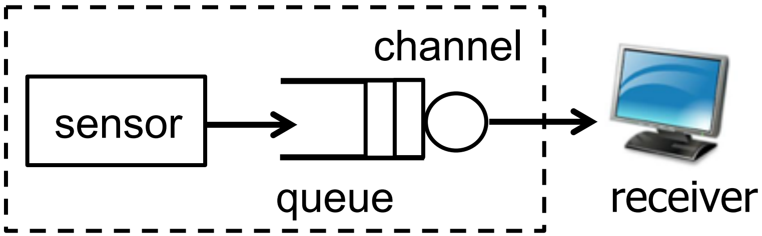
- Systems Research
  - Defined in real-time databases [Song,Liu'90] or earlier
  - Real-time search [Twitter Search'08]
  - Real-time ads bidding [Google'09]
  - …

- Theoretical Research (a new area)
  - Average age analysis of certain policies [Kaul, Yates, Gruteser'12, Kam, Kompella, Ephremides'13, etc.]
  - Low-age source and channel coding [Zhong, Yates'16, etc.]

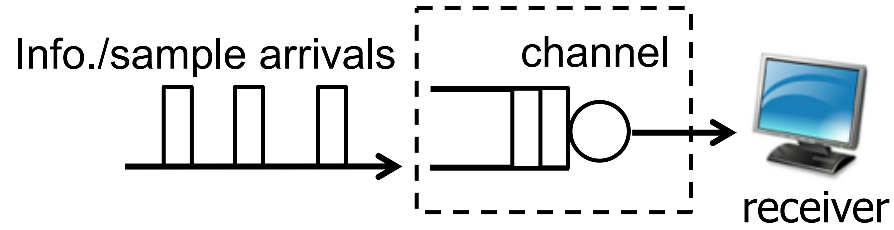
Analysis, but not optimization

Our Research: Minimizing the Age-of-Information

Overview of Our Research on Age-of-Information



Joint sampling and transmission control



Transmission control only

- **Age-optimal sampling & trans.**

Counter-Intuitive Result: ⟵ **Next**

Sampling ASAP may not be age-optimal.

[Yates ISIT15; Sun, Uysal, Yates, Koksal, Shroff, Infocom16]

- **Real-time signal transmissions**

Age-of-Information is not a perfect metric

Age-of-Information ≠ Change-of-Information

[Sun, Polyanskiy, Uysal, ISIT17]

- **Age-optimal transmission**

An intuitive policy (Last Generated First Served) is age-optimal in a quite strong sense (stochastic ordering of age process) in

- Multi-channel networks

  [Bedewy, Sun, Shroff, ISIT16]

- Multi-hop networks

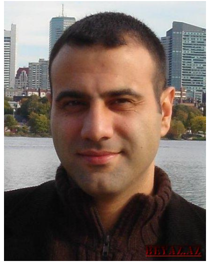
  [Bedewy, Sun, Shroff, ISIT17]

Optimal result on minimizing general Age-of-Information metrics in each of these models.

Age-optimal Sampling & Transmissions

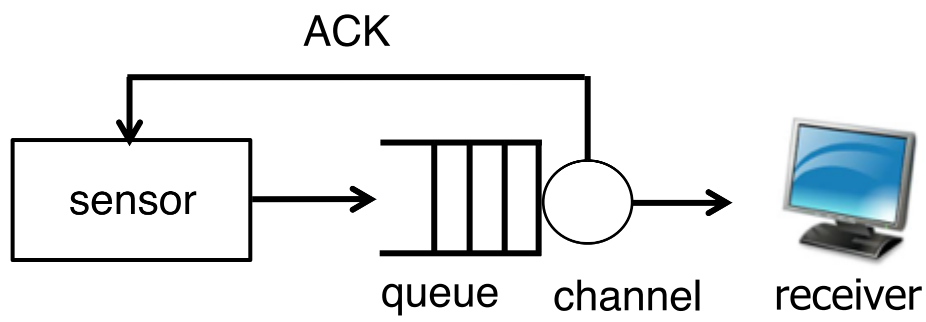
Joint work with






Elif Uysal-Biyikoglu, Roy D. Yates, C. Emre Koksal, Ness B. Shroff

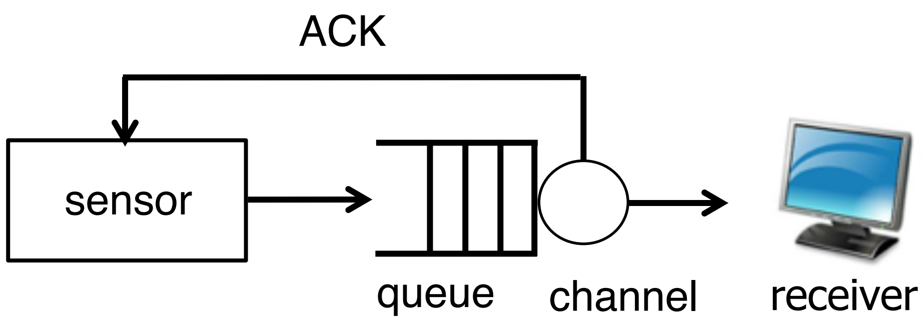
Model: Age-Optimal Sampling



- Two terminals: **Sensor** sends data samples to **Receiver**
- **Channel**: FIFO queue with *i.i.d.* transmission times
- **Feedback:** ACK
  - A sample is deemed delivered after ACK

- Sample i is generated at S_i, with channel transmission time Y_i
  - $Y_i \geq 0$ has **general** distribution
  - Can be discrete/continuous random variable (finite moments)
  - More general than prior studies

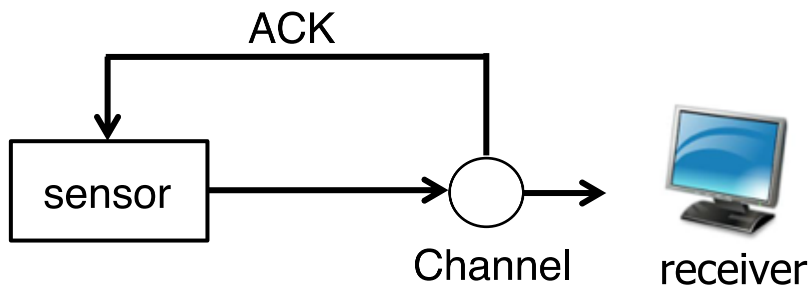
Model: Age-Optimal Sampling



- The samples may wait in the queue for its transmission opportunity, and become stale while waiting!

- Better idea:
  - Take a new sample after all previous samples are delivered
  - Samples are fresh ➔ It is better to keep the queue empty

Natural Choice: Zero-wait Policy



- **Zero-wait policy: (Sampling ASAP)**

Take a new sample once the previous sample is delivered, i.e.,

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

- Zero-wait is **throughput-optimal & delay-optimal**
  - Throughput-optimal: Channel is always busy, throughput $= 1/\mathbb{E}[Y_i]$
  - Delay-optimal: Waiting time is zero, delay $= \mathbb{E}[Y_i]$

Counter-Intuitive Result: Zero-wait is NOT always Age-optimal!

Example: Channel transmission time = 0 or 2 with Prob 0.5

0, 0, 2, 0, 2, 2, 0, 2, 0, 0, …

If Sample 1 has zero transmission time, when to take Sample 2?

- **Zero-wait policy:**
  - Samples 1 & 2 are taken at the same time.
  - After Sample 1 is delivered, Sample 2 carries no new information.
  - However, Sample 2 has a long avg. transmission time

  No Gain, Only Pain

Which policy is better than zero-wait?

R. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *IEEE ISIT*, 2015.

Sun, Uysal-Biyikoglu, Yates, Koksal, Shroff, "Update or Wait: How to Keep Your Data Fresh", *IEEE Infocom*, 2016. Submitted to *IEEE Transactions on Information Theory.*

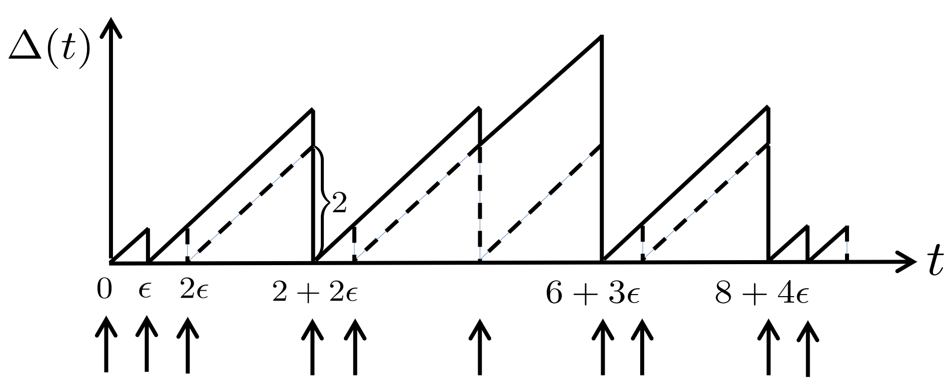
Counter-Intuitive Result: Zero-wait is NOT always Age-optimal!

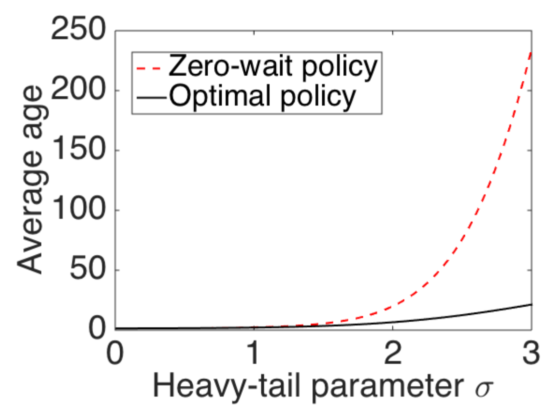
Example: Channel transmission time = 0 or 2 with Prob 0.5

0, 0, 2, 0, 2, 2, 0, 2, 0, 0, …

If Sample 1 has zero transmission time, when to take Sample 2?

- ϵ **-wait policy:**
  - If the previous sample has zero transmission time, wait for ϵ seconds
  - If the previous sample has 2 seconds of transmission time, do not wait
  - Average age: $$\bar{\Delta}(\epsilon) = (\epsilon^2/2 + \epsilon^2/2 + 2\epsilon + 4^2/2)/(4 + 2\epsilon)$$ $$= (\epsilon^2 + 2\epsilon + 8)/(4 + 2\epsilon)$$
  - Zero-wait: $\bar{\Delta}(0) = 2$ seconds, ϵ-wait: $\bar{\Delta}(0.5) = 1.85$ seconds





Counter-Intuitive Result: Zero-wait is NOT always Age-optimal!

Example: Channel transmission time = 0 or 2 with Prob 0.5

0, 0, 2, 0, 2, 2, 0, 2, 0, 0, …

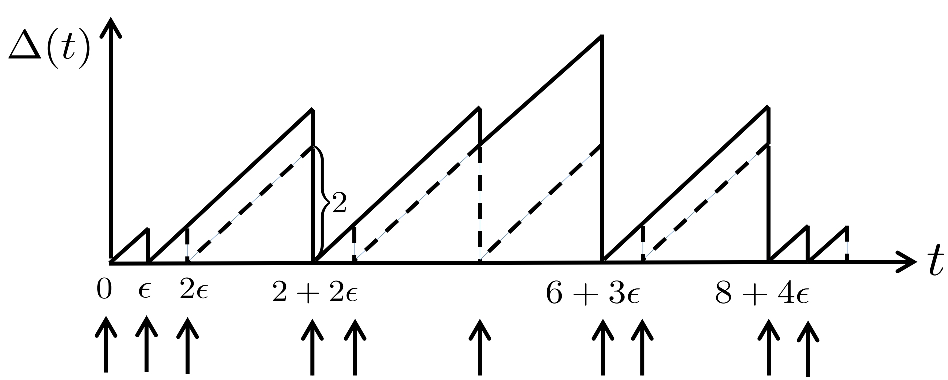
If Sample 1 has zero transmission time, when to take Sample 2?

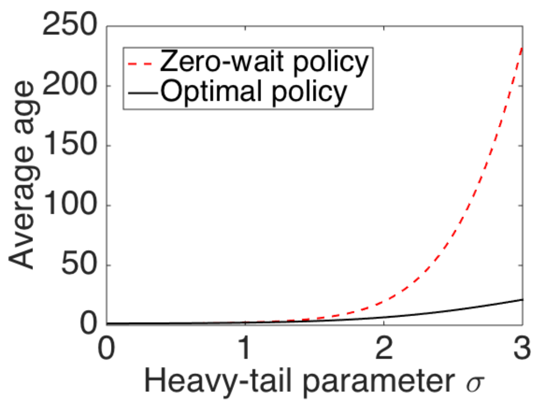
- ϵ **-wait policy:**

Data communication vs **information updates**

1. Data communication: All packets are important

1. Information updates: A packet is important only if it brings fresh info.





Age Minimization Problem

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

Expected time-average age

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

Avg. sampling-rate constraint

- Sampling policy: $\pi = (S_1, S_2, \ldots)$ sequence of sampling times
- Space of causal policies: Π

Sampling time S_i decided by the transmission times $Y_1, \ldots, Y_{i-1}$ of previous samples (stopping time)

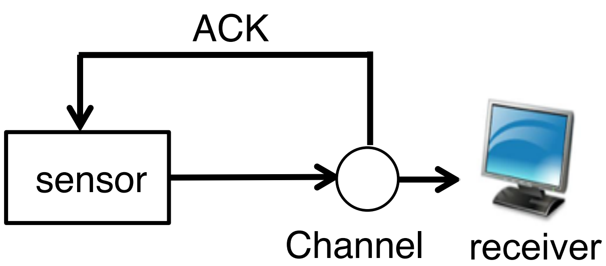
Questions: Which policy minimizes the age?

If there is no constraint, when is zero-wait age-optimal?

Difficulty of Age Minimization

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

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



- Markov Decision Process
  - Continuous time:

    Sampling times $(S_1, S_2, \ldots)$ are real numbers
  - Continuous state space:

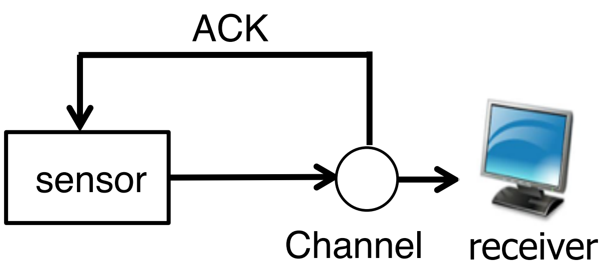
    Age $\Delta(t)$ and transmission time Y_i are real numbers
  - With a constraint:

    Bellman's equation works for MDP with no constraint
- Difficult to solve in closed-form

Which policy minimizes the age? Simple Age-Optimal Solution

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

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



Theorem 1. The age-optimal solution is given by

$$S_{i+1} = S_i + \max\{Y_i, \beta\}$$

The optimal β is determined by solving:

$$\mathbb{E}[\max(\beta, Y)] = \max\left(\frac{1}{f_{\max}}, \frac{\mathbb{E}[\max(\beta^2, Y^2)]}{2\beta}\right)$$

where Y has the same distribution with transmission time Y_i.

Low-complexity Bisection Alg: (10-20 iterations)

Algorithm 3 Bisection method for solving β

given $l = 0$, u, tolerance ϵ.
repeat
  $\beta := (l + u)/2$.
  $o := \mathbb{E}[\max(\beta, Y)] - \max\left(\frac{1}{f_{\max}}, \frac{\mathbb{E}[(\max(\beta^2, Y^2)]}{2\beta}\right)$.
  if $o \geq 0$, $u := \beta$; **else**, $l := \beta$.
until $u - l \leq \epsilon$.
Compute $z(\cdot)$ by (21).
return $z(\cdot)$.

Sun, Uysal-Biyikoglu, Yates, Koksal, Shroff, "Update or Wait: How to Keep Your Data Fresh", *IEEE Infocom*, 2016. Submitted to *IEEE Transactions on Information Theory.*

Which policy minimizes the age? Simple Age-Optimal Solution

Low complexity age-optimal solution for general transmission time distributions.

Theorem 1. The age-optimal solution is given by

$$S_{i+1} = S_i + \max\{Y_i, \beta\}$$

The optimal β is determined by solving:

$$\mathbb{E}[\max(\beta, Y)] = \max\left(\frac{1}{f_{\max}}, \frac{\mathbb{E}[\max(\beta^2, Y^2)]}{2\beta}\right)$$

where Y has the same distribution with transmission time Y_i.

Low-complexity Bisection Alg: (10-20 iterations)

Algorithm 3 Bisection method for solving β

given $l = 0$, u, tolerance ϵ.
repeat
  $\beta := (l + u)/2$.
  $o := \mathbb{E}[\max(\beta, Y)] - \max\left(\frac{1}{f_{\max}}, \frac{\mathbb{E}[(\max(\beta^2, Y^2)]}{2\beta}\right)$.
  if $o \geq 0$, $u := \beta$; **else**, $l := \beta$.
until $u - l \leq \epsilon$.
Compute $z(\cdot)$ by (21).
return $z(\cdot)$.

Sun, Uysal-Biyikoglu, Yates, Koksal, Shroff, "Update or Wait: How to Keep Your Data Fresh", *IEEE Infocom*, 2016. Submitted to *IEEE Transactions on Information Theory.*

When is Zero-Wait Optimal? A Sufficient and Necessary Condition

Theorem 2. If there is no constraint, zero-wait policy is age-optimal if and only if

$$\mathbb{E}[Y^2] \leq 2y_{\min}\mathbb{E}[Y]$$

where $y_{\min}$ is the minimum possible transmission time, i.e.,

$$y_{\min} = \inf\{y \in [0, \infty) : \Pr[Y \leq y] > 0\}$$

Corollary. (a) If trans. time Y is constant, then zero-wait is optimal.
(b) If $\mathbb{E}[Y] > 0,\ y_{\min} = 0$, then zero-wait is not optimal.

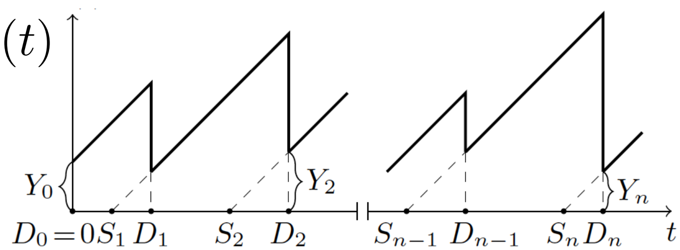
By **Corollary(b)**, if transmission times follow exponential, geometric, gamma distributions etc., zero-wait policy is not optimal.

Sun, Uysal-Biyikoglu, Yates, Koksal, Shroff, "Update or Wait: How to Keep Your Data Fresh", *IEEE Infocom*, 2016. Submitted to *IEEE Transactions on Information Theory.*

Extension: Minimizing "Dissatisfaction" of Stale Data

- **General** age penalty function $g(\Delta)$:
  - Any positive and non-decreasing function of age, chosen by application
  - Low-complexity optimal solution: Two-layer nested bisection

Age: $\Delta(t)$

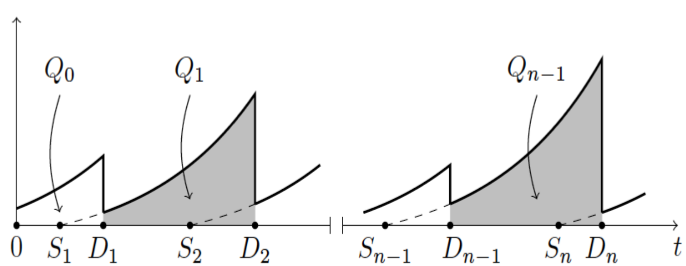


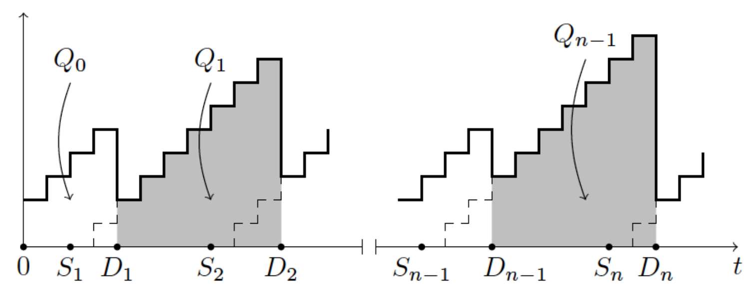
Age penalty: $g_1(\Delta(t)) = e^{0.2\Delta(t)} - 1$

(Impatient user)

$g_2(\Delta(t)) = \lfloor \Delta(t) \rfloor$ Non-convex, discontinuous

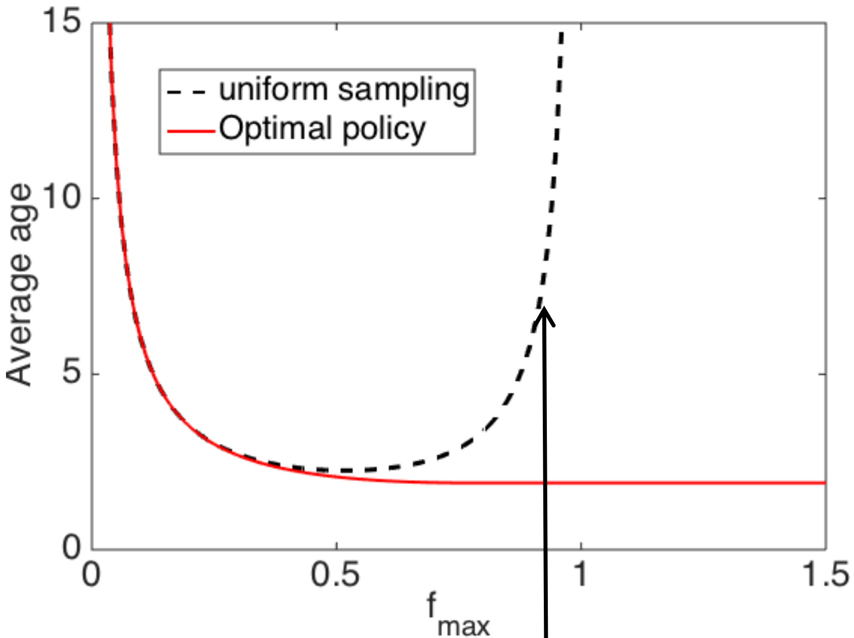
(Check periodically)



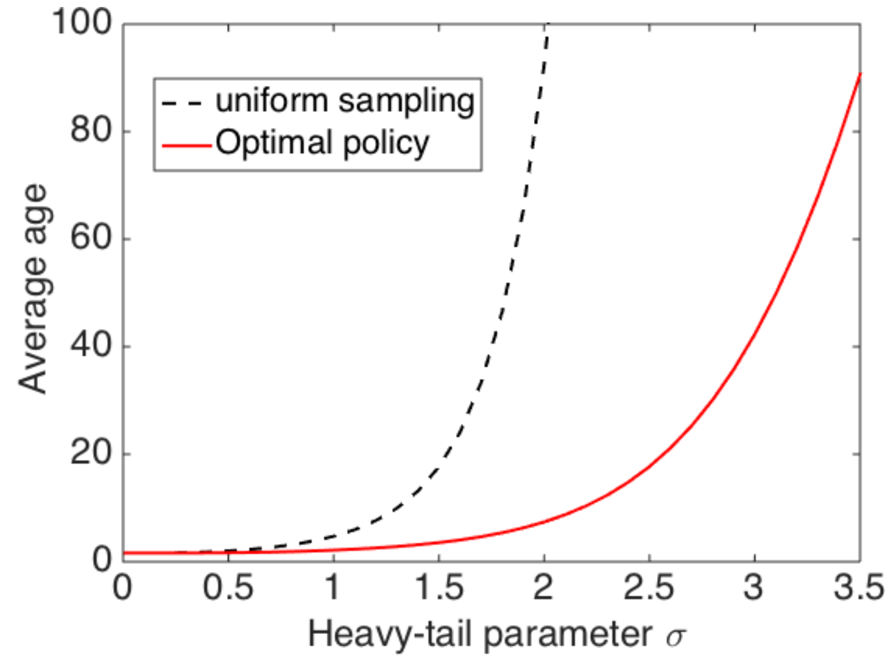


Sun, Uysal-Biyikoglu, Yates, Koksal, Shroff, "Update or Wait: How to Keep Your Data Fresh", *IEEE Infocom*, 2016. Submitted to *IEEE Transactions on Information Theory.*

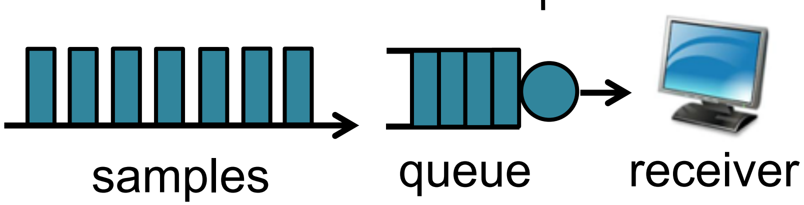
Uniform Sampling can be FAR from Optimal



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



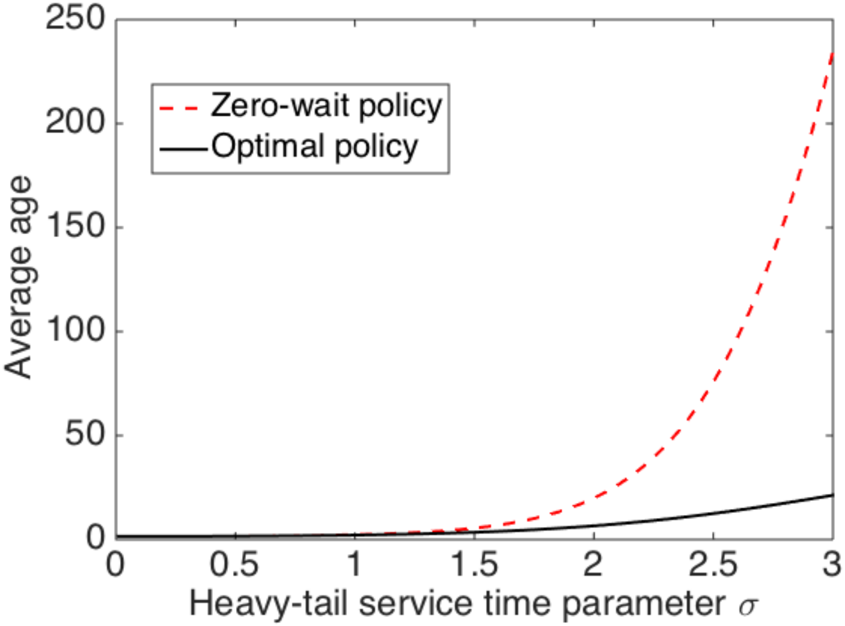
i.i.d. log-normal trans. times $\mu = 1$



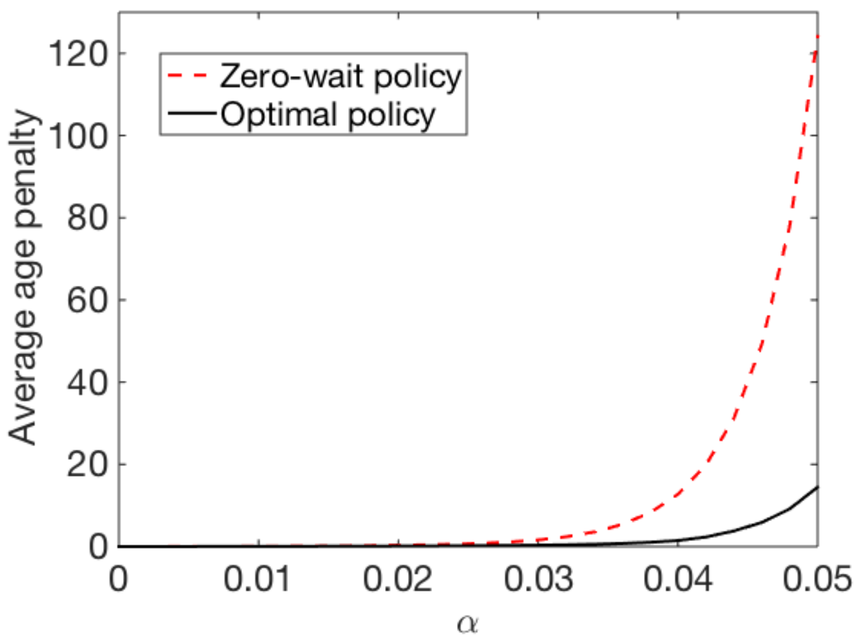
Become stale while waiting

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 $\mu = 1$



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

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

Zero-wait is far from optimal if

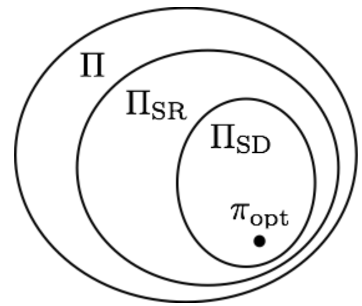
(1) Transmission time is highly random

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

How to Find Age-Optimal Policy? A Divide & Conquer Approach

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

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



- **Step 1:** Prove that there exists a **stationary randomized** policy that is **optimal**.
  - Π_{SR} : Inter-sample time $S_{i+1} - S_i$ depends only on trans. time Y_i of the immediate previous sample, with $\Pr[S_{i+1} - S_i | Y_i]$ fixed for all i
  - Proof idea: Sufficient statistics of MDP

    Expected time-averages form a convex and compact set
- **Step 2:** Prove that there exists a **stationary deterministic** policy that is **optimal**.
  - Π_{SD} : $S_{i+1} = S_i + f(Y_i)$
  - Proof idea: Jensen's inequality
- **Step 3:** Find the **optimal** function f
  - π_{opt} : $f(y) = \max\{y, \beta\}$ ➔ $S_{i+1} = S_i + \max\{Y_i, \beta\}$
  - Proof idea: Functional (quasi)-convex optimization, calculus of variations

Summary

- Optimal solution for minimizing Age-of-Information
  - Developed new methodology
- Low complexity age-optimal solution, general transmission time distribution, general age-penalty functions
- Sufficient and necessary condition to characterize when zero-wait is age-optimal
- Uniform sampling and zero-wait can be far from optimal
  - Useful for real-time monitoring and control applications

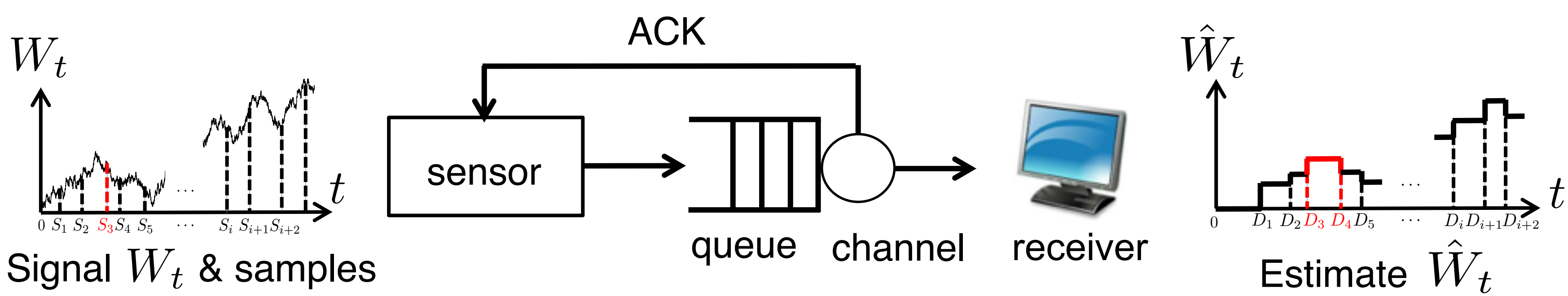
Real-time Signal Transmissions

Joint work with

Yury Polyanskiy, Elif Uysal-Biyikoglu

Age-of-Information is not a perfect metric

- **Age-of-Information ≠ Change-of-Information**
  - If the information has not changed, “old” information is useful.
  - If the information changes quickly, “young” information is not useful.
- **Effective Age-of-Information**
  - Open problem suggested by Tony Ephremides at ITA 2015



Information in the form of a Gauss-Markov **signal** W_t

- Wiener, Ornstein-Uhlenbeck process (physics, finance, etc.)

Goal: Minimize **estimation error** between signal W_t and estimate $\hat{W}_t$

Estimation Error Minimization Problem

Question: Which sampling policy minimizes estimation error?

$$\min_{\pi\in\Pi} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T (W_t - \hat{W}_t)^2 dt\right]$$ **Avg. estimation error (MMSE)**

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

More challenging:

- Sampling time S_i decided by the history of signal and channel (i.e., stopping time)
- If sampling times are independent of signal, then

$$\frac{1}{T}\mathbb{E}\left[\int_0^T (W_t - \hat{W}_t)^2 dt\right] = \frac{1}{T}\mathbb{E}\left[\int_0^T \Delta(t) dt\right]$$ **(MMSE = Age-of-Information)**

Contributions: Optimal solution with nice structure and low complexity

Powerful methodology: *Optimal Stopping Rules*, A.N. Shiryaev, 1978.

Sun, Polyanskiy, Uysal-Biyikoglu, "Remote Estimation of the Wiener process over a channel with random delay", *IEEE ISIT*, 2017. https://arxiv.org/abs/1701.06734

Age-Optimal vs. MMSE-Optimal Sampling

Theorem: (Age-Optimal Sampling)

The age-optimal sampling policy is

$$S_{i+1} = \inf\{t \geq S_i + Y_i : \underbrace{t - S_i}_{\text{Time difference}} \geq \beta\}$$

optimal β is determined by solving

$$\mathbb{E}[\max(\beta, Y)] = \max\left(\frac{1}{f_{\max}}, \frac{\mathbb{E}[\max(\beta^2, Y^2)]}{2\beta}\right)$$

Y: channel transmission time

$f_{\max}$: max. sampling rate

Simple optimal solution determined by channel and sampler.

Sun, Uysal-Biyikoglu, Yates, Koksal, Shroff, "Update or Wait: How to Keep Your Data Fresh", *IEEE Infocom*, 2016.

Submitted to *IEEE Transactions on Information Theory.*

Theorem: (MMSE-Optimal Sampling)

The MMSE-optimal sampling policy is

$$S_{i+1} = \inf\left\{t \geq S_i + Y_i : |\underbrace{W_t - W_{S_i}}_{\text{Signal difference}}| \geq \sqrt{\beta}\right\}$$

optimal β is determined by solving

$$\mathbb{E}[\max(\beta, W_Y^2)] = \max\left(\frac{1}{f_{\max}}, \frac{\mathbb{E}[\max(\beta^2, W_Y^4)]}{2\beta}\right)$$

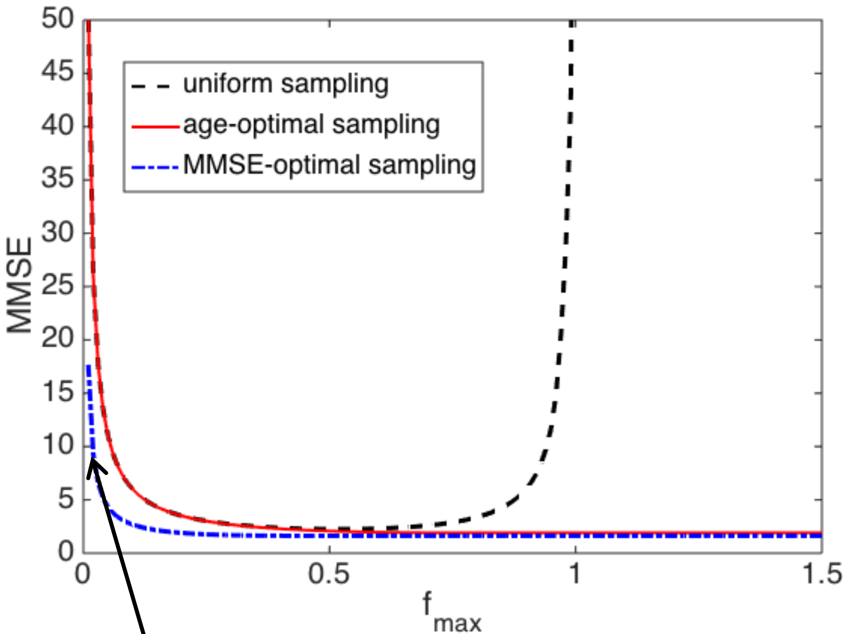
W_Y : Signal variation in channel trans. time

$f_{\max}$: max. sampling rate

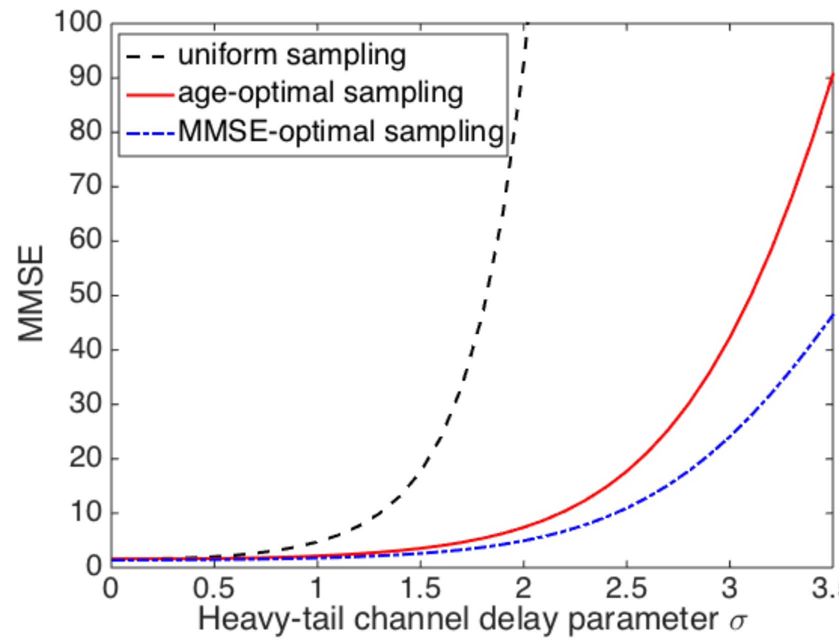
Simple optimal solution determined by signal, channel, and sampler.

Sun, Polyanskiy, Uysal-Biyikoglu, "Remote Estimation of the Wiener process over a channel with random delay", *IEEE ISIT*, 2017. https://arxiv.org/abs/1701.06734

How Good is Age-Optimal Sampling?



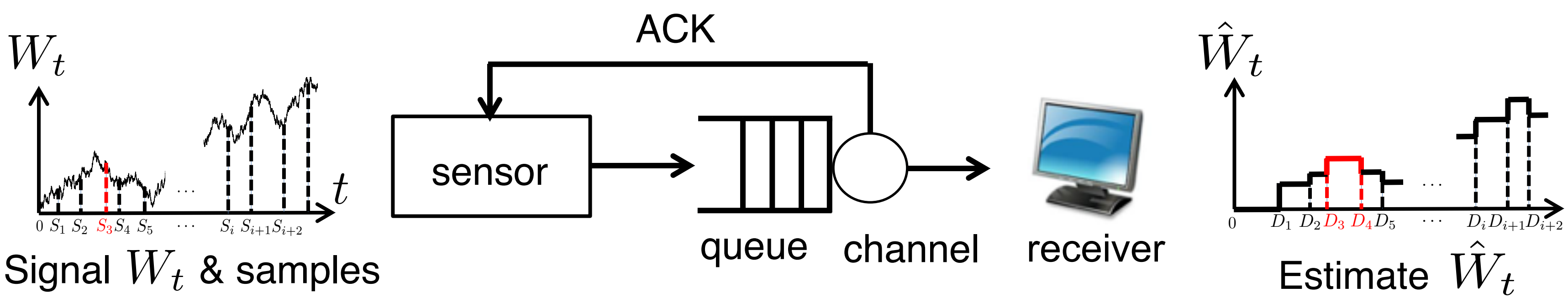
i.i.d. exponential service times $\mu = 1$



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

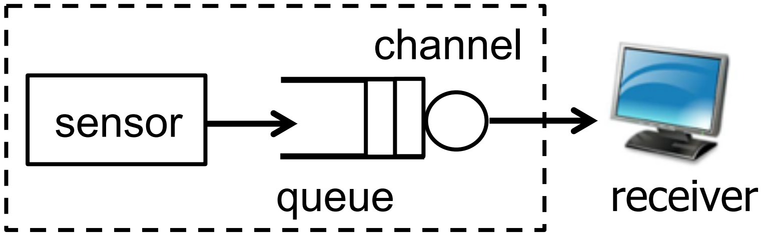
Sub-optimality gap of age-optimal sampling is large if
(1) Sampling rate is low
(2) Transmission time is highly random

Summary



- Real-time signal transmissions
  - Optimal solution with nice structure and low complexity
  - Interesting relationship between Age-of-Information and MMSE (Signal Processing)
- More practical signal models
  - Wiener process ➔ Ornstein-Uhlenbeck process ➔ ??
    (Continuous-time AR 1 model)
- Related to remote estimation & Event-triggered estimation (difference in model and solution)

Conclusions



Joint sampling and transmission control

Age metric:

- Any non-decreasing function of age
- MMSE of real-time signal transmissions

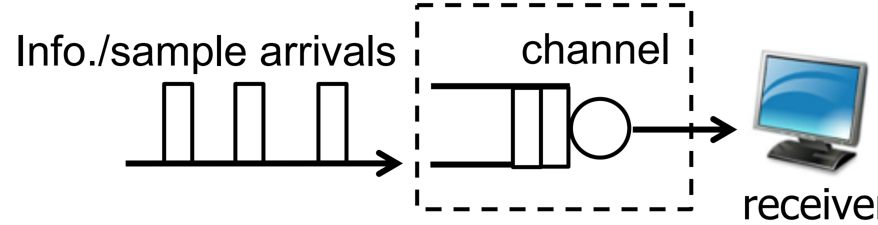
Contribution:

- Age-optimal results
- Relationship between Age-of-Info and signal processing

[Sun, Uysal, Yates, Koksal, Shroff, Infocom16], [Sun, Polyanskiy, Uysal, ISIT17]

Methodology:

- Special solution techniques of MDP



Transmission control only

Age metric:

- Any non-decreasing functional of age process **(most general)**

Contribution:

- For arbitrary arrival process, an intuitive policy (Last Generated First Served) minimizes the age process (in the sense of stochastic ordering of stochastic process)

  [Bedewy, Sun, Shroff, ISIT16&17]

Methodology:

- Sample-path ordering + coupling


Thank You!