Goal Oriented Communication for the Scale-up of MTC

Elif Uysal

Middle East Technical University (METU),Turkey

CS Distinguished Seminar
Chalmers University, Sweden
November 10, 2023

METU

CNG
Communication Network Group

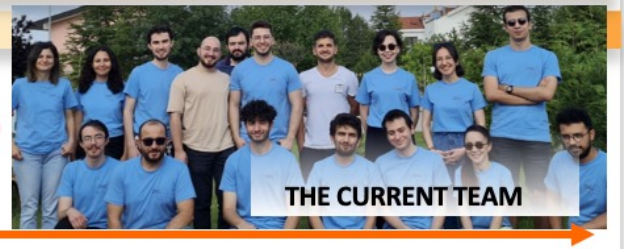
Communication Networks Research Group
http://cng-eee.metu.edu.tr

Where do I come from

PROF. ELİF UYSAL
Fellow, IEEE Fellow, AAIA

S.M 1997-1999 | PhD 1999-2003 | Lecturer 2003-2005 | Assistant Prof 2005-2007 | Assoc Prof 2007-2014 | Visiting Researcher 2014-2015 | 2015-2016 | 2017-

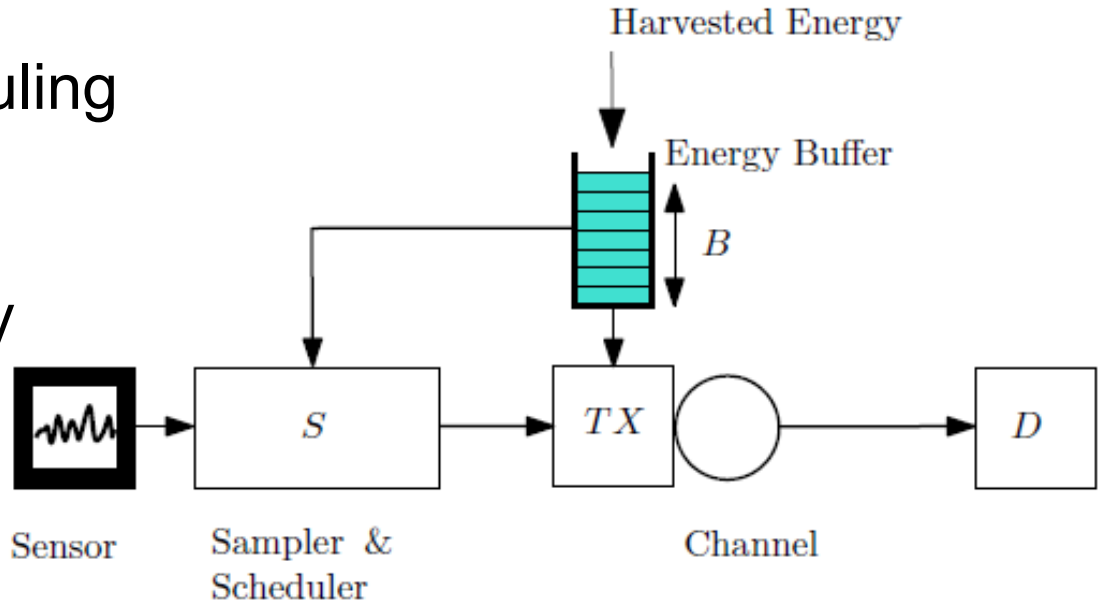
Energy-efficient Transmission Scheduling and WSNs | Wireless Multiuser Scheduling | Energy Harvesting Networks | Age of Information--->GoC



http://cng.eee.metu.edu.tr/

Our exploration of Machine Type Communication:

- 2000s: Energy Efficient Transmission Scheduling
  -Exogeneous "bursty" arrivals of data
- 2010s: Energy-harvesting, "Zero-Energy"
  -Exogeneous "bursty" arrivals of energy
- 2015-2020: Age of Information FRESHDATA
  - "Generate-at-will"
- 2021-today: Goal-Oriented Communication
  - "Selection-from-buffer"



Communication Networks Research Group (CNG)

CNG Team

- 2 faculty
- 1 Postdoc
- Grad/undergrad researchers
- 2 admins
- >35 alumni

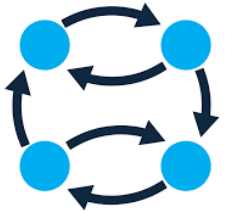
FRESHDATA Technologies

Spin-off, 2022

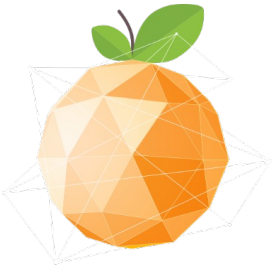
- SUIT (Sustainable Urbanization through Innovative Technologies) (2022)
- Consortium of universities, research labs, companies

General Outline

- Why Goal-oriented Communication?



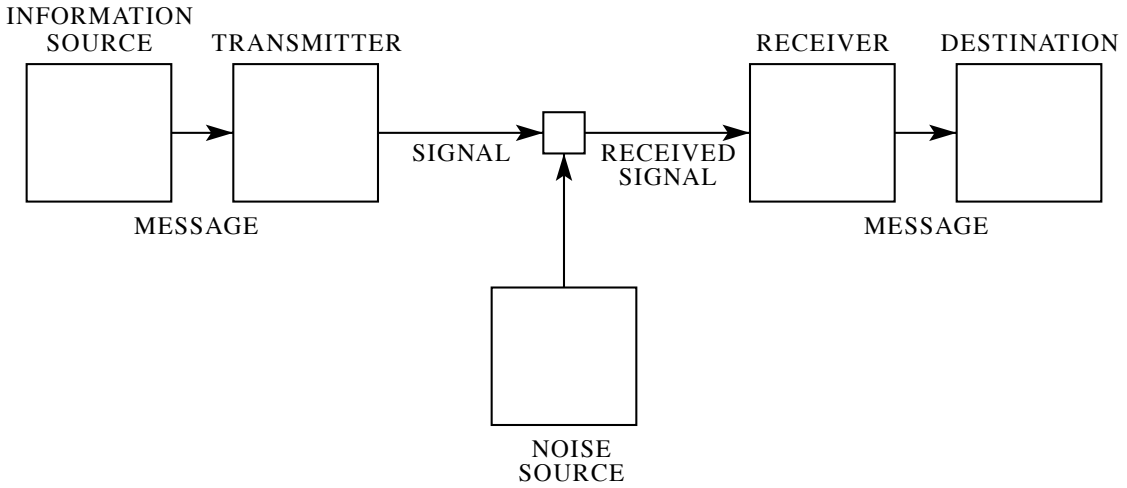
- What are some goal-oriented KPIs that can guide the design of protocols today



- Re-design of networks through these metrics

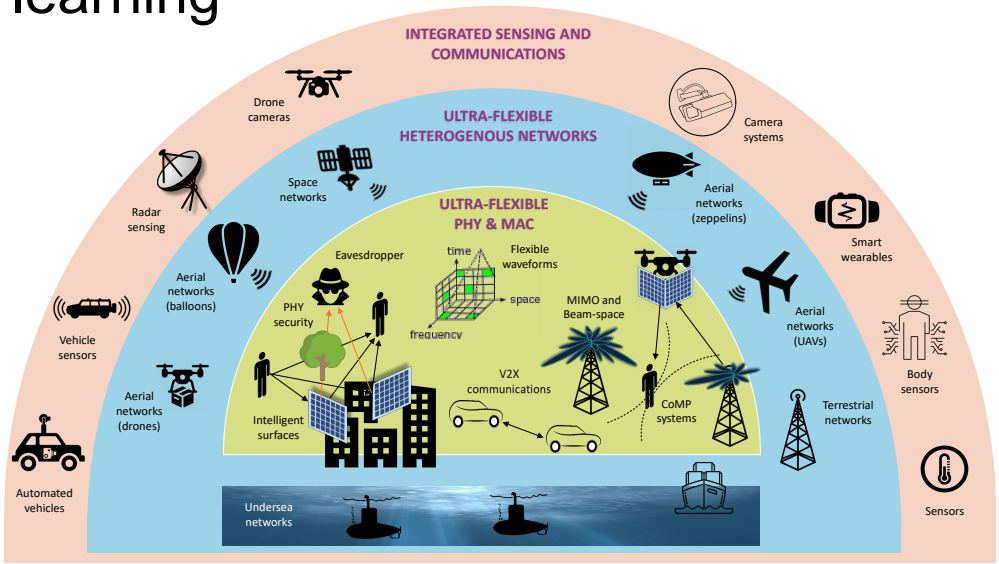
Classical Communication Systems vs Machine Type Communications

- Humans choose the data
- The network ensures its correct, to some extent timely delivery



Shannon's point to point communication model

- *Networked applications*, machine-type data.
- Real-time systems; automated decisions in a sense-compute-actuate cycle.
- Massive access e.g. IoT
- Learning for communications/communications for learning



Source: Yazar et al, 6G Vision: An ultra flexible perspective.

Transmission Problem vs Effective Communication

Recent Contributions to
The Mathematical Theory of Communication

Warren Weaver

September, 1949

Claude Shannon

Warren Weaver

Foundations of the Theory of Signs
Charles W. Morris

Charles W. Morris

1.2 Three Levels of Communications Problems

Relative to the broad subject of communication, there seem to be problems at three levels. Thus it seems reasonable to ask, serially:

LEVEL A. How accurately can the symbols of communication be transmitted? (The technical problem.)

LEVEL B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)

LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)

Semiosis and Semiotic

1) syntactics – the study of the methods by which signs may be combined to form compound signs,
2) semantics – the study of the signification of signs,
3) pragmatics – the study of the origins, uses, and effects of signs.

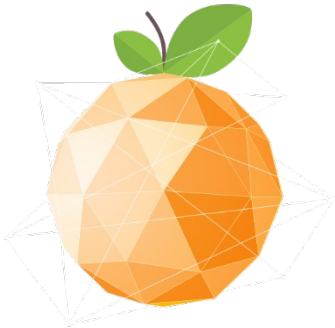
The science of relation of signs to their interpreters

R Seising. 60 years 'A Mathematical Theory of Communication' - Towards a 'Fuzzy Information Theory', 2009.

GAP current protocols and effective communication

Available Communication Protocols

Semantic/Effective Communication



Traditional protocols

Optimized for high throughput/low delay,
Low loss (transmit all the data)

Real time monitoring/decision making

Goal-oriented performance criteria

- **MTC is Loss resilient !**

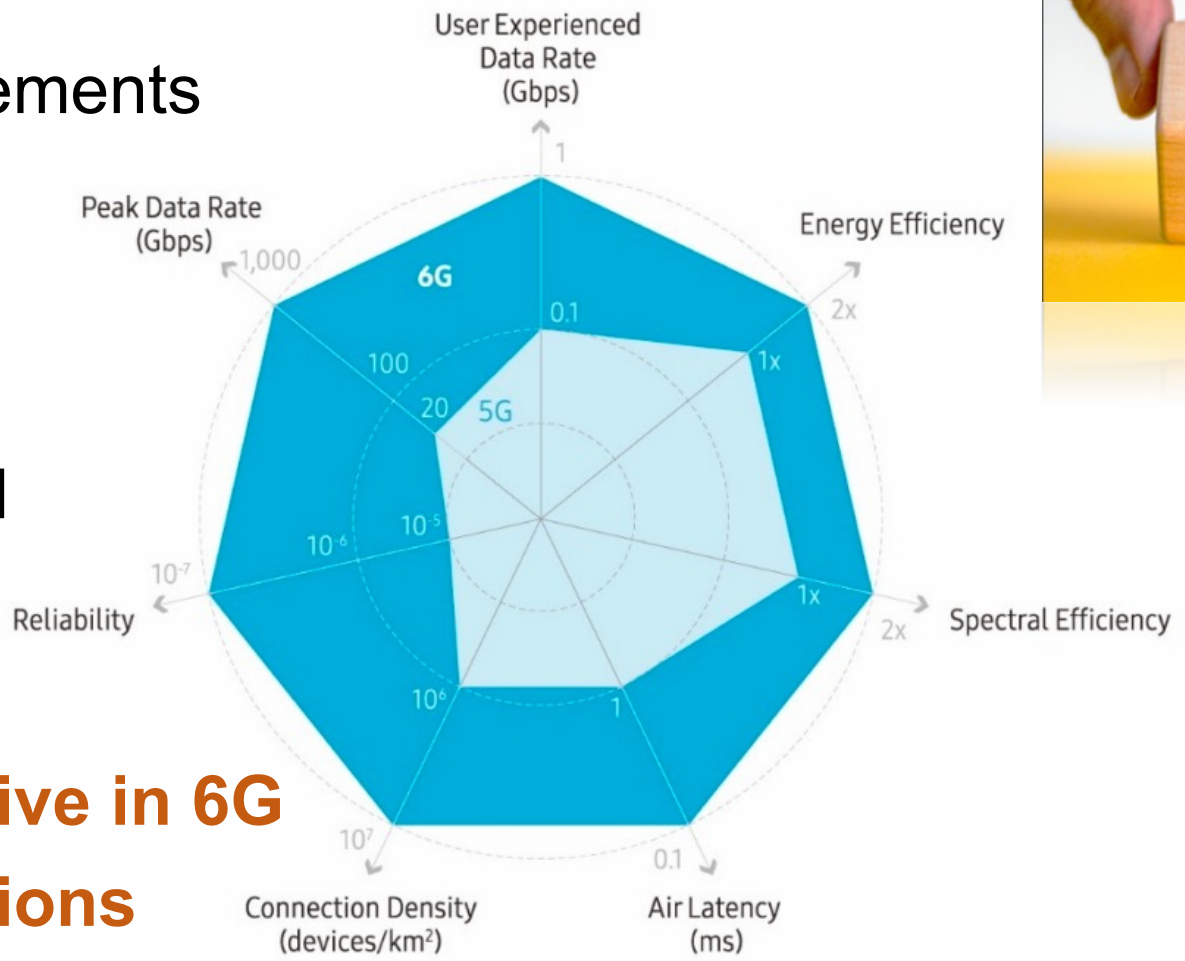
e,g. irrelevant data can be dropped from queues

- **Future networks: AI-native**

Exploit cognition introduced into the communication system for resilient and robust networking

MTC in 6G: we cannot ignore the gap

- Conflicting ambitious requirements
- One size fits all not possible
- Must rethink the **end-to-end** communication system design



Sensing and NTN will be native in 6G

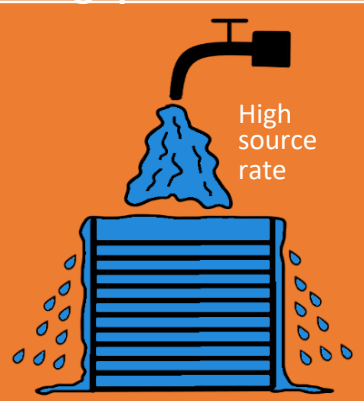
Further: Space Communications

https://news.samsung.com/global/samsungs-6g-white-paper-lays-out-the-companys-vision-for-the-next-generation-of-communications-technology

Effectiveness KPI Example: Age of Information

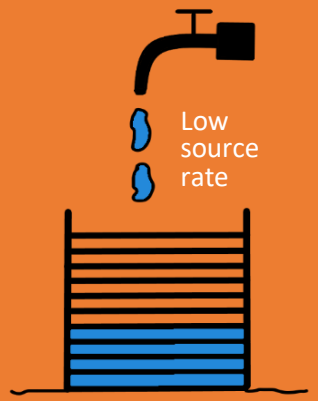
Traditional Approach

Throughput Maximization



Minimum idling: full buffers

Delay Minimization



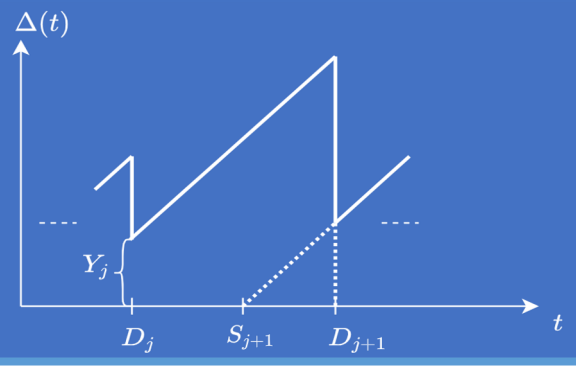
No congestion
Minimum queuing: light buffers

Age is a composite measure of Throughput and Delay.

Age of Information

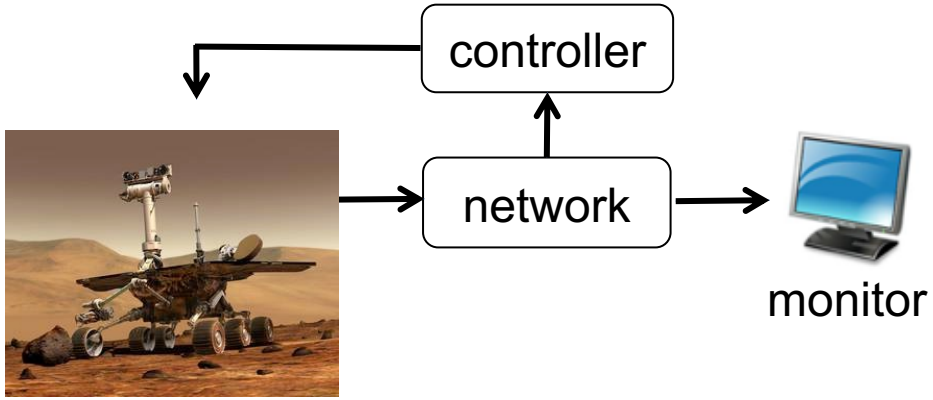
- A measure of freshness for an information flow
- Let $u(t)$ be the timestamp of the newest packet that the destination has received by time t.
- The age of this flow at t is

$$\Delta(t) = t - u(t)$$

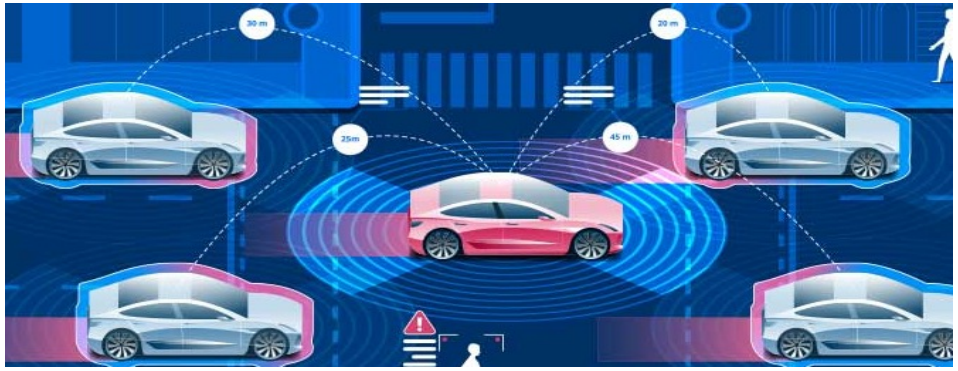


Effectiveness KPI Example: Age of Information

Satellite
IoT



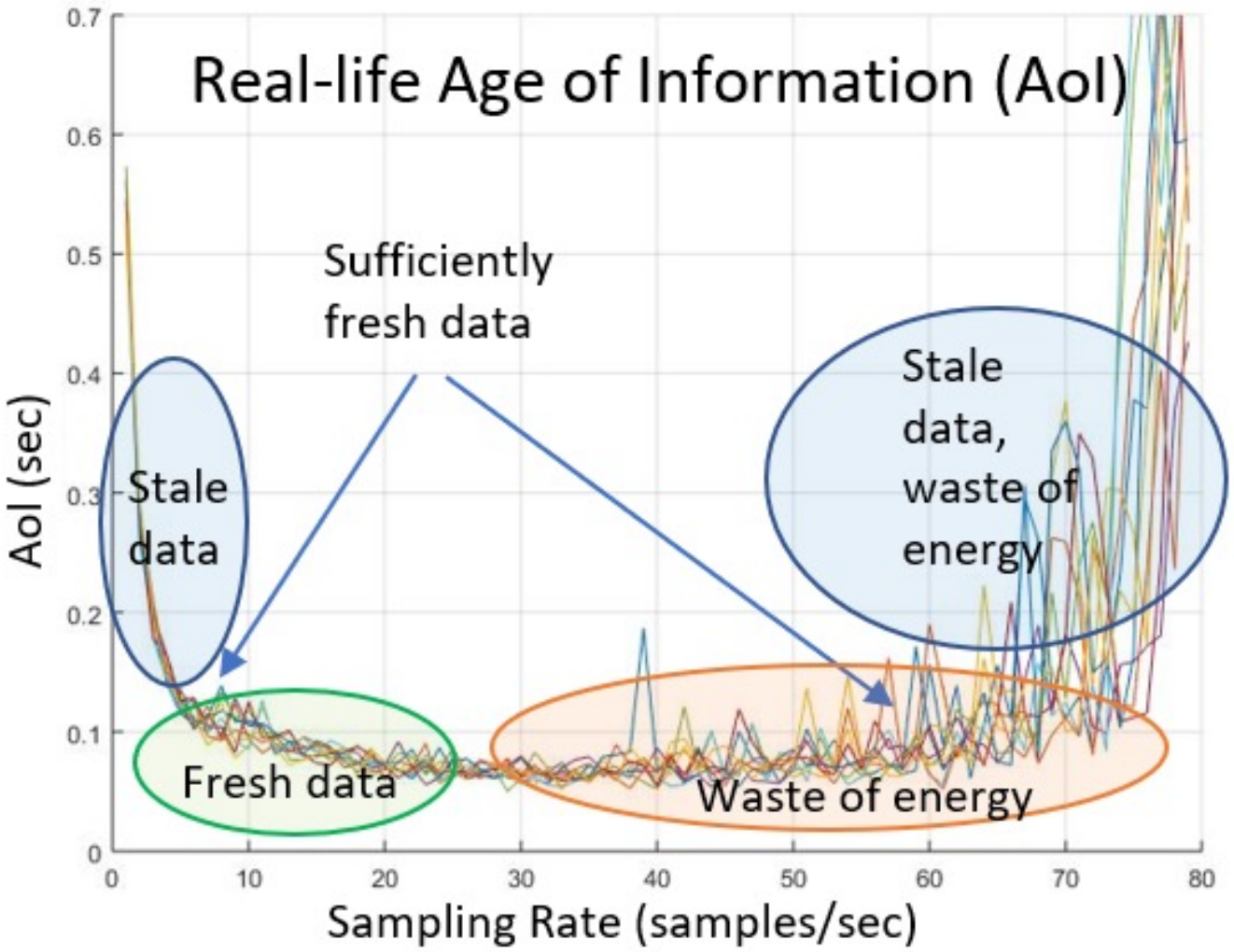
Remote Monitoring and Control



Automated Vehicles

Terrestrial
localization
and tracking

Effectiveness KPI example: Age of Information



Book Chapter:

Age of Information In Practice

[Uysal, Kaya, Baghaee, Beytur, 2023]

Conference presentations:

[Guloglu, Baghaee , Uysal 2021]

[Beytur, Baghaee, Uysal 2020]

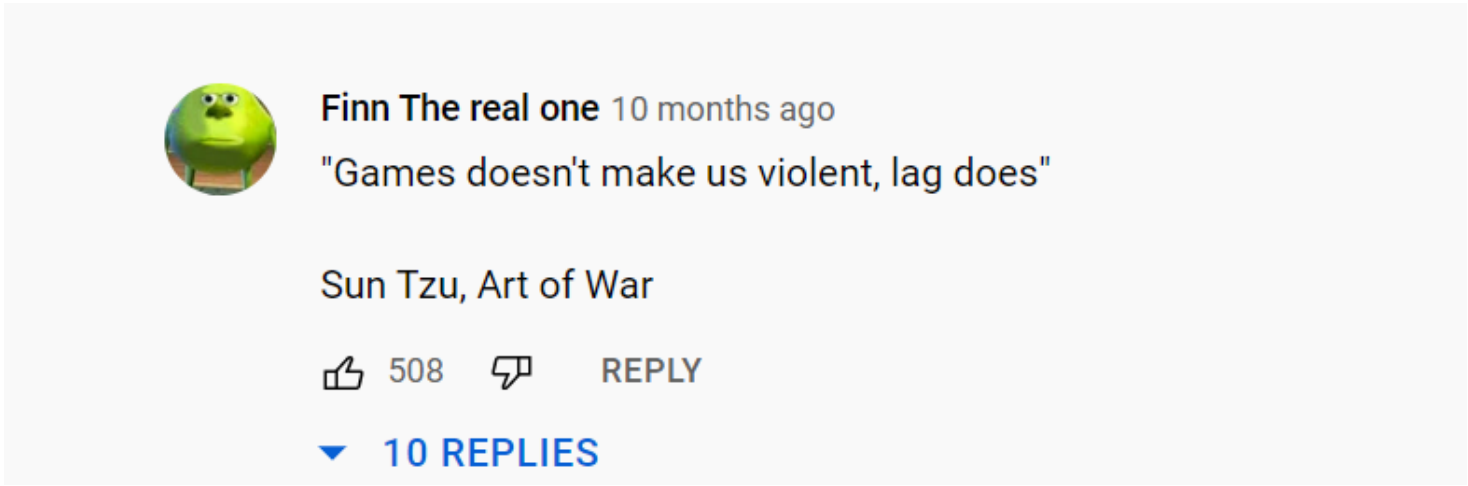
[Beytur, Baghaee, Uysal 2019]

[Sonmez, Baghaee, Ergisi, Uysal 2018]

[Sert, Sönmez, Baghaee, Uysal 2018]

[Baghaee, Beytur, Uysal 2019]

URLLC – The Right Approach?



- Ultra-Reliable Low Latency Communication (URLLC) :
  - high reliability (e.g., > 99.999%)
  - 1 ms delay

- One “umbrella” to satisfy almost all MTC applications currently envisioned

- Same requirements to be followed in 6G?

Low Latency: Neither necessary nor sufficient!

Latency	1.1 ms	1.1 ms
Sampling period	1 ms	0.1 ms
Peak Age	2.1 ms	1.2 ms

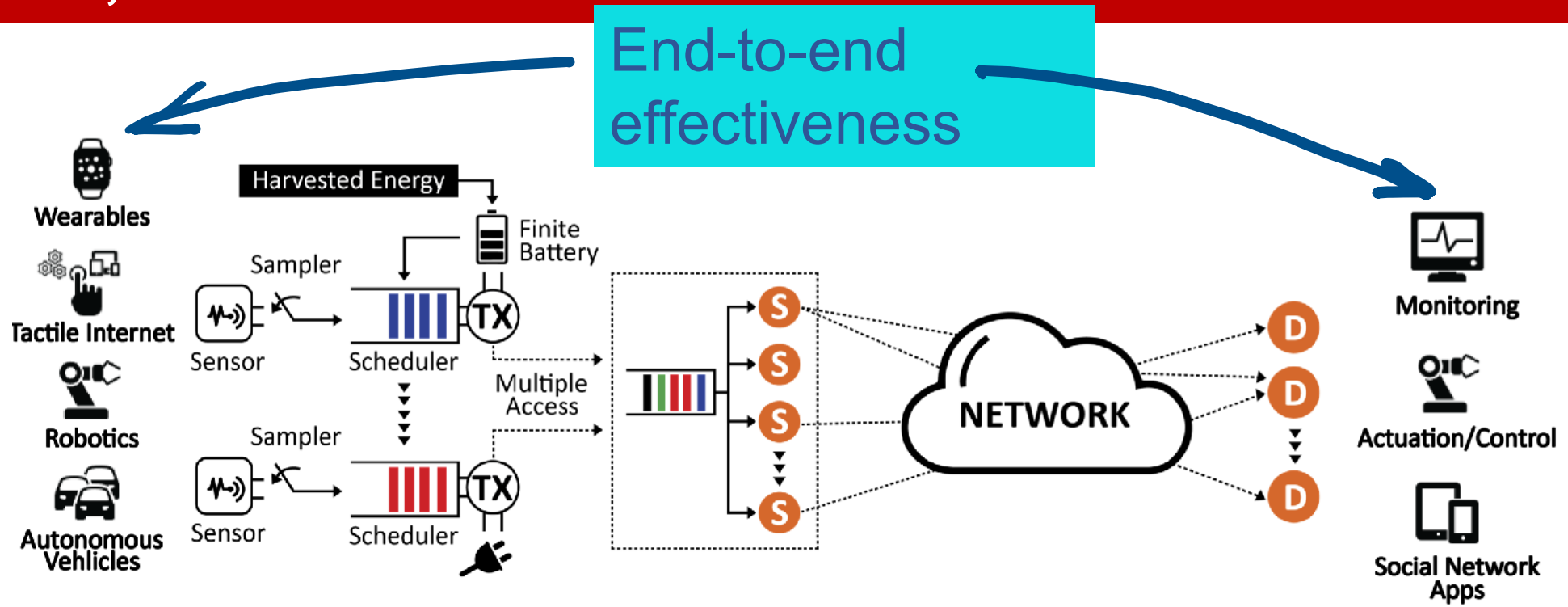
Not sufficient!

- In cyber-physical systems timeliness cannot be captured with low latency alone

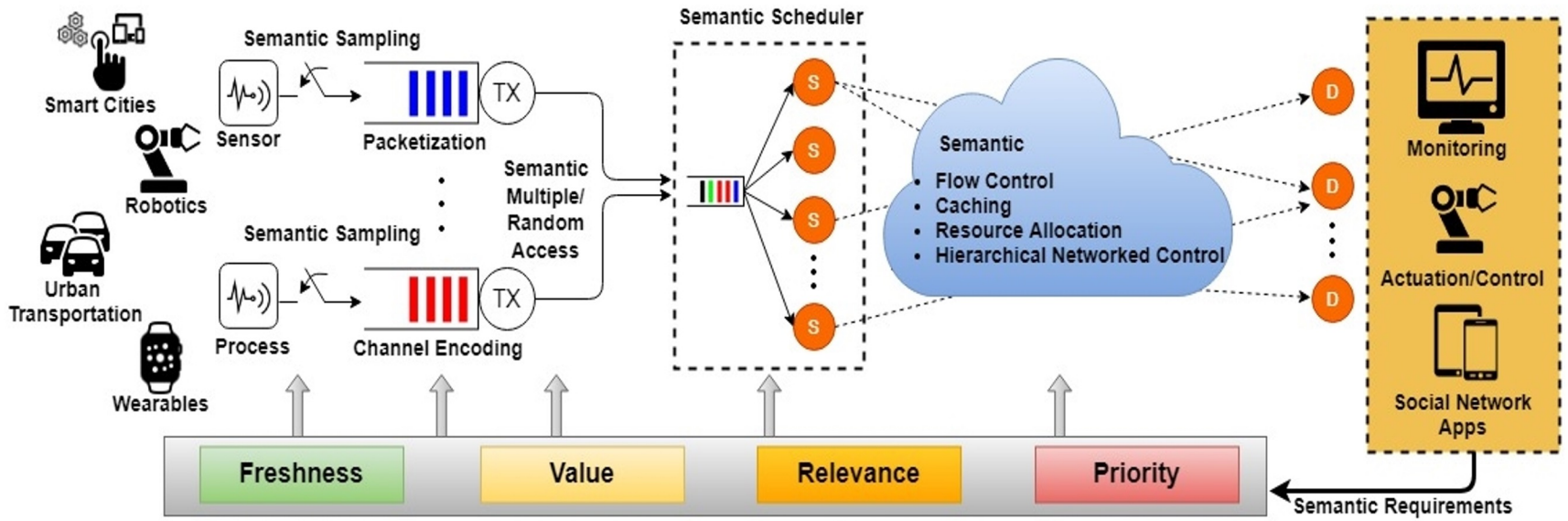
Unnecessary!

- Consider eMBB and URLLC coexistence
- High rate vs intermittent critical updates
- Low latency ⇒ reserved slots ⇒ **over-provisioning**

Semantic communication: a data significance perspective, Uysal et al., IEEE Network 2022



End-to-end Goal-Oriented Communication Architecture



- New effectiveness measures and metrics that define them.
- Develop goal-oriented link, transport and application layer principles in concert
- Relax the exogeneous data arrival assumption
  - **Non-uniform process-aware sampling**.
- New communication protocol principles tailored for information flow in networked **control systems**.

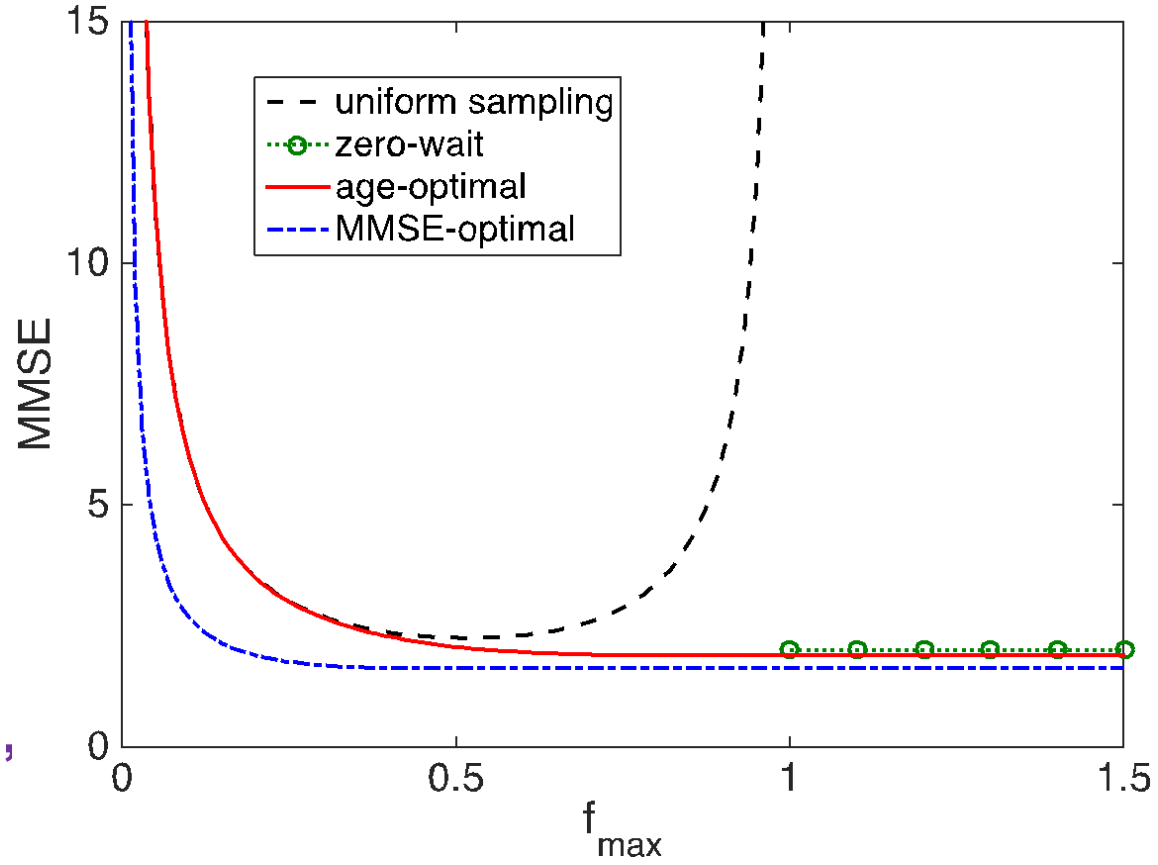
Semantic Attribute: Relevance

- Remote Tracking from Samples sent over a network with delay
- Measurements of a process/images/video

sent for remote estimation/ AI, etc

- Separate handling of sampling, encoding and transmission -> highly suboptimal
- **Optimal sampling and transmission:**

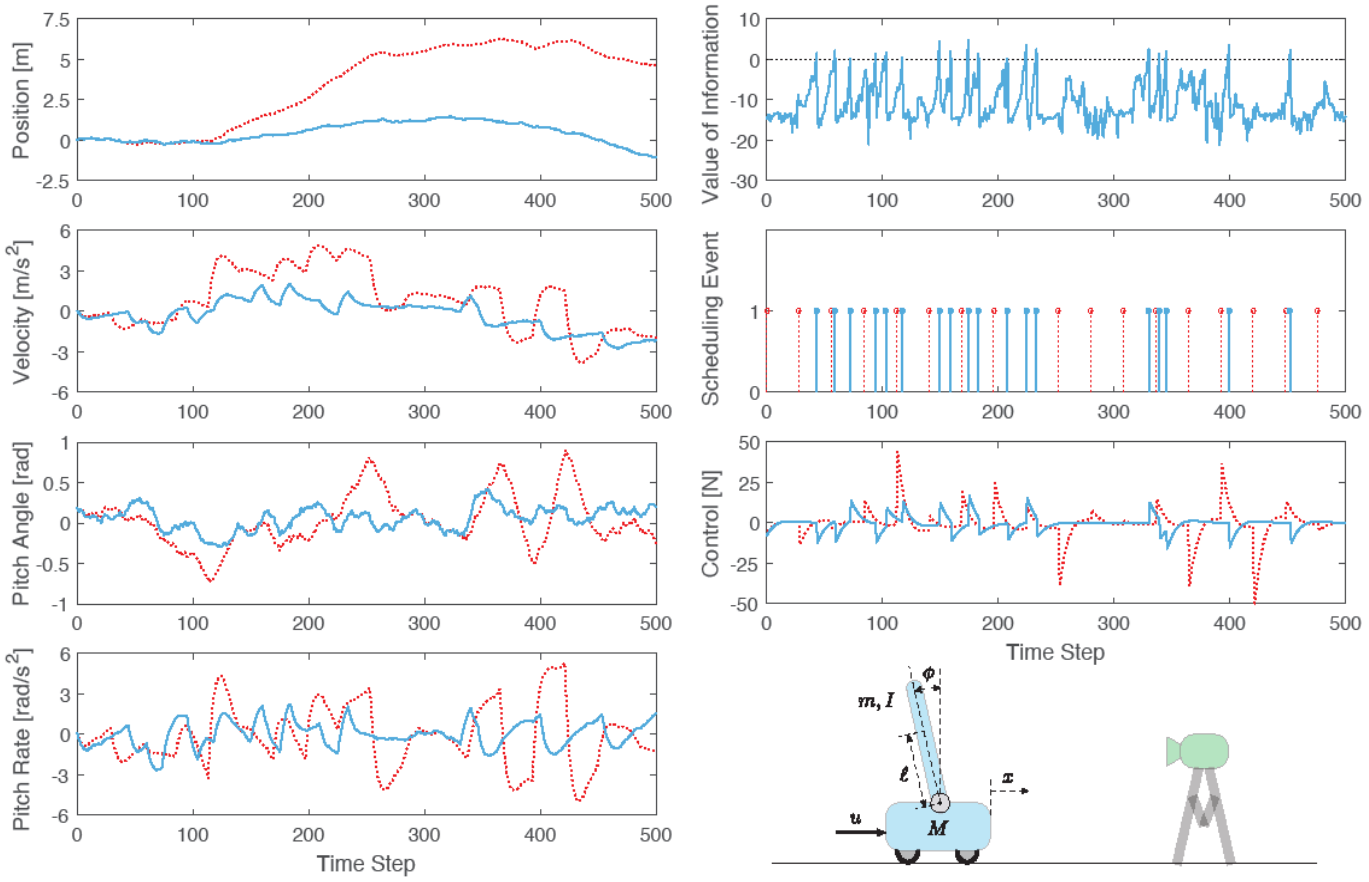
Generate At Will, based on Age, delay statistics, and the state of the process



[Sun Polyanskiy Uysal 2019]

Semantic Attribute: Value- VoI

- Relevance: source based
- Value: the *value* of the next source sample to the point of computation.
- (VoI) : difference between the benefit of having this sample and the cost of its transmission.

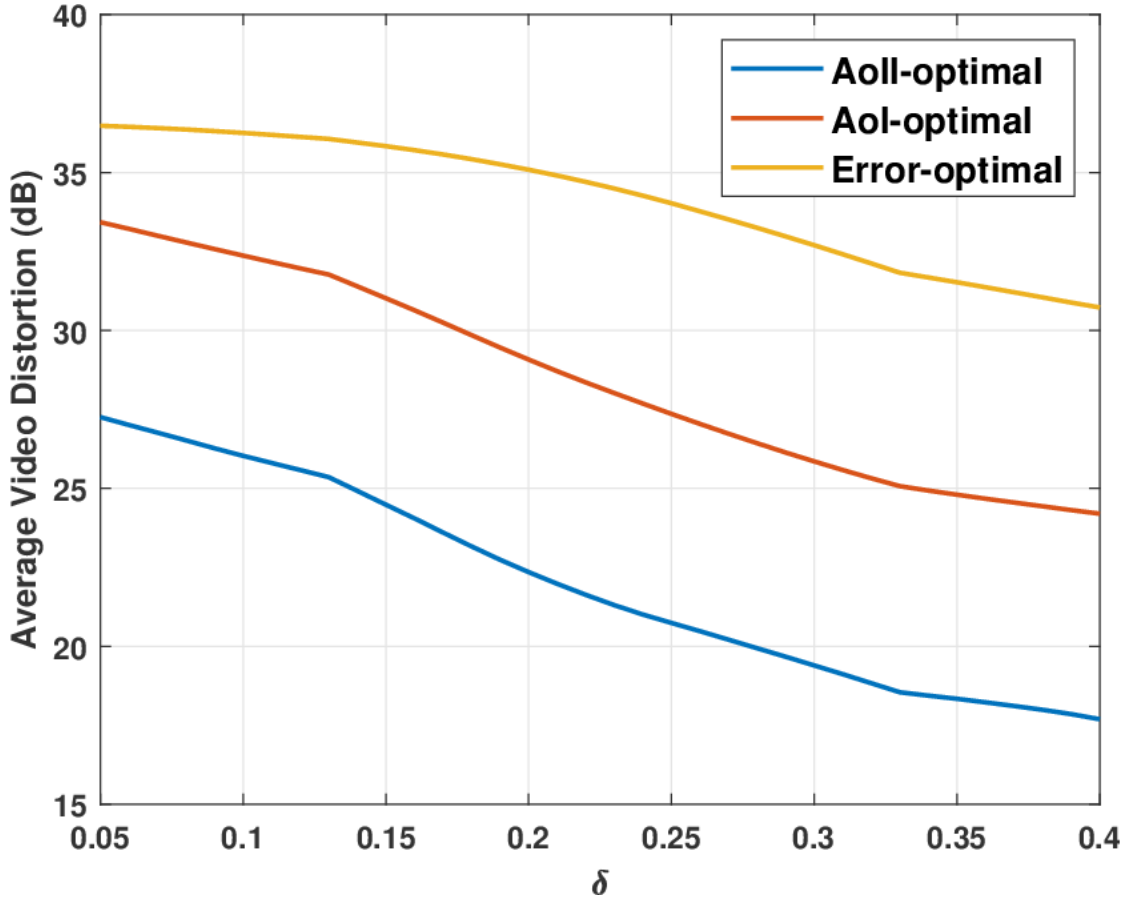


- VoI >0 18 times out of 500 -> transmit control signal.
- VoI based (blue), periodic with same number of transmissions (red).

T. Soleymani, Value of Information Analysis in Feedback Control. Ph.D. Thesis, Technical University of Munich, 2019.

Semantic Attribute: Freshness & Relevance -AoII

- Combines aging and accuracy
- E.g. application in video streaming: an AoII-based transmission policy reduces the distortion with respect to AoI-based and conventional error-based policies.



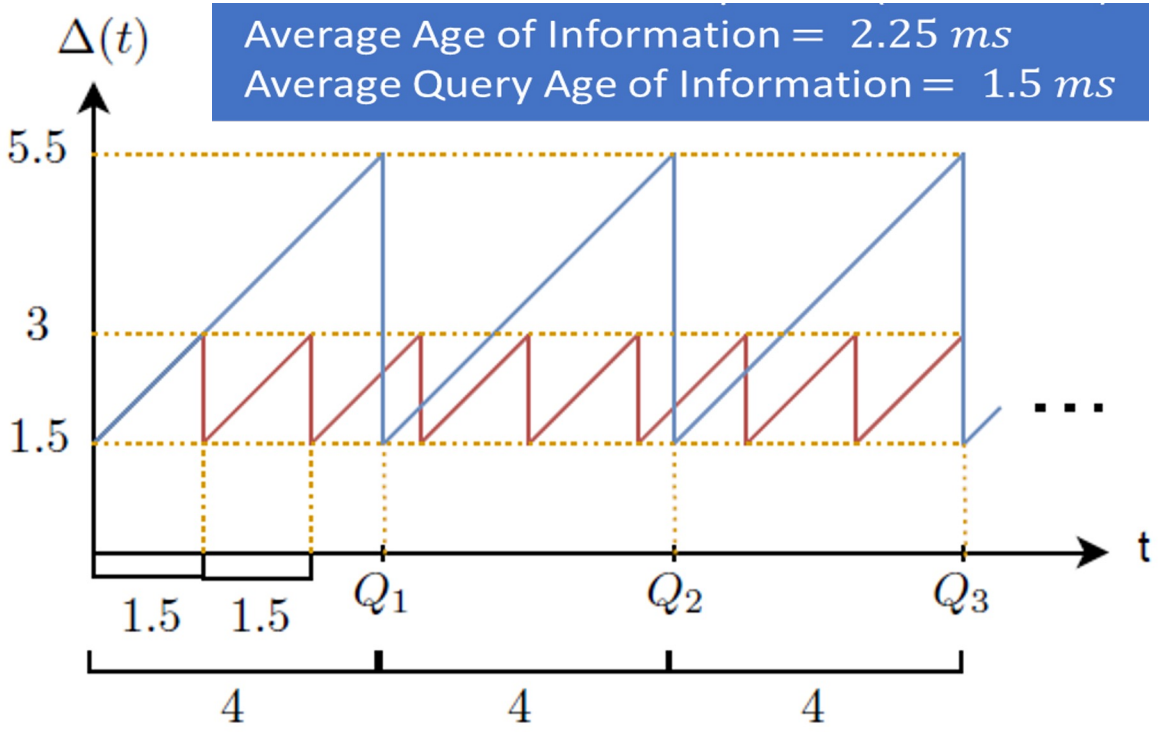
A. Maatouk, M. Assaad, T. Ephremides. The Age of Incorrect Information: an Enabler of Semantics-Empowered Communication, 2020

Semantic Attribute: Freshness & Value: QAoI

➢ **Query Age of Information (QAoI)**: AoI at query instants Q_k (Chiarotti et. al. 2021)

- Pull based systems.
- Eg. Satellite IoT
  - GEO: Periodic Query instants, constant coverage
  - LEO: Intermittent connectivity with some blind slots, yet predictable query times.
  - Best to send right before query time, but perhaps allowing enough time for retransmission.

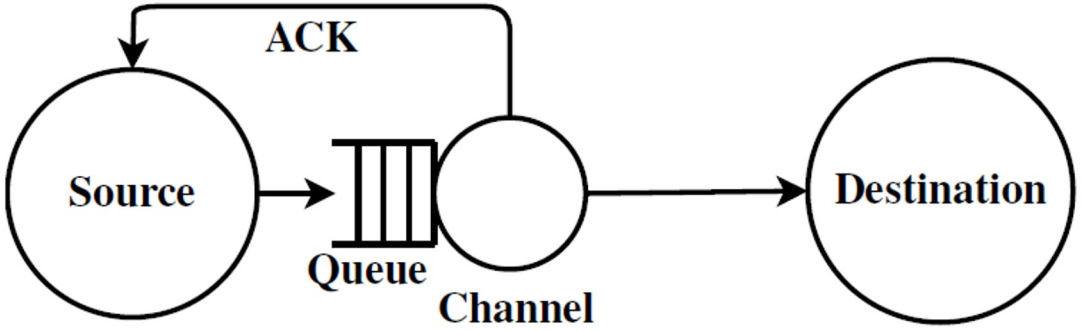
[Ildız et al 2021]

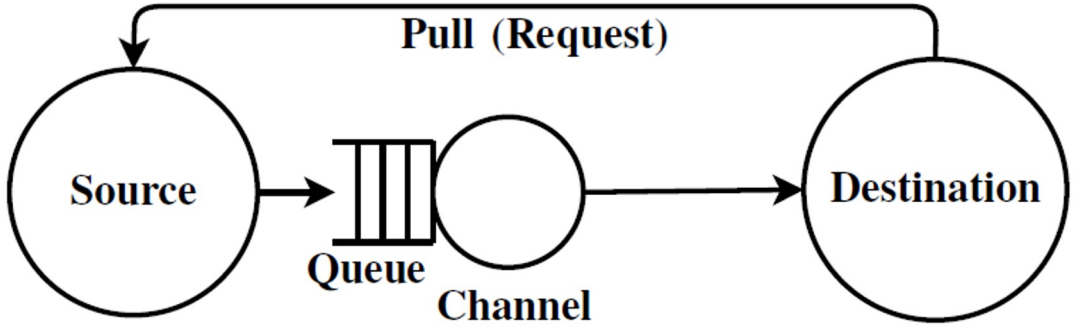


AoI Optimization
- Worse Performance
- Wasteful

QAoI Optimization
- Better freshness
- Fewer transmissions

“Update-or-Wait” vs “Pull-or-Wait”





$$\bar{g}_{opt} = \min_{\pi \in \Pi} \lim_{n\to\infty} \sup \frac{E\left[\int_0^{D_n} g\left(\Delta(t)\right)dt\right]}{E[D_n]}$$

$$s.t. \lim_{n\to\infty} \inf \frac{1}{n} E\left[\sum_{i=1}^{n} \left(Y_j + Z_j\right)\right] \geq \frac{1}{f_{max}}$$

*

$$\bar{h}_{opt} = \min_{\pi \in \Pi} \lim_{n\to\infty} \sup \frac{1}{n} E\left[\sum_{k=1}^{n} g\left(\Delta(Q_k)\right)\right]$$

$$s.t. \lim_{n\to\infty} \inf \frac{1}{n} E\left[\sum_{i=1}^{n} \left(Y_j + Z_j\right)\right] \geq \frac{1}{f_{max}}$$

- Equal, for Poisson queries (Ildız et al 2021)
- PoW dominates (Ildız et al 2022) for
- periodic queries, or
- Constant delay
- General problem open

with the assumptions

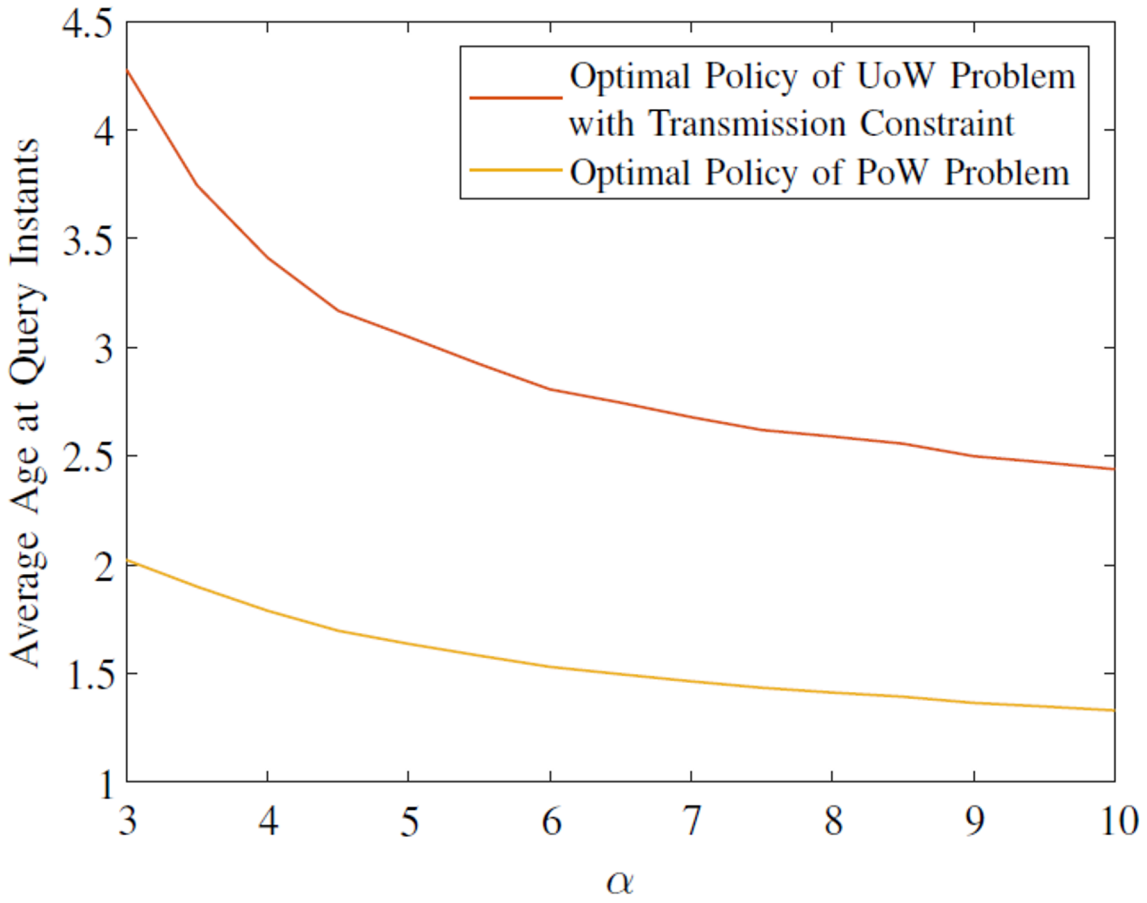
- The penalty function g is continuous.
- Y_j and Z_j is lower and upper bounded i.e.

$$P\left(Y_j \in [B_L, B_U]\right) = 1$$

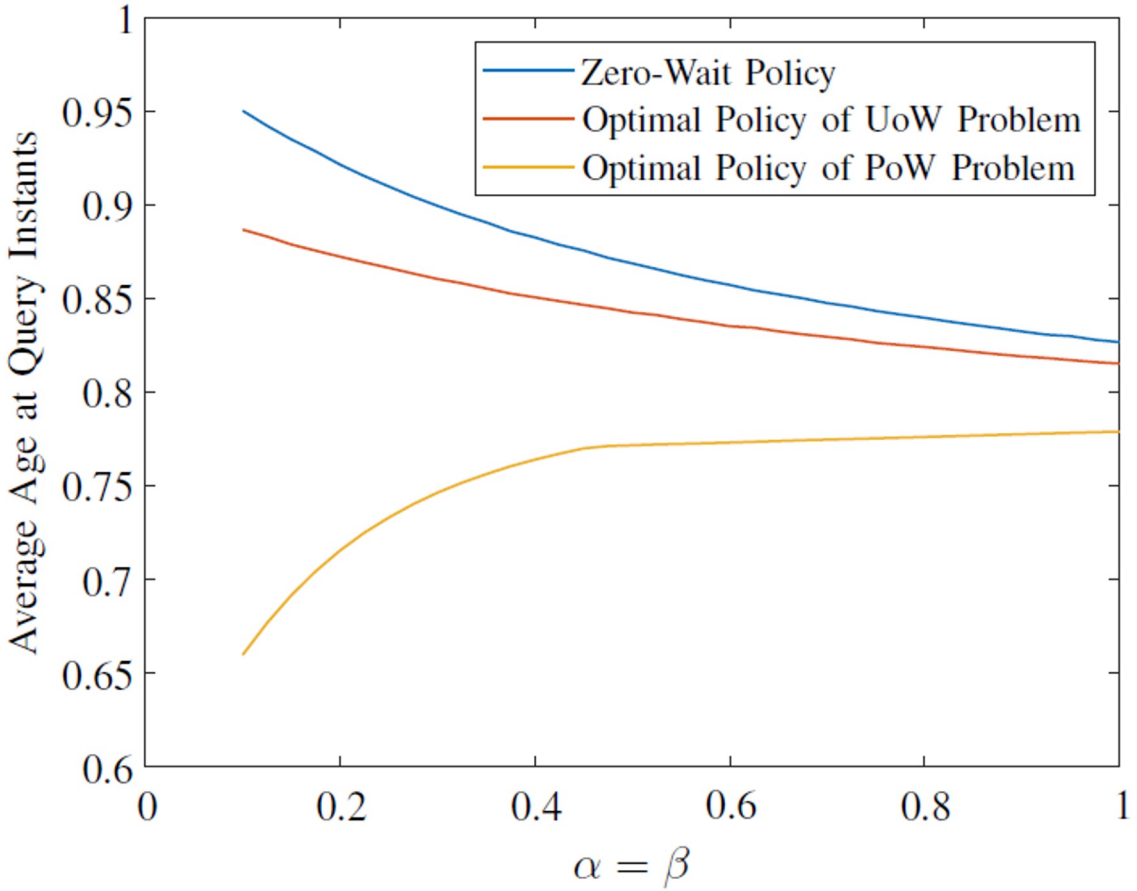
$$Z_j \in [0, M]$$

PoW dominates UoW in terms of performance

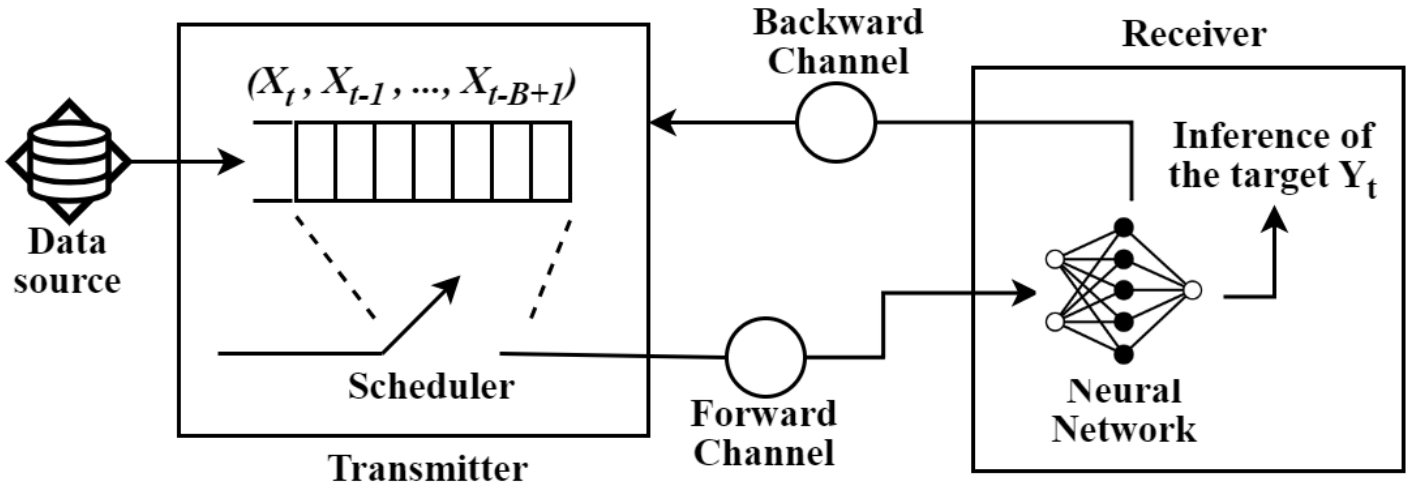
Pareto distributed transmission delays (IID with paremeters $x_m = 1$ and α)



Transmission delays are i.i.d. beta random variables with parameters (α, β).



Selection From Buffer (Shisher & Sun 2022)



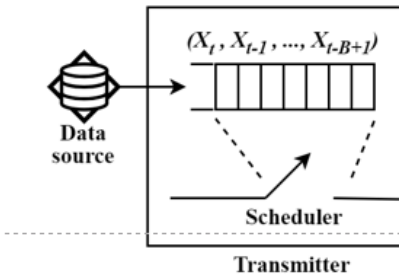
$$h_{opt} = \inf_{(f,g)\in\Pi} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_{(f,g)}\left[\sum_{t=0}^{T-1} h(\Delta(t))\right]$$

The function h is **task-specific** and not necessarily monotone

Scheduling policy optimizes inference performance using AoI as an **intermediate parameter**.

*Shisher, M. K. C., & Sun, Y. (2022, October). How does data freshness affect real-time supervised learning?. In *Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing* (pp. 31-40).

Selection from Buffer (SFB)



Buffer contains the most recently sampled B packets $(X_t,\ X_{t-1},\ \dots,\ X_{t-B+1})$ at any time slot t.

An optimal scheduler on the transmitter side must determine (i) when to submit a packet to the forward channel and (ii) which packet in the buffer to submit.

In case of IID transmission delay and immediate feedback, index-based threshold policy:

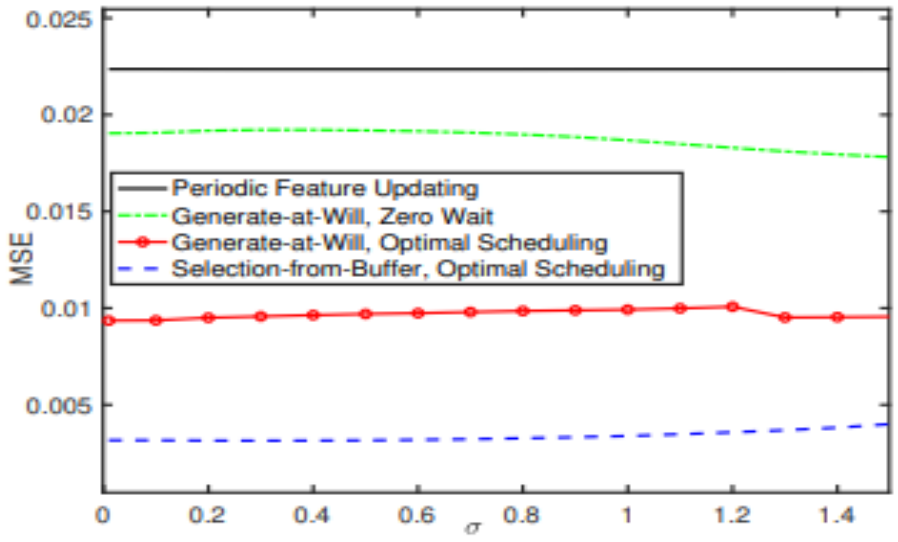
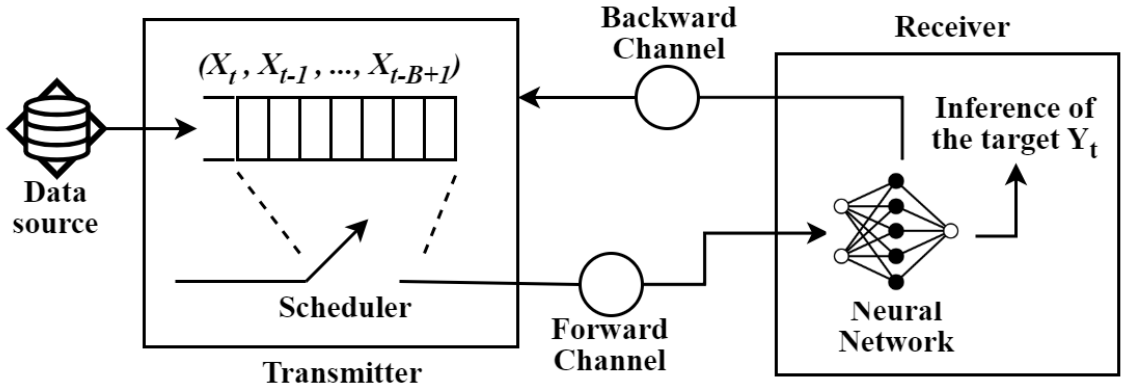


Figure 5: Time average inference error (MSE) vs. the scale parameter σ of discretized i.i.d. log-normal transmission time distribution for single-source scheduling (in robot state prediction task).

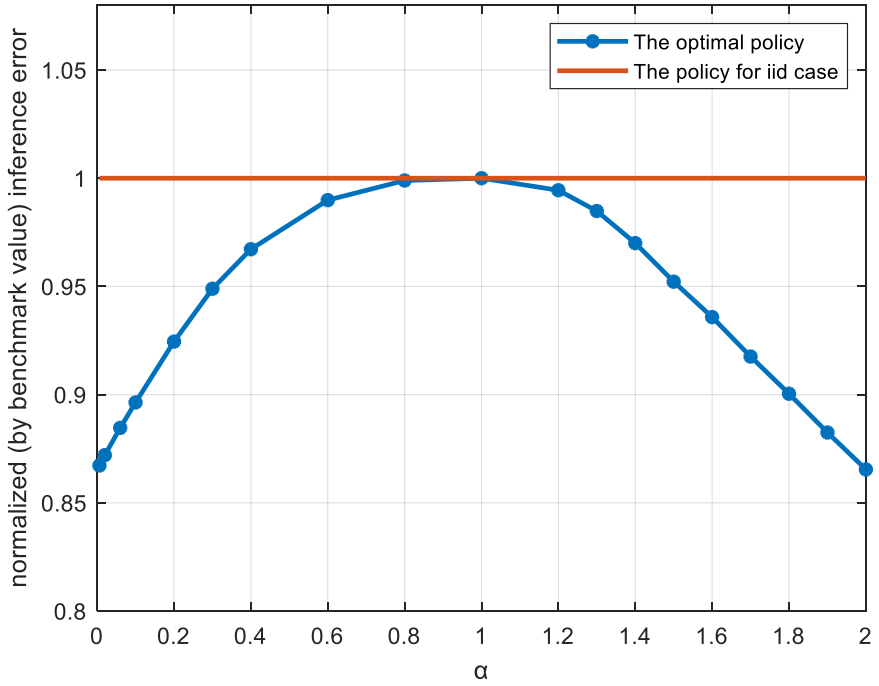
*Shisher, M. K. C., & Sun, Y. (2022, October). How does data freshness affect real-time supervised learning?. In *Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing* (pp. 31-40).

SFB under variable delay [Ari, Shisher, Uysal, Sun 2023]



In this case, the solution is an index-based threshold policy where the index varies based on the scheduler's knowledge of the memory of the delay:

$$\gamma(\delta, c) = \inf_{\nu \in \mathbb{Z}^+} \frac{1}{\nu} \sum_{k=0}^{\nu-1} \mathbb{E}\left[h(\delta + k + T_{i+1}) \middle| c_i = c \right]$$



$$\alpha = p_{01} + p_{10}$$

$$p_{01} = p_{10}$$

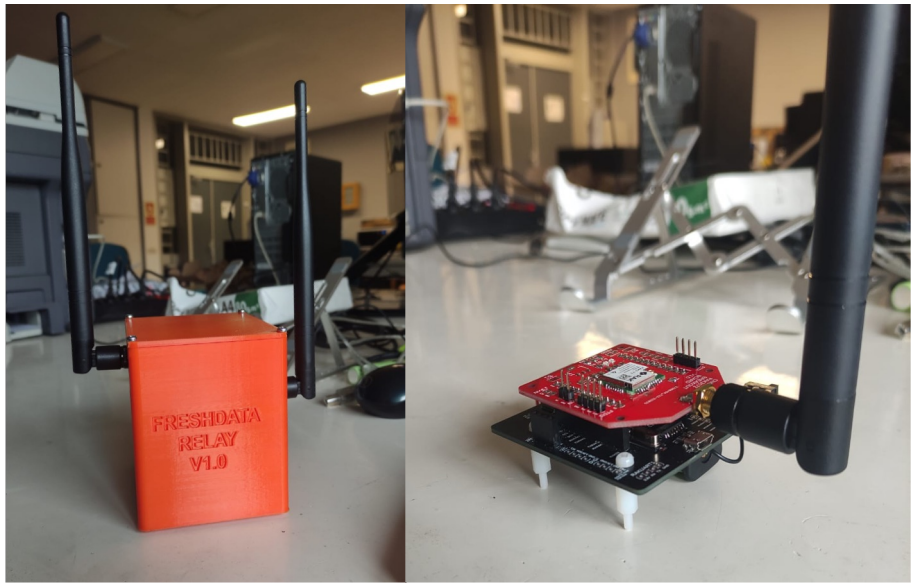
$\alpha = 1$ implies that the forward channel is iid.

Two forward channel states 0 and 1.

As α gets away from 1, the memory increases.

Goal Oriented Communication entering products

- Startup: FRESHDATA Technology
- Products and IPR development targeting LoRa Alliance, 3GPP



Satellite IoT

NTN

Terrestrial IoT

Broader Applications: Sustainable Urbanization

- SUIT (Sustainable Urbanization through Innovative Technologies) http://suit.metu.edu.tr/

- Consortium of universities, research labs, companies

- 11 projects – all spinning off from FRESHDATA

Thank you, from the CNG team

CNG
Communication Network Group

Est. 2007

http://cng-eee.metu.edu.tr