Native AI as enabler of Semantic or Goal-Oriented Communication in 6G and beyond

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

FRESHDATA

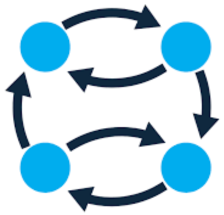
Elif Uysal

Professor, Middle East Technical University,Turkey
Founder, FRESHDATA Technologies

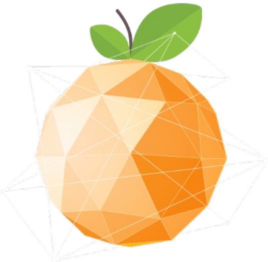
PIMRC 2023 WS08: Native AI and Semantic Communication
Toronto, ON, Canada
September 5, 2023

Outline

- Why do we need "Semantic" or Goal-oriented Communication



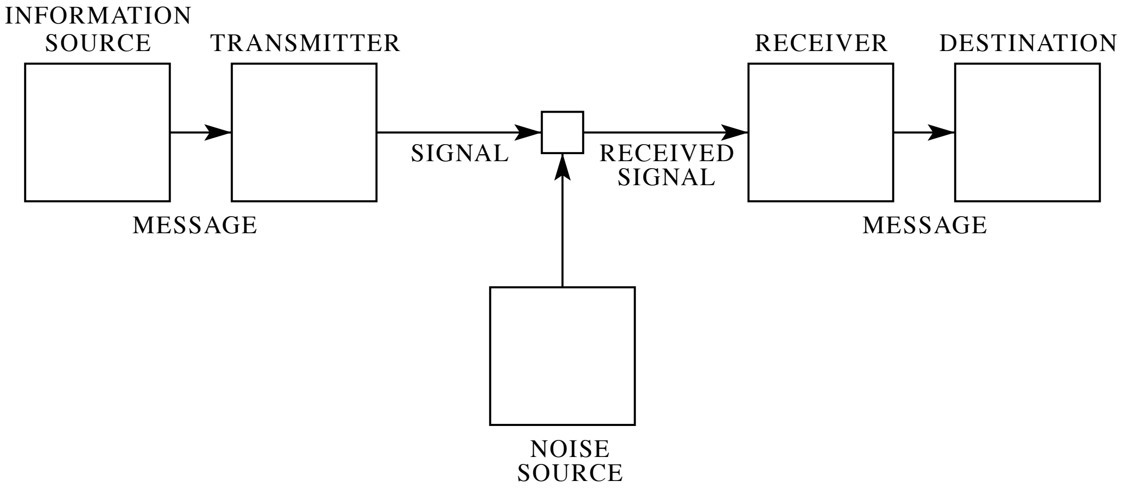
- What are some goal-oriented KPIs that can guide the design of protocols today



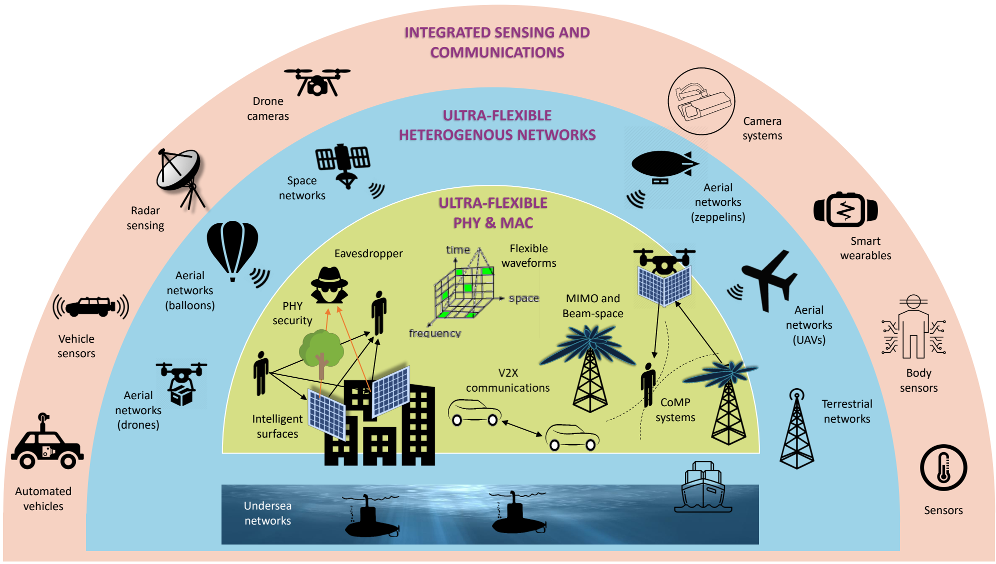
- AI/ML approaches to cater semantic and goal oriented objectives

Classical Communication Systems and Emerging Communication Systems

- Humans choose the data
- Network ensures **correct, timely delivery of ALL of the data**
- Shannon's definition of the transmission problem perfectly optimized this technical problem



- Machine-type Communications (MTC)
- Network caters timely & useful data for **correct decision making/actuation**
- Inefficient for communication system to ignore the **sense-compute-actuate cycle in many applications**



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

Semantic/Effectiveness Problems

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

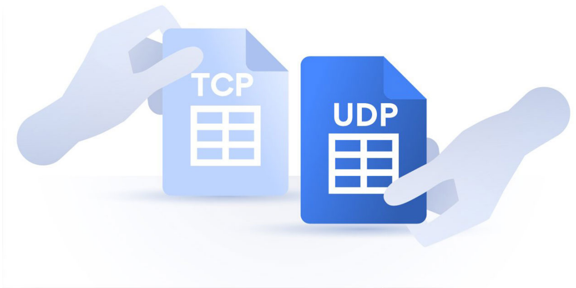
1) syntactics – the study of the methods by which signs may be combined to form compound signs,
2) semantics – the study of the signification of signs,
3) pragmatics – the study of the origins, uses, and effects of signs.

The science of relation of signs to their interpreters

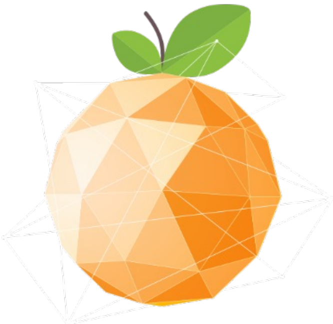
R Seising. 60 years 'A Mathematical Theory of Communication' - Towards a 'Fuzzy Information Theory', 2009.

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 !**

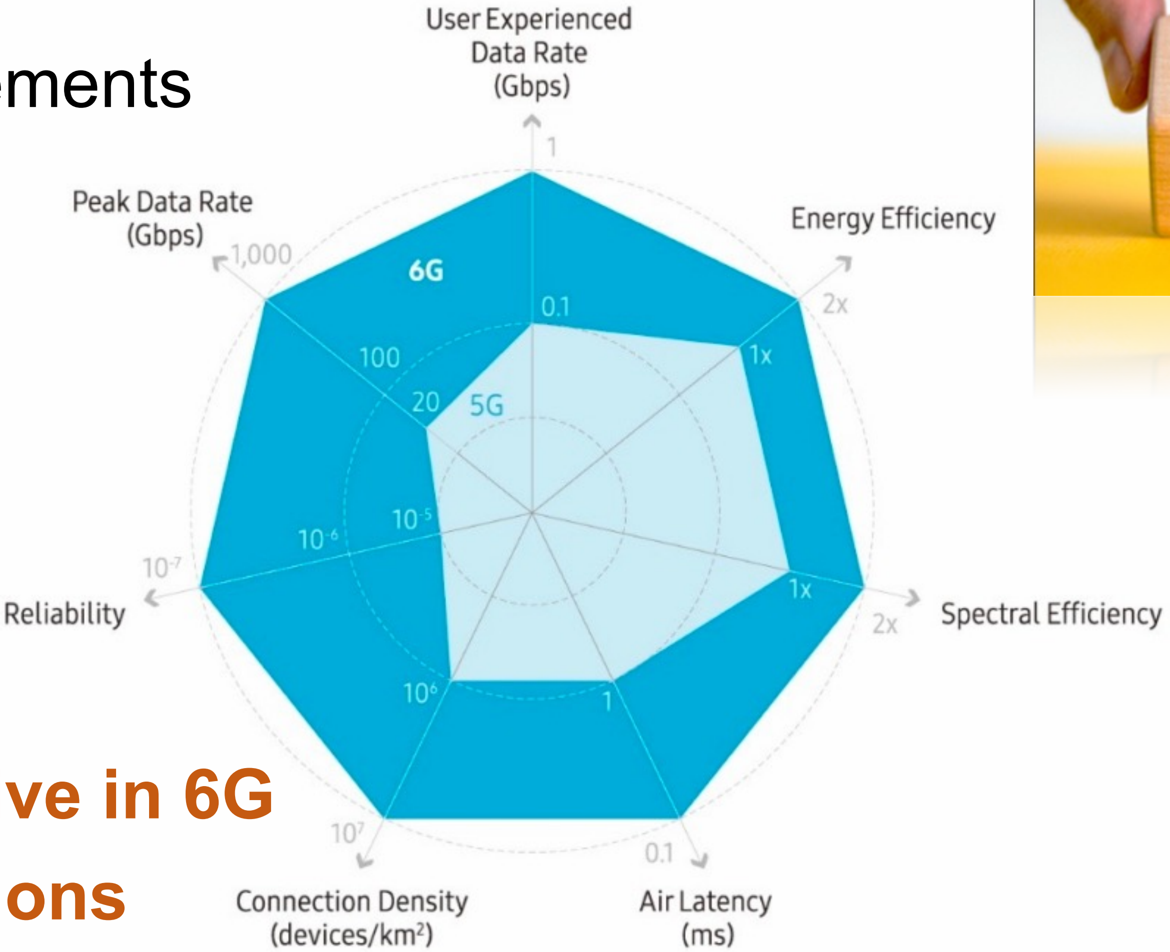
e,g. irrelevant data can be dropped from queues

- **Future networks: AI-native**

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

6G Evolution: we cannot ignore the gap

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

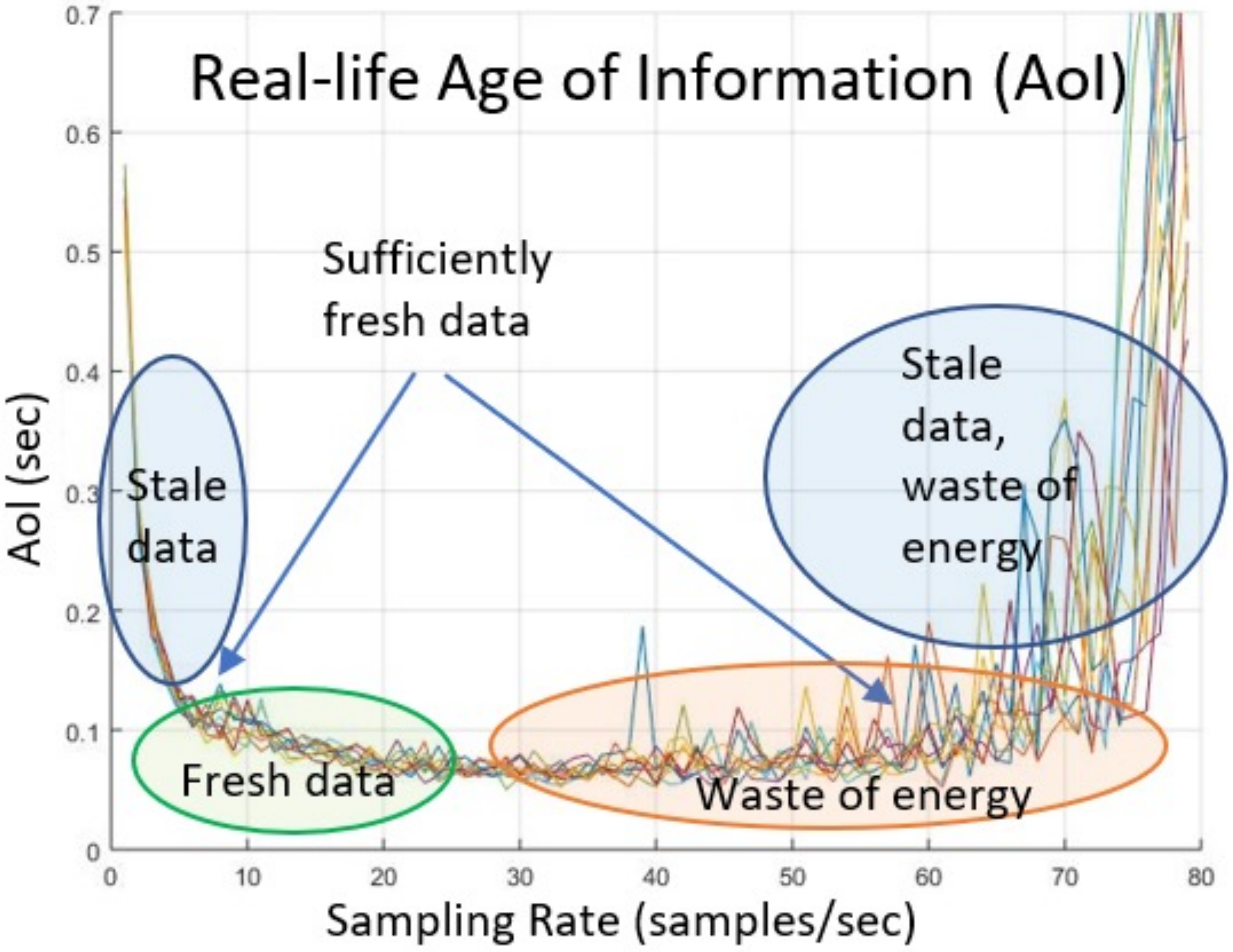


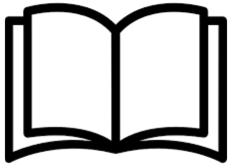
Sensing and NTN will be native in 6G

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





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]

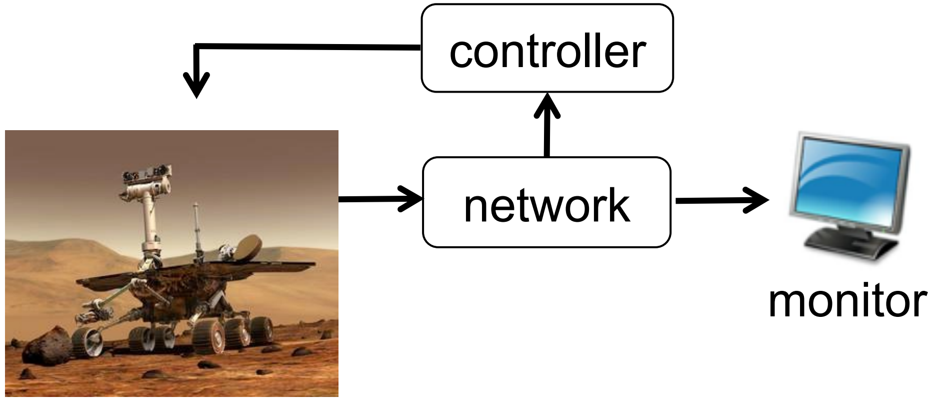
[Sonmez, Baghaee, Ergisi, Uysal 2018]

[Sert, Sönmez, Baghaee, Uysal 2018]

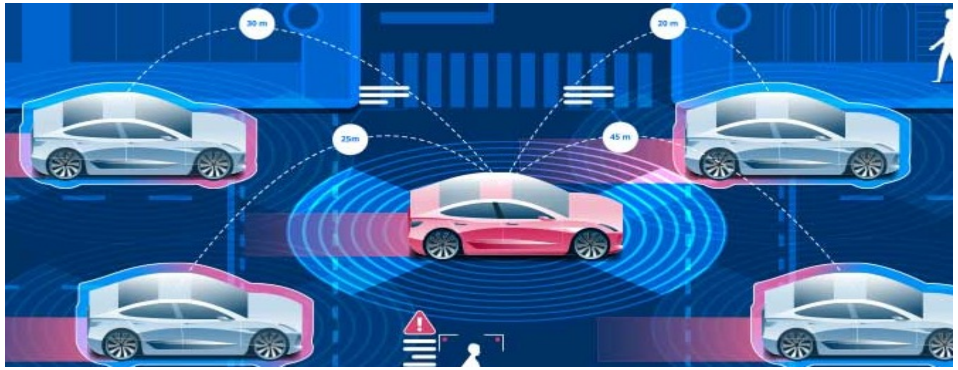
[Baghaee, Beytur, Uysal 2019]

Freshness as an effectiveness KPI for IoT

Satellite
IoT



Remote Monitoring and Control

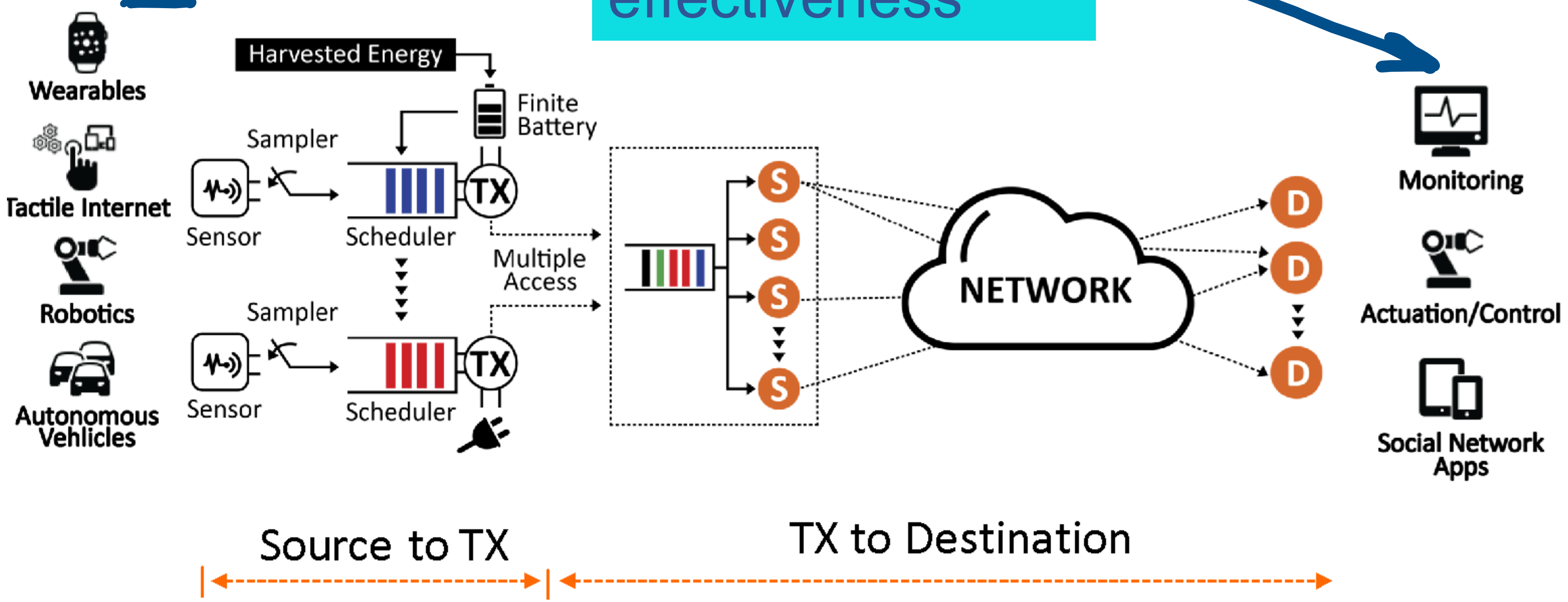


Automated Vehicles

Terrestrial
localization
and tracking

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

End-to-end effectiveness

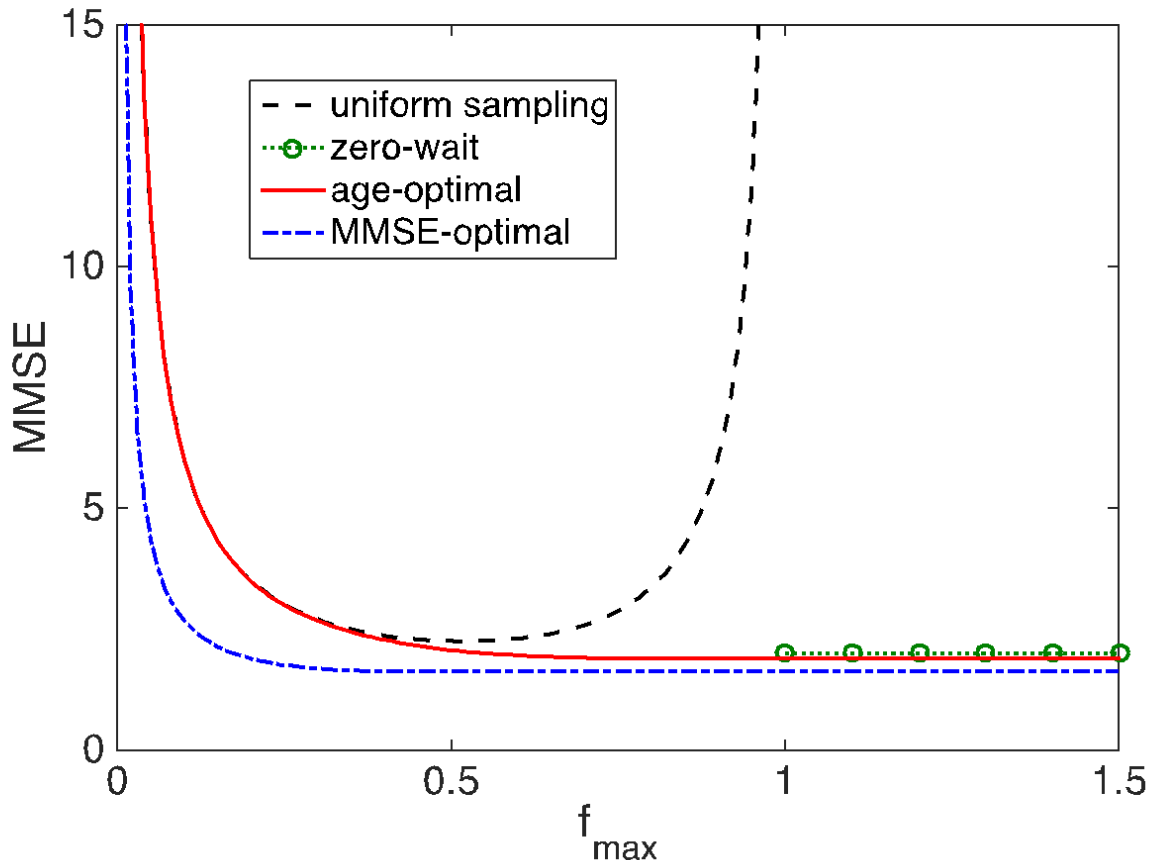


Relevance

- Send the significant bits of information

- Eg. Measurements of a process/images/video sent for remote estimation/ AI, remote training of Models, digital twins, etc

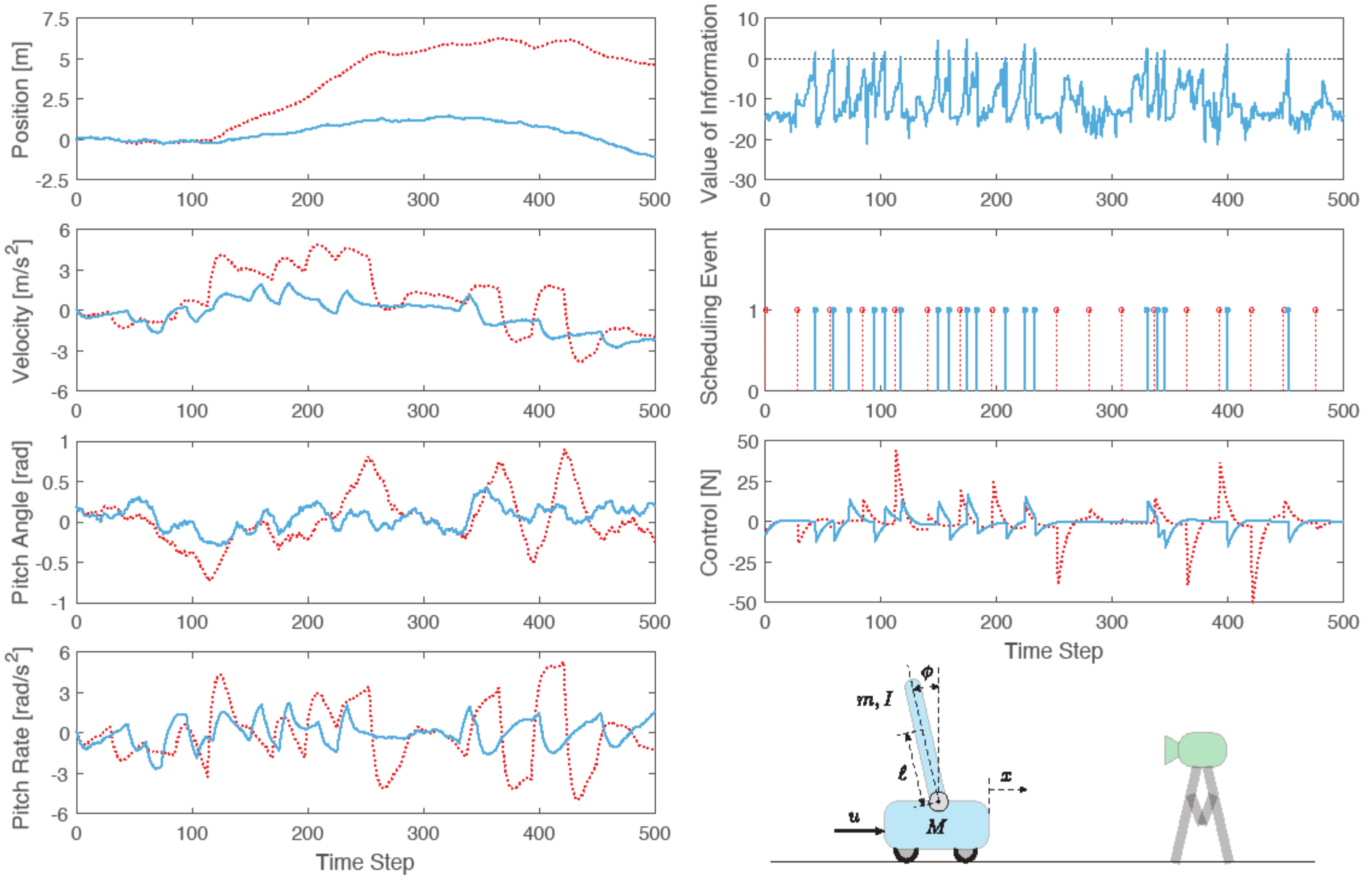
- Separate handling of sampling, encoding and transmission -> highly suboptimal
- Non-uniform/semantics aware sampling and JSCC



[Sun Polyanskiy Uysal 2019]

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.

Freshness / Relevance Age of Incorrect Information (AoII)

$$\Delta_{AoII}(t) = f(t) \times g\left(X(t), \hat{X}(t)\right)$$

time penalty information penalty [Maatouk et al 2020]

$$f_{linear}(t) = t - V(t),$$

$$f_{exponential}(t) = \exp\left(\alpha\left(t - V(t)\right)\right)$$

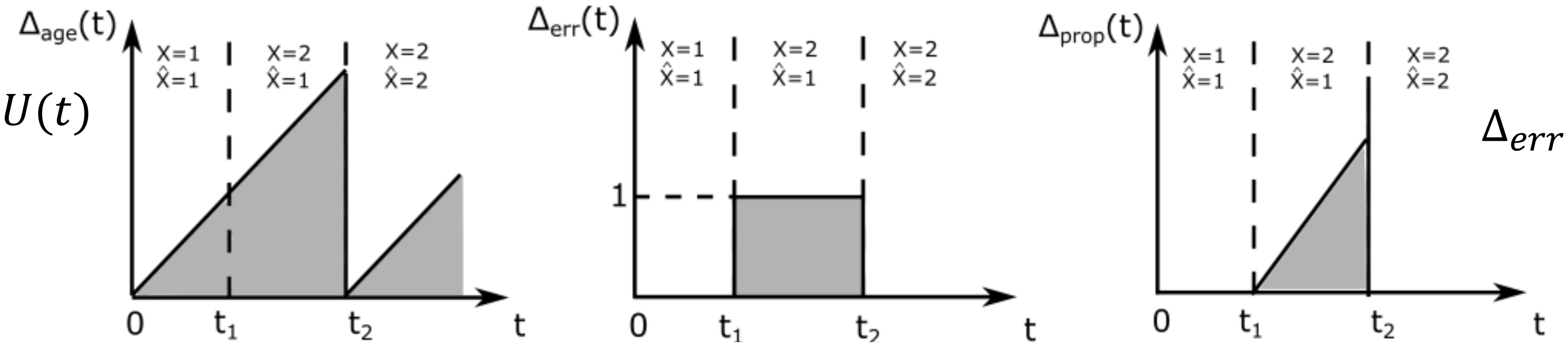
$$f_{threshold}(t) = \mathbb{1}_{\{t - V(t) \geq d\}}$$

$$g_{ind}\left(X(t), \hat{X}(t)\right) = \mathbb{1}_{\{X(t) \neq \hat{X}(t)\}}$$

$$g_{sq}\left(X(t), \hat{X}(t)\right) = \left(X(t) - \hat{X}(t)\right)^2$$

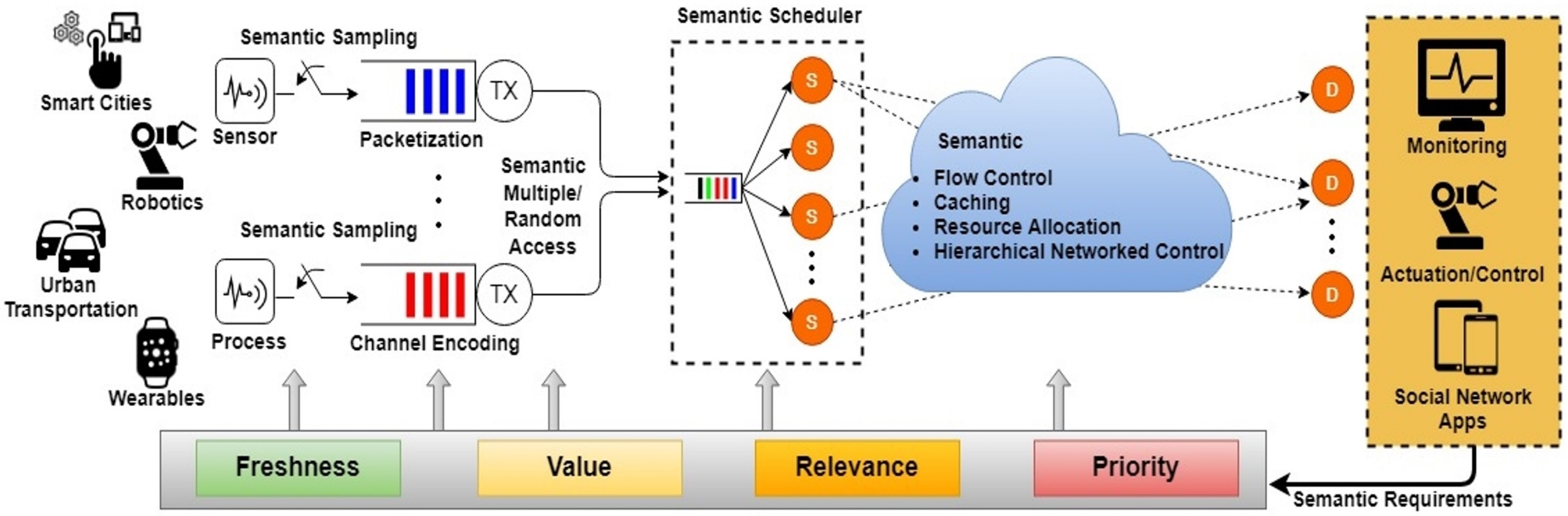
$$g_{threshold}\left(X(t), \hat{X}(t)\right) = \mathbb{1}_{|X(t) - \hat{X}(t)| \geq c}$$

$$\Delta_{age}(t) = t - U(t)$$



$$\Delta_{err}(t) = \mathbb{1}_{\{X(t) \neq \hat{X}(t)\}}$$

End-to-end Semantic Communication Architecture

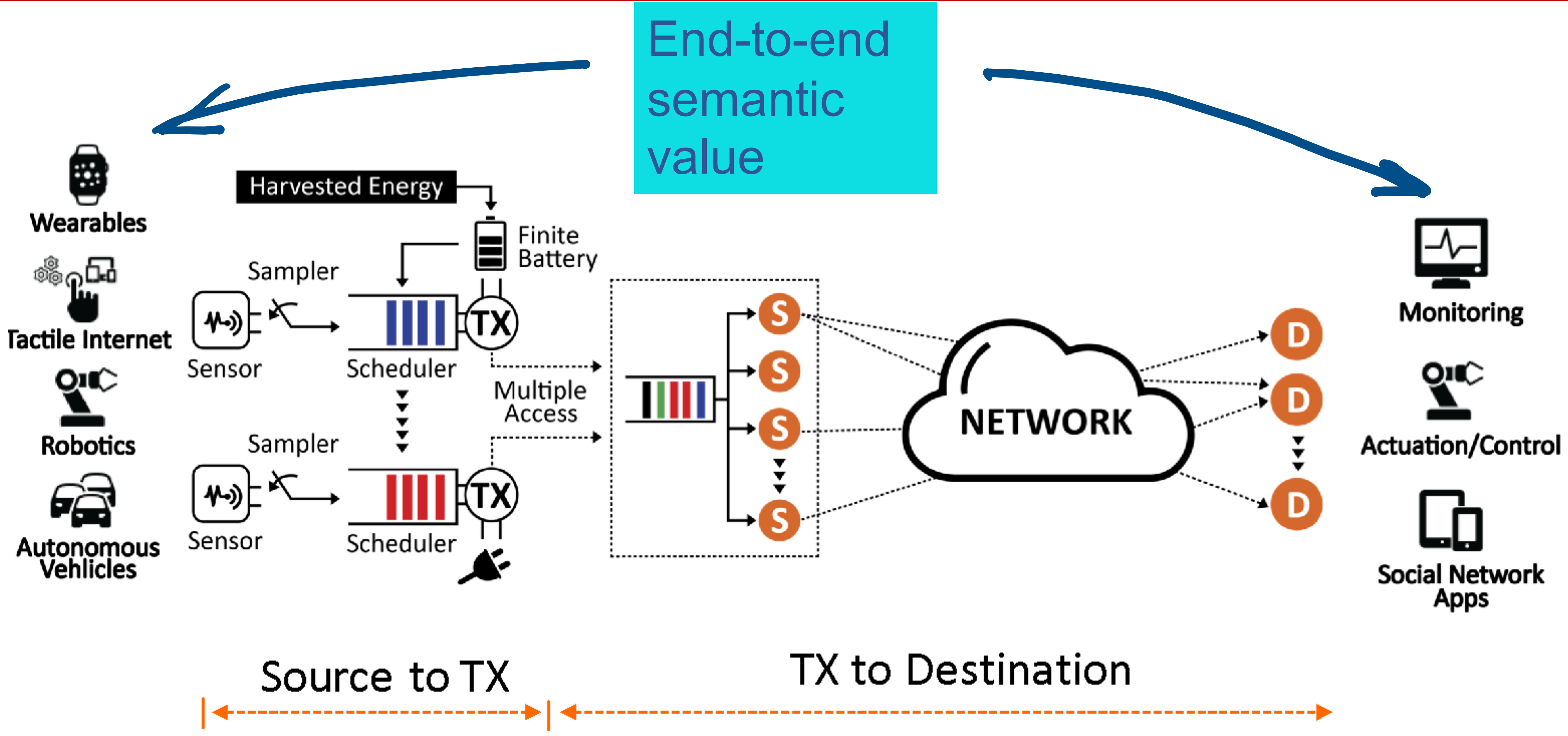


- New semantic/effectiveness measures and metrics that define them.
- Develop link, transport and application layer principles in concert to fulfil semantic-related targets
- Relax the exogeneous data arrival assumption
  - **Non-uniform process/ semantics aware sampling/JSCC**
- New communication protocol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

AI/ML Enablers of Semantic/Goal Oriented Comm

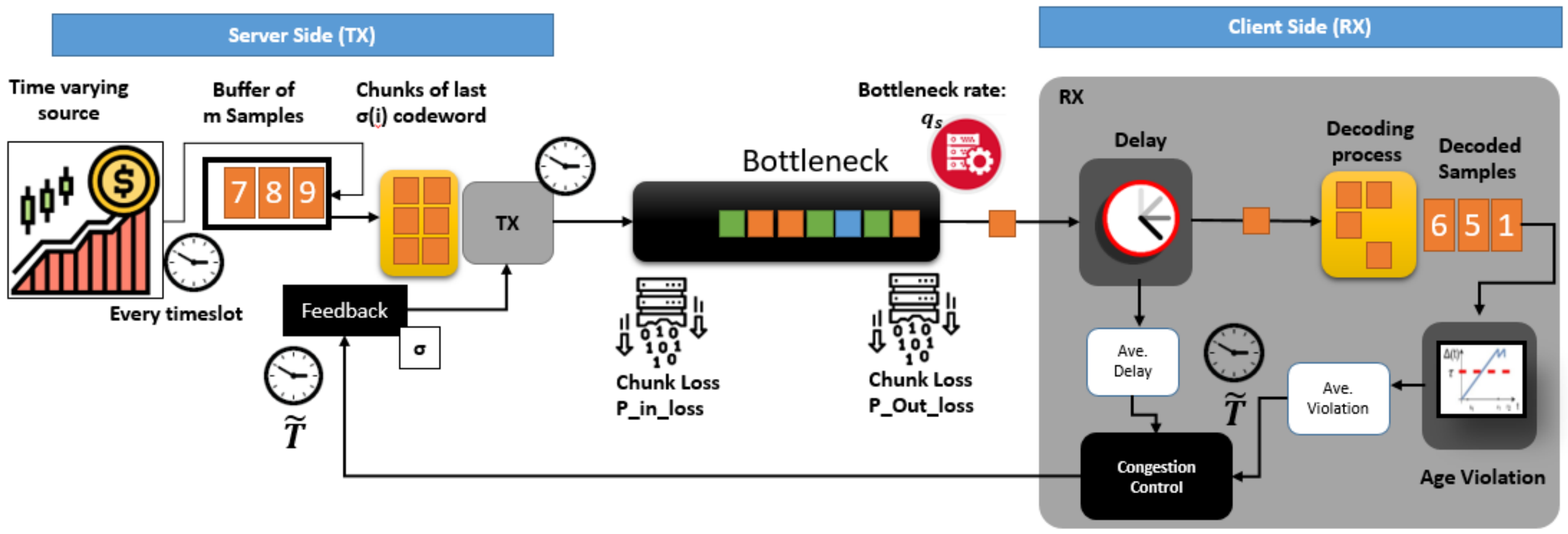
- Communications for Learning. <=> Learning for Communications
- Interesting results by Saad et al, Gunduz et al, Bennis et al, and others
- Examples: training of the digital twin of the communication system, MARL for cooperative/multiple access communication, joint optimization of federated (edge) learning and communications…

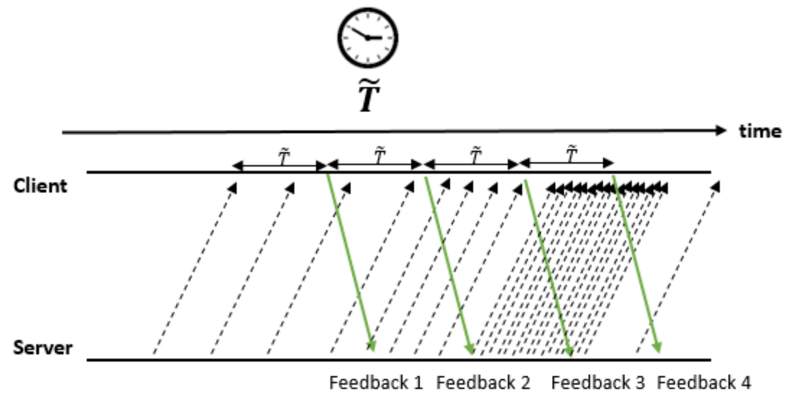
Freshness via Transport and Higher Layer Mechanisms

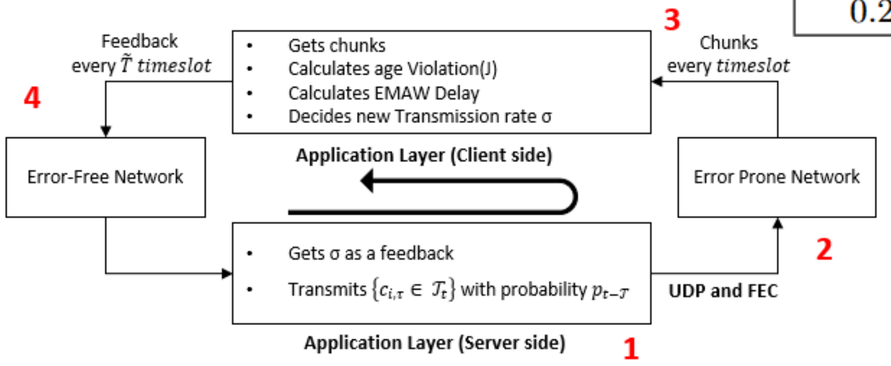


A3L–FEC protocol: Age–Aware Application Layer FEC

system model







$A^3L - FEC$ and ACP+ performance:
(Amount of time the violation happens)

- under different packet loss probabilities
- Age violation threshold equal to 5.

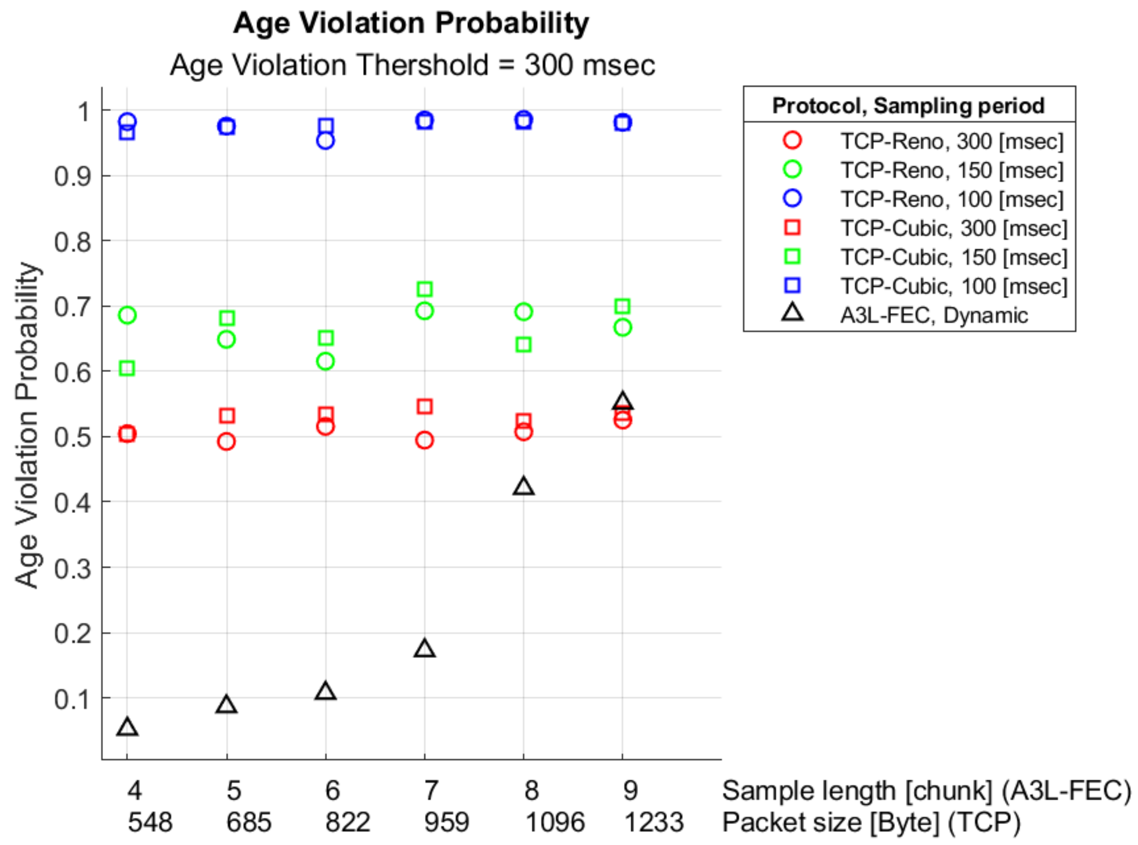
Pinloss	Poutloss	ACP+	A3L-FEC	k	n
0	0	0.66	0	3	3
0	0.1	0.02	0.0001	3	3
0	0.2	0.02	0.0067	3	3
0.1	0	0.004	0.00005	3	3
0.1	0.1	0.014	0.003	3	3
0.1	0.2	0.048	0.0037	3	4
0.2	0	0.014	0.0004	4	5
0.2	0.1	0.046	0.0028	3	4
0.2	0.2	0.11	0.0078	3	4

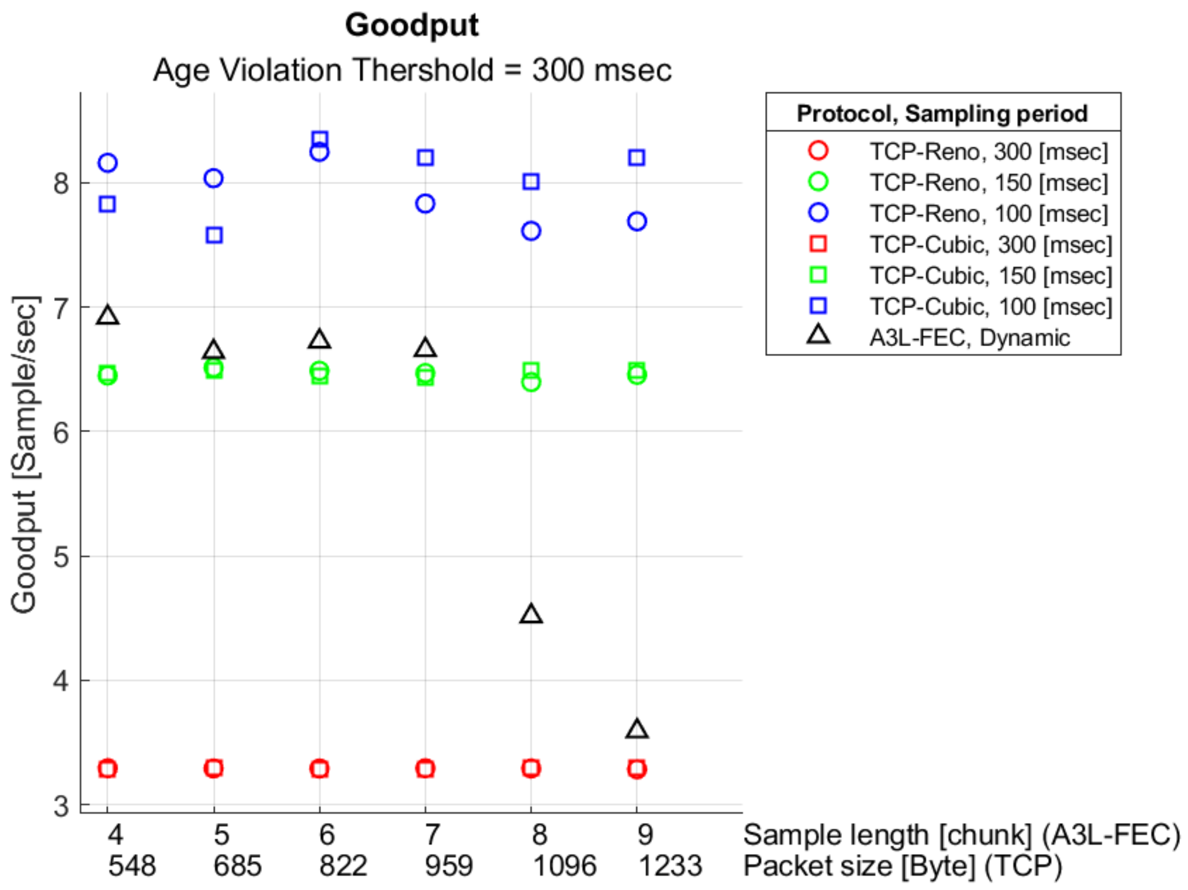
Average of 100 seeds

Patent application

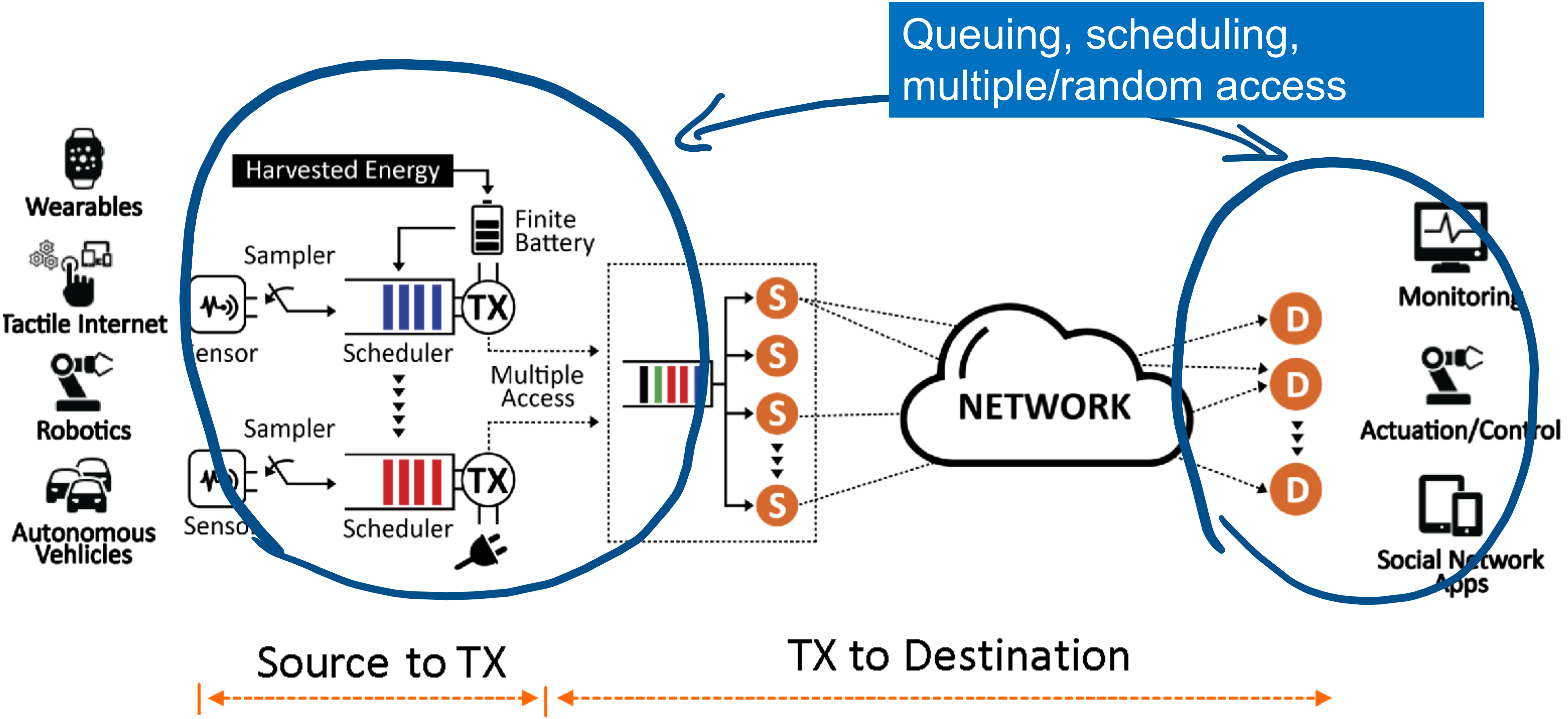
[Baghaee, Bacınoğlu , Shakiba-Herfeh, Uysal 2023]

Results: A^3L–FEC–VSVB Vs. TCP-Cubic and TCP-Reno (2000 samples)





Freshness via Link Layer Mechanisms

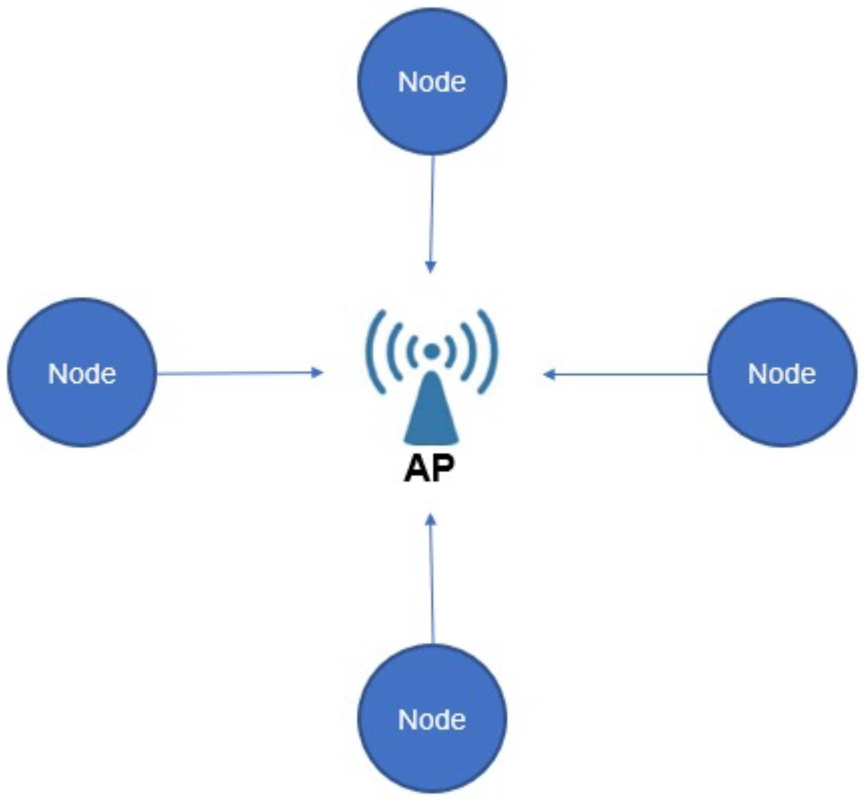


Freshness in Random Access

- IoT/MTC
  - Short Packets
- CSMA Types not suitable:
  - Significant overhead with large populations

Distributed Policy

- Generate-at-will model
- Slotted time, no collision resolution
- Each source makes independent decisions
- Sources keep track of their age



The goal is to minimize average AoI across time (symmetric users)

Threshold ALOHA [Atabay, Kaya & Uysal, Infocom AoI Wksp 2020]

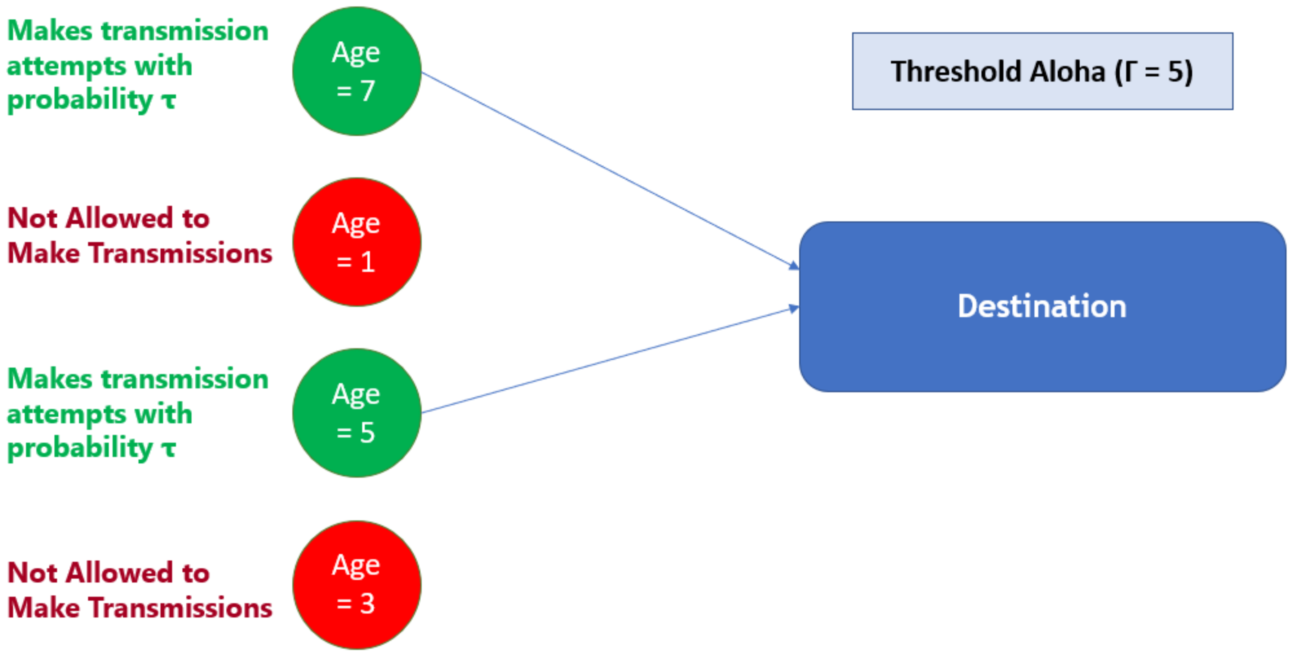
SLOTTED ALOHA

All users can transmit at any time

THRESHOLD ALOHA

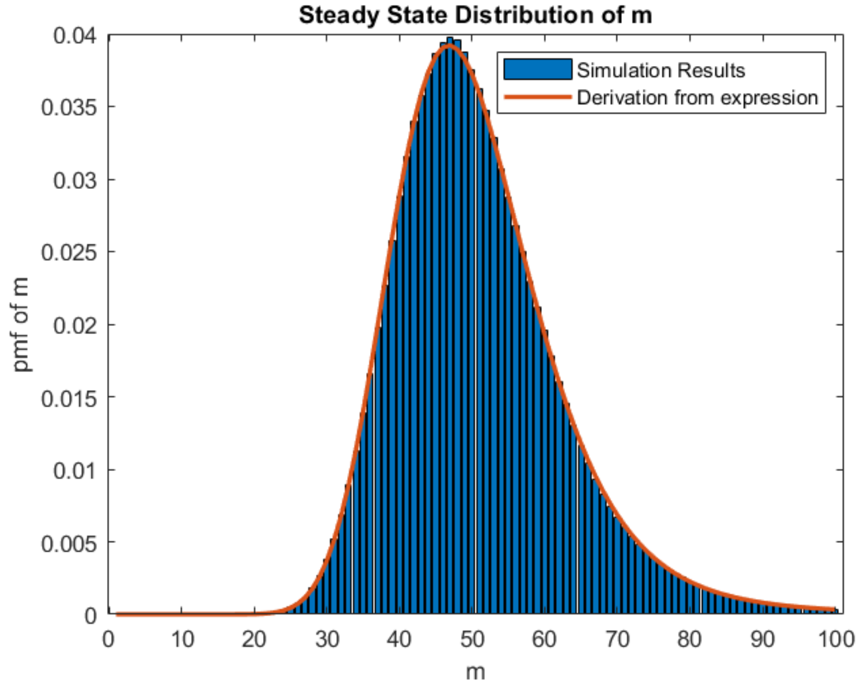
- Users idle for Γ time slots after successful transmission (Passive, sleeping) before becoming active
- Active users attempt transmission with probability τ in each time slot

Γ = 1 → Slotted ALOHA (Special Case)

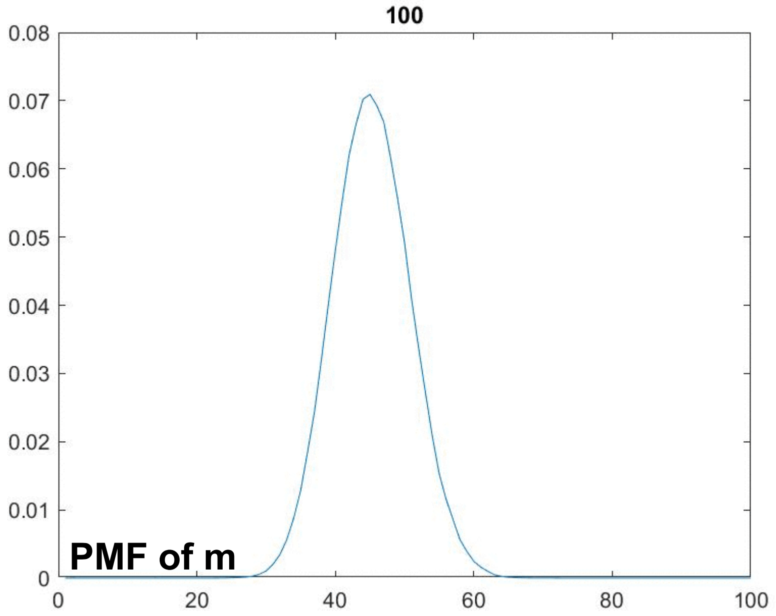


Fresh flows stay silent → Network thinning → Greater reduction of AoI

Network Thinning [Yavascan & Uysal JSAC 2020]



Number of Sources

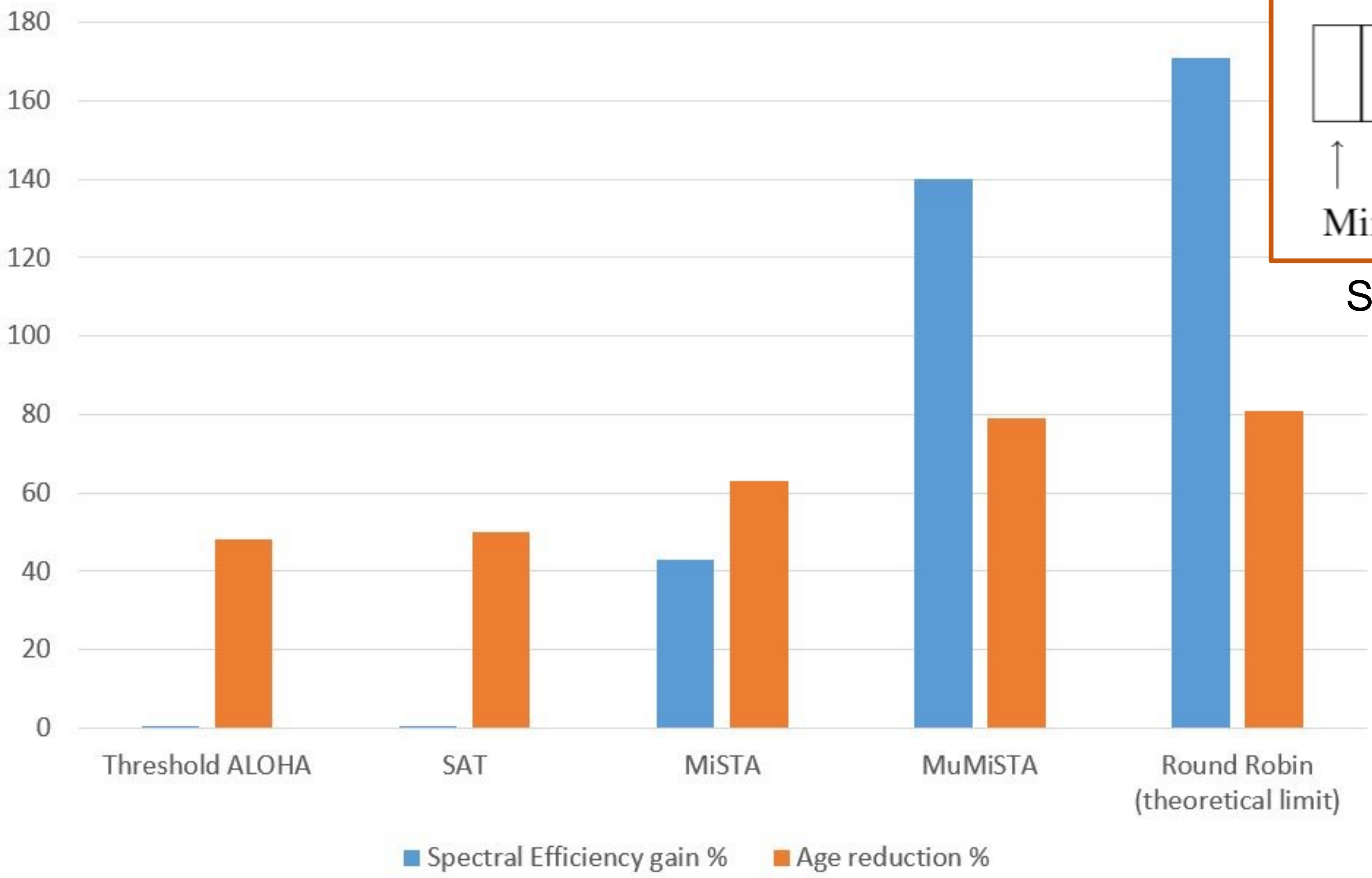


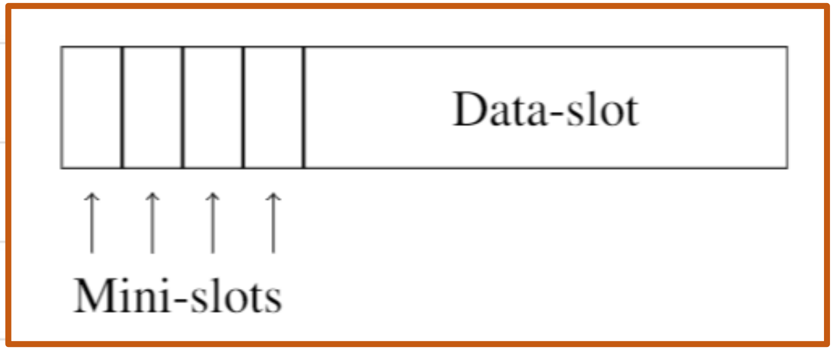
Only *20% of all users* are active at the steady state under optimal parameters and average AoI is reduced by 48%.

Steady state distribution of the number of active nodes in the network can be derived explicitly.

System converges to a slotted ALOHA with *fewer* number of users as the number of users grows.

MiSTA [Ahmetoglu, Yavascan & Uysal 2020]
MuMiSTA, patented 2021





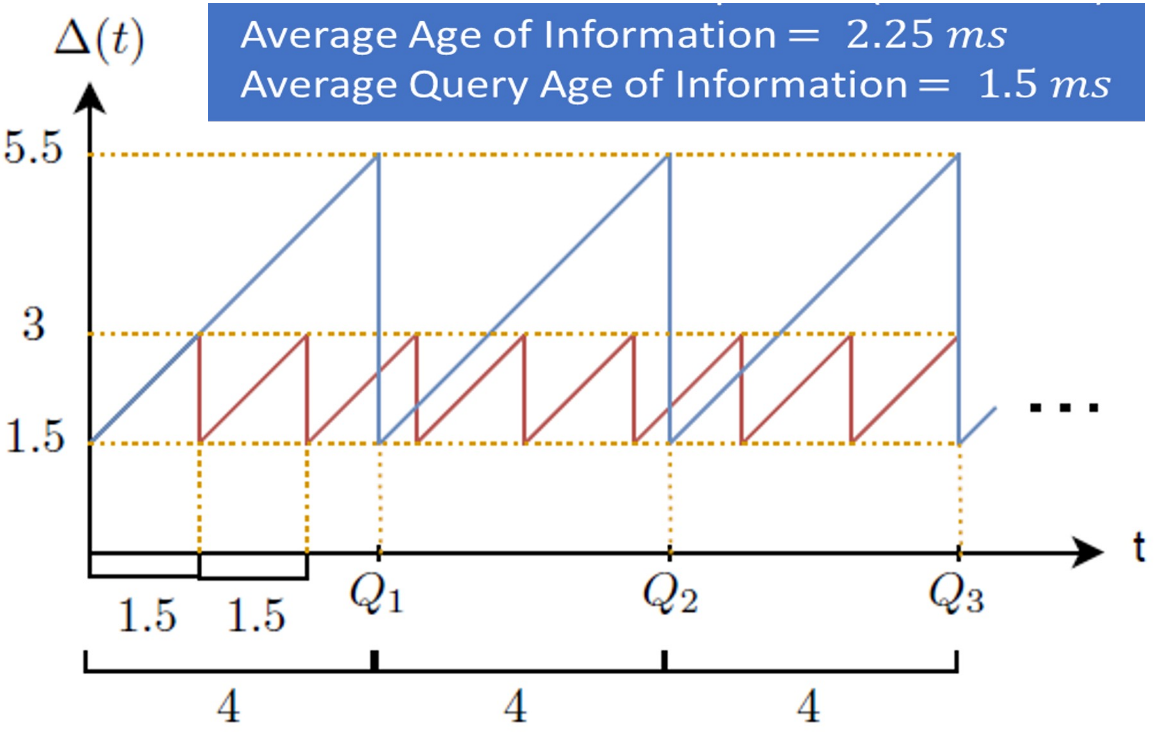
Slot Structure of MuMiSTA

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.

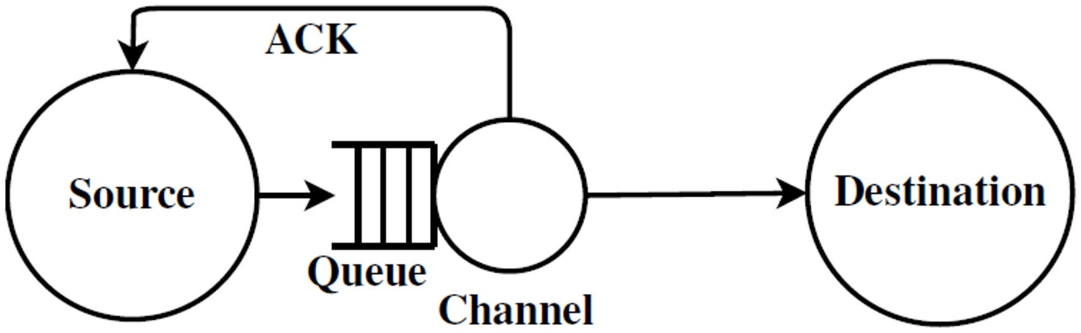
[Ildız et al 2021]

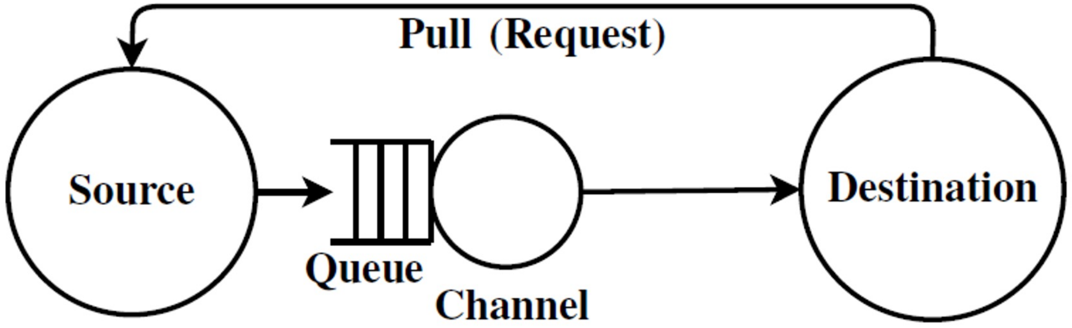


AoI Optimization
- Worse Performance
- Wasteful

QAoI Optimization
- Better freshness
- Fewer transmissions

“Update-or-Wait” vs “Pull-or-Wait”





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

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

*

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

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

- 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

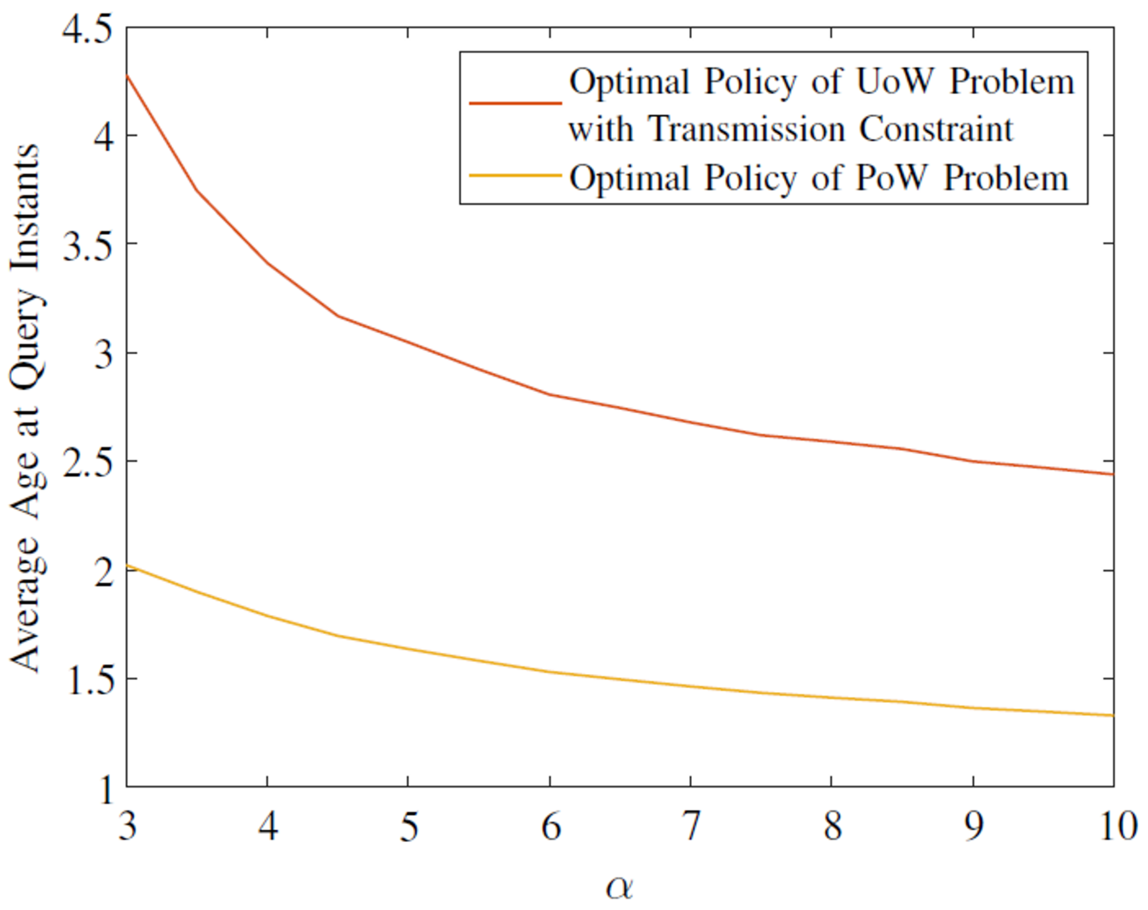
- The penalty function g is continuous.
- Y_j and Z_j is lower and upper bounded i.e.

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

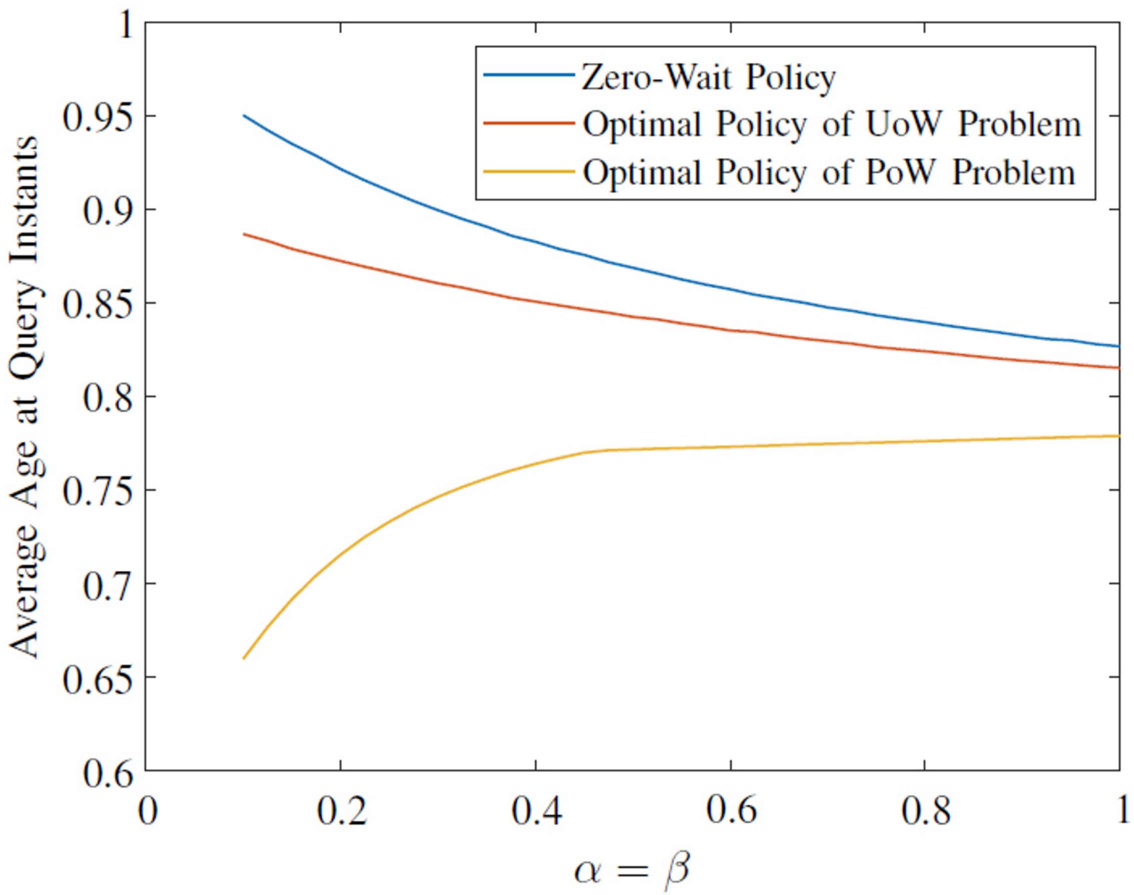
$$Z_j \in [0, M]$$

PoW vs UoW

Pareto distributed transmission delays
(IID with paremeters $\boldsymbol{x}_{\mathbf{m}} = \mathbf{1}$ and $\boldsymbol{\alpha}$)



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

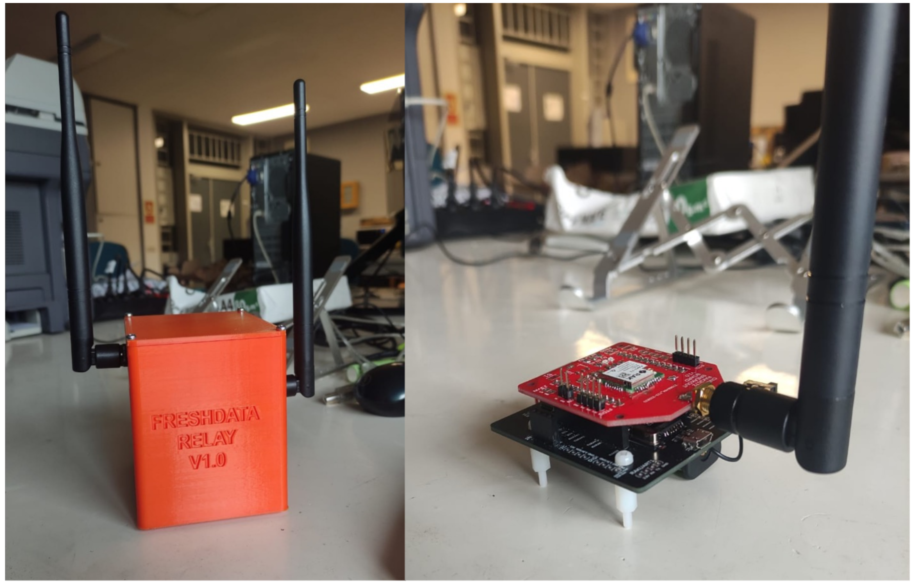


Goal Oriented Communication entering products

- New startup: FRESHDATA Technology
- Contributions to LoRaWAN standardization in progress

LoRa Alliance®

6G

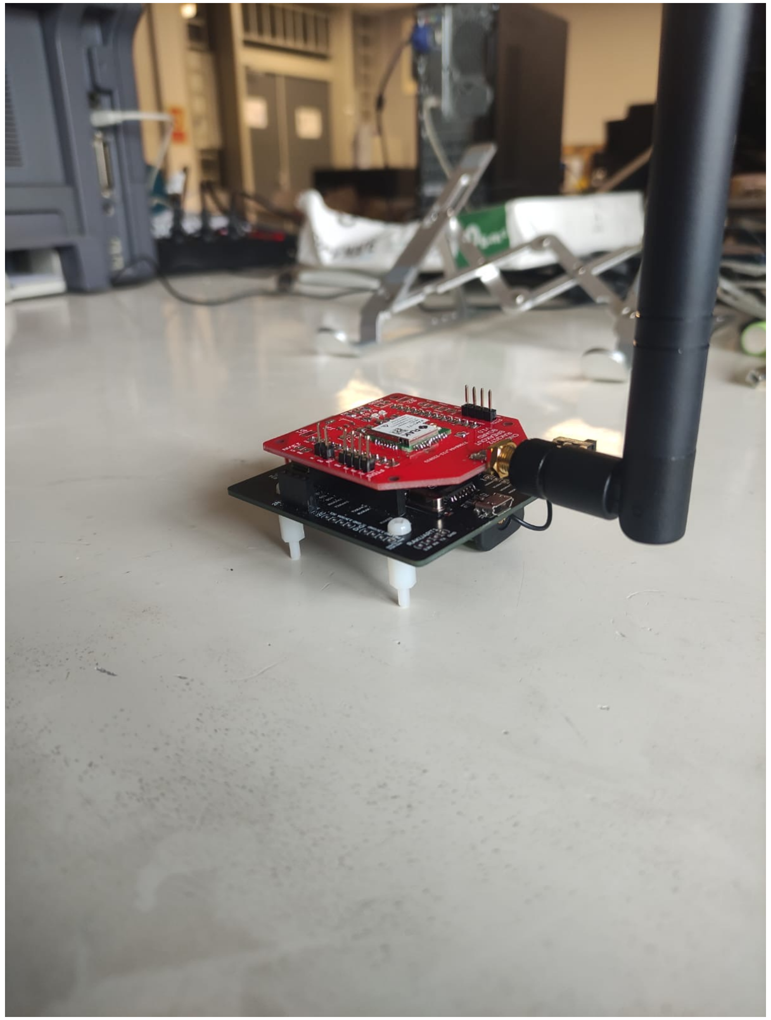


Satellite IoT

NTN

Terrestrial IoT

FRESHDATA Relay



Broader Applications

- SUIT (Sustainable Urbanization through Innovative Technologies)

- Consortium of universities, research labs, companies

- 11 projects – all spinning off from FRESH-IoT

Thank you, from the CNG team

Est. 2007

Currently:

2 Faculty members; 2 PhD, 8 MS, 13 Undergraduate (STAR) students; 34+ Alumni