Semantics Communications for Future Wireless Communications

Nikolaos Pappas, Associate Professor, Docent

Linköping University, Sweden

LINKÖPING UNIVERSITY

ELLIIT

CENIIT

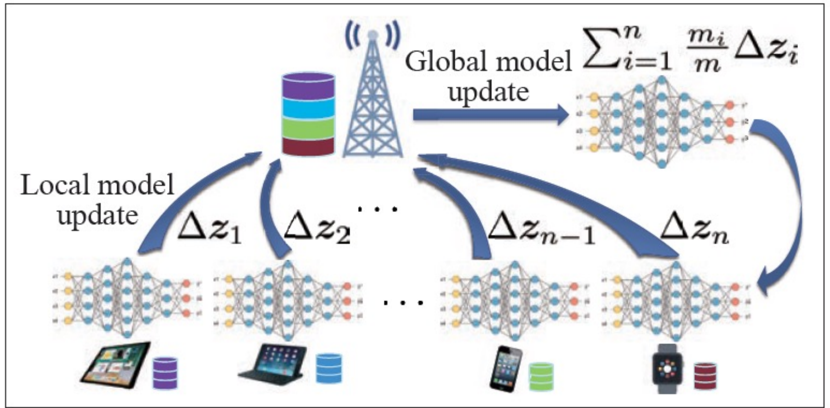
Swedish Research Council

Outline

- Introduction and motivation
- Current approaches for measuring importance
  - Age of Information (AoI)
  - Value of Information and Cost of Update Delay
- Ongoing research
  - Goal-Oriented Communication for Real-Time Tracking in Autonomous Systems
  - Semantics-aware Source Coding
  - Fault detection in IoT
- Concluding remarks

Emerging wireless ecosystem

- Networked Intelligent Systems:
  - real time autonomous systems
- Sensor fusion, on time status updates, real time information reconstruction, network and device computation, traffic flows with synced requirements, autonomous interactions
- Distributed ML over wireless
  - Exchange of large datasets in a timely manner



Emerging wireless ecosystem

The envisioned use cases and applications will stress future networks to deliver an unprecedented number of highly demanding requirements.

- All wireless systems are built upon fundamental principles of reliable communications over noisy channels.

- In existing communication paradigms, the main objective is to optimize performance metrics, such as throughput, delay, or packet loss. Quality of service (QoS) is provisioned through network over-provisioning and resource reservation control.

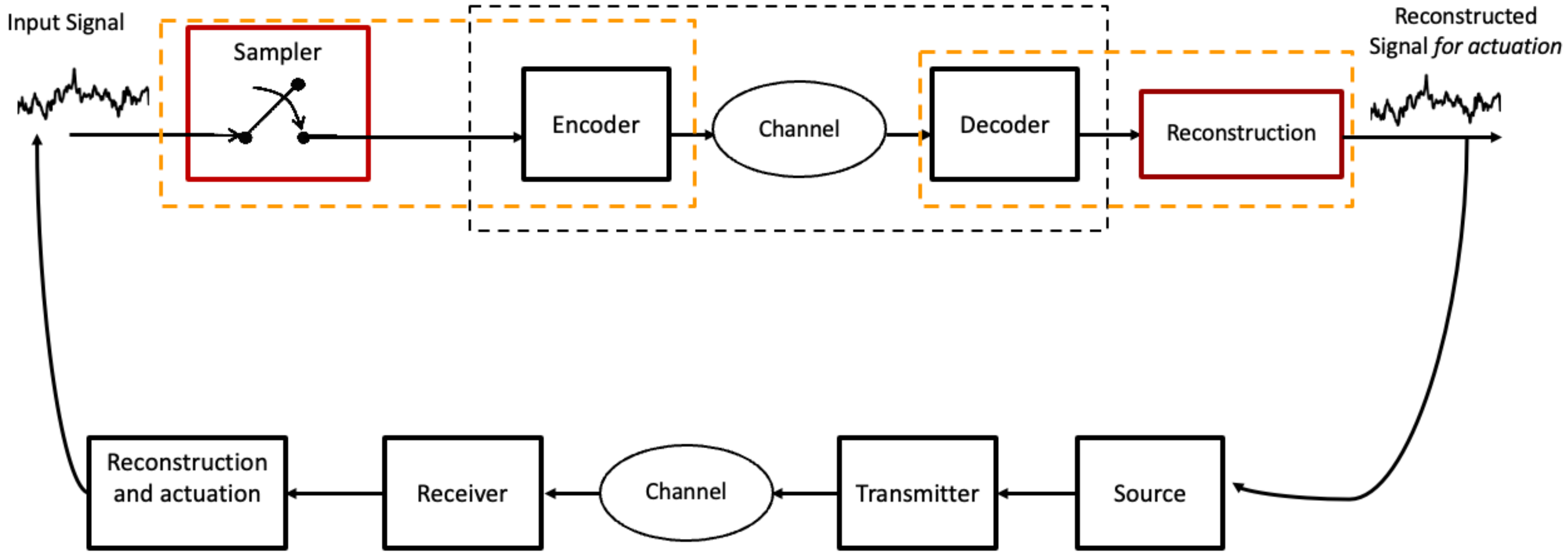
- *Is not only about understanding the throughput-reliability-delay tradeoff.*

- *Maximizing throughput or minimizing delay is not enough for optimal operation in applications based on timely status updates, remote computations, and/or real-time event detection.*

Towards Goal-oriented Semantic Communication

- Communication is about achieving specific goals.
- *Semantics: the semantic value of information is its usefulness in attaining a certain goal (pragmatics).*
- *Influence the relevance and effectiveness of the information we communicate, depending on the applications objectives.*
- Utilize *innate* and *contextual* attributes of information.
- *A holistic redesign of the entire process of information generation, processing, transmission, and reconstruction.*

Goal-oriented Semantic Communication

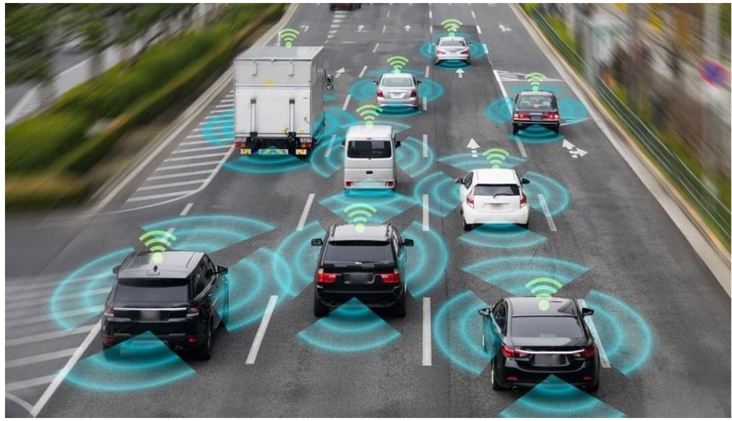


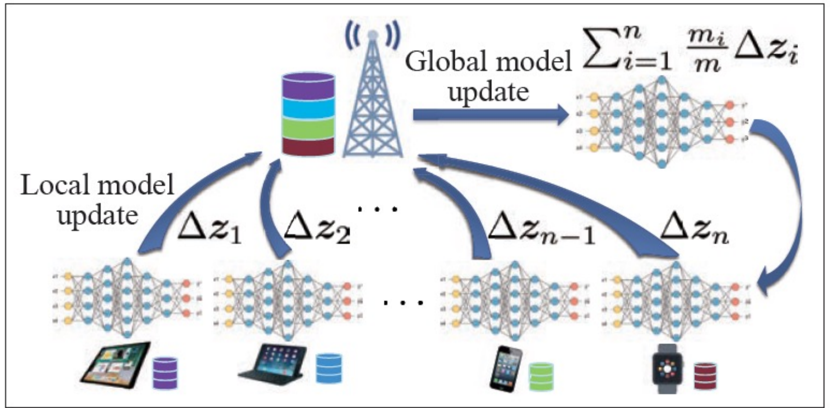
<u>Relevant use cases</u>

- **control-oriented** (e.g., remote control, actuation, stochastic process tracking,…)
- **computation-oriented** (e.g., function computation, labelling, feature extraction)
- **learning-oriented** (e.g., distributed/federated learning, …)
- **sensing/perception-oriented** (e.g., multi-view cameras, situational awareness,…)

How to quantify importance of information

- **Real-time / time sensitive systems**: *Information usually has the highest value when it is fresh! (e.g., autonomous driving: info about location/speed/sensors)*
- *Age of Information (AoI):*
  - *AoI and its variants: simple, quantitative proxy metrics of information semantics*
  - ***Instrumental in establishing suboptimality of separate handling of sampling and communication***
- Other cases such as
  - Quality of Information (QoI)
  - Value of Information (VoI)





Why we need fresh data

- Performance metrics used in the literature to characterize time sensitive information:
  - **Packet delay** tracks the time that elapsed from the generation of the packet until its delivery,
  - **inter-delivery time** is the time between two successive deliveries.
- These metrics are not sufficient to maintain fresh information at the destination.

Why we need fresh data

- Timeliness of information has emerged as a new field of network research.
- Even in the simplest queueing systems, timely updating is not the same as maximizing the utilization of the system that delivers these updates, nor the same as ensuring that updates are received with minimum delay.
  - While utilization is maximized by sending updates as fast as possible, this strategy will lead to a monitor receiving delayed updates that were backlogged in the communication system.
  - In this case, the timeliness of status updates at the receiver can be improved by reducing the update rate.
  - Reducing the update rate will cause outdated status information at the receiver due to the lack of updates.

S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" IEEE INFOCOM 2012.
A. Kosta, N. Pappas, V. Angelakis, "Age of Information: A New Concept, Metric, and Tool", Foundations and Trends in Networking: Vol. 12, No. 3, 2017.

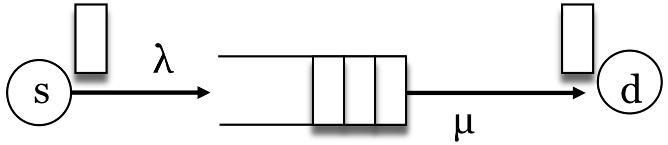
Definition and Modeling of AoI

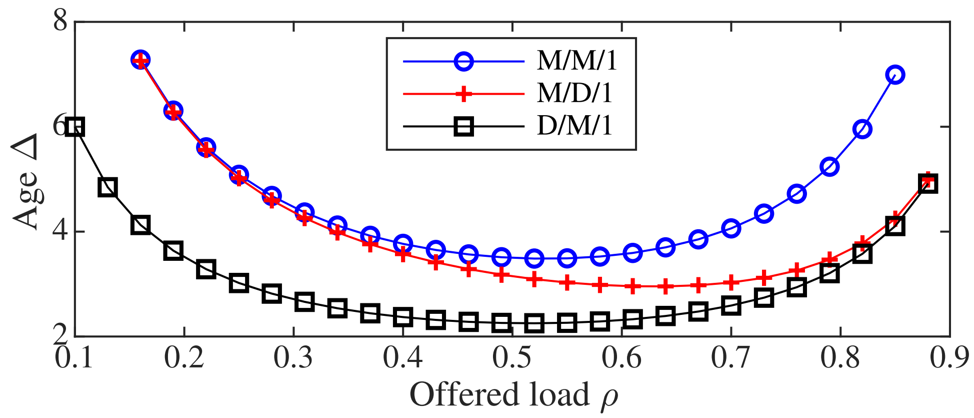
Definition of Age of Information (AoI)

- AoI is an end-to-end metric that can characterize latency in status updating systems and applications and captures the timeliness of the information.
- An update packet with timestamp u has age $t-u$ at a time t.
- An update is fresh if its age is zero.
- When the monitor's freshest received update at time t has timestamp $u(t)$, the age is the random process $\Delta(t) = t - u(t)$.

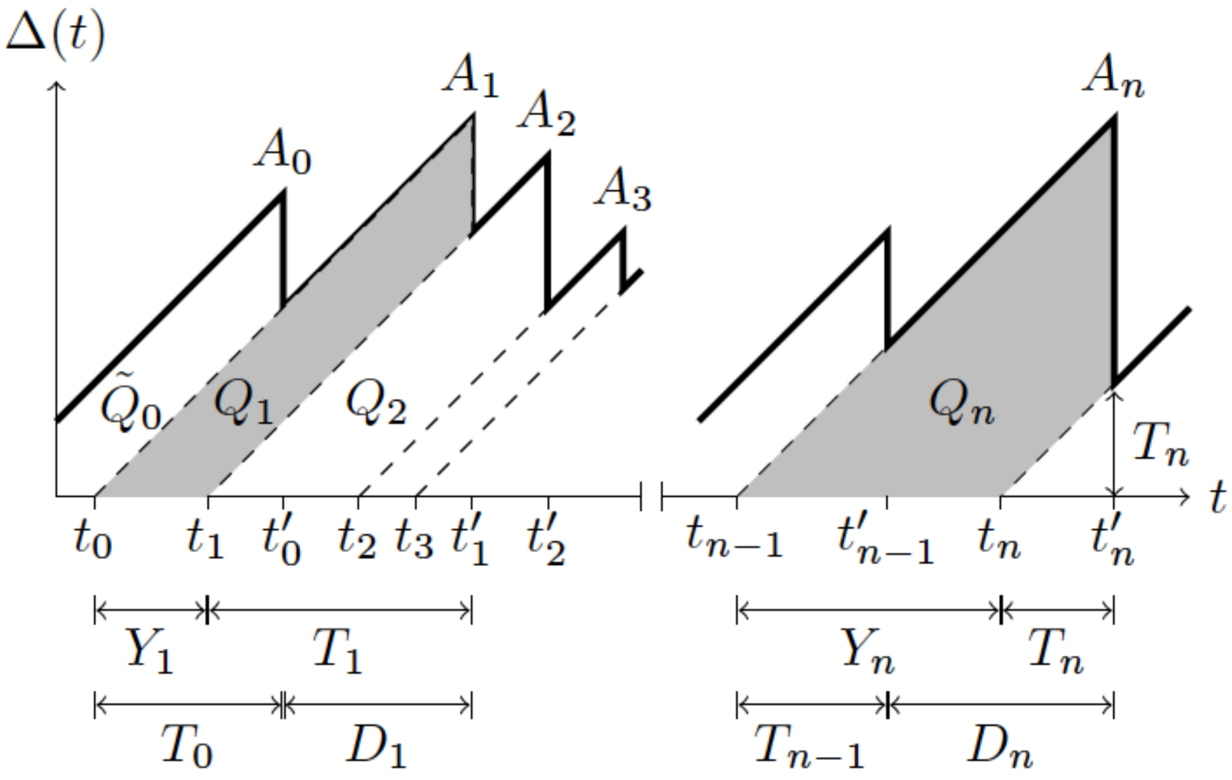
Definition of Age of Information (AoI)

- AoI is an end-to-end metric that can characterize latency in status updating systems and applications and captures the timeliness of the information.
- An update packet with timestamp *u* has age *t-u* at a time *t*.
- An update is fresh if its age is zero.
- When the monitor's freshest received update at time *t* has timestamp *u(t)*, the age is the random process $\Delta(t) = t - u(t)$.





Time Average AoI – Sawtooth Sample path

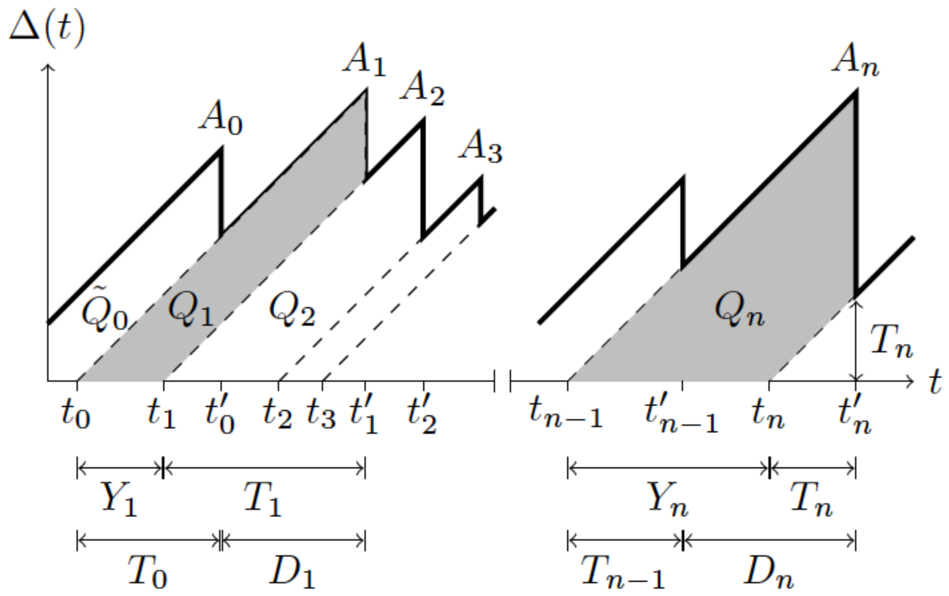


- $t_0, t_1, t_2, \ldots$ times that are updates are generated
- $t_0', t_1', t_2', \ldots$ times that updates are received at the monitor

- For the n-th received update
  - $Y_n = t_n - t_{n-1}$ interarrival time
  - T_n system time
  - $D_n = t_n' - t_{n-1}'$ interdeparture time
  - A_n corresponding peak age

- A. Kosta, N. Pappas, V. Angelakis, "Age of Information: A New Concept, Metric, and Tool", Foundations and Trends in Networking: Vol. 12, No. 3, 2017.
- R. D. Yates, Y. Sun, D. R. Brown III, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of Information: An Introduction and Survey", IEEE JSAC SI AoI, May 2021.

Time Average AoI $\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}} \Delta(t)dt \qquad \mathcal{T} = t'_n$



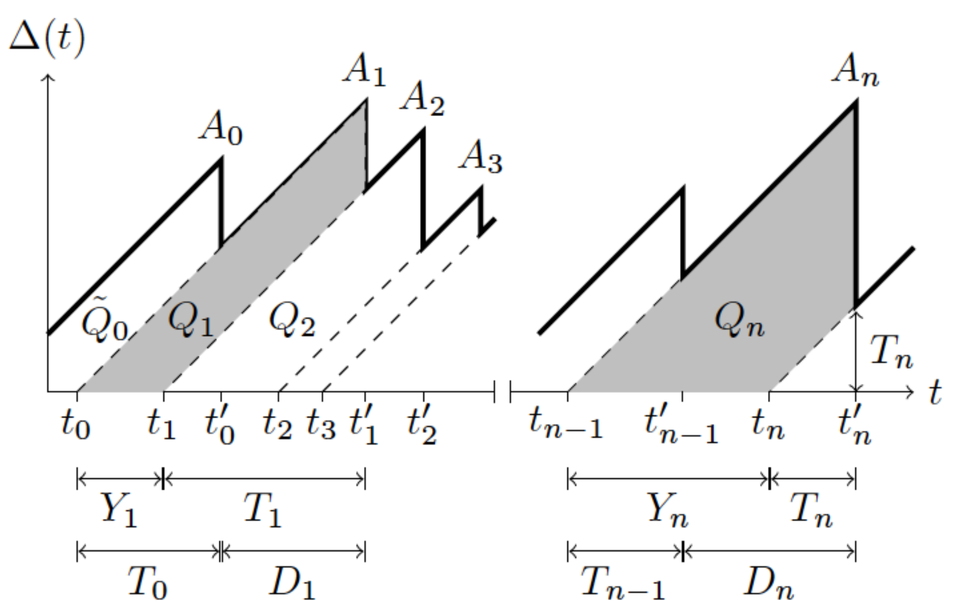
$$Q_n = \frac{1}{2}(T_n + Y_n)^2 - \frac{1}{2}T_n^2 = Y_n T_n + Y_n^2/2$$

$$N(\mathcal{T})/\mathcal{T} \to 1/\mathrm{E}[Y]$$

$$\frac{1}{N(\mathcal{T})}\sum_{j=1}^{N(\mathcal{T})} Q_j \to \mathrm{E}[Q] \quad \mathcal{T} \to \infty$$

$$\Delta = \frac{\mathrm{E}[Q_n]}{\mathrm{E}[Y_n]} = \frac{\mathrm{E}[Y_n T_n] + \mathrm{E}\left[Y_n^2\right]/2}{\mathrm{E}[Y_n]}$$

Time Average AoI $\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}} \Delta(t)dt \quad \mathcal{T} = t'_n$



$$Q_n = \frac{1}{2}(T_n + Y_n)^2 - \frac{1}{2}T_n^2 = Y_n T_n + Y_n^2/2$$

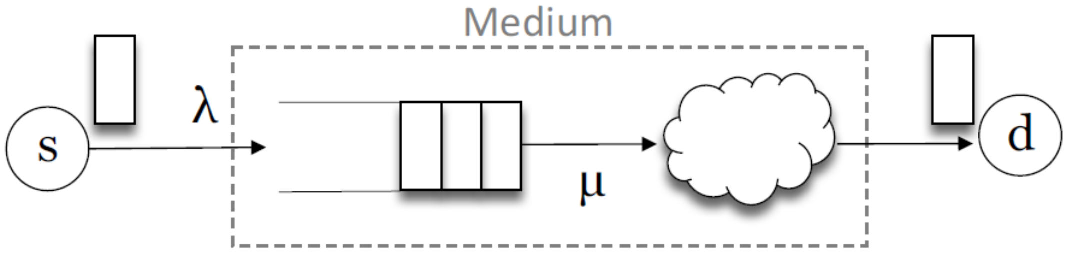
$$N(\mathcal{T})/\mathcal{T} \to 1/\mathrm{E}[Y]$$

$$\frac{1}{N(\mathcal{T})}\sum_{j=1}^{N(\mathcal{T})} Q_j \to \mathrm{E}[Q] \quad \mathcal{T} \to \infty$$

$$\Delta = \frac{\mathrm{E}[Q_n]}{\mathrm{E}[Y_n]} = \frac{\mathrm{E}[Y_n T_n] + \mathrm{E}\left[Y_n^2\right]/2}{\mathrm{E}[Y_n]}$$

Large interarrival time allows queue to be empty, thus, the waiting time can be small, causing small system time T_n.
Y_n and T_n are negatively correlated which complicates the calculation of $E[Y_n T_n]$

Single-source and single-server systems



i.i.d interarrival times with expected value E[Y]
λ=1/ E[Y] : arrival rate
E[S] : expected service time
μ=1/ E[S] : service rate
ρ = λ/ μ : offered load

For FCFS M/M/1 queue the average is $\Delta_{\text{M/M/1}} = \frac{1}{\mu}\left(1 + \frac{1}{\rho} + \frac{\rho^2}{1-\rho}\right)$
The optimal age is achieved for $\rho^* \approx 0.53$

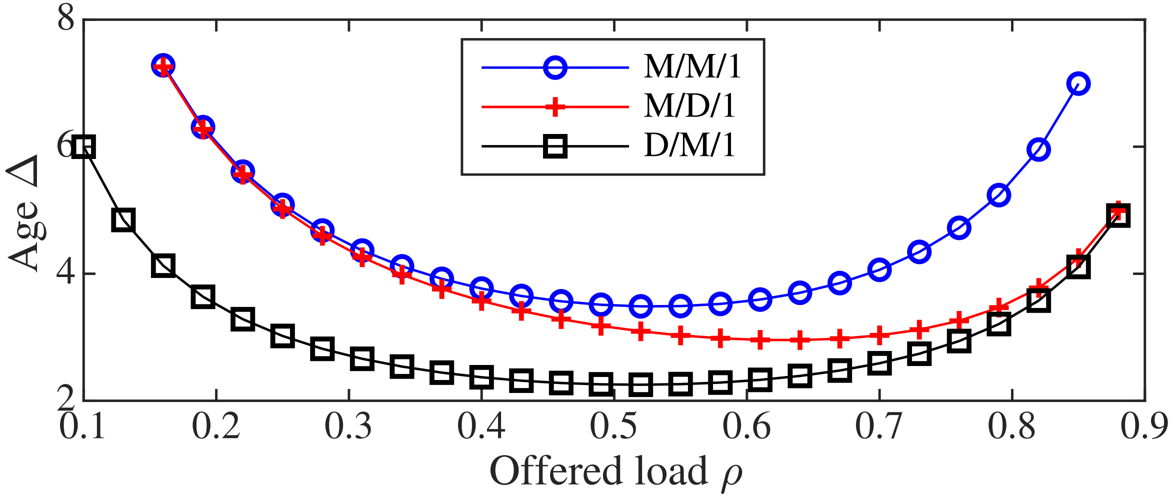
- Optimal age is achieved by choosing a λ which makes the server being slightly busy than being idle.
- If ρ is close to 1 we maximize the throughput.
- If ρ is close to 0, we minimize the delay.

- S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" IEEE INFOCOM 2012.
- Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "A general formula for the stationary distribution of the age of information and its application to single-server queues," IEEE Trans. Info. Theory 2019.

Single-source and single-server systems

For M/D/1 and D/M/1 queues the average AoI are given by

$$\Delta_{\mathrm{M/D/1}} = \frac{1}{\mu}\left(\frac{1}{2(1-\rho)} + \frac{1}{2} + \frac{(1-\rho)\exp(\rho)}{\rho}\right) \qquad \Delta_{\mathrm{D/M/1}} = \frac{1}{\mu}\left(\frac{1}{2\rho} + \frac{1}{1-\gamma(\rho)}\right) \qquad \gamma(\rho) = -\rho\mathcal{W}\left(-\rho^{-1}e^{(-1/\rho)}\right)$$



- At low load, randomness in the interarrivals dominates the average age.
- At high load, M/D/1 and D/M/1 outperform M/M/1 because the determinism in either arrivals or service helps to reduce the average queue length.
- Unique value of ρ that minimizes the average age.

Single-source and single-server systems – Packet management

- The arrival rate can be optimized to balance frequency of updates against congestion.
  - Departure from the external arrivals assumption.
- Study of lossy queues that may discard an arriving update while the server was busy or replace an older waiting update with a fresher arrival.
- ***Packet management inherently prioritizes some packets over others which is a first indication of different value of the packets thus the prioritization!***

- S. Kaul, R. Yates, M. Gruteser, "Status updates through queues", CISS 2012.
- N. Pappas, J. Gunnarsson, L. Kratz, M. Kountouris, V. Angelakis, "Age of Information of Multiple Sources with Queue Management", IEEE ICC 2015.
- M. Costa, M. Codreanu, A. Ephremides, "On the age of information in status update systems with packet management", IEEE Trans. Info. Theory 2016.
- A. Kosta, N. Pappas, A. Ephremides, V. Angelakis, "Age of Information Performance of Multiaccess Strategies with Packet Management", IEEE/KICS JCN, June 2019.

Non-linear Ageing

- AoI grows over time linearly
  - the performance degradation caused by information aging may not be a linear function of time.
- One way to capture the nonlinear behavior of information aging is to *define freshness and staleness as nonlinear functions of AoI*.
- A penalty function of the AoI is non-decreasing. *Outdated data is usually less desirable than fresh data.*

- Y. Sun, E. Uysal-Biyikoglu, R. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh", IEEE Trans. Inf. Theory, 2017.
- A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case", IEEE ISIT 2017.
- Y. Sun and B. Cyr, "Sampling for data freshness optimization: Nonlinear age functions", IEEE/KICS JCN 2019.
- A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The cost of delay in status updates and their value: Non-linear ageing", IEEE Trans. Comm., 2020.

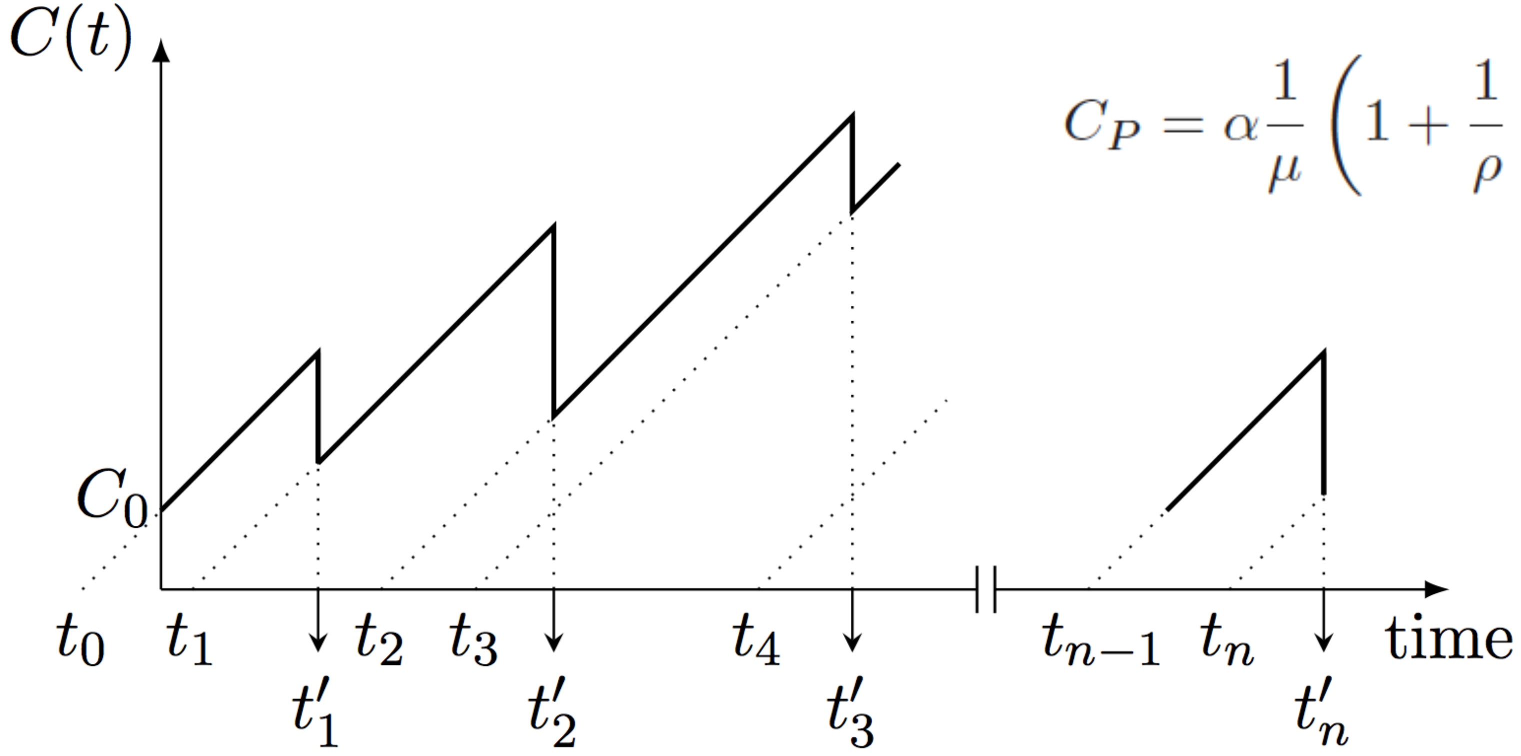
Cost of Update Delay (CoUD)

- CoUD metric associates the cost of staleness with the statistics of the source
- $C(t) = f_s(t-u(t))$
  - $f_s(t)$ is a monotonically increasing function
  - $u(t)$ timestamp of the most recently received update
- Different cost functions can represent different utilities

- A. Kosta, N. Pappas, A. Ephremides, V. Angelakis, "Age and Value of Information: Non-linear Age Case", *IEEE ISIT* 2017.
- A. Kosta, N. Pappas, A. Ephremides, V. Angelakis, "The Cost of Delay in Status Updates and their Value: Non-linear Ageing", IEEE Trans. Comm., 2020.

Cost of Update Delay (CoUD): The linear case

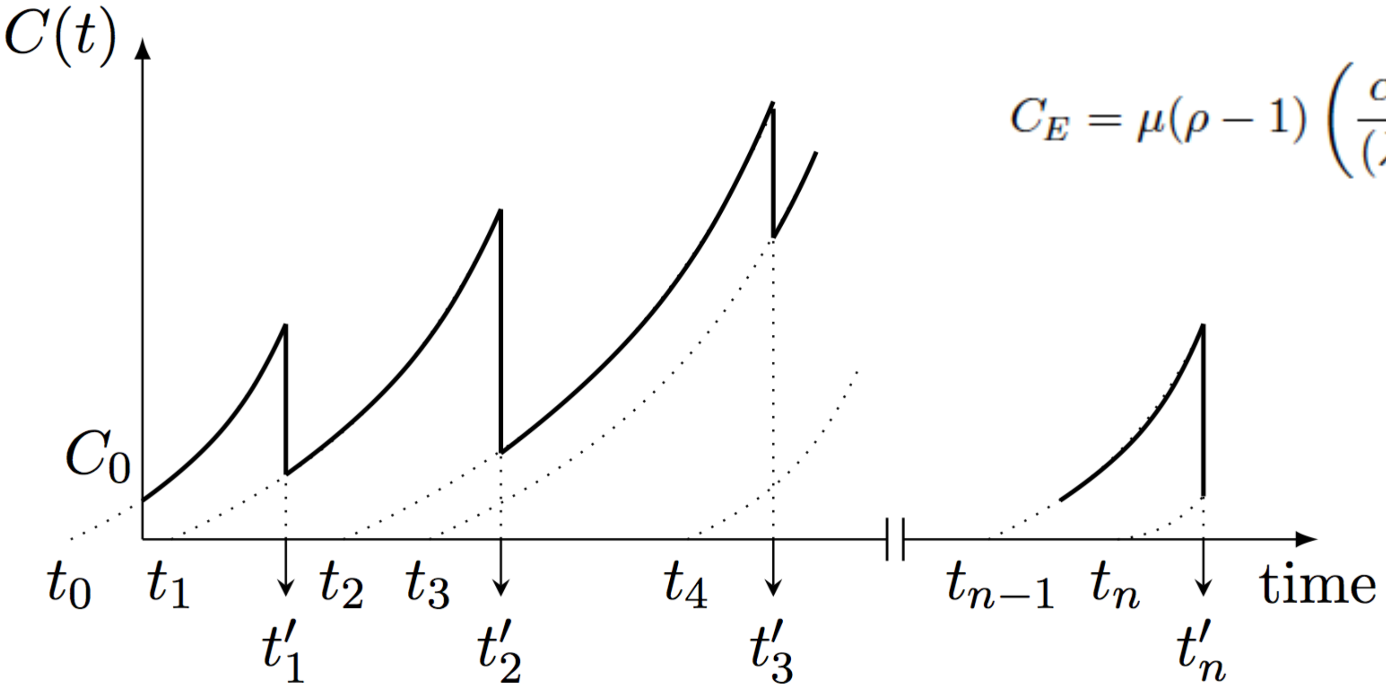
$$f_s(t) = \alpha t$$



$$C_P = \alpha \frac{1}{\mu}\left(1 + \frac{1}{\rho} + \frac{\rho^2}{1-\rho}\right)$$

Cost of Update Delay (CoUD): The exponential case

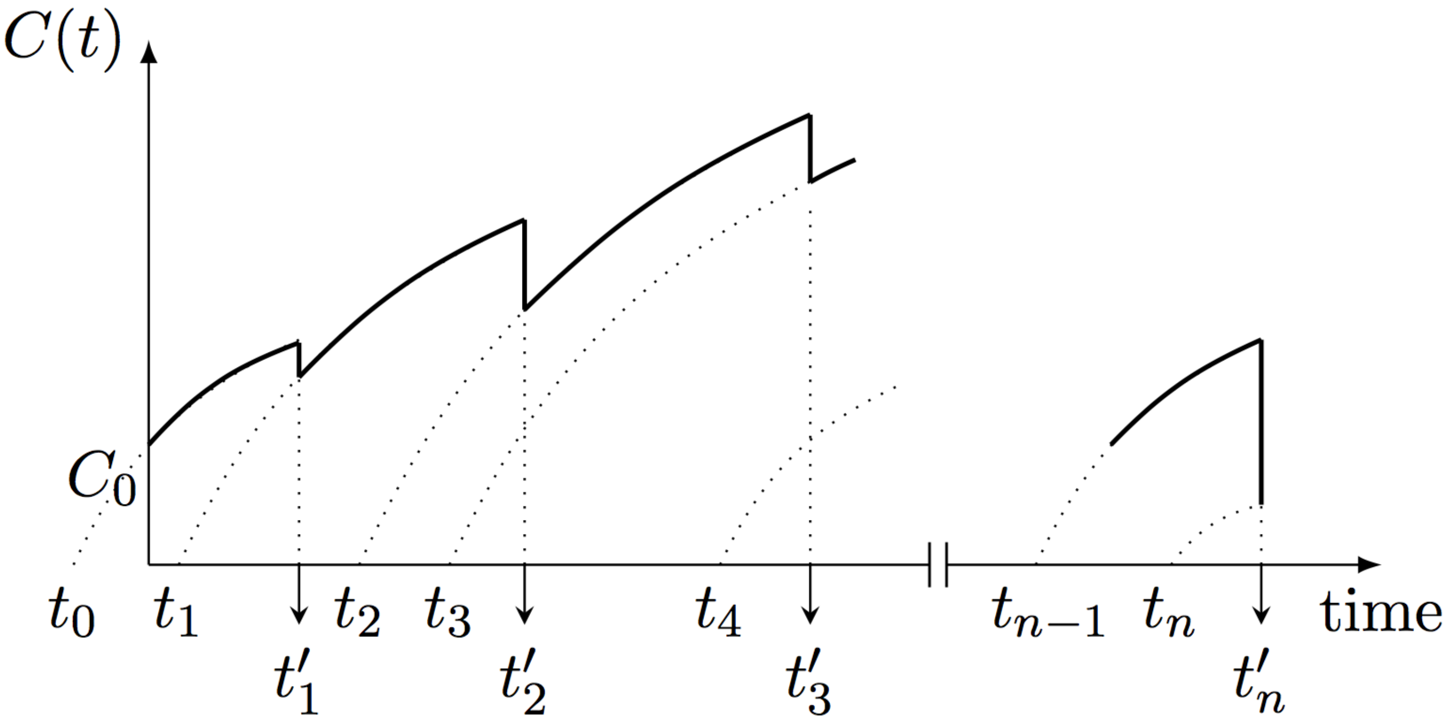
$f_s(t) = e^{\alpha t} - 1 \longleftrightarrow$ low autocorrelation



$$C_E = \mu(\rho - 1)\left(\frac{\alpha(\alpha - (\lambda + \mu))}{(\lambda - \alpha)(\alpha - \mu)^2} + \frac{1}{\alpha - \mu(1-\rho)} + \frac{1}{\mu(1-\rho)}\right)$$

Cost of Update Delay (CoUD): The logarithmic case

$$f_s(t) = \log(\alpha t + 1) \longleftrightarrow \text{high autocorrelation}$$

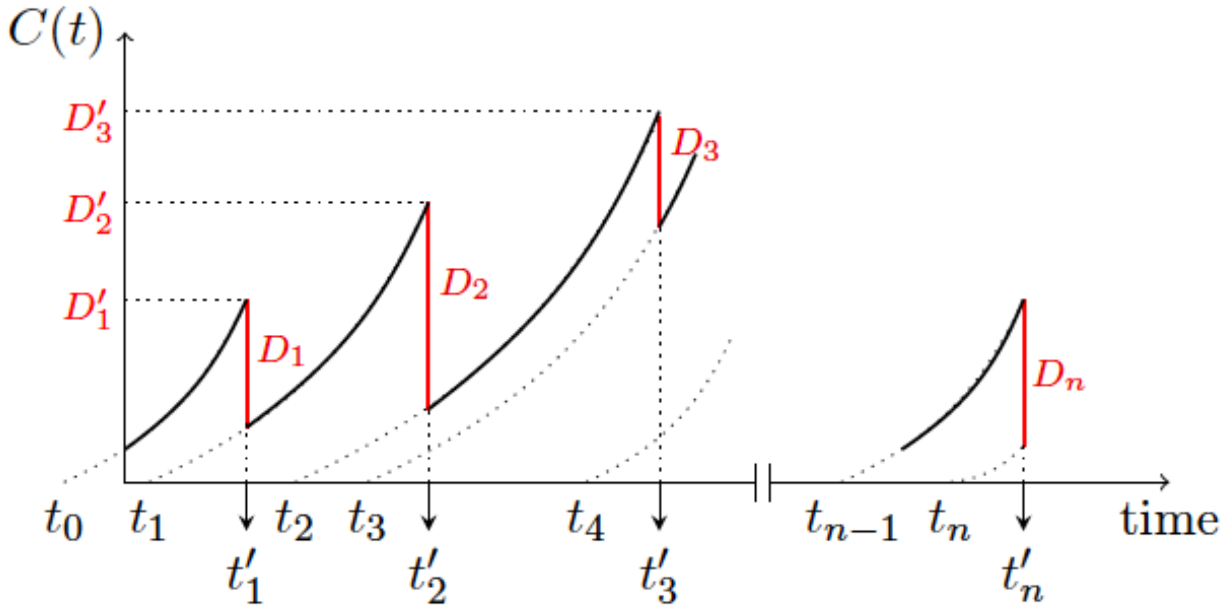


$$C_L = \frac{1}{\alpha(\lambda-\mu)^2}\left(e^{-\frac{\mu\rho}{\alpha}}\left(\mu(1-\rho)Ei\left[-\frac{\mu}{a}\right](\alpha\mu+\lambda^2-\lambda\mu)\,e^{\frac{\mu(\rho+1)}{\alpha}} - \alpha\mu^2(1-\rho)Ei\left[-\frac{\lambda}{\alpha}\right]e^{\frac{\lambda+\mu\rho}{\alpha}} - \alpha e^{\mu/\alpha}(\lambda-\mu)^2 Ei\left[-\frac{\mu(1-\rho)}{\alpha}\right]\right) - \alpha\lambda(1-\rho)(\mu-\lambda)\right)$$

Value of Information of Update (VoIU)

- It captures *the degree of importance of an update*

$$V_i = \frac{f_s(t'_i - t_{i-1}) - f_s(t'_i - t_i)}{f_s(t'_i - t_{i-1})} = \frac{D_i}{D'_i}.$$



- In the linear CoUD case, VoIU is independent of the cost assigned per time unit →the Value is independent of the slope.

$$V_i = \lim_{t'_i \to t_i} \frac{f_s(t'_i - t_{i-1}) - f_s(t'_i - t_i)}{f_s(t'_i - t_{i-1})} = 1$$

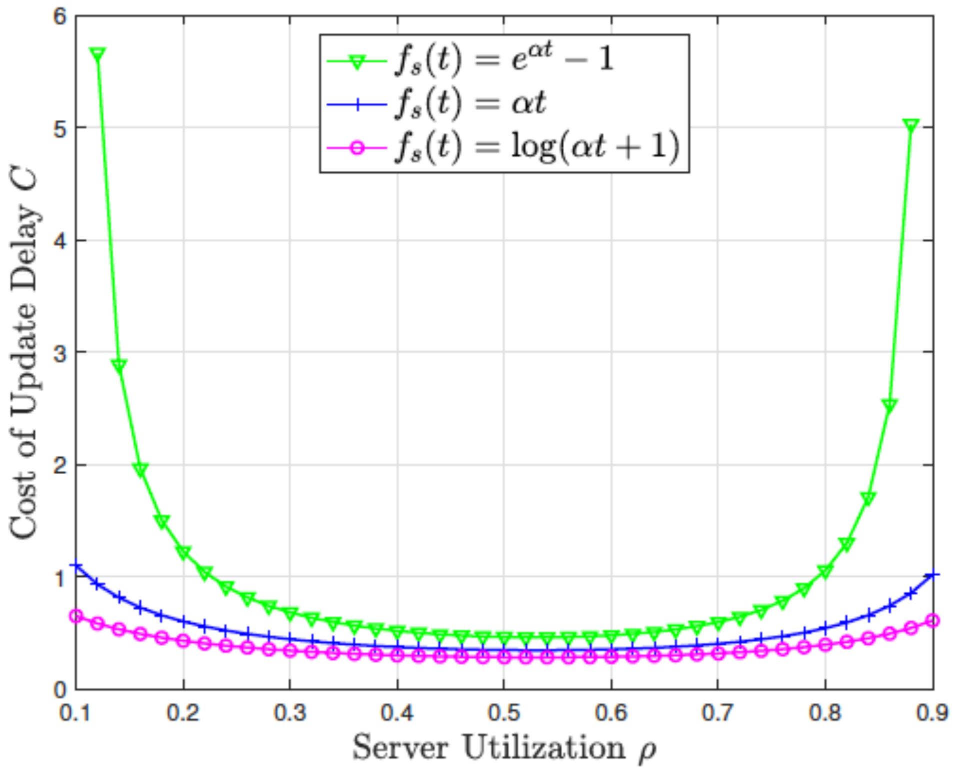
- Linear case, the average VoIU for the M/M/1 system with an FCFS queue discipline.

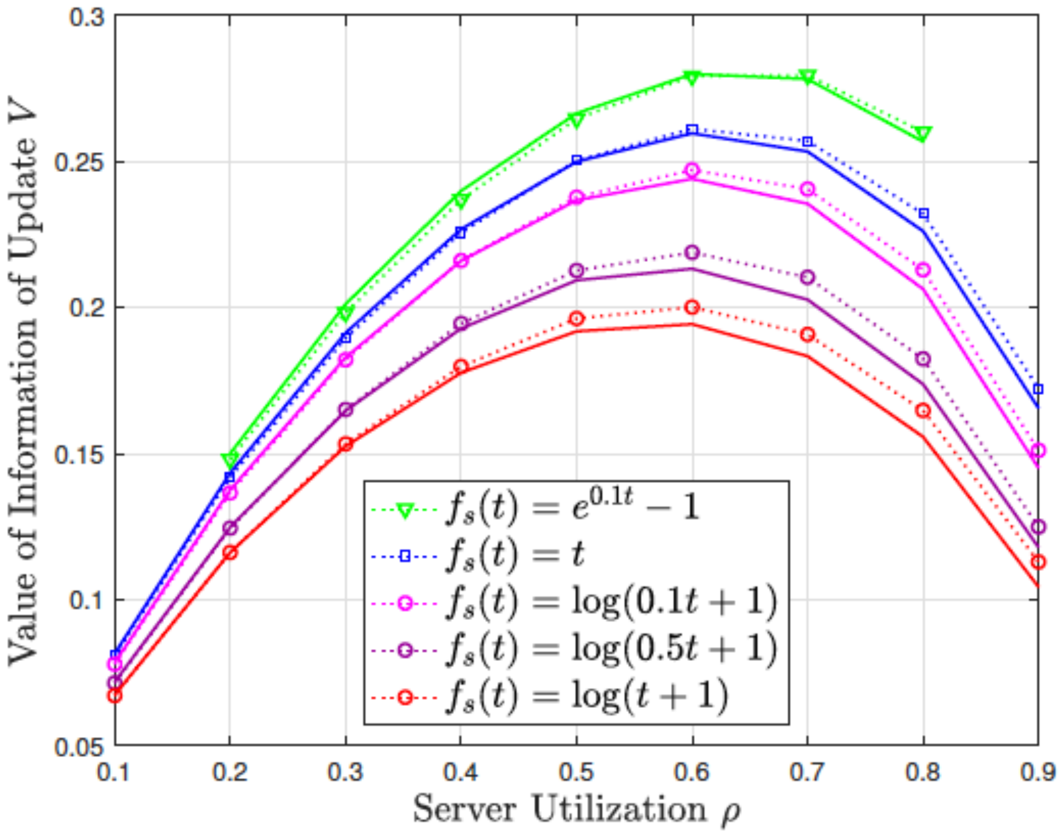
$$V_P = \lambda \frac{(1-\rho)}{2\rho} {}_2F_1\left(1, 2; 3; 2 - \frac{1}{\rho}\right)$$

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}$$

Pochhammer symbol

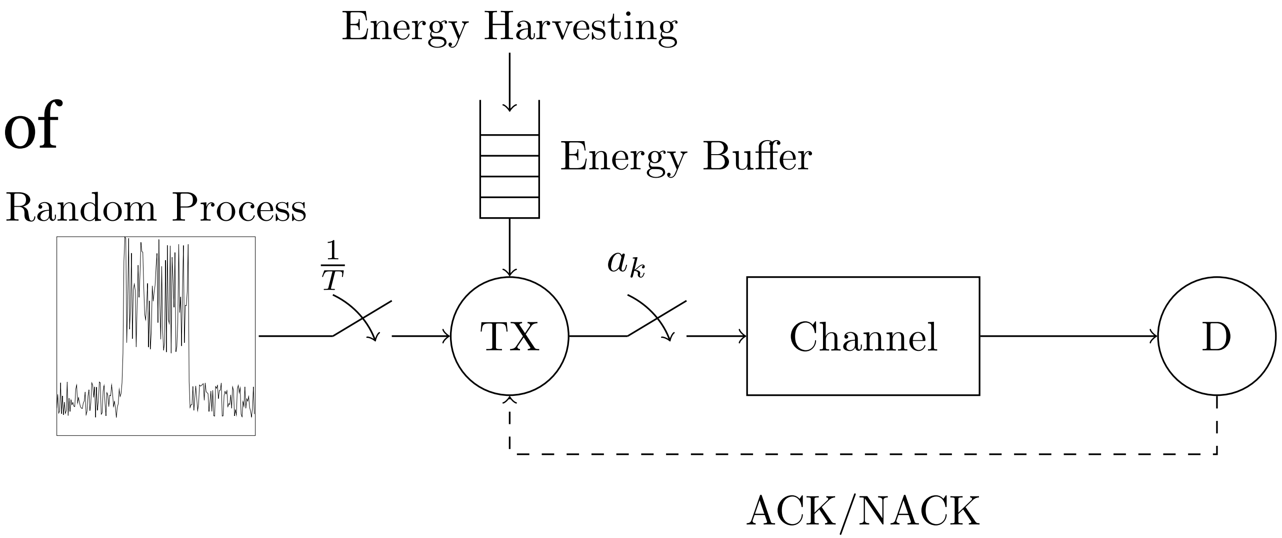
Numerical evaluation





Extending AoI

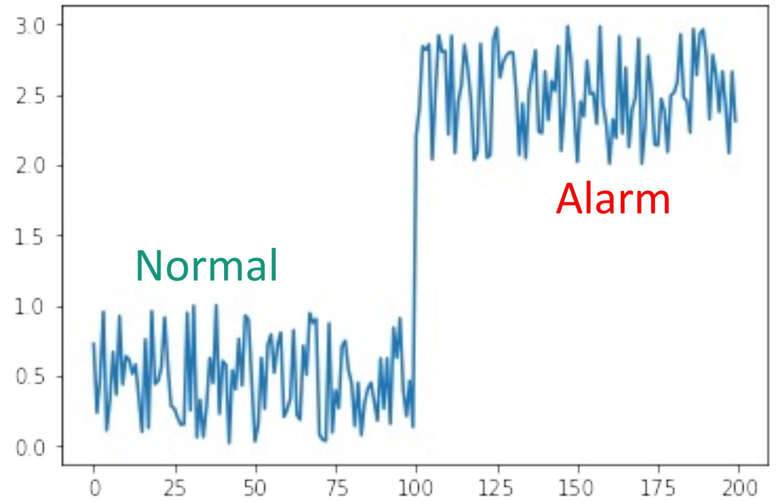
- The classical AoI does not capture properties of the source
  - except timeliness itself which is a semantic property.
- With non-linear AoI and the VoI, we can go a step further.
- Here we will discuss another extension of AoI.

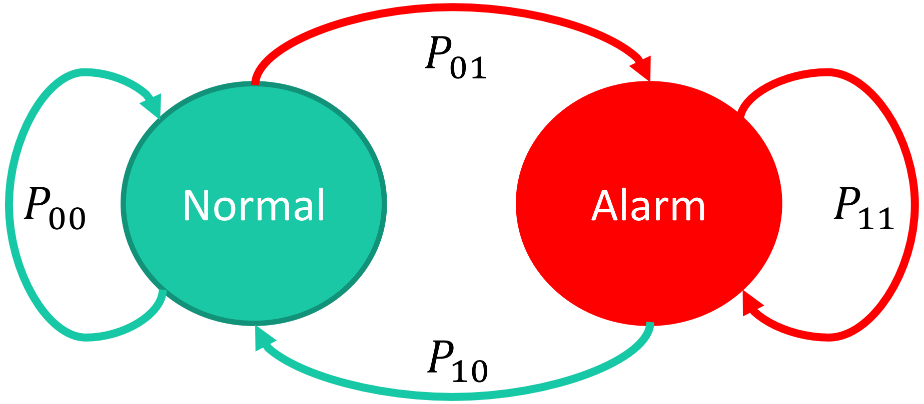


G. Stamatakis, N. Pappas, A. Traganitis, "Control of Status Updates for Energy Harvesting Devices that Monitor Processes with Alarms", IEEE GLOBECOM Workshops, Dec. 2019.

Stochastic process with alarms (or a two-state)

- Short timescale: Stochastic process Z_k evolves over discrete time k.
- Long timescale: A two state Markov Process
  - States: Normal, Alarm,
  - Geometrically distributed sojourn times in each state.
- The alarm state indicates the need for more frequent updates
  - Closely follow/track the evolution of Z_k to make informed decisions.
- Examples of Z_k
  - the network load under normal operation and under a DoS attack.
  - Physical phenomena such as temperature, water levels, and air pollution.
  - *Tracking of a process in general*.
- Objective: Optimize the freshness of status updates at the destination while considering the energy resources currently available as well as future demands for energy (especially during alarm periods).



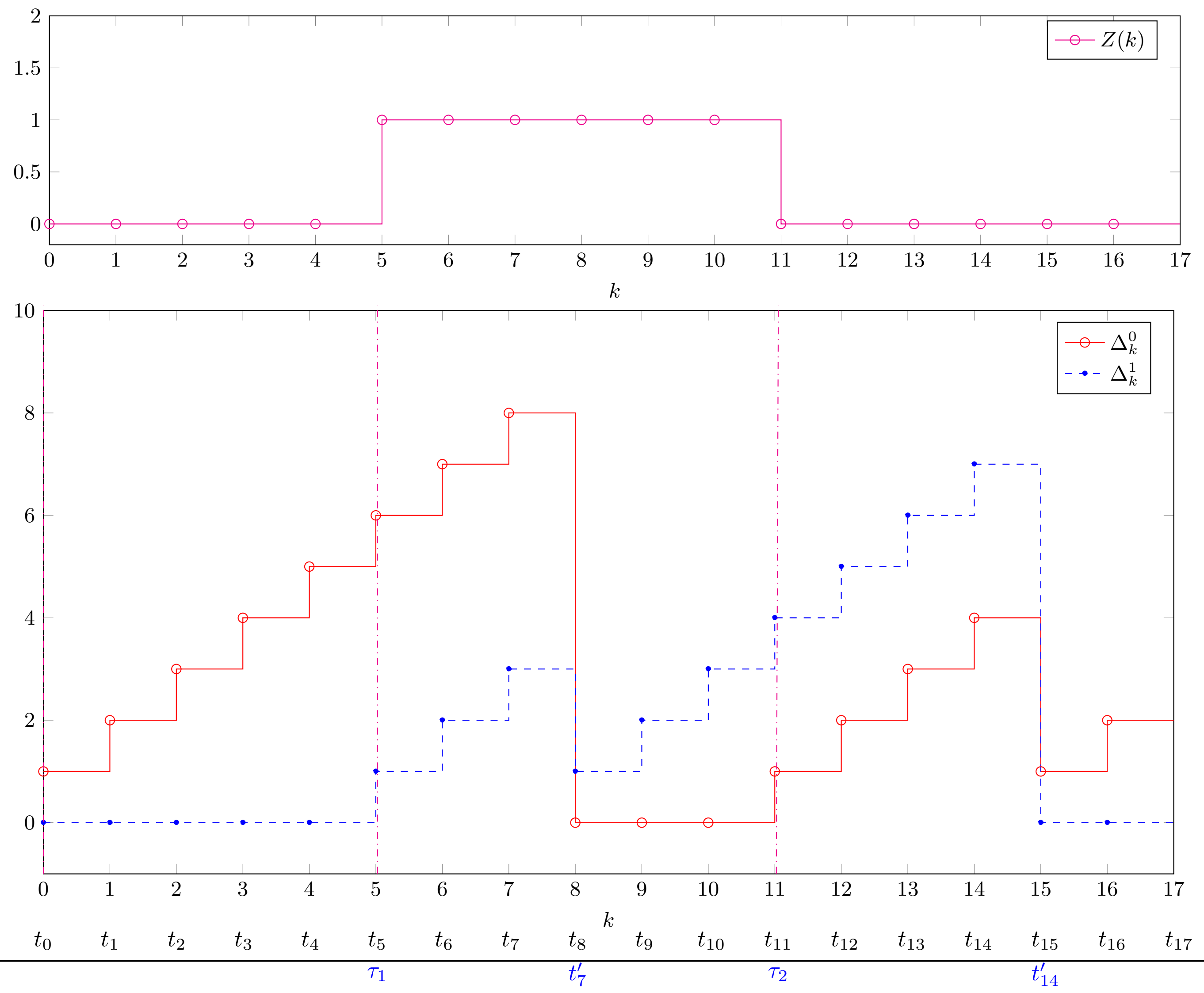


G. Stamatakis, N. Pappas, A. Traganitis, "Control of Status Updates for Energy Harvesting Devices that Monitor Processes with Alarms", IEEE GLOBECOM Workshops, Dec. 2019.

AoI for stochastic processes with alarms

- Extend the definition of AoI
  - the amount of time that has elapsed since the generation the last status update that has been successfully received by the monitor (typical)
  - the amount of time that has elapsed since the last state change of the stochastic process for which the destination in uninformed (new)
- Use two AoI variables, one for each state of the process $\Delta_{\mathrm{k}}^{z}, z \in \{0,1\}$.
- The destination knows the stochastic process to be in state Z_k^d.
  - Not necessarily the actual state of the stochastic process indicated with Z_k.
- Sequence of time indices where a state change has occurred
  - $\{\tau_n : Z_{\tau_n} \neq Z_{\tau_{n-1}}, n = 1,2, \dots\}$
- Extended definition of AoI $\Delta_{\mathrm{k}}^{z} = \begin{cases} k - U_k, & \text{if } z = Z_k^d \\ k - \tau_n, & \text{if } z \neq Z_k^d \text{ and } z = Z_k \\ 0, & \text{if } z \neq Z_k^d \text{ and } z \neq Z_k \end{cases}$

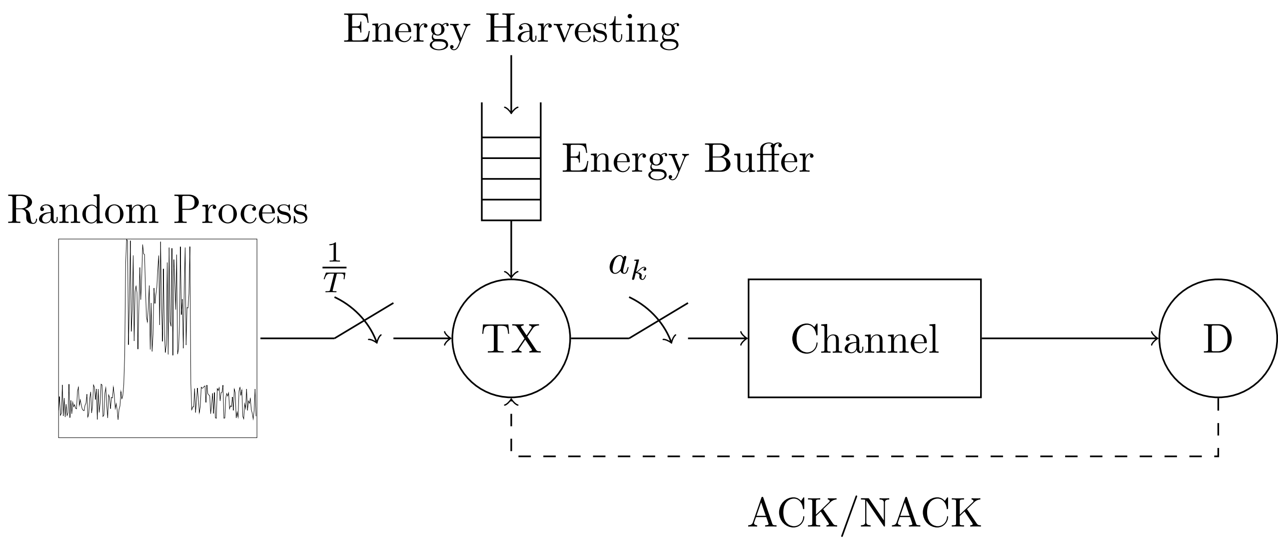
Extended AoI - illustration for the two-state process



$$\Delta_{\mathrm{k}}^{z} = \begin{cases} k - U_k, & \text{if } z = Z_k^d \\ k - \tau_n, & \text{if } z \neq Z_k^d \text{ and } z = Z_k \\ 0, & \text{if } z \neq Z_k^d \text{ and } z \neq Z_k \end{cases}$$

System model (State & Action spaces)

- At the beginning of the k-th timeslot the sensor samples the stochastic process in order to assess Z_k.
- The sensor also considers
  - The state of the process known at the destination Z_k^d.
  - The energy stored at the energy buffer E_k
  - The values of both AoI variables $\Delta_\mathrm{k}^0, \Delta_\mathrm{k}^1$
- These features constitute the **state** of the system
  - $s_k = [Z_k, Z_k^d, E_k, \Delta_\mathrm{k}^0, \Delta_\mathrm{k}^1]$
- Given s_k the sensor must choose whether to transmit a fresh status update or not, $a_k \in \{0,1\}$.



System model (Stochastics & Dynamics)

- By the end of the k-th timeslot
  - An energy unit may have been harvested as indicated by random variable $W_k^e \in \{0,1\}$ with P_e
  - The state of the stochastic process will change randomly $W_k^z \in \{0,1\}$
    - $P_z = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix}$
  - If a transmission is attempted, it will succeed with probability
    - $P_s = P\{W_k^s = 1\}, W_k^s \in \{0,1\}$

- Determine the state of the system at the beginning of the $(k+1)$-th time-slot
  - $Z_{k+1} = W_k^z$
  - $Z_{k+1}^d = \begin{cases} Z_k^d, \text{if } W_k^s = 0 \\ Z_{k,} \text{if } W_k^s = 1 \end{cases}$
  - $E_{k+1} = \begin{cases} E_k + W_k^e - 1, \text{if } a_k = 1 \\ E_k + W_k^e, \text{if } a_k = 0 \end{cases}$
  - $\Delta_{\mathrm{k+1}}^0, \Delta_{\mathrm{k+1}}^1$ are given by a recursive expression equivalent to the definition presented.

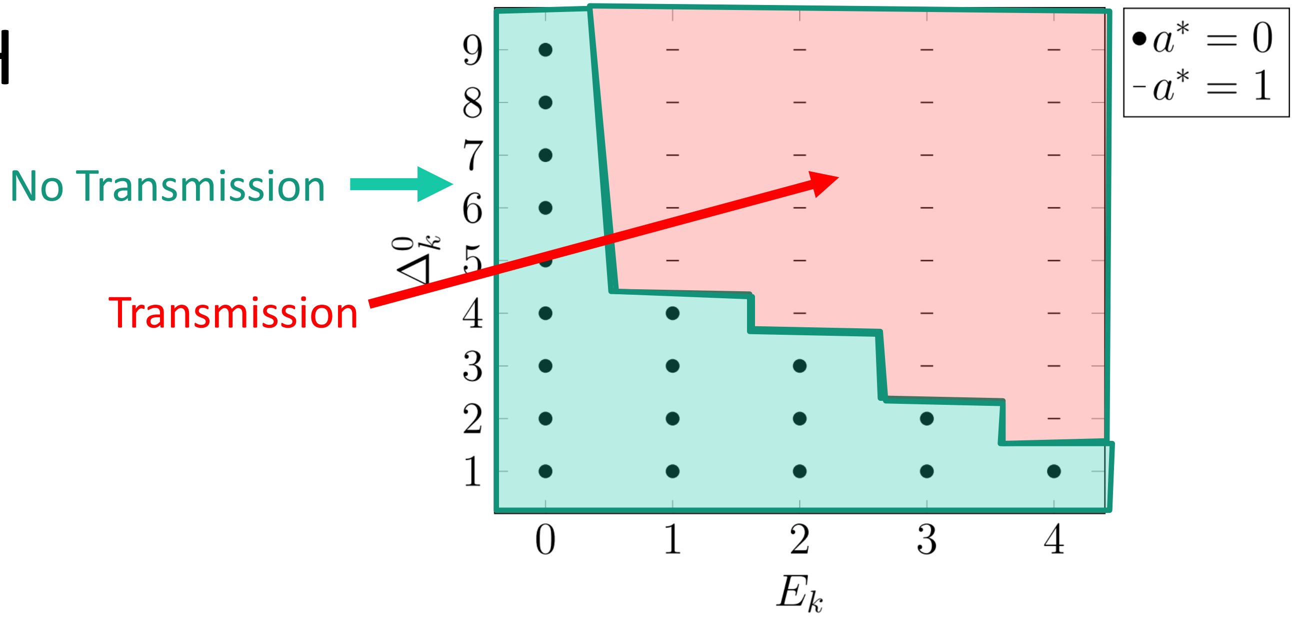
System model (Transition & Total Cost, Optimal policy)

- At the end of each time-slot a cost is paid by the sensor.
- The transition cost is an increasing function of Δ_k^0 and Δ_k^1.
- $g(\Delta_k^0, \Delta_k^1) = g_0(\Delta_k^0) + g_1(\Delta_k^1)$
  - **$g_1(\cdot)$ increases faster than $g_0(\cdot)$.**
  - ***This expresses the need for frequent status updates when in alarm state →The value of information in that case is higher!***
- Examples:
  - $g(\Delta_k^0, \Delta_k^1) = (1 - Z_k)\Delta_k^0 + Z_k(\Delta_k^1)^2$
  - i.e., cost is a function of the true state of the stochastic process and not the one perceived by the destination.
  - $g(\Delta_k^0, \Delta_k^1) = \Delta_k^0 + (\Delta_k^1)^2$
  - cost considers both AoI variables simultaneously (Upcoming work)

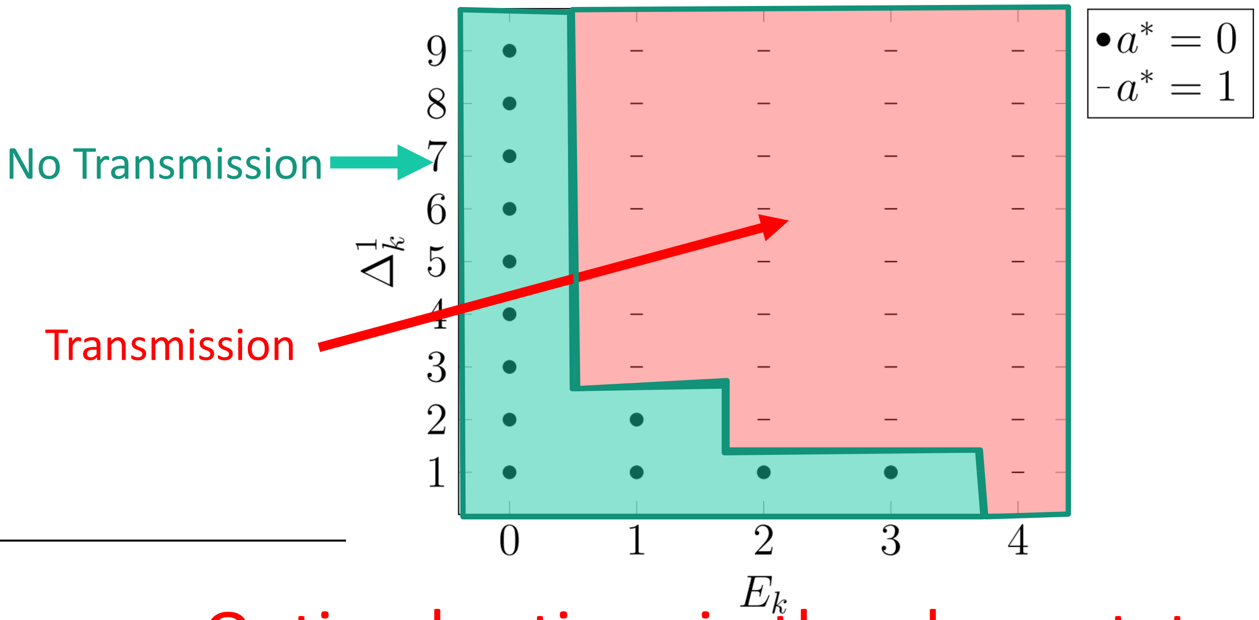
- Objective: Find an optimal policy that, given s_k, decides whether to transmit a status update to minimize the discounted transition costs accumulated over an infinite horizon.
- The problem is a Markov Decision Process, and the optimal policy can be found via the Value Iteration algorithm.
- *The curse of dimensionality can be circumvented by utilizing structural results for the optimal policy.*

Optimal Policy - Low Probability EH ($P_e = 0.4$) - $P_s = 0.8$

- Scenario
  - the process spends most of its time in normal mode with relative short periods of alarm states.
    - State transition matrix $P_z = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$
- *The optimal policy will save energy in the normal state in order to be able to transmit in the alarm state*
  - Threshold structure: Transmissions occur when Δ_k^0 and Δ_k^1 is larger than a threshold value given E_k.



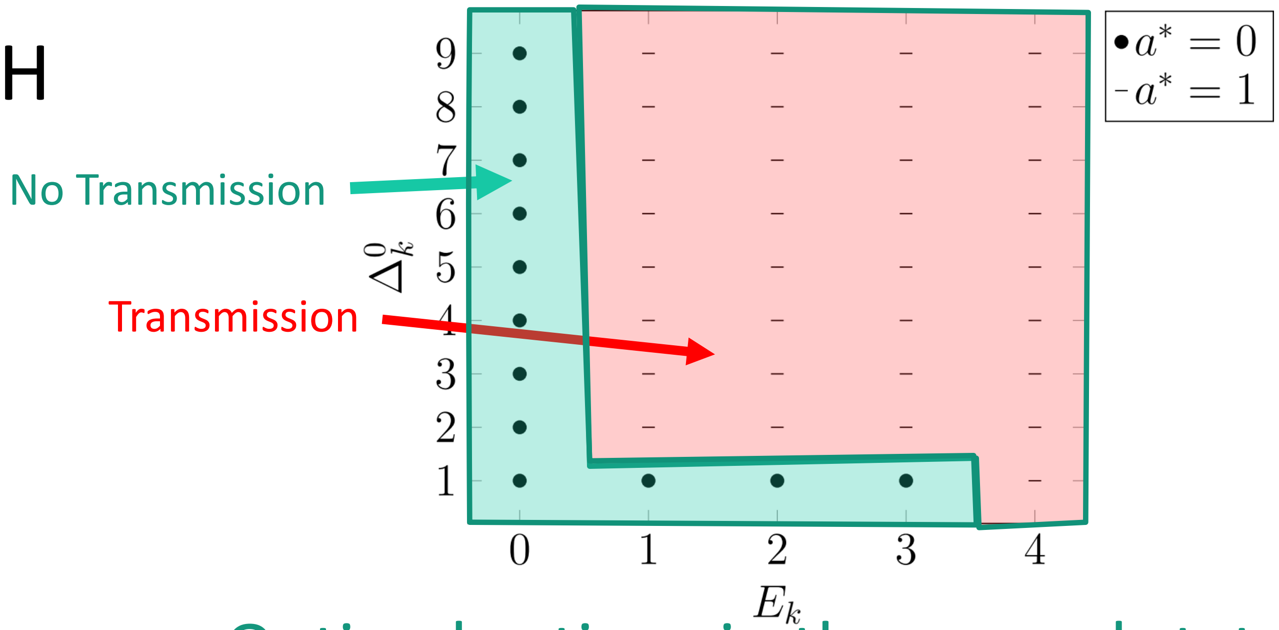
Optimal actions in the normal state



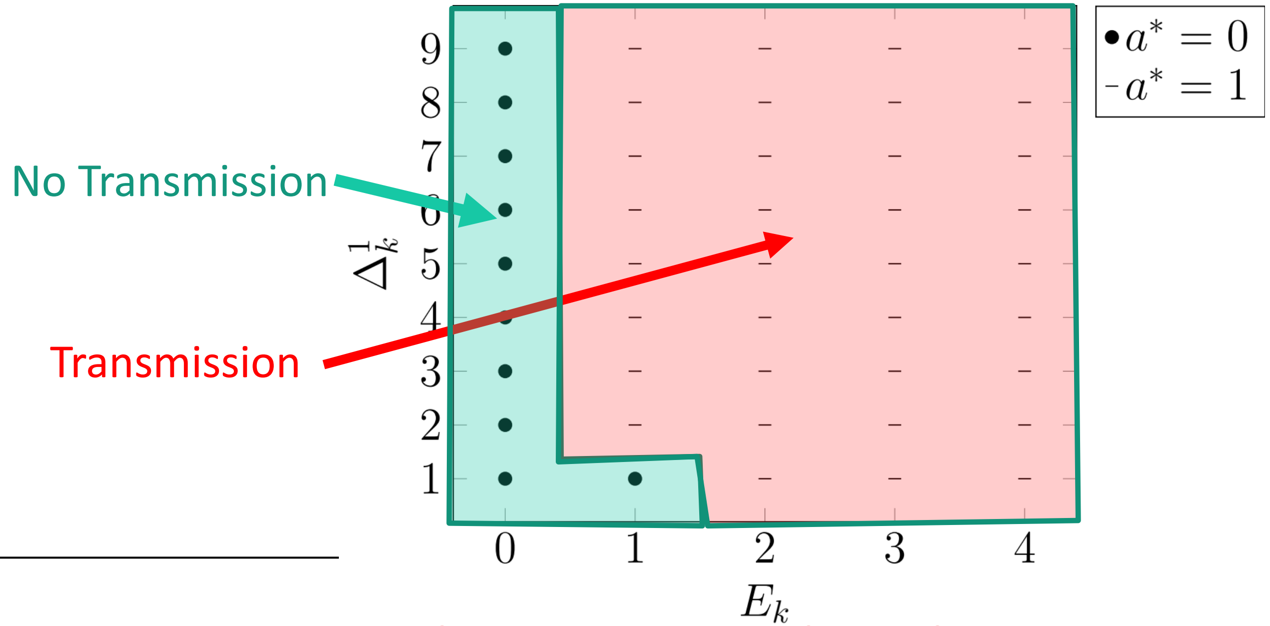
Optimal actions in the alarm state

Optimal Policy - High Probability EH ($P_e = 0.8$) - $P_s = 0.8$

- Scenario
  - the process spends most of its time in normal mode with relative short periods of alarm states.
    - State transition matrix $P_z = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$
- *Energy saving is less important when EH occurs with high probability*
  - Threshold structure: Transmissions occur when Δ_k^0 and Δ_k^1 is larger than a threshold value given E_k.



Optimal actions in the normal state

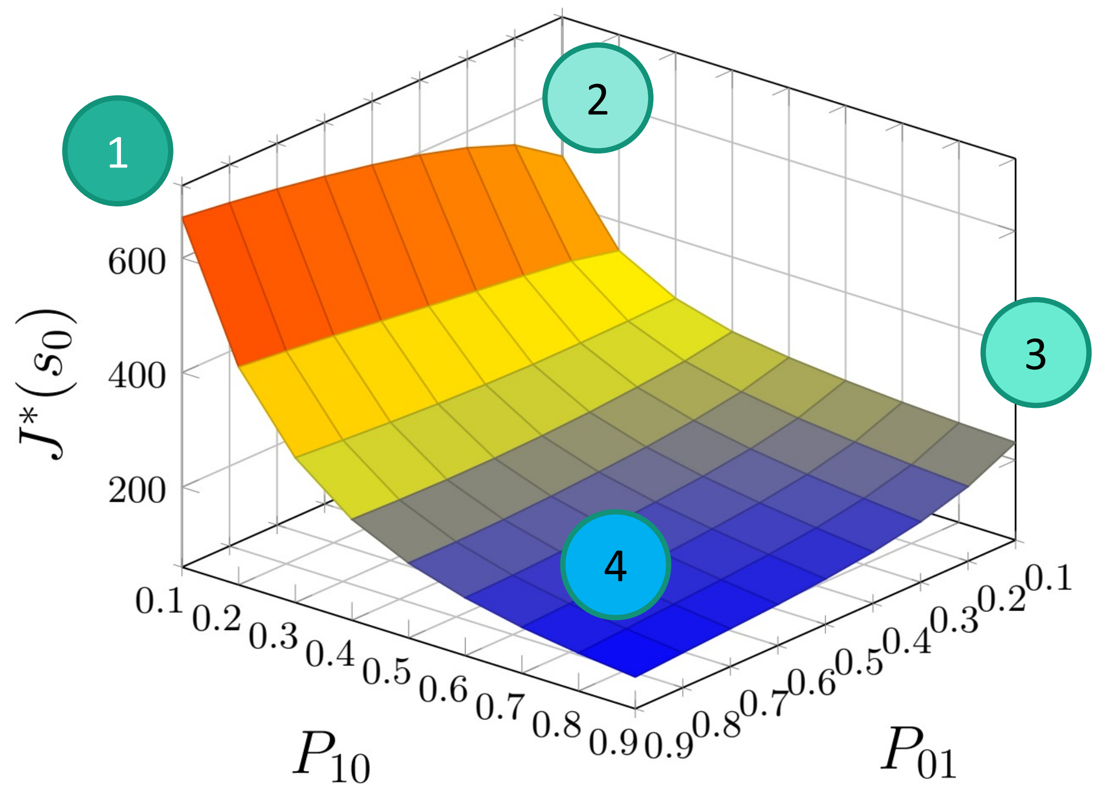


Optimal actions in the alarm state

Cost-to-go for different state transition probabilities, $P_e = 0.8, P_s = 0.8$

- Cost-to-go starting from a state s_0 where the stochastic process is in normal mode, the destination is aware of it, the energy buffer is empty and $\Delta_k^0 = 1, \Delta_k^1 = 0$.
- Various combinations of P_{10}, P_{01}
  - $P_z = \begin{bmatrix} 1 - P_{01} & P_{01} \\ P_{10} & 1 - P_{10} \end{bmatrix}$

1. Large sojourn time in the alarm state.
   - Large probability to enter the alarm state and small probability to leave out of it.
2. Large sojourn times in both states
   - Small probability to leave a state once in it.
3. Large sojourn time in the normal state.
   - AoI may increase up to large values.
4. The stochastic process oscillates between normal and alarm state with small sojourn times in each state.
   - AoI and costs remain small due to the short sojourn times.



Discussion

- *Some other metrics appeared after that work*
  - Age of Incorrect Information,
  - Pull based AoI.
- *Later we will discuss the case of real-time tracking a source with the purpose of remote actuation in real-time.*

- A. Maatouk, S. Kriouile, M. Assaad and A. Ephremides, "The Age of Incorrect Information: A New Performance Metric for Status Updates", IEEE/ACM Trans. on Networking 2020.
- J. Holm, A. E. Kalør, F. Chiariotti, B. Soret, S. Jensen, T. Pedersen, and P. Popovski, "Freshness on demand: Optimizing Age of Information for the query process", IEEE ICC 2021.
- F. Li, Y. Sang, Z. Liu, B. Li, H. Wu, and B. Ji, "Waiting but not aging: Optimizing information freshness under the pull model", IEEE/ACM Trans. on Networking 2021.
- X. Zheng, S. Zhou, and Z. Niu, "Urgency of Information for context aware timely status updates in remote control systems", IEEE Trans. on Wir. Comm. 2020.

AoI and VoI in Control

- AoI considers only the timeliness!
- It has been shown that AoI alone does not capture the requirements of networked control loops.
- *Introduction of non-linear AoI facilitated the adoption in networked-control systems (NCS).*
- *VoI can reduce the estimation error in an NCS setup!*
- Very active research area that started recently.

O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, "Age-of-Information vs. Value-of-Information Scheduling for Cellular Networked Control Systems", 10th ACM/IEEE ICCPS 2019.

Towards a complete characterization of the AoI distribution

- Stochastic hybrid systems are utilized to analyze AoI moments and the moment generating function of AoI in networks
  - R. D. Yates, "The Age of Information in Networks: Moments, Distributions, and Sampling," IEEE Trans. Info. Theory 2020.

- A general formula of the stationary distribution of AoI is obtained and applied to a wide class of continuous-time single server queues with different disciplines
  - Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "A general formula for the stationary distribution of the age of information and its application to single-server queues," IEEE Trans. Info. Theory 2019.

- The distribution of AoI for the GI/GI/1/1 and GI/GI/1/2* systems, under non-preemptive scheduling
  - J. P. Champati, H. Al-Zubaidy, and J. Gross, "On the distribution of aoi for the GI/GI/1/1 and GI/GI/1/2* systems: Exact expressions and bounds," IEEE INFOCOM 2019.

Towards a complete characterization of the AoI distribution

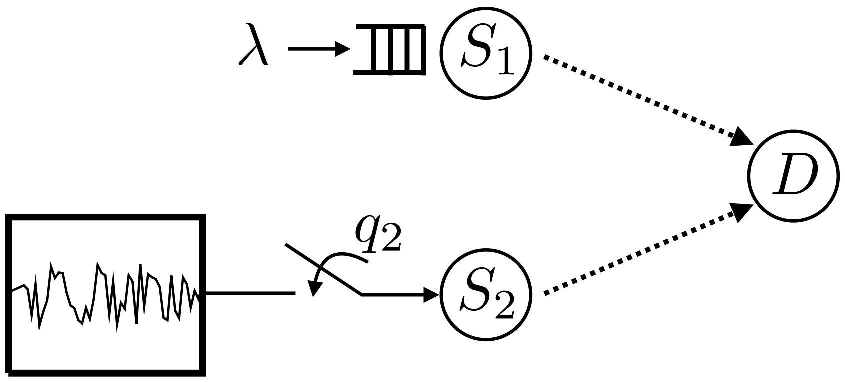
- The AoI distribution in bufferless systems
  - G. Kesidis, T. Konstantopoulos, M.A. Zazanis, "The distribution of age-of-information performance measures for message processing systems", Queueing Systems 2020.

- Complete characterization of the AoI stationary distribution in a discrete time queueing system for: FCFS, preemptive LCFS, a bufferless system with packet dropping.
- *A methodology for analyzing general non-linear age functions, using representations of functions as power series*.
  - A. Kosta, N. Pappas, A. Ephremides, V. Angelakis, "The Age of Information in a Discrete Time Queue: Stationary Distribution and Non-linear Age Mean Analysis", IEEE JSAC SI on AoI, 2021. (shorter version in IEEE ICC 2020).

Interplay between AoI and other metrics

AoI and Delay Violation Probability Interplay in the Two-user MAC

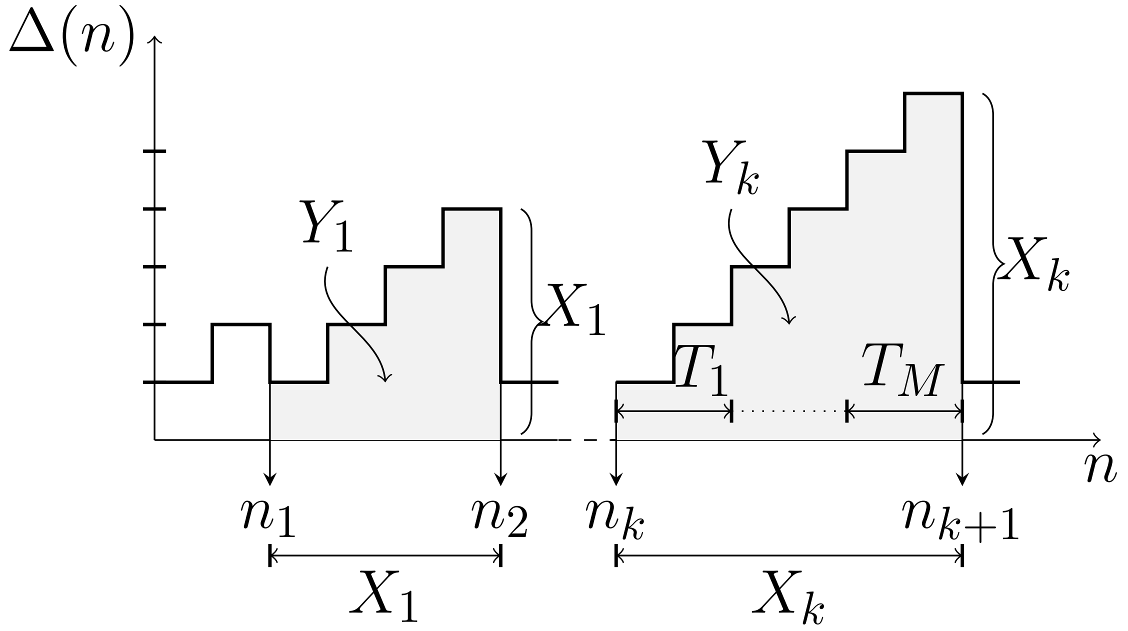
- Two sources sending packets to a common destination.
- Source S_1 has external traffic with stringent delay requirements.
- Source S_2 monitors a sensor and samples a status update on each slot w.p. q_2. (Departure from the classical model of external updates that was common in the early studies of AoI).
  - Then, transmits the update to the destination through a channel with success probability p_2.
  - If the transmission of a status update fails, then it is dropped.
- Time is slotted.
- Instantaneous and error-free ACK/NACK.



N. Pappas, M. Kountouris, "Delay Violation Probability and Age-of-information Interplay in the Two-user Multiple Access Channel", *IEEE SPAWC* 2019.

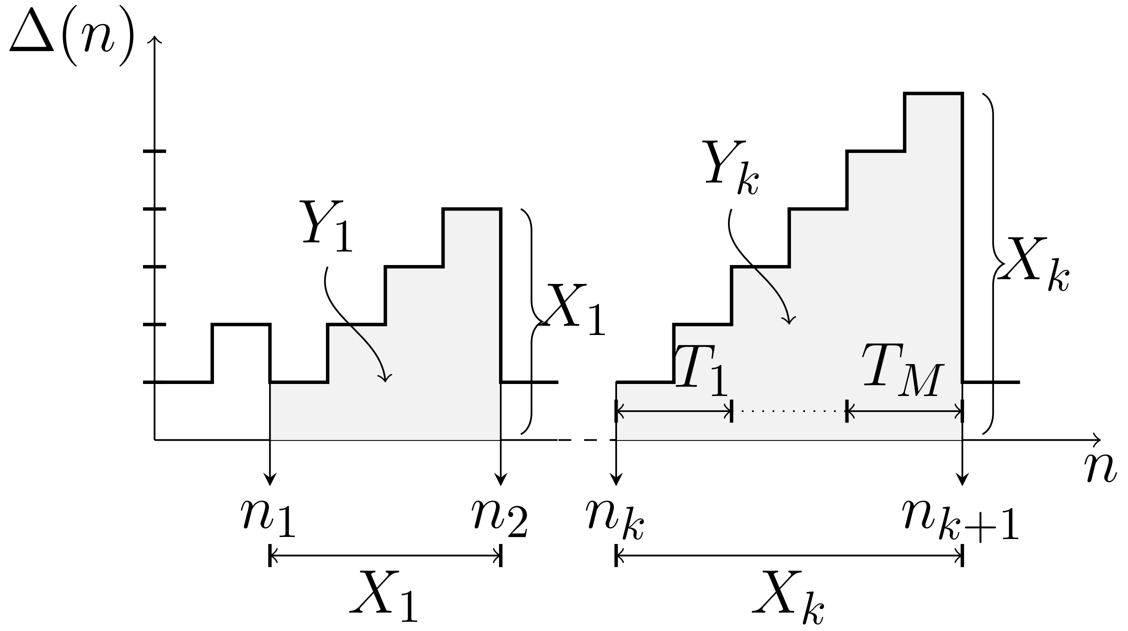
Average AoI

T_i: time between two consecutive attempted transmissions
X_k: elapsed time at the destination between successful reception of k-th and the $(k+1)$-th status updates
M: number of attempted transmissions between two successfully received status updates at D



Average AoI

T_i: time between two consecutive attempted transmissions
X_k: elapsed time at the destination between successful reception of k-th and the $(k+1)$-th status updates
M: number of attempted transmissions between two successfully received status updates at D

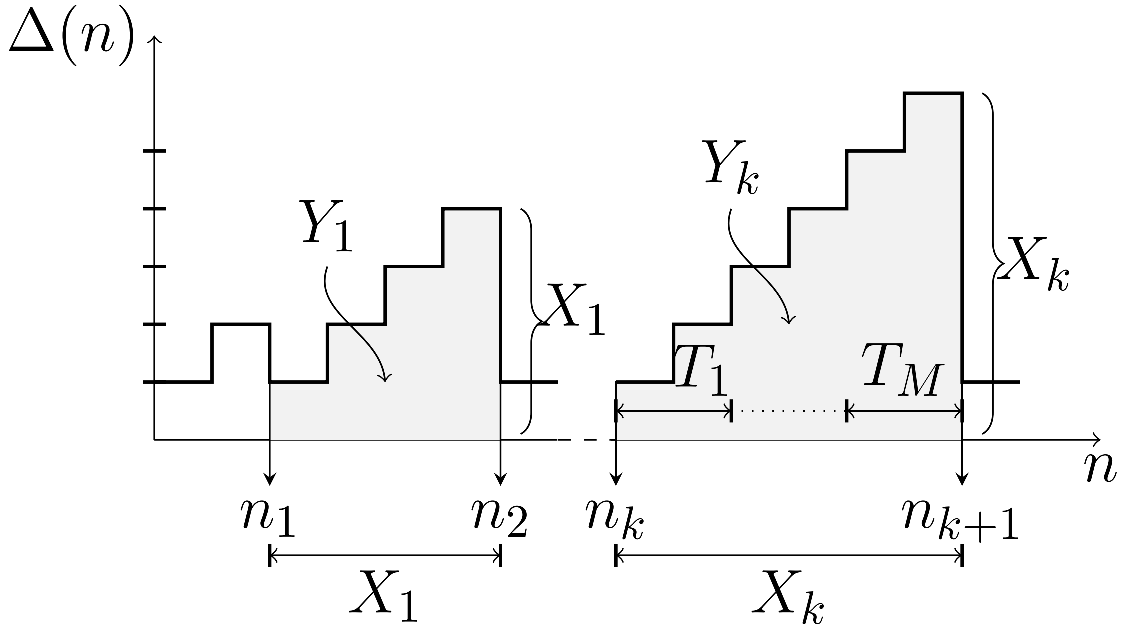


$$X_k = \sum_{i=1}^{M} T_i \qquad \Delta_N = \frac{1}{N}\sum_{n=1}^{N}\Delta(n) = \frac{1}{N}\sum_{k=1}^{K} Y_k = \frac{K}{N}\frac{1}{K}\sum_{k=1}^{K} Y_k$$

$$\Delta = \lim_{N\to\infty}\Delta_N = \frac{\mathbb{E}[Y]}{\mathbb{E}[X]} \qquad Y_k = \sum_{m=1}^{X_k} m = \frac{X_k(X_k+1)}{2}$$

Average AoI

T_i: time between two consecutive attempted transmissions
X_k: elapsed time at the destination between successful reception of k-th and the $(k+1)$-th status updates
M: number of attempted transmissions between two successfully received status updates at D



$$X_k = \sum_{i=1}^{M} T_i$$

$$\Delta_N = \frac{1}{N}\sum_{n=1}^{N}\Delta(n) = \frac{1}{N}\sum_{k=1}^{K} Y_k = \frac{K}{N}\frac{1}{K}\sum_{k=1}^{K} Y_k$$

$$\Delta = \lim_{N\to\infty}\Delta_N = \frac{\mathbb{E}[Y]}{\mathbb{E}[X]} \qquad Y_k = \sum_{m=1}^{X_k} m = \frac{X_k(X_k+1)}{2}$$

$$\Delta_N = \frac{K}{N}\frac{1}{K}\sum_{k=1}^{K} Y_k = \frac{\mathbb{E}\left[\frac{X_k^2}{2}+\frac{X_k}{2}\right]}{\mathbb{E}[X]} = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]} + \frac{1}{2}$$

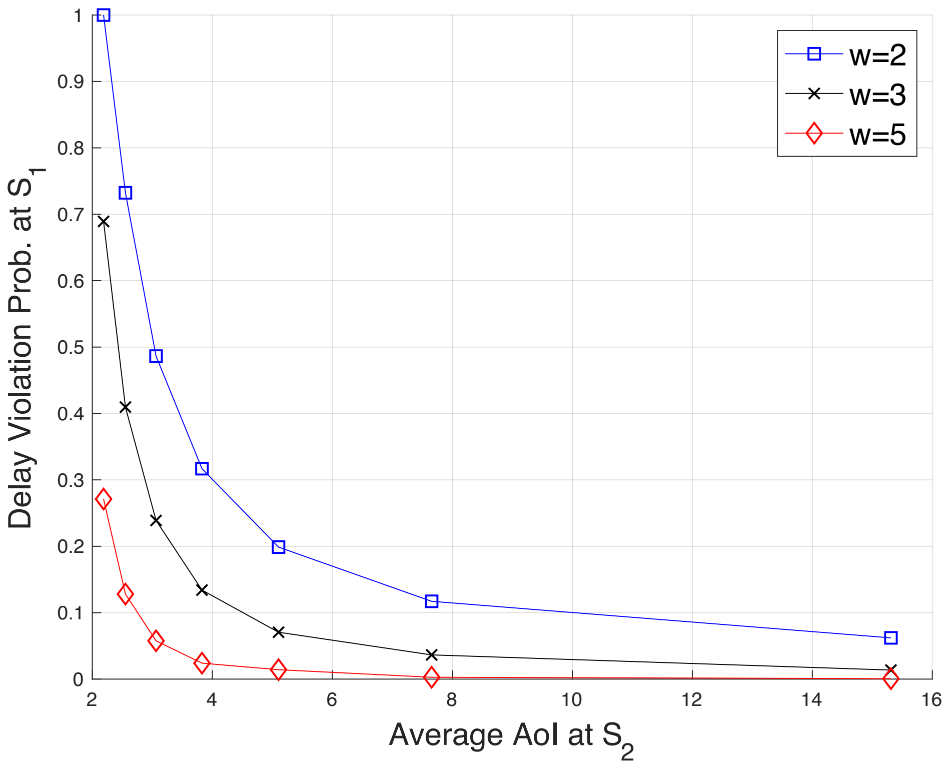
$$\mathbb{E}[X] = \sum_{M=1}^{\infty} M\mathbb{E}[T](1-p_2)^{M-1}p_2 = \frac{\mathbb{E}[T]}{p_2}$$

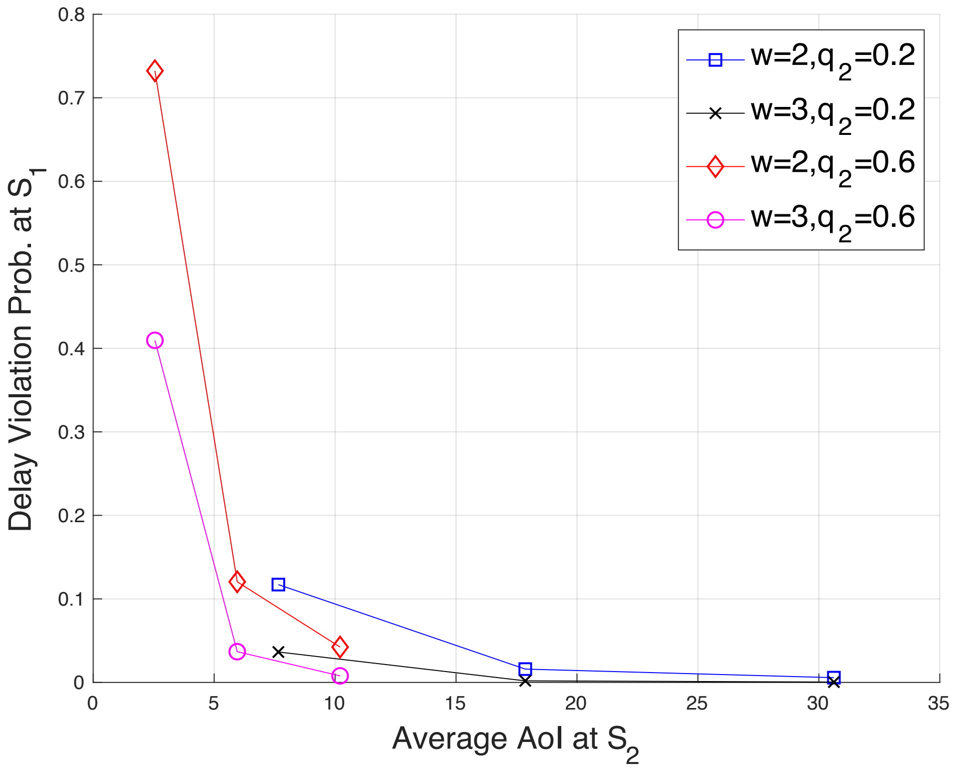
$$\mathbb{E}[X^2] = \sum_{M=1}^{\infty}\mathbb{E}[X^2|M](1-p_2)^{M-1}p_2$$

$$\overset{p_2>0}{=} \frac{\mathbb{E}[T^2]}{p_2} + \frac{2(1-p_2)\mathbb{E}[T]^2}{p_2^2}$$

$$\Delta = \frac{\mathbb{E}[T^2]}{2\mathbb{E}[T]} + \frac{\mathbb{E}[T](1-p_2)}{p_2} + \frac{1}{2}$$

$$= \frac{1}{q_2 p_2}$$



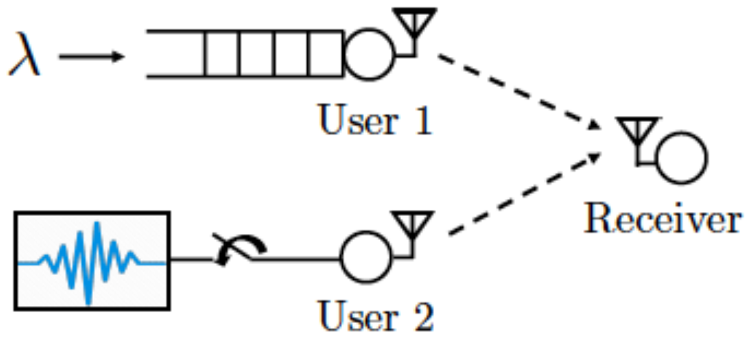


- As w increases, the delay violation probability decreases since S_1 becomes more delay tolerant.
- Increasing the transmit power of S_1 results in significant decrease of the delay violation probability and an increase of AoI due to larger interference.

Both delay violation probability and AoI can be kept low even for stringent delay constraints if the sampling rate is properly adapted.

AoI and Packet Drop Rate Interplay

- The first user has deadline-constrained traffic and access the channel with probability q_1 when there is a packet in its queue
- User 2 (AoI-oriented) accesses the channel only if samples an update with a probability q_2
- If the transmission of a status update by user 2 fails, then is dropped (avoid transmitting outdated information)
- *For the AoI-oriented user, we provide the distribution of the AoI, the average AoI, and the probability the AoI to be larger than a value for each time slot.*



E. Fountoulakis, T. Charalambous, N. Nomikos, A. Ephremides, N. Pappas, "Information Freshness and Packet Drop Rate Interplay in a Two-User Multi-Access Channel", IEEE ITW, Apr. 2021.

AoI and Packet Drop Rate Interplay

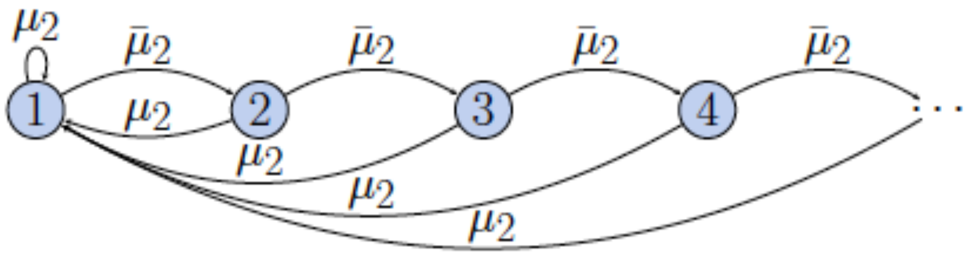
- We model the evolution of AoI as a Discrete Time Markov Chain
- The probability that AoI has value i is given by

$$\pi_i^A = (1 - \mu_2)^{(i-1)}\mu_2, \ \forall i$$

$$\mu_2 = q_2(1 - q_1 P\{Q > 0\})P_{2/2} + q_2 q_1(\Pr\{Q > 0\}P_{2/2,1})$$

- The average AoI is $\bar{A} = \frac{1}{\mu_2},$
- We can also obtain the *AoI violation probability* as

$$P\{A > x\} = (1 - \mu_2)^x,$$



E. Fountoulakis, T. Charalambous, N. Nomikos, A. Ephremides, N. Pappas, "Information Freshness and Packet Drop Rate Interplay in a Two-User Multi-Access Channel", IEEE ITW, Apr. 2021.

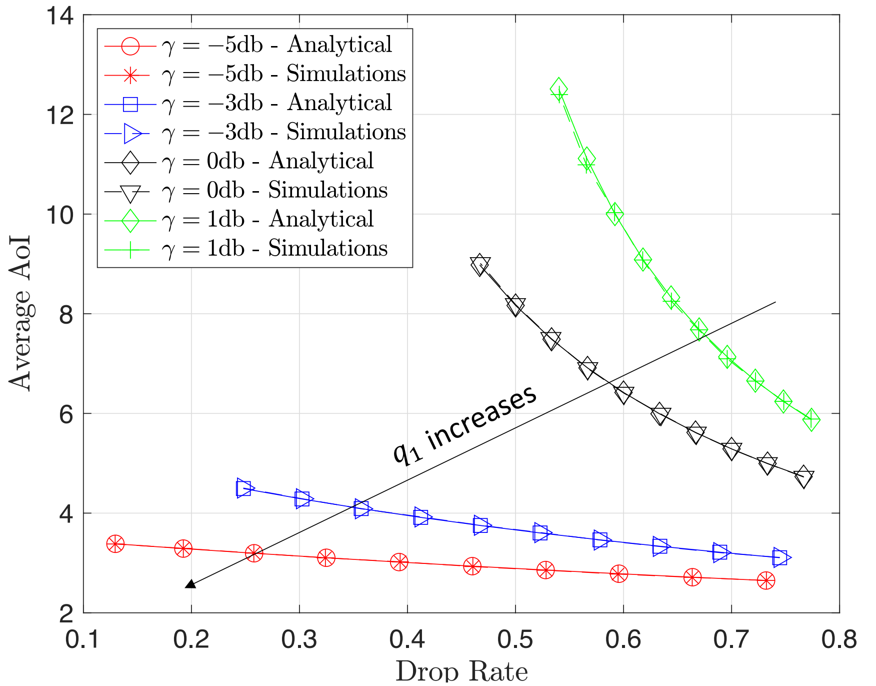
AoI and Packet Drop Rate Interplay

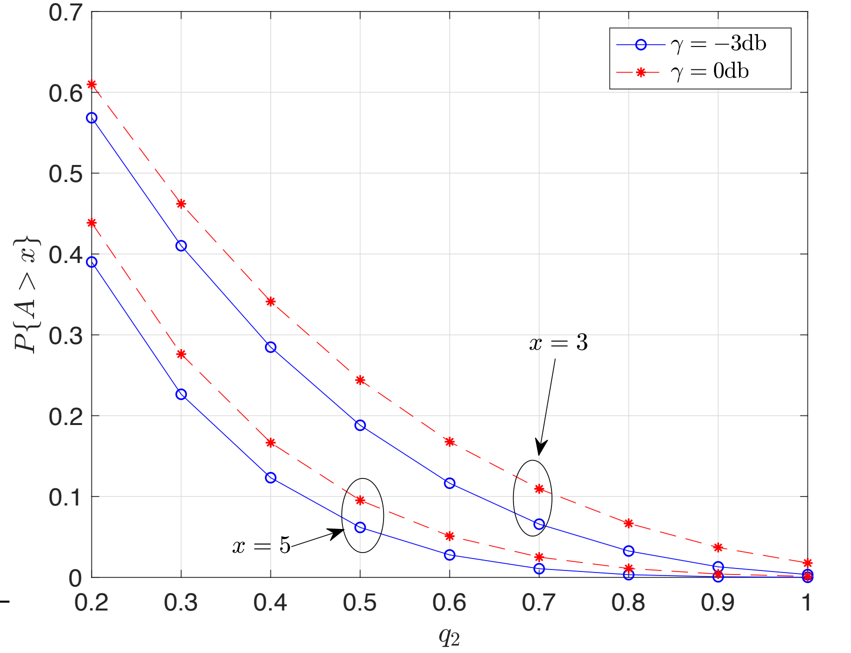
- We model the evolution of AoI as a Discrete Time Markov Chain
- The probability that AoI has value i is given by

$$\pi_i^A = (1-\mu_2)^{(i-1)}\mu_2, \ \forall i$$

- The average AoI is $\bar{A} = \frac{1}{\mu_2}$,
- We can also obtain the AoI violation probability as

$$P\{A > x\} = (1-\mu_2)^x,$$





E. Fountoulakis, T. Charalambous, N. Nomikos, A. Ephremides, N. Pappas, "[Information Freshness and Packet Drop Rate Interplay in a Two-User Multi-Access Channel](#)", IEEE ITW, Apr. 2021.

Remarks and future directions

- AoI has emerged as an end-to-end performance metric for systems that employ status updates.

- Introduction of information freshness requirements *will create systems that work smarter than harder*, so they will be more *effective*.
    - The updating process should not underload nor overload the system.
    - The system should process new updates rather than old.
    - The system should avoid processing updates without sufficient novelty.

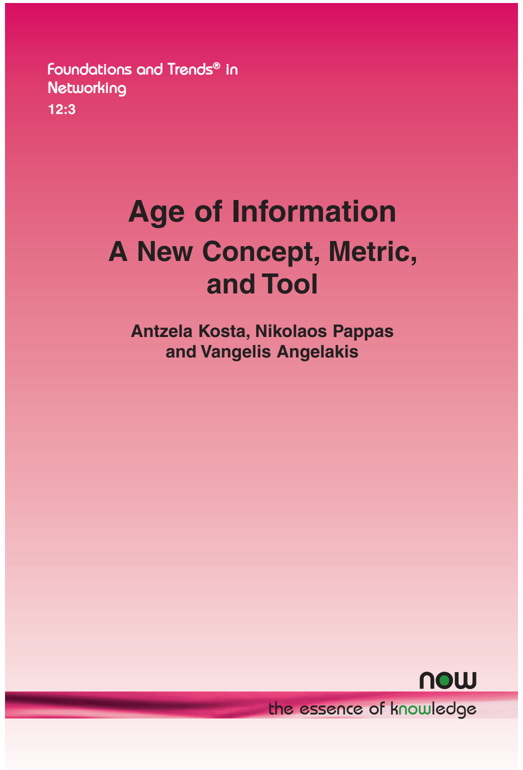
Remarks and future directions

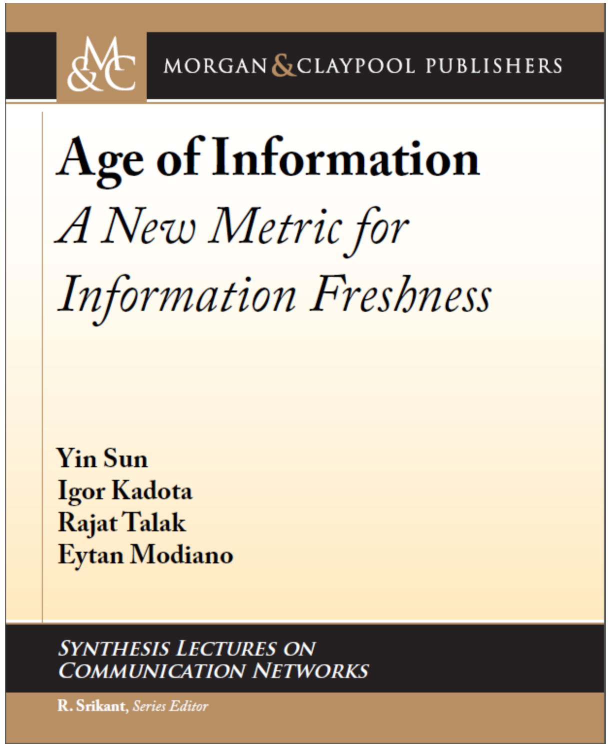
- There are still many interesting research directions
  - Definition of effective age (term coined by Prof. Ephremides in ITA 2015)
  - Sampling and remote reconstruction
  - Deploying of AoI in machine learning

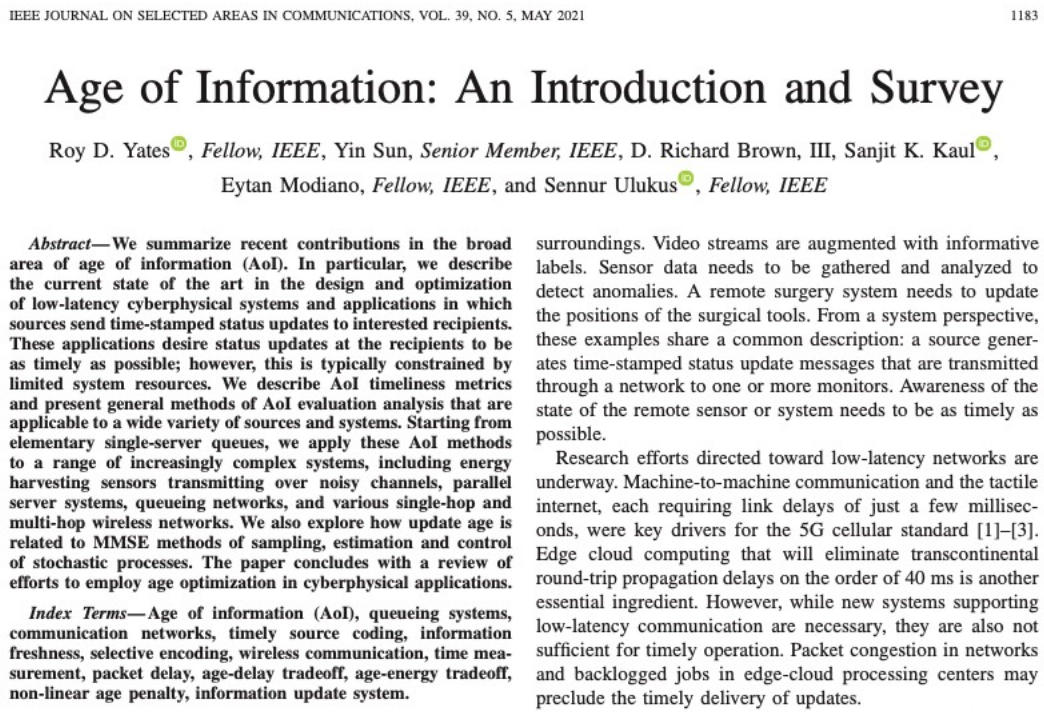
- It provides stronger connections with areas such as Signal Processing
- Metrics that can capture the requirements of Wireless Networked Control Systems

- AoI is one of the dimensions of *semantics-empowered communications*!

M. Kountouris, N. Pappas, "Semantics-Empowered Communication for Networked Intelligent Systems", *IEEE Communications Magazine, June 2021.*







IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 39, NO. 5, MAY 2021 1183

Age of Information: An Introduction and Survey

Roy D. Yates, *Fellow, IEEE*, Yin Sun, *Senior Member, IEEE*, D. Richard Brown, III, Sanjit K. Kaul, Eytan Modiano, *Fellow, IEEE*, and Sennur Ulukus, *Fellow, IEEE*

***Abstract*—We summarize recent contributions in the broad area of age of information (AoI). In particular, we describe the current state of the art in the design and optimization of low-latency cyberphysical systems and applications in which sources send time-stamped status updates to interested recipients. These applications desire status updates at the recipients to be as timely as possible; however, this is typically constrained by limited system resources. We describe AoI timeliness metrics and present general methods of AoI evaluation analysis that are applicable to a wide variety of sources and systems. Starting from elementary single-server queues, we apply these AoI methods to a range of increasingly complex systems, including energy harvesting sensors transmitting over noisy channels, parallel server systems, queueing networks, and various single-hop and multi-hop wireless networks. We also explore how update age is related to MMSE methods of sampling, estimation and control of stochastic processes. The paper concludes with a review of efforts to employ age optimization in cyberphysical applications.**

***Index Terms*—Age of information (AoI), queueing systems, communication networks, timely source coding, information freshness, selective encoding, wireless communication, time measurement, packet delay, age-delay tradeoff, age-energy tradeoff, non-linear age penalty, information update system.**

surroundings. Video streams are augmented with informative labels. Sensor data needs to be gathered and analyzed to detect anomalies. A remote surgery system needs to update the positions of the surgical tools. From a system perspective, these examples share a common description: a source generates time-stamped status update messages that are transmitted through a network to one or more monitors. Awareness of the state of the remote sensor or system needs to be as timely as possible.

Research efforts directed toward low-latency networks are underway. Machine-to-machine communication and the tactile internet, each requiring link delays of just a few milliseconds, were key drivers for the 5G cellular standard [1]–[3]. Edge cloud computing that will eliminate transcontinental round-trip propagation delays on the order of 40 ms is another essential ingredient. However, while new systems supporting low-latency communication are necessary, they are also not sufficient for timely operation. Packet congestion in networks and backlogged jobs in edge-cloud processing centers may preclude the timely delivery of updates.

Journals & Magazines > IEEE Internet of Things Journal > Volume: 8 Issue: 19

Guest Editorial Special Issue on Age of Information and Data Semantics for Sensing, Communication, and Control Co-Design in IoT

Publisher: IEEE Cite This PDF

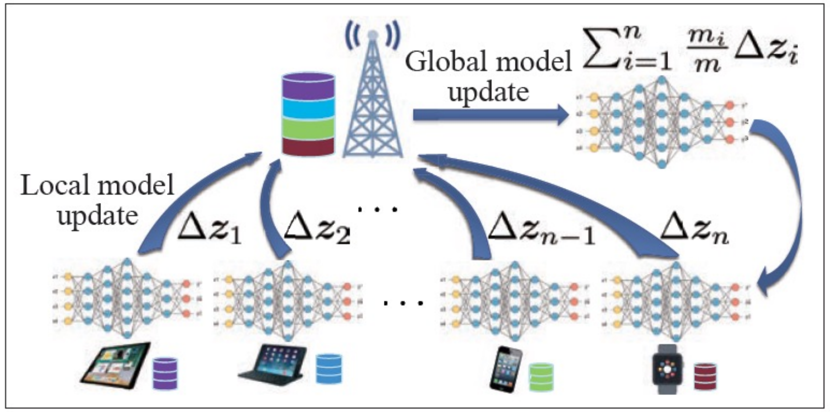
Sheng Zhou ; Zhiyuan Jiang ; Nikolaos Pappas ; Anthony Ephremides ; Luiz A. DaSilva All Authors

N. Pappas, M. A. Abd-Elmagid, B. Zhou, W. Saad, H. S. Dhillon, Age of Information: Foundations and Applications. Cambridge University Press, in press. Expected publication: Nov. 2022. (Edited)

Wireless ecosystem in the near future

- Networked Intelligent Systems:
  - real time autonomous systems
- Sensor fusion, on time status updates, real time information reconstruction, network and device computation, traffic flows with synced requirements, human in the loop
- Distributed ML over wireless
  - Exchange of large datasets in a timely manner over wireless

- ***AoI is a proxy towards semantics communications***
- ***Value and importance of information, accuracy***



General description of Semantics of Information (SoI) metric

- A comprehensive system metric, Semantics of Information (SoI), which captures the significance and usefulness of information w.r.t the goal of data exchange and the application requirements.
- Information attributes, which can be decomposed into *innate* (objective) and *contextual* (subjective).
- **Innate** are the attributes inherent to information regardless of its use, such as AoI, precision, correctness.
- **Contextual** are attributes that depend on the particular context or application for which information is being used.
  - For example, timeliness - a function of AoI, accuracy (distortion), perception via divergence or distance function.

- M. Kountouris, N. Pappas, "Semantics-Empowered Communication for Networked Intelligent Systems", *IEEE Communications Magazine, June 2021.*
- N. Pappas, M. Kountouris, "Goal-Oriented Communication for Real-Time Tracking in Autonomous Systems", *IEEE International Conference on Autonomous Systems (ICAS), Aug. 2021.*

Goal-Oriented Communication for Real-Time Tracking in Autonomous Systems

N. Pappas, M. Kountouris, "Goal-Oriented Communication for Real-Time Tracking in Autonomous Systems", IEEE International Conference on Autonomous Systems (ICAS), Aug. 2021.

LINKÖPING UNIVERSITY

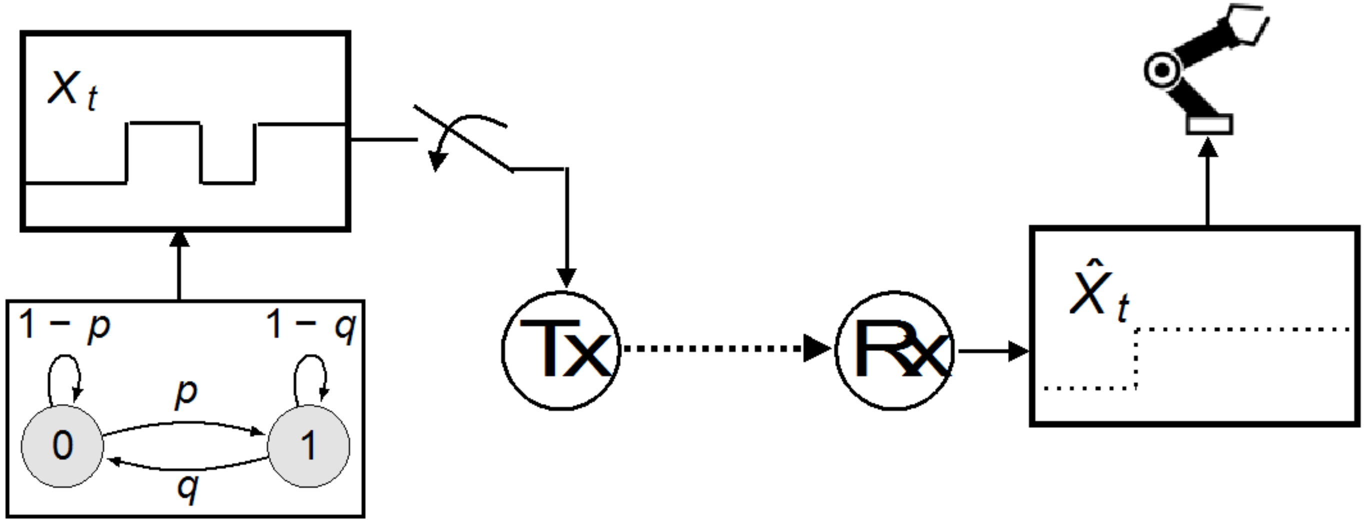
Introduction

- We consider real-time tracking and reconstruction of an information source.
- Real-time reconstruction is performed at the destination for remote actuation.
- This setting is relevant for real-time applications in autonomous networked systems.

- We introduce *new goal-oriented, semantics-empowered sampling and communication policies*, which account for the temporal evolution of the source/process and the semantic and application-dependent value of data being generated and transmitted.

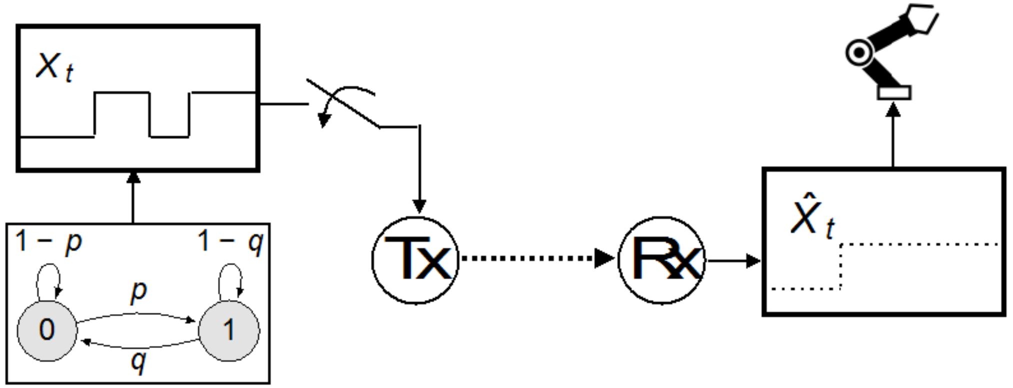
System Model

- A device monitors a two-state random process.
- The source initiates certain actions to the robotic object.
- The monitoring device samples and transmits status updates regarding the evolution of the source.
- *The application objective is to perform/maintain the actions of the original object in real-time.*



System Model

- Time is slotted.
- Wireless erasure communication channel with success probability:
  $p_s = \mathbb{P}\left(h_t = 1\right)$ where h_t is the channel realization.
- ACK/NACK instantaneous and error free.
- Information source, X_t, is modelled by a two state Markov Chain.
- X_t is reconstructed at the destination, $\hat{X}_t$, to perform actuation.
- The action of transmitting a sample is $\alpha_t^{\mathrm{tx}} = 1$, otherwise, the transmitter remains silent $\alpha_t^{\mathrm{tx}} = 0$.



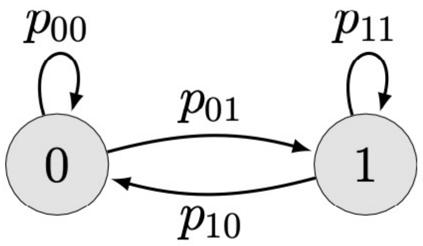
Key Performance Metrics

- **Real-time reconstruction error (innate)**: measures the discrepancy between the original and the reconstructed source in a timeslot

$$E_t = \mathbb{1}\left(X_t \neq \hat{X}_t\right) = \left|X_t - \hat{X}_t\right|$$

Time-averaged

$$\bar{E} = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} E_t}{T} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(X_t \neq \hat{X}_t\right)$$



$$p_{ji} = \mathbb{P}\left(E_{t+1} = j \,|\, E_t = i\right)$$

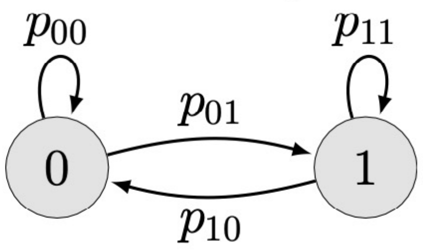
Key Performance Metrics

- **Real-time reconstruction error (innate)**: measures the discrepancy between the original and the reconstructed source in a timeslot

$$E_t = \mathbb{1}\left(X_t \neq \hat{X}_t\right) = \left|X_t - \hat{X}_t\right|$$

Time-averaged

$$\bar{E} = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} E_t}{T} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(X_t \neq \hat{X}_t\right)$$



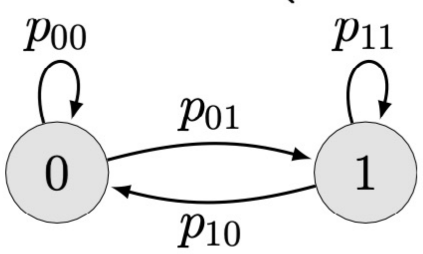
$$p_{ji} = \mathbb{P}\left(E_{t+1} = j | E_t = i\right)$$

$$p_{00} = \sum_{i=0}^{1} \mathbb{P}\left(E_{t+1} = 0 | E_t = 0, X_t = i\right) \mathbb{P}\left(X_t = i\right), \; \mathbb{P}\left(X_t = 0\right) = \frac{q}{p+q}, \; \mathbb{P}\left(X_t = 1\right) = \frac{p}{p+q}.$$

Key Performance Metrics

- **Real-time reconstruction error (innate)**: measures the discrepancy between the original and the reconstructed source in a timeslot

$$E_t = \mathbb{1}\left(X_t \neq \hat{X}_t\right) = \left|X_t - \hat{X}_t\right|$$



$$p_{ji} = \mathbb{P}\left(E_{t+1} = j | E_t = i\right)$$

Time-averaged

$$\bar{E} = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} E_t}{T} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(X_t \neq \hat{X}_t\right)$$

$$p_{00} = \sum_{i=0}^{1} \mathbb{P}\left(E_{t+1} = 0 | E_t = 0, X_t = i\right) \mathbb{P}\left(X_t = i\right), \ \mathbb{P}\left(X_t = 0\right) = \frac{q}{p+q}, \ \mathbb{P}\left(X_t = 1\right) = \frac{p}{p+q}.$$

$$\mathbb{P}\left(E_{t+1} = 0 | E_t = 0, X_t = 0\right) = 1 - p + p\mathbb{P}\left(\alpha_{t+1}^{\mathrm{s}} = 1, \alpha_{t+1}^{\mathrm{tx}} = 1, h_{t+1} = 1\right).$$

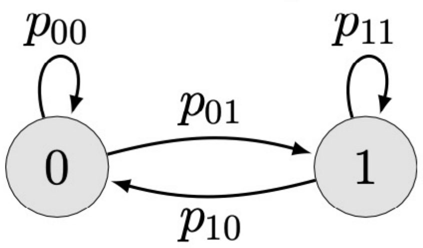
Key Performance Metrics

- **Real-time reconstruction error (innate)**: measures the discrepancy between the original and the reconstructed source in a timeslot

$$E_t = \mathbb{1}\left(X_t \neq \hat{X}_t\right) = \left|X_t - \hat{X}_t\right|$$

Time-averaged

$$\bar{E} = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} E_t}{T} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(X_t \neq \hat{X}_t\right)$$



$$p_{ji} = \mathbb{P}\left(E_{t+1} = j | E_t = i\right)$$

$$p_{00} = \sum_{i=0}^{1} \mathbb{P}\left(E_{t+1} = 0 | E_t = 0, X_t = i\right)\mathbb{P}\left(X_t = i\right), \ \mathbb{P}\left(X_t = 0\right) = \frac{q}{p+q}, \ \mathbb{P}\left(X_t = 1\right) = \frac{p}{p+q}.$$

$$\mathbb{P}\left(E_{t+1} = 0 | E_t = 0, X_t = 0\right) = 1 - p + p\mathbb{P}\left(\alpha^{\mathrm{s}}_{t+1} = 1, \alpha^{\mathrm{tx}}_{t+1} = 1, h_{t+1} = 1\right).$$

$$\mathbb{P}\left(E_{t+1} = 1 | E_t = 1, X_t = 0\right) =$$
$$(1-p)\left[\mathbb{P}\left(\alpha^{\mathrm{s}}_{t+1} = 1, \alpha^{\mathrm{tx}}_{t+1} = 1, h_{t+1} = 0\right) + \mathbb{P}\left(\alpha^{\mathrm{s}}_{t+1} = 0\right)\right].$$

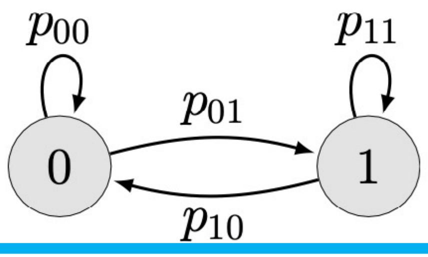
Key Performance Metrics

- **Real-time reconstruction error (innate)**: measures the discrepancy between the original and the reconstructed source in a timeslot

$$E_t = \mathbb{1}\left(X_t \neq \hat{X}_t\right) = \left|X_t - \hat{X}_t\right|$$

Time-averaged

$$\bar{E} = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} E_t}{T} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left(X_t \neq \hat{X}_t\right)$$



$$p_{ji} = \mathbb{P}\left(E_{t+1} = j \middle| E_t = i\right)$$

- **Cost of actuation error (contextual)**: captures the significance of the error at the point of actuation. *Some errors may have larger impact than others.*
  - $C_{i,j}$ denotes the cost of being in state i at the original source and in j at the reconstructed, when E_t=*1*. In general $C_{0,1} \neq C_{1,0}$.

Average cost of actuation $\bar{C}_A = \pi_{(0,1)} C_{0,1} + \pi_{(1,0)} C_{1,0}$

Sampling and communication policies

- Uniform: sampling is performed periodically, independently of the temporal evolution of the source.
  - *It is a process-agnostic policy that could result in missing several state transitions during the time interval between two collected samples.*
- Age-aware: the receiver triggers the acquisition and transmission of a new sample, once the AoI reaches a predefined threshold A_{th}.
  - *This policy is source-agnostic regarding the value of information but takes into account the timeliness.*

Sampling and communication policies

- Change-aware: sample generation is triggered at the transmitter whenever a change at the state of the source, with respect to the previous sample, is observed.
- Semantics-aware: extends the Change-aware into that the amount of change is not solely measured at the source but is also tracked by the difference in state *between receiver and transmitter*.

Sampling and communication policies

- <u>Semantics-aware</u>: extends the Change-aware into that the amount of change is not solely measured at the source but is also tracked by the difference in state *between receiver and transmitter*.

The transition matrix P_E for the DTMC that models the system state is given by

$$\mathrm{P}_E = \begin{bmatrix} 1 - \frac{2pq(1-p_s)}{p+q} & \frac{2pq(1-p_s)}{p+q} \\ p_s + \frac{2pq(1-p_s)}{p+q} & 1 - p_s - \frac{2pq(1-p_s)}{p+q} \end{bmatrix}.$$

The probability that the system is in an erroneous (not synced) state

$$\bar{E} = \frac{2pq(1-p_s)}{p_s(p+q) + 4pq(1-p_s)}.$$

Sampling and communication policies

- <u>Semantics-aware</u>: extends the Change-aware into that the amount of change is not solely measured at the source but is also tracked by the difference in state *between receiver and transmitter*.

The transition matrix P_E for the DTMC that models the system state is given by

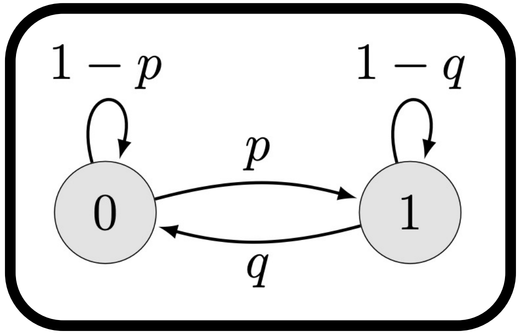
$$\mathrm{P}_E = \begin{bmatrix} 1 - \frac{2pq(1-p_s)}{p+q} & \frac{2pq(1-p_s)}{p+q} \\ p_s + \frac{2pq(1-p_s)}{p+q} & 1 - p_s - \frac{2pq(1-p_s)}{p+q} \end{bmatrix}.$$

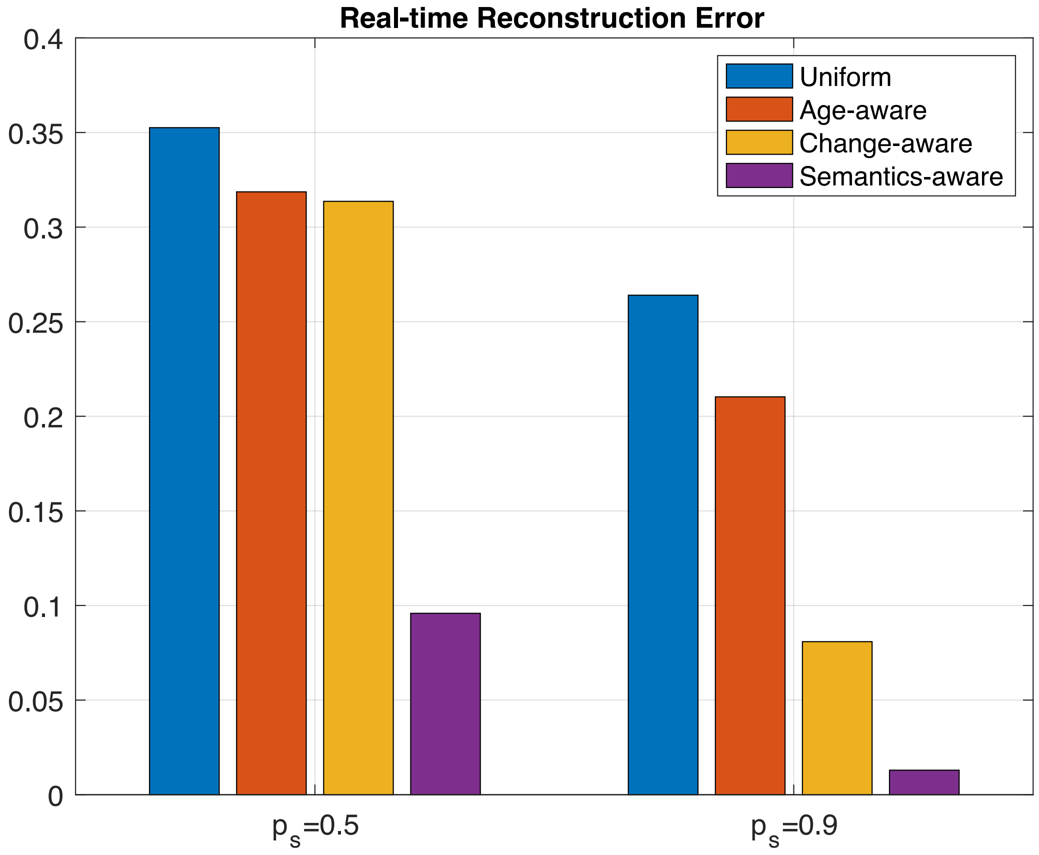
The probability that the system is in an erroneous (not synced) state

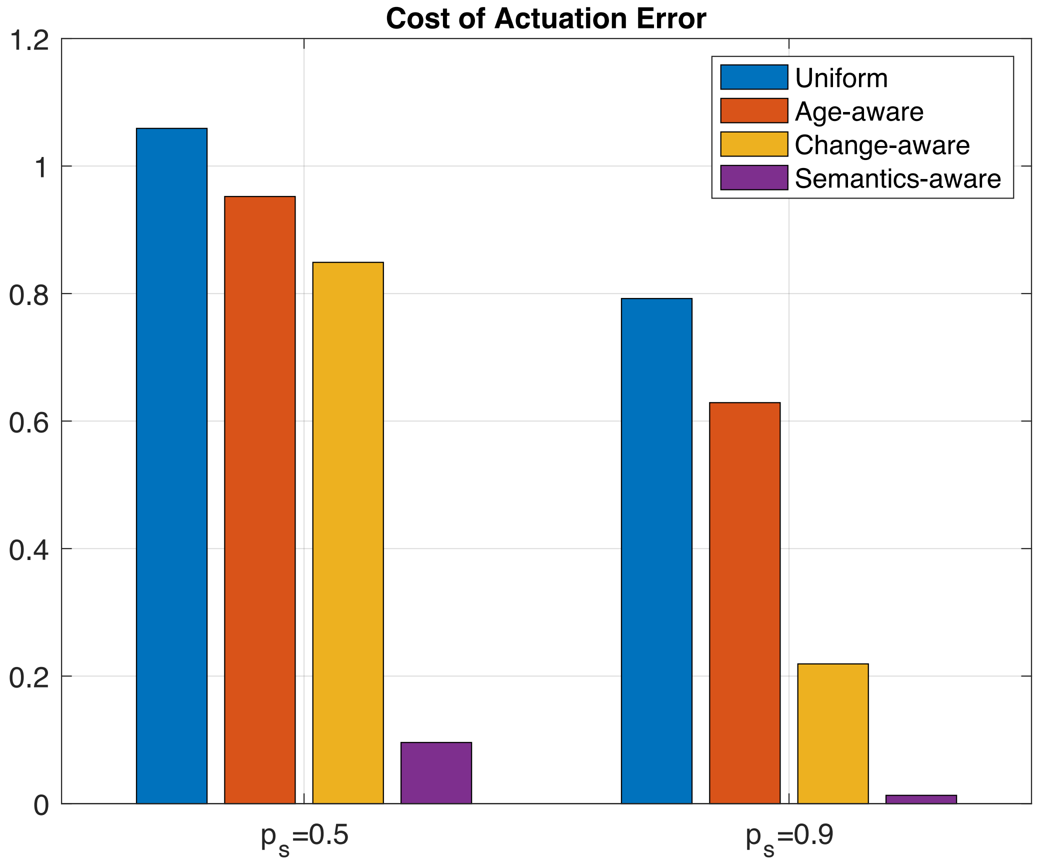
$$\bar{E} = \frac{2pq(1-p_s)}{p_s(p+q) + 4pq(1-p_s)}.$$

- Sampling and transmission at every timeslot could provide the best performance for real-time reconstruction. It requires a very large number of samples, which are not necessarily useful and require excessive resources.
- The semantics-empowered policies *reduce or even eliminate* the generation of uninformative sample updates, thus *improving network resource usage*.

Slowly-varying source – (p = 0.1, q = 0.15)

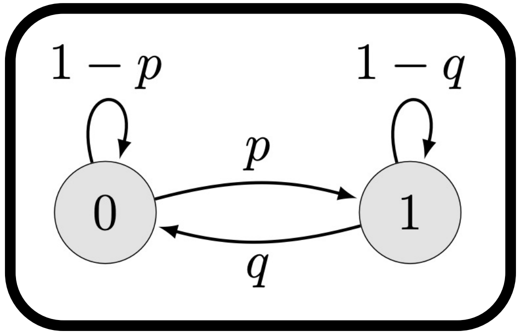


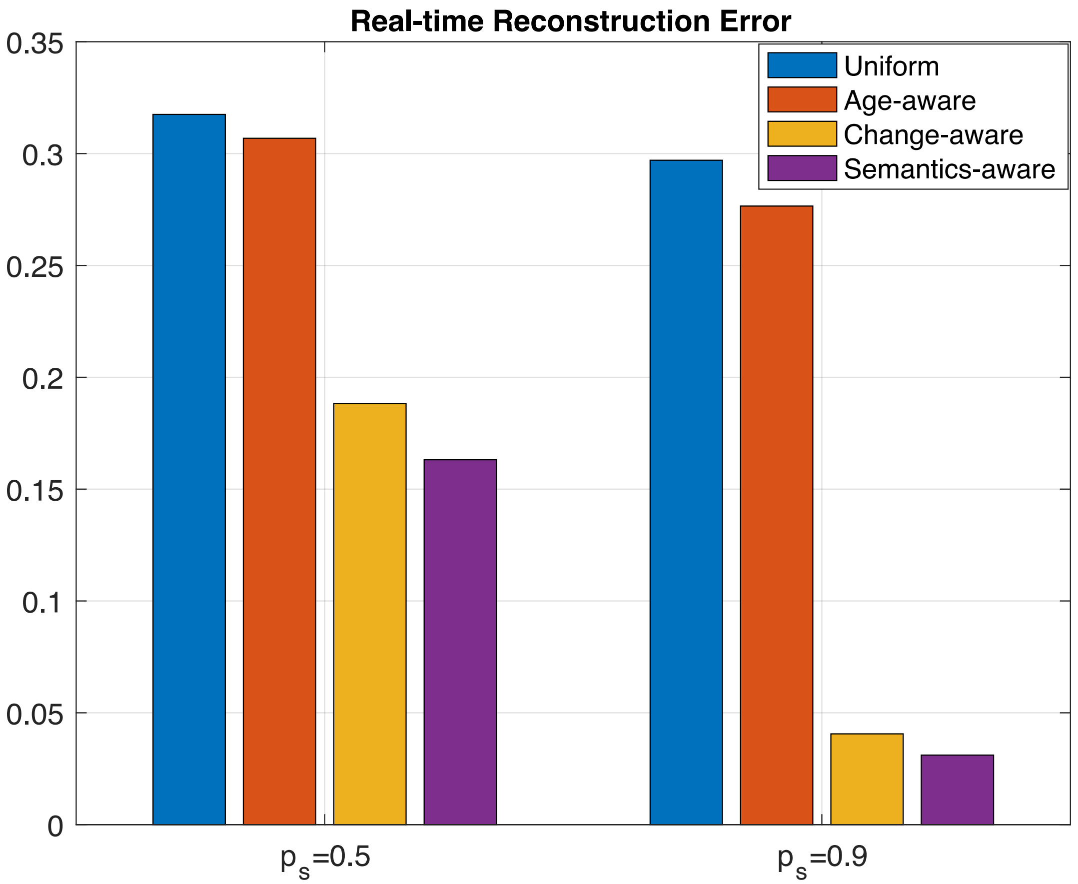


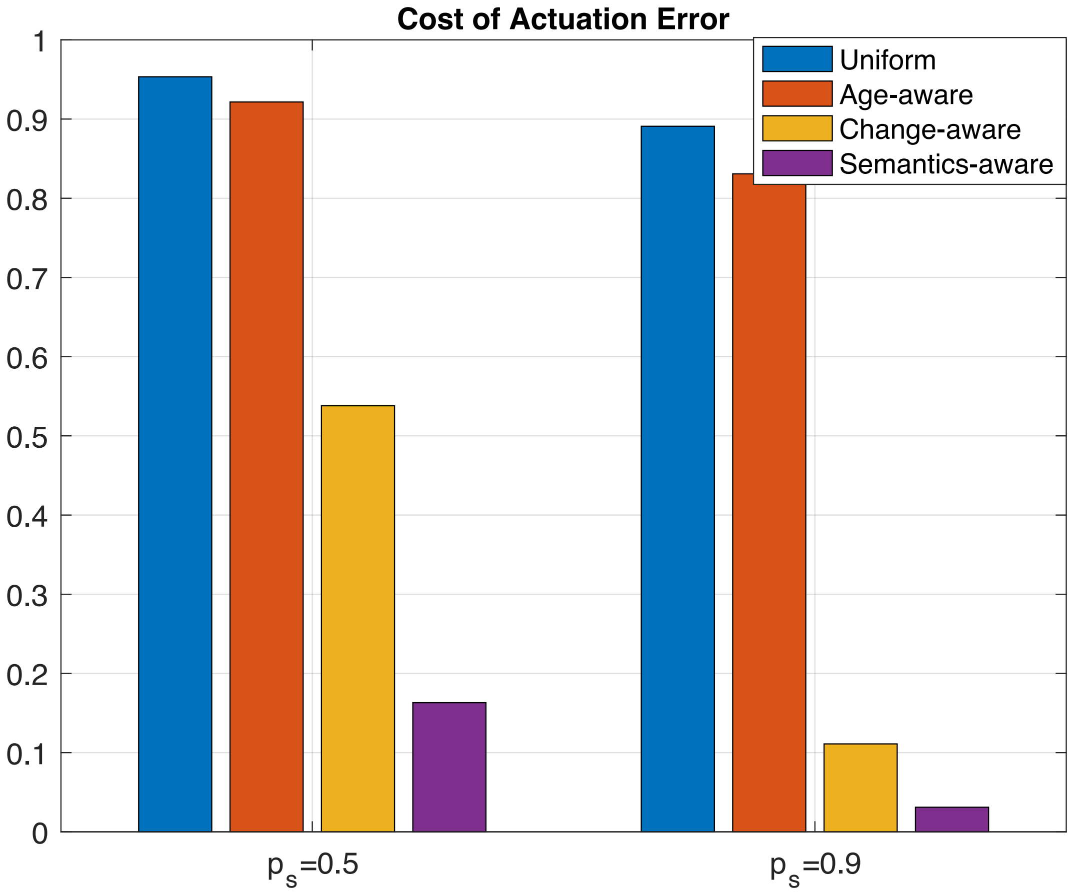


$C_{0,1}$=5, $C_{1,0}$=1

Rapidly-varying source – (p = 0.2, q = 0.7)







$C_{0,1}$=5, $C_{1,0}$=1

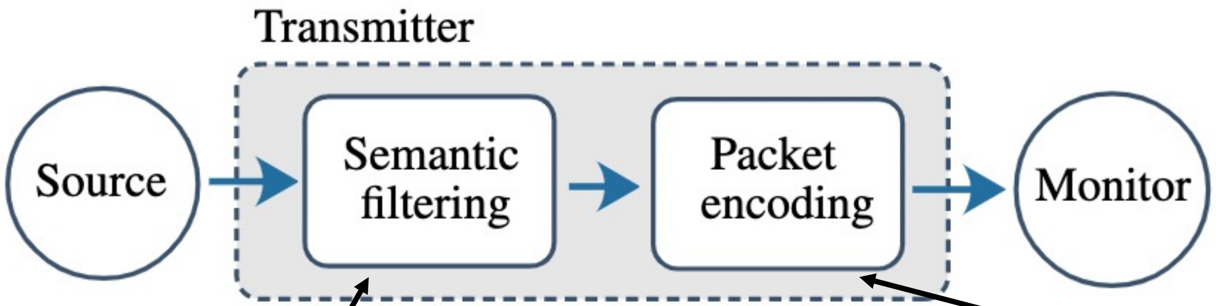
Semantics-aware Source Coding

P. Agheli, N. Pappas and M. Kountouris, “Semantics-Aware Source Coding in Status Update Systems”, arXiv, March 2022

Semantics-aware source coding

$\mathcal{X} = \{x_1, x_2, \cdots, x_n\}$
Discrete symbols

Probability of realization
$\widetilde{p_i} = P_X(x_i)$ – known PMF
wlog $\widetilde{p_i} \geq \widetilde{p_j}, \forall\, i \leq j$



Importance-based Selective Encoding

Importance/semantic value
only k *most/least* probable realizations selected
packets from remaining $n-k$ realizations discarded

- TX encodes an admitted packet from the i-th realization using a prefix-free code based on the truncated distribution with conditional probabilities $p_i = \frac{\widetilde{p_i}}{q_k}, \forall\, i \in \mathcal{I}_k \subset \mathcal{I}$, where $q_k = \sum_{i \in \mathcal{I}_k} \widetilde{p_i}$

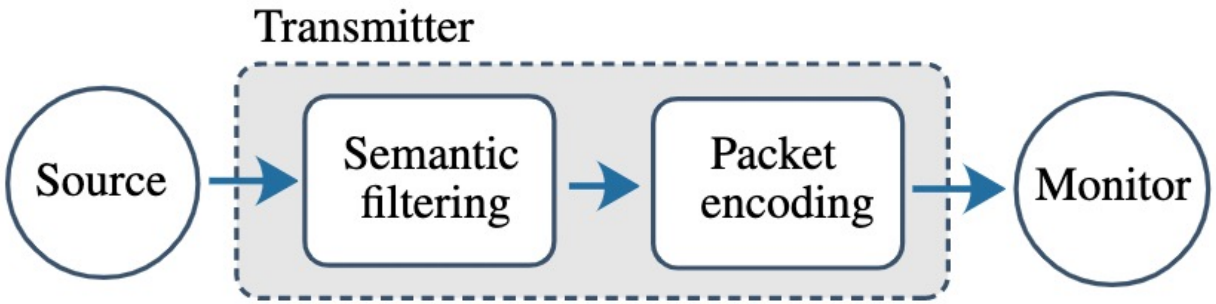
- Information source generates status updates (packets) and forwards them to a TX
- i.i.d. sequence of observations
- Packets generated according to Poisson(λ)
- TX encodes the packets and sends them to a remote monitor (RX)
- TX is bufferless (a newly-admitted packet is blocked when the channel is busy)
- Error-free channel

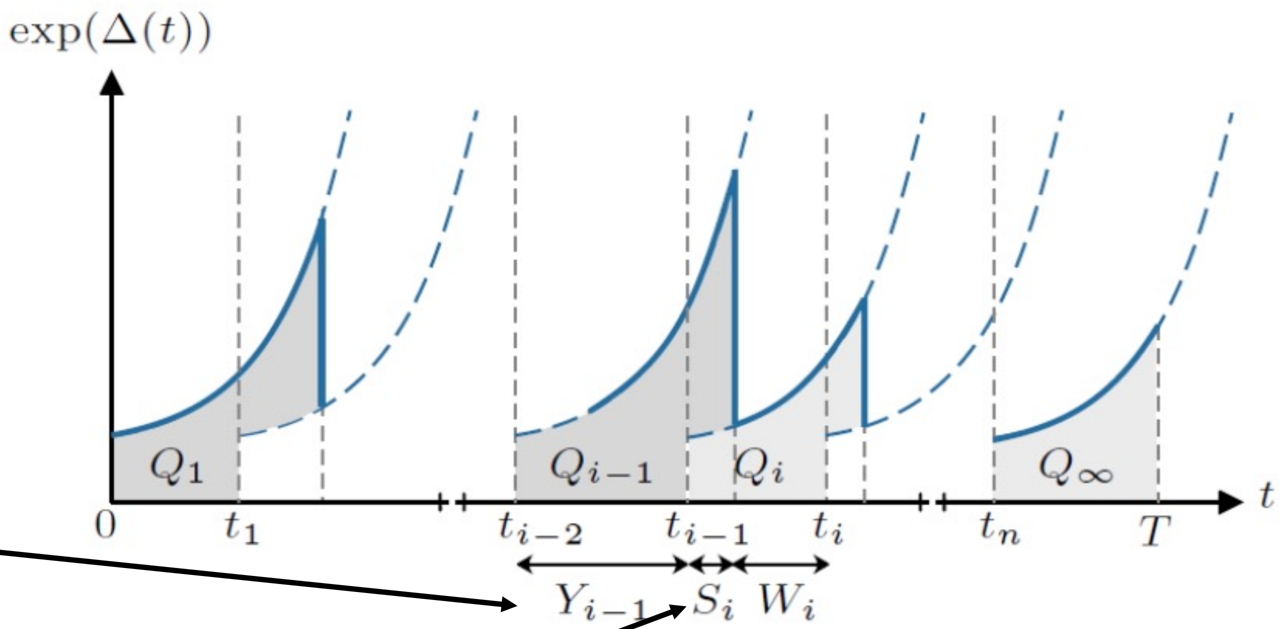
M.Bastopcu, B.Buyukates, S.Ulukus,"Optimal selective encoding for timely updates," CISS 2020.

Average SoI

- Timeliness (SoI): $\mathcal{S}(t) = g(\Delta(t))$ (time-varying stoch. process)
  - $g: \mathbb{R}_0^+ \to \mathbb{R}$ a non-increasing utility function of information freshness
  - AoI: $\Delta(t) = t - u(t)$

$$L(\Delta) = \lim_{T\to\infty} \frac{1}{T}\int_0^T g(\Delta(t))\mathrm{d}t = \lim_{T\to\infty} \frac{1}{T}\left\{ \sum_{i=1}^{\mathcal{N}(T)} Q_i + Q_\infty \right\} = \eta\mathbb{E}[Q].$$





$Y_i = t_i - t_{i-1}$
update interval between the i-th successive arrival and its next one

service time
$S_i = \ell_i$ time units
ℓ_i: length of the codeword assigned to x_i

waiting time $W_i = \sum_{k=1}^{N} Z_k$
N: # admitted arrivals generated before finding the channel idle
Z_k: time between two admitted arrivals, $\sim \exp(\gamma)$, $\gamma = 1/(\lambda q_k)$

- **Aim**: Find the codeword lengths ℓ_i that optimize a weighted sum of the average SoI and the average length for a cost function $\varphi(\ell_i)$, i.e., $\sum_i p_i \varphi(\ell_i)$.

$$\min_{\{\ell_i\}} \quad L(\Delta) + w \sum_{i \in \mathcal{I}_k} p_i \phi(\ell_i)$$
$$\text{s.t.} \quad \sum_{i \in \mathcal{I}_k} 2^{-\ell_i} \leq 1$$
$$\ell_i \in \mathbb{Z}^+$$

$$L(\Delta) = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(\Delta(t)) \mathrm{d}t$$

$f : \mathbb{R}_0^+ \to \mathbb{R}$ a non-decreasing function

$$\phi(x) = \alpha x + \beta x^2, \, \alpha, \beta \geq 0$$

- Quadratic cost function for the codeword length under binary alphabetic
- φ convex: longer (shorter) codewords are penalized more (less) than the linear case (e.g., Huffman coding)

Relaxation

$$\min_{\{\ell_i\}} \quad \underbrace{\mathbb{E}[Q] + w \sum_{i \in \mathcal{I}_k} p_i (\alpha \ell_i + \beta \ell_i^2)}_{\mathcal{J}_{\mathrm{SoI}}}$$
$$\text{s.t.} \quad \sum_{i \in \mathcal{I}_k} 2^{-\ell_i} \leq 1,$$
$$\ell_i \in \mathbb{R}^+.$$

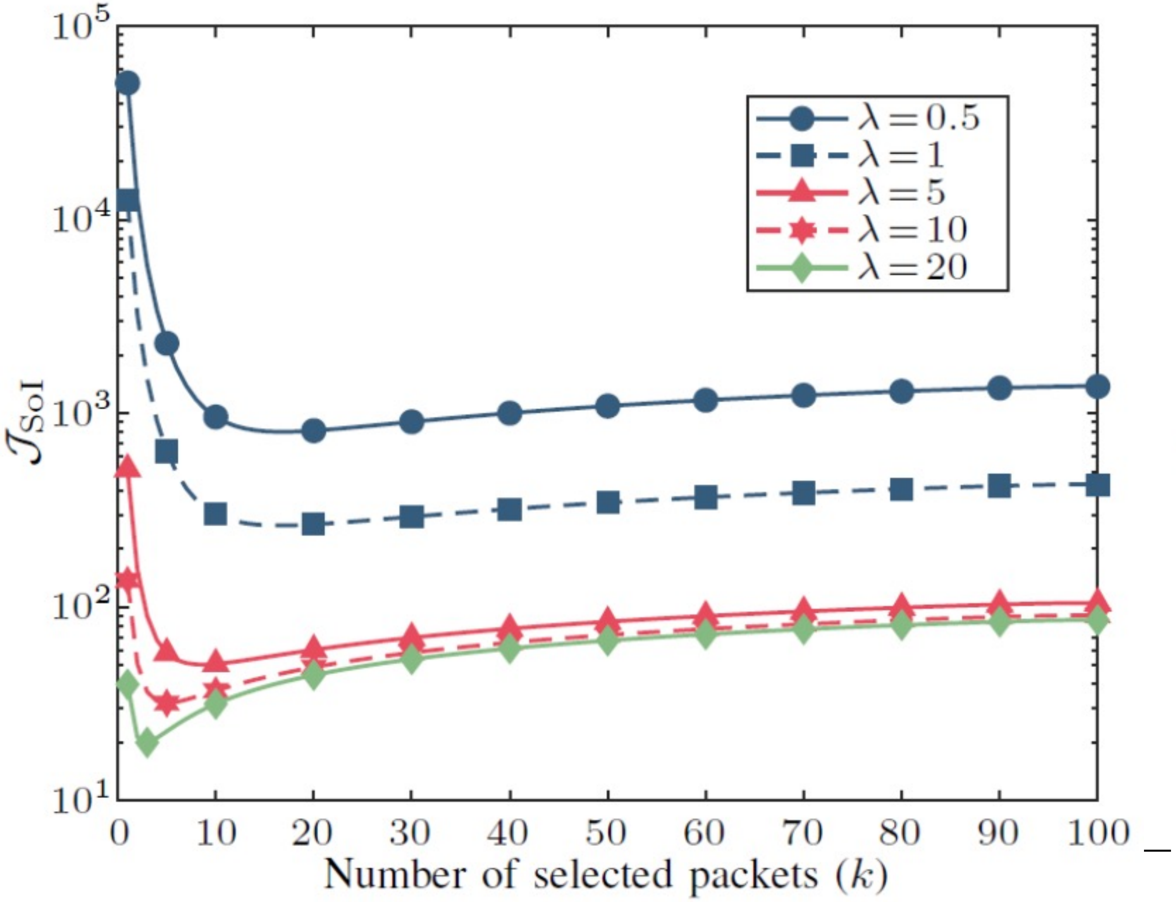
Kraft-McMillan inequality: existence of a uniquely decodable code for a given set of codeword lengths

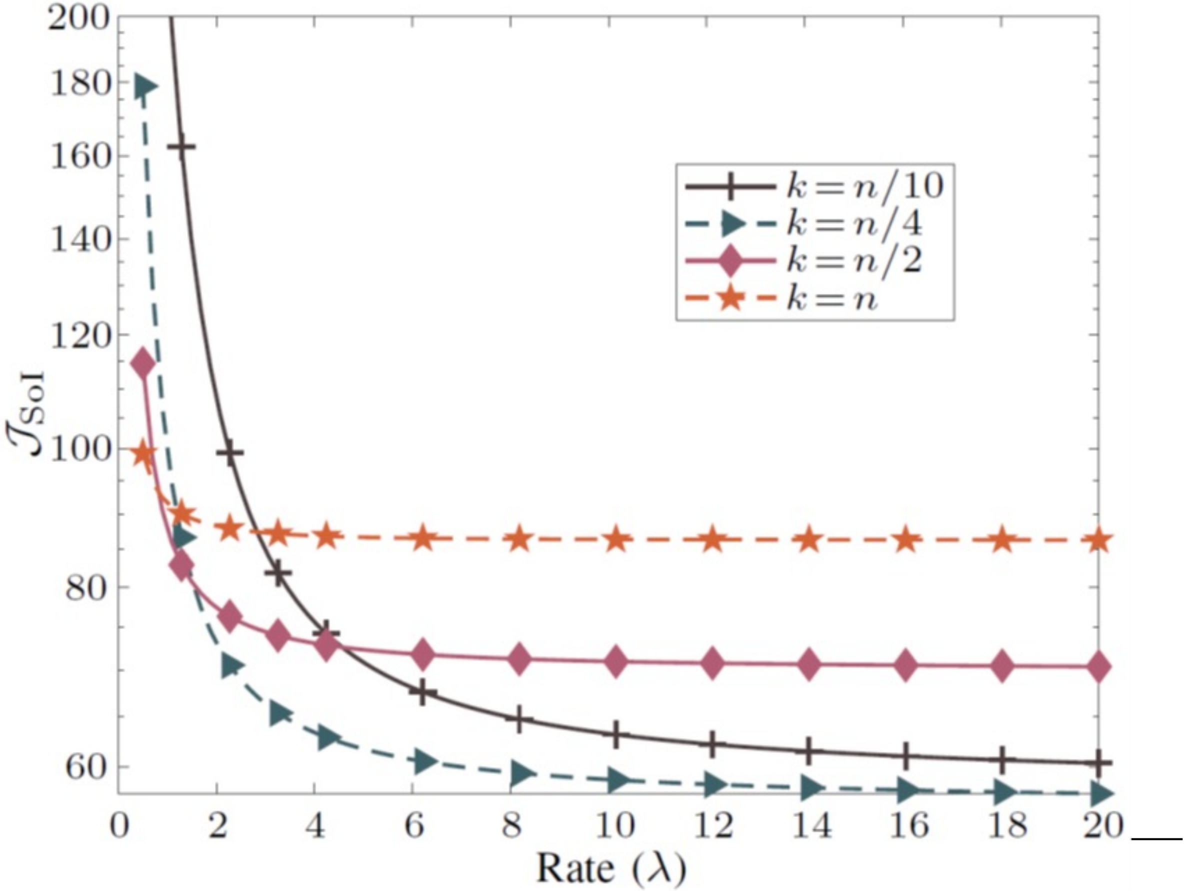
Numerical evaluation

Zipf(n, s) distribution with pmf $P_X(x) = \frac{1/x^s}{\sum_{j=1}^{n} 1/j^s}$

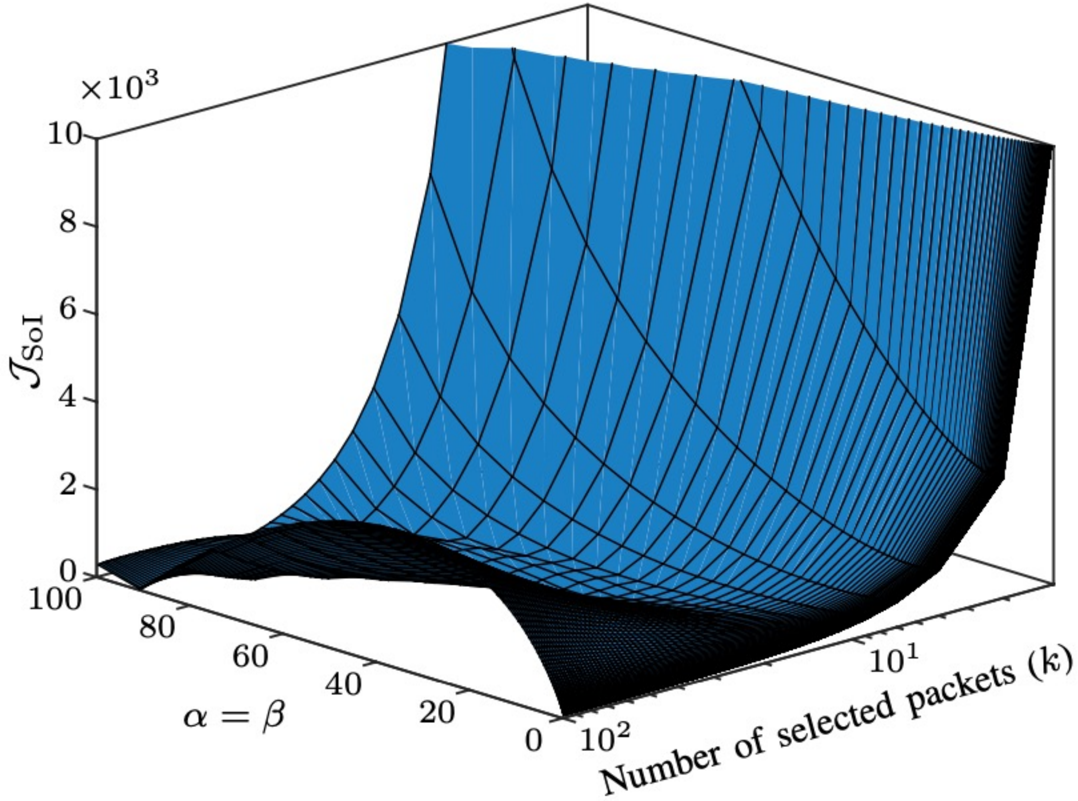
$n = |\mathcal{X}| = 100$ and exponent $s = 0.4$

$s = 0$ uniform, $\nearrow s$ "peaky distribution"





Interplay among SoI, semantic filtering (k), and codeword length



OPTIMAL PARAMETERS UNDER THE EDT SCENARIO

λ	k	$\alpha=\beta$	λ	k	$\alpha=\beta$
0.5	20	1.26	10	5	2.5
1	18	1.58	20	2	12.59
5	10	1.99			

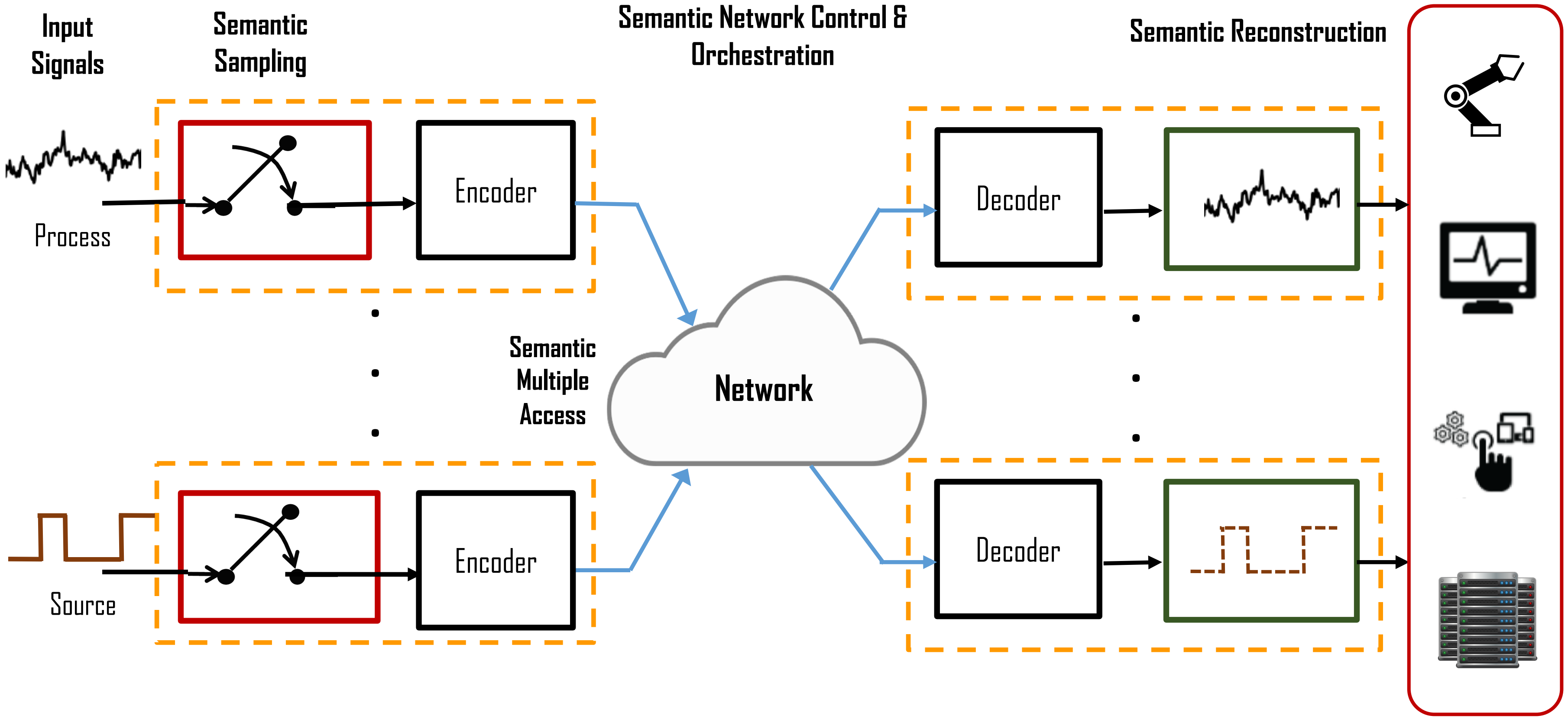
- Objective function continuously increases as cost parameters increase for small k
- *For large* k: increasing cost parameters causes the objective function to increase then decrease
- *Increasing the input rate* (hence, decreasing k^*), optimal cost parameters increase
- *When input rate is high*: larger penalties for the codeword length must be assigned to ensure transmitting the most important data.

Other applications of SoI

- G. Stamatakis, N. Pappas, A. Fragkiadakis, A. Traganitis, “Semantics-Aware Active Fault Detection in Status Updating Systems", arXiv, Feb. 2022.
- G. Stamatakis, N. Pappas, A. Fragkiadakis, A. Traganitis, “Autonomous Maintenance in IoT Networks via AoI-driven Deep Reinforcement Learning", IEEE INFOCOM - 4th Age of Information Workshop, May 2021.

Concluding remarks

Where we go from here



Communication process extends up to goal-oriented signal reconstruction and information exploitation
A monitored signal: a physical phenomenon/event distributed in space and evolving in time

Key semantic operations - Prioritize information and goal-driven representation of it

Projects funding this research

- "Semantics-Empowered Communication for Networked Intelligent Systems", Swedish Research Council, (2022-2025)
- "Information Handling in Industrial IoT", ELLIIT, (2021-2025)
- "Low Latency Communications for Wireless Networks: Exploiting Traffic Characteristics", CENIIT, (2018-2023)

Swedish Research Council

ELLIIT

CENIIT

Thank you!
Reach me at
nikolaos.pappas@liu.se