INDUSTRIAL IOT IN 5G NETWORKS AND BEYOND:
FROM URLLC TO HAPTIC COMMUNICATIONS

Salah El Ayoubi

Professor at CentraleSupélec – Université Paris Saclay

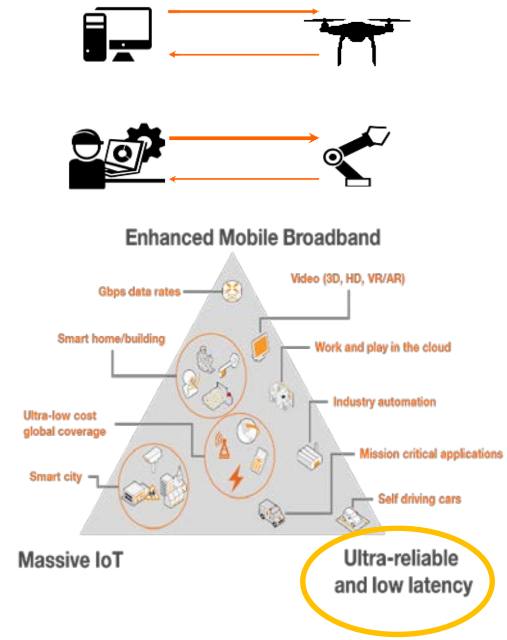
Head of ILOCOS research team, Laboratory of Signals and Systems
(Information, Learning, Optimization and COmmunication Sciences)

Holder of the Sustainable 6G chair funded by Orange
salaheddine.elayoubi@centralesupelec.fr

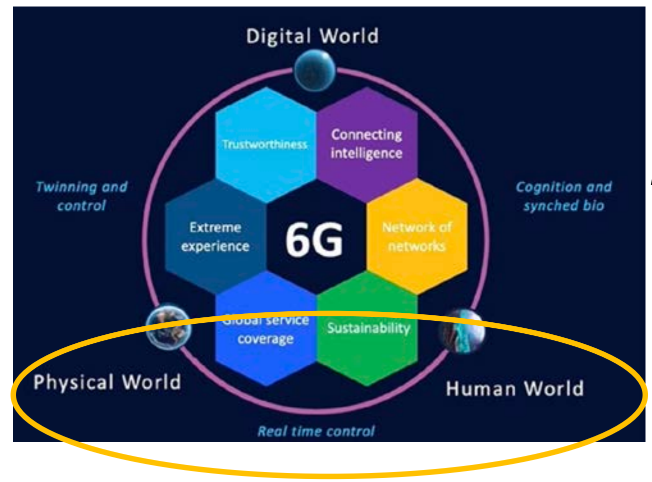
13ème atelier en évaluation des performances – IRIT– December 2024

Industrial IoT in 5G/6G

- **IoT networks allow more than low rate sensor connectivity**
- **Applications requiring reliability belong to the IIoT (Industrial IoT) world:**
  - machines communicating in a factory,
  - tele-operation of drones, etc.
- **5G URLLC service intends to serve IIoT:**
  - 1 ms latency, 99.999% reliability
- **6G networks aim to bridge the physical, digital and human worlds:**
  - 0.1ms latency, 99.99999% reliability??
- **Objective of this talk:**
  - See how these targets can be achieved
  - explore a more sustainable, approach
  - 6G focus: haptic service for real-time control



Source: Orange



Source: HEXA-X

Outline

- **Network optimization for URLLC**
  - Resource dimensioning by queuing theoretical models
  - URLLC evolution with 6G

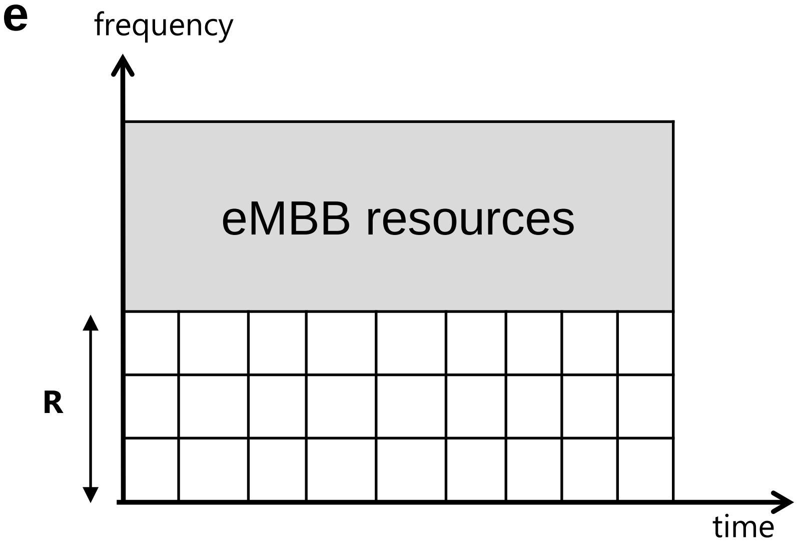
- **From URLLC to haptic communications**
  - Tactile Internet with haptic feedback
  - Joint control and communications

Outline

- **Network optimization for URLLC**
  - Resource dimensioning by queuing theoretical models
  - URLLC evolution with 6G
- **From URLLC to haptic communications**
  - Tactile Internet with haptic feedback
  - Joint control and communications

Resource allocation for URLLC traffic

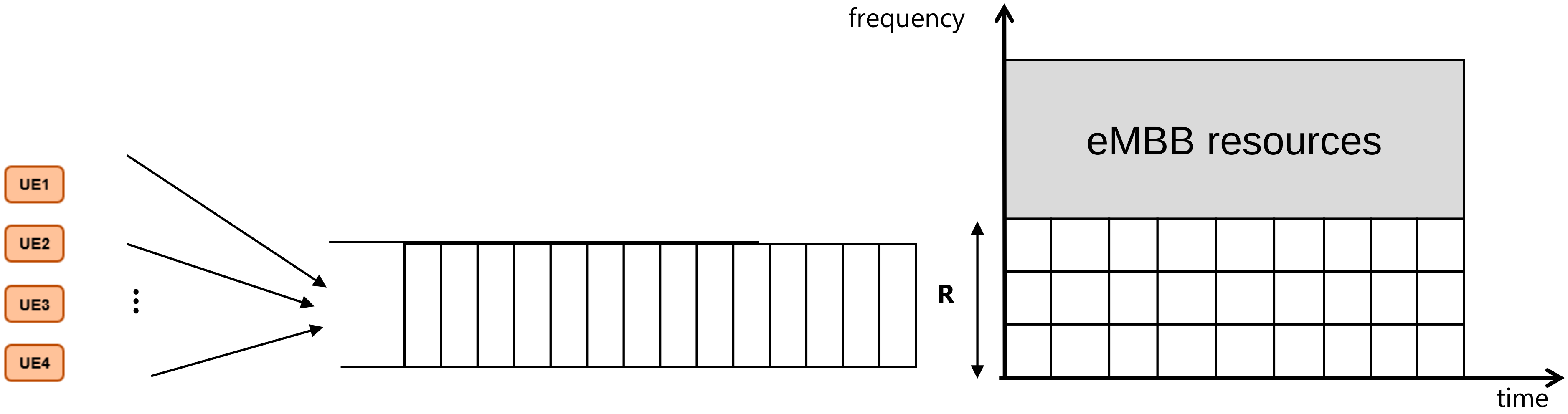
- **Illustration for the 5G time/frequency frame**
  - Spectral dimension:
    - *Subcarrier of 30 KHz*
    - *packets of 320 bytes*
    - *QPSK ½ modulation (1 bit/symbol)*
    - *1 packet occupies 80 subcarriers=2.4 MHz*
  - Time constraint:
    - *1 slot=0.144ms*
    - *Target delay=1 ms*
    - *1 slot for alignment and 1 slot for processing*
    - *the packet has a queuing budget of δ=4 slots*



- **Scenario: large number of users, sporadic traffic**
- **Stringent quality of service (QoS) guarantee:**
  - 3GPP target: more than 99,999% of packets correctly received within 1 ms.
- **Focus of the standard: grant-free scheduling, preemption, early decoding**
  - for ensuring that a packet can be correctly received and decoded within 1 ms.
- **But this does not ensure that the delay requirement is ensured:**
  - Queuing delay should be modeled.

Centralized scheduling approach (downlink)

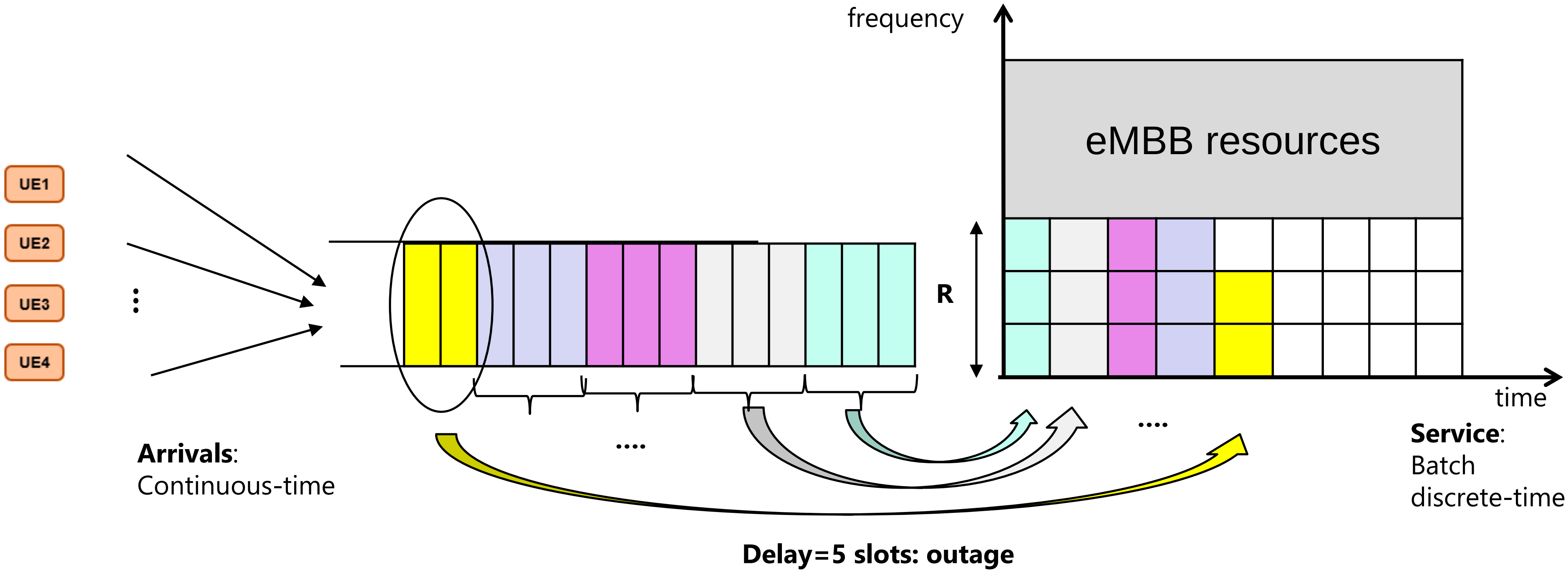
- **We consider a classical grant-based scheduling approach**



Arrivals:
Continuous-time

Centralized scheduling approach

- **We consider a classical grant-based scheduling approach**



- **Objective: ensure that the probability of delay exceeding threshold is low**

Formulating the outage probability

- ***U*** **users, probability** ***f*** **of generating a packet in a given slot**
- **Users have different radio conditions, and different spectral efficiencies:**
  - With probability β_k, a packet consumes α_k resources
- **Number of resources requested by packets generated in slot** ***i*****:**

$$a(i) = \sum_{u=1}^{U} X_{u,i} \quad \text{, with} \quad X_{u,i} = \begin{cases} 0, & \text{with prob. } (1-f) \\ \alpha_k & \text{with prob. } f\beta_k \end{cases}$$

- **Define the overflow at time slot** ***i*** **as the amount of resource demands that cannot be satisfied in the slot, and must be queued:**

$$B(i) = (a(i) + B(i-1) - R)^+$$

- **Outage: current overflow cannot be served in the next (δ-1) slots:**

$$O = \lim_{i \to \infty} Pr[B(i) > (\delta - 1)R]$$

Queuing model – M/D/C equivalent in discrete time

- **Example: packet of size *s*, belongs to a user whose MCS efficiency is e bit/s/Hz, time slot of τ milliseconds, RB of w Hz:**
  - one packet occupies $r = \left\lceil \frac{s}{e\tau\omega} \right\rceil$ Resource Blocks
- **If there are *R* RBs at the base station, each slot can serve:**
  - $C = \left\lfloor \frac{R}{\left\lceil \frac{s}{e\tau\omega} \right\rceil} \right\rfloor$ packets
- **For Poisson arrivals with intensity λ packets/slot, the equivalent continuous time queue model is M/D/C:**
  - Efficient algorithms known for computing steady-state (Tjims, 2003).
- **Two drawbacks:**
  - The modulation scheme depends on the radio conditions, and there is a mix of modulations over the cell: M/D/C where *C* varies with time depending on radio conditions (efficiencies *e*)
  - The system is discrete time, as service is done slot per slot.

A discrete-time low complexity queuing model

- **We developed a discrete-time queuing model for the system:**
  - q_b is the probability that the queue length is equal to *b* (RBs)
  - Let z_i be the limiting probability that new packets request *a(t)* = *i* RBs
  - The transition matrix for the system is derived by:

$$Q_{jb} = \begin{cases} z_{b+R-j}, & \text{if } b \in]0, B_{max}[\\ \sum_{i \geq b+R-j} z_i, & \text{if } b = B_{max} \\ \sum_{i \leq R-j} z_i, & \text{if } b = 0 \\ 0, & \text{otherwise} \end{cases}$$

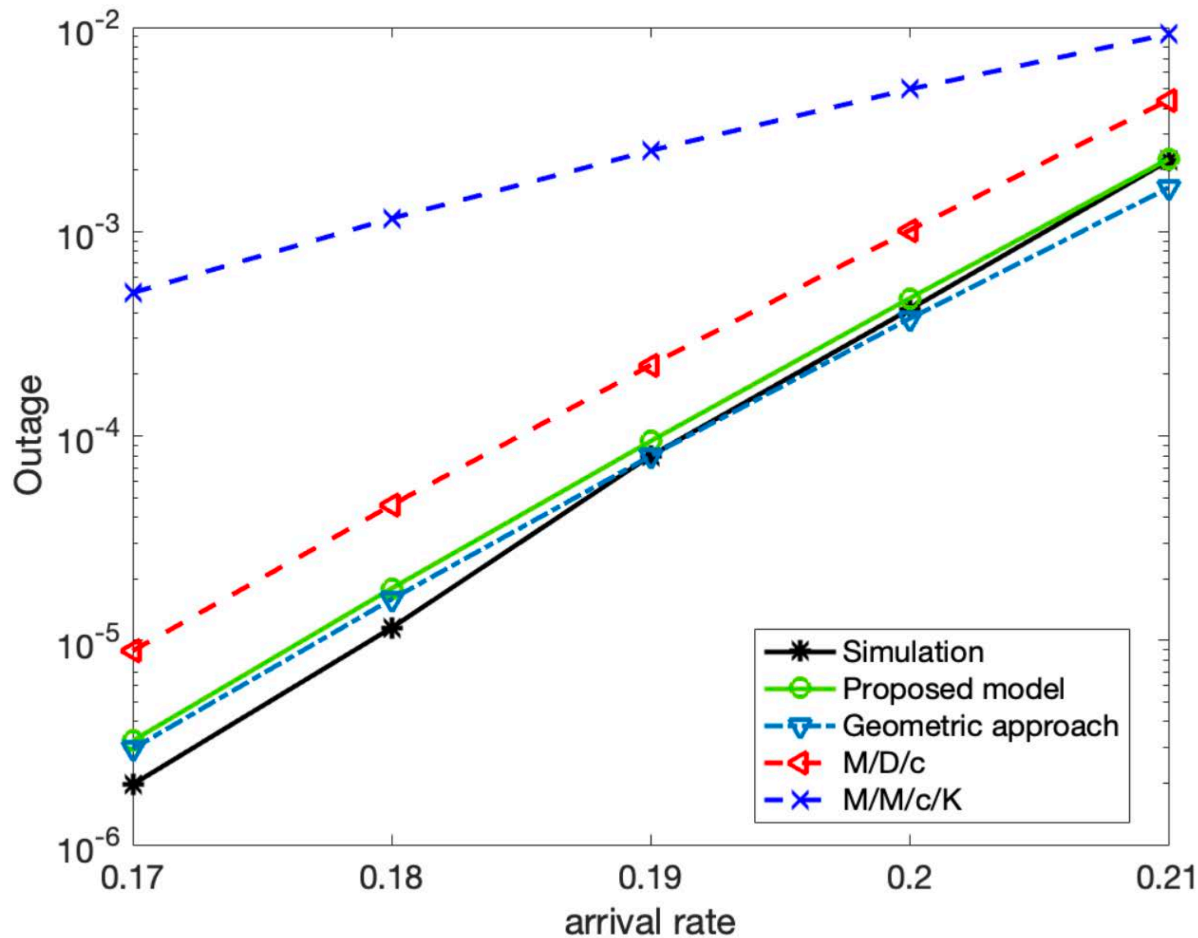
- **Theorem: For binomial and Poisson arrivals, the equilibrium probabilities exhibit the geometric tail behavior:**

$$q_j \sim \gamma \eta^j \;\; as \;\; j \to \infty$$

for some constant $\gamma > 0$ and $0 < \eta < 1$. For sufficiently large M:

$$q_j = q_M \eta^{j-M}, \;\; j \geq M.$$

Results: outage probability



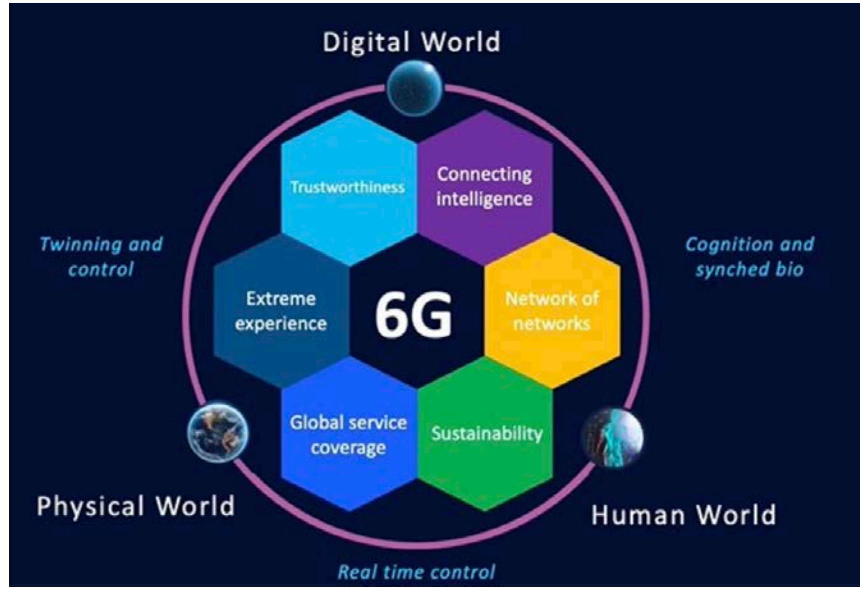
Outline

- **Network optimization for URLLC**
  - Resource dimensioning by queuing theoretical models
  - URLLC evolution with 6G

- **From URLLC to haptic communications**
  - Tactile Internet with haptic feedback
  - Joint control and communications

Extreme URLLC for 6G

- **URLLC becomes xURLLC for extreme URLLC**
- **0.1 ms latency target**
- ***10^{-6}* to *10^{-8}* loss rate**
- **No waiting**
- **No queuing**
- **One-shot transmission**



- **Optimization under chance constraint:**
  - Minimize R
  - s.t. $\Pr\left[\sum_{u=1}^{U} X_u > R\right] < \varepsilon$
- **X_u is the amount of resources requested by user *u* at a given slot**
  - i.i.d. random variables

Chebychev bound formulation

$$\Pr\left[\sum_{u=1}^{U} X_u > R\right] < \varepsilon$$

- **Simplest approach: use Bienaymé- Chebychev bound**
- **Step 1: compute the moments for the amount of packets generated by a single user in a slot *i*:**

$$X_{u,i} = \begin{cases} 0, & \text{with prob. } (1-f) \\ \alpha_k & \text{with prob. } f\beta_k \end{cases} \Rightarrow \begin{array}{l} \mu_0 = E[X_{u,i}] = f\sum_k \alpha_k\beta_k \\ \sigma_0^2 = E[X_{u,i}^2] - \mu_0^2 = f\sum_k \alpha_k\beta_k^2 - f^2(\sum_k \alpha_k\beta_k)^2 \end{array}$$

- **Step 2: compute the moments of the accumulated traffic:**

$$\hat{\sigma} = \sqrt{U}\sigma_0, \hat{\mu} = U\mu_0$$

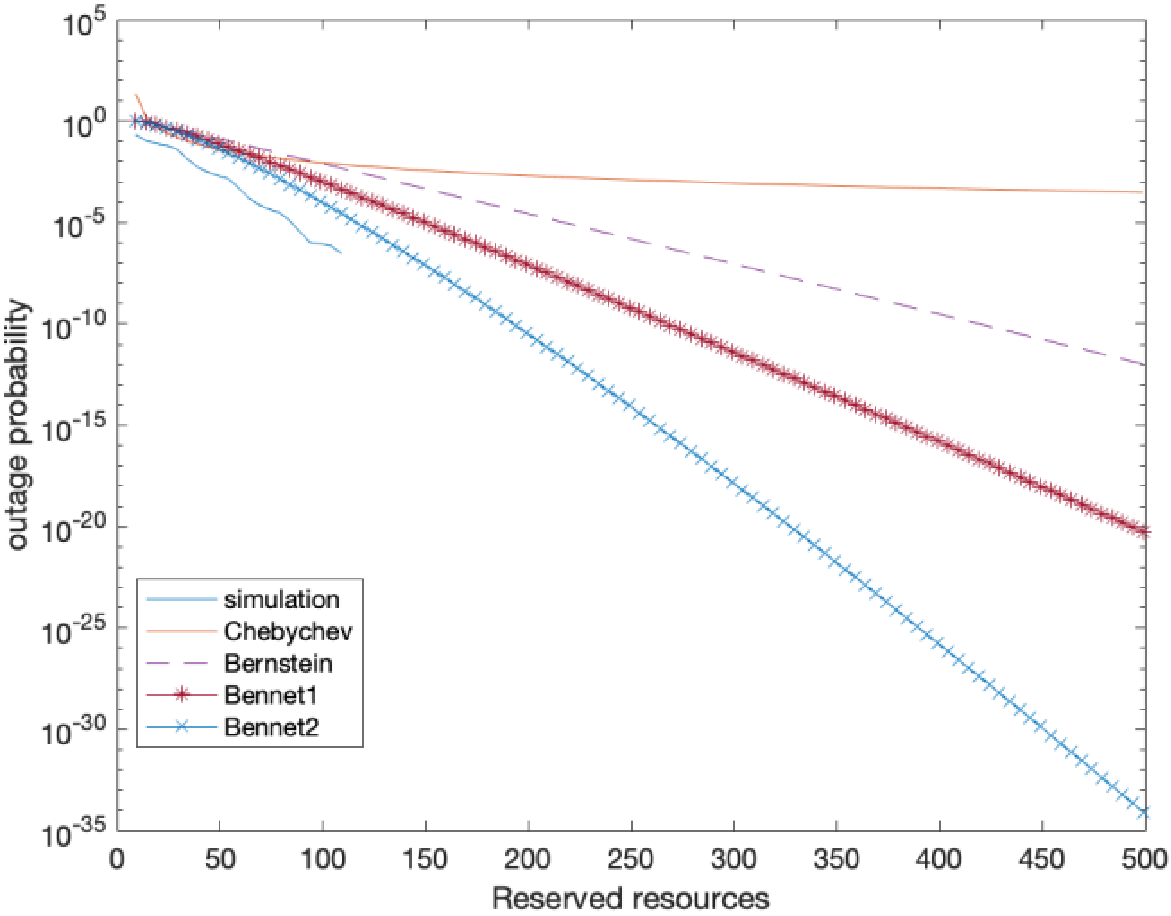
- **Apply the Bienaymé- Chebychev bound:**
- $\Pr\left[\sum_{u=1}^{U} X_u > \hat{\sigma}s\right] < \frac{1}{s^2}$ with $s = \frac{R - U\mu_0}{\hat{\sigma}}$

Developing tighter bounds

$$\Pr\left[\sum_{u=1}^{U} X_u > R\right] < \varepsilon$$

- **Bienaymé- Chebychev bound is known to be loose**
- **Idea: exploit the fact that the accumulated traffic is the sum of independent random variables**
- **Some works propose tighter bounds in this case:**
  - Bernstein, S.: Theory of probability. Moscow. MR0169758 (1927)
  - Bennett, G.: Probability inequalities for the sum of independent random variables. Journal of the American Statistical Association 57(297), 33–45 (1962)
- **In summary: for independent random variables x_u bounded by a value M:**
  - Bernstein: $$Pr(\sum_{u=1}^{U} x_u > s\sigma) \leq \exp\left[-\frac{s^2}{2 + \frac{2}{3}\frac{M}{\sigma}s}\right]$$
  - Bennet: $$Pr(\sum_{u=1}^{U} x_u > s\sigma) \leq \exp\left[-\frac{s^2}{1 + \frac{1}{3}\frac{M}{\sigma}s + \sqrt{1 + \frac{2}{3}\frac{M}{\sigma}s}}\right]$$

Performance results (Binomial arrivals)



Outline

- **5G network optimization for URLLC**
  - Resource dimensioning by queuing theoretical models
  - URLLC evolution with 6G

- **From URLLC to haptic communications**
  - Tactile Internet with haptic feedback
  - Joint control and communications

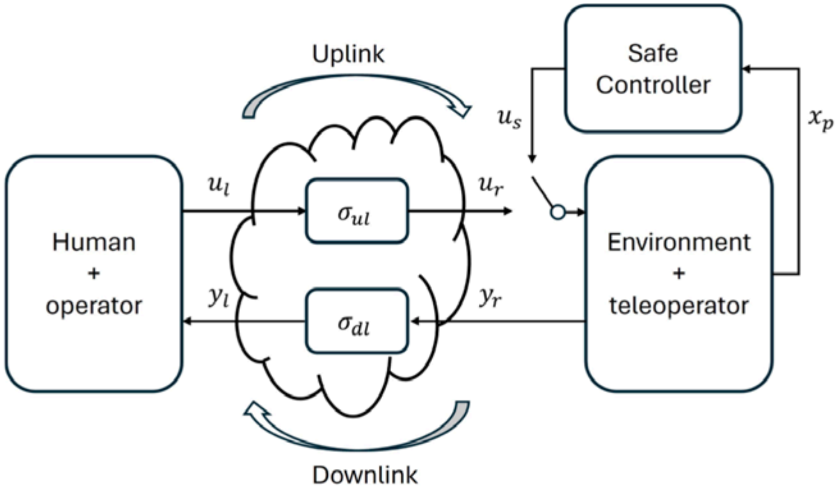
We have solved the probem, but is this THE problem to solve?

- **Back to the starting point: we solved a problem defined by 3GPP, the organism that standardizes 4G/5G and 6G…**
  - the **proportion** of packets,
  - **correctly received** by the controller
  - within the **delay budget** (e.g. 1 ms for 5G, 0.1 ms for 6G)
  - has to be larger than a **reliabilty target** (e.g. loss probability < 10^{-5}, or 10^{-7} for 6G)

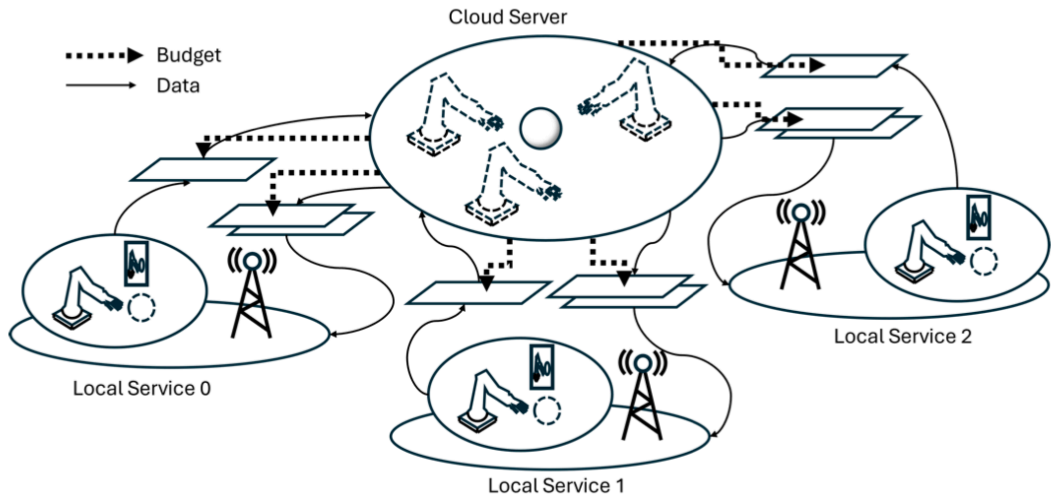
- **But why 1ms? or 0.1ms?**
- **Is it only a matter of delay, or throughput also matters?**
- **What happens if some packets are lost? Can the application compensate for it?**

Remote control with haptic feedback: a challenging use case

- **Remote control with haptic feedback over wireless networks**



A human controlling a machine
Example: remote surgery

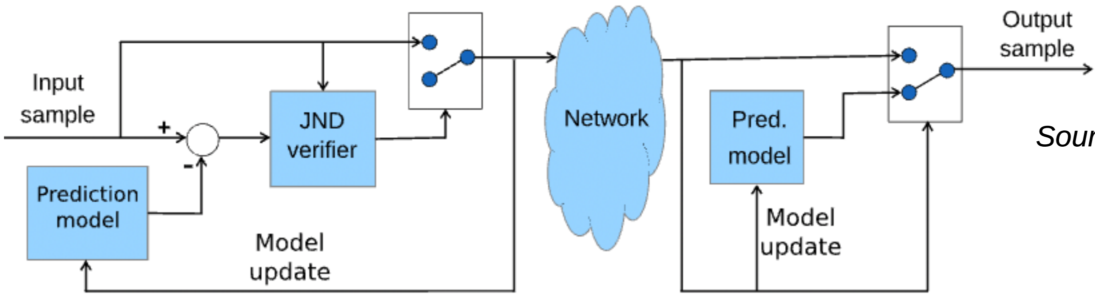


A cloud-based virtual remote touch
Example: commercial product examination

  - **Needs URLLC-like**: errors and delays may have safety impacts
  - **May be multi-modal**: haptic, but also visual and audio
  - **Should be optimized**: continuously sending packets may be too energy and resource consuming

Modeling axis 1: JND-based codec and consecutive losses

- **Just-Noticeable Difference (JND) codec is standardized:**
  - JND is the minimum amount of change in stimulus intensity needed for a perceptible increment in sensory experience
- **JND codecs are generally associated with predictors**



Source: Antonakoglou et al., 2018

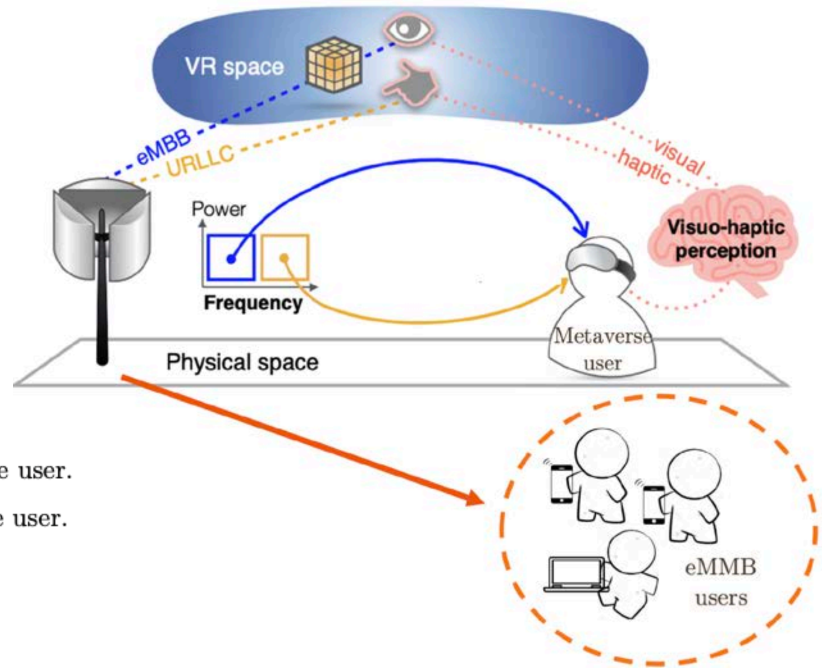
  - the predictor is able to compensate some losses
  - what is really annoying is **consecutive losses**
  - **No need for 0.1 ms and 10^{-7} loss probability.**
- **Ongoing work:**
  - incorporate the JND codec into the queuing model and compute the consecutive losses as a performance indicator
  - Develop adapted network control schemes

Modeling axis 2: multi-modal haptic-visual communications

- **Haptic communications are usually coupled with visual flows:**
  - The brain merges optimally the two flows*

$$\gamma_{n,12}^{-2} = (1-\eta_{n,1})^{-2} + (1-\eta_{n,2})^{-2}$$

  - γ is the JND
  - η_1 the haptic loss , η_2 the visual loss



- **Optimal joint resource allocation:**

$$\begin{aligned} \underset{w_i,\eta_i}{\text{maximize}} \quad & \alpha R_2 + (1-\alpha)(C_E - \lambda_E) \\ \text{s.t} \quad & R_1 = \hat{R}_1 \\ & R_2 \geq \hat{R}_2 \\ & 0 < \gamma_{12} \leq \hat{\gamma}_{12} \\ & 0 < \hat{\eta}_2 \leq \eta_2 < \eta_1 < 1 \\ & \frac{\lambda_E}{C_E} < 1 \end{aligned}$$

R_1 Haptic throughput of metaverse user.
R_2 Visual throughput of metaverse user.
$\hat{R}_1$ Required haptic throughput.
$\hat{R}_2$ Minimal visual throughput.
γ_{12} JND of metaverse user.
$\hat{\gamma}_{12}$ Maximal JND.
$\hat{\eta}_2$ Minimal visual decoding success probability.
λ_E eMBB users traffic.
C_E Cell capacity for eMBB users.

**Ernst, Marc O., and Martin S. Banks. "Humans integrate visual and haptic information in a statistically optimal fashion." Nature, 2002.*

Outline

- 5G network optimization for URLLC
  - Resource dimensioning by queuing theoretical models
  - URLLC evolution with 6G

- **From URLLC to haptic communications**
  - Tactile Internet with haptic feedback
  - Joint control and communications: work within Doctoral Network TOAST

TOAST

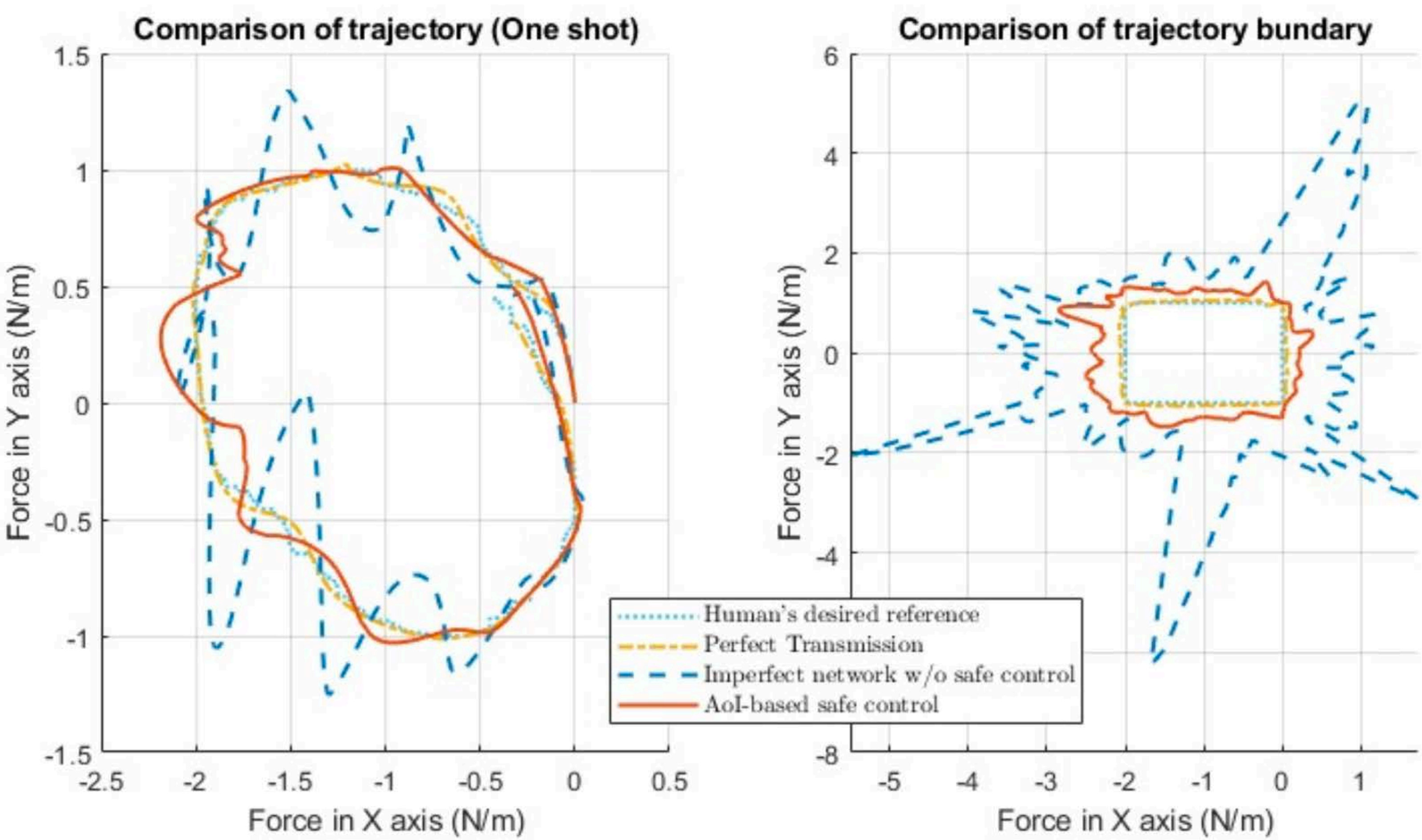
Age of Information based control

- **Define a switching policy, ζ(t) = 1 or ζ(t) = 0, between local (safe) mode and haptic-based human control**
  - when there is a burst of packet losses, the classical Hold The Last Sample (HLS) technique may lead to safety issues.
- **Age of Information (AoI)-based control**
- **We model the closed loop system as a Markov jump linear system (MJLS)**

$$\mathcal{G}_{AoI} : \chi(t+1) = \mathcal{A}_{\hat{\theta}(t)}\chi(t) + \mathcal{B}_{\hat{\theta}(t)}$$

  - $\chi(t)$ is the state at time *t*: (state of the plant and controller, command sent and received, haptic feedback)
  - $\hat{\theta}(t)$ is the AoI at time *t*, for both links
  - $\mathcal{A}$ and $\mathcal{B}$ are matrices describing the dynamics of the system, under the AoI and the associated control policy.

Result: safe control



Conclusion

- **URLLC service is an essential enabler for several industrial applications**
  - It is a costly service in terms of resources and should be carefully designed
  - Queuing theoretical models should be revisited with focus on discrete time packet level models

- **The adopted approach by 3GPP is not sufficient for integrating industrial and Metaverse applications in 6G**
  - Inconvenient: Too energy and resource costly for ensuring the same stringent requirements for all services

- **Proposed approach: within the global framework of Goal-oriented communications**
  - Difficulty: Each application has its own goal and you need to delve into the control system
  - Competencies: optimal control, signal processing, image processing, etc.
- **Example for haptic communications:**
  - New shape of traffic: JND-based codec, multi-modal flows
  - New metrics for performance: consecutive packet losses, safety of the controller
  - Ongoing works: joint design of the controller and the network for ensuring safety and efficiency

Many thanks to my co-authors on this topic

- **From Orange:**
  - Antonia Masucci
- **From Telecom SudParis**
  - Tijani Chahed
- **From CentraleSupelec:**
  - Richard Combes, Abdel Lisser
- **From CNRS:**
  - Vineeth Varma
- **Current PhD students:**
  - Mohammed Abdullah, Jorge Mirande, Marc Pierre, Yu Yeh
- **Former PhD students**
  - Abdellatif Chagdali, Nathalie Naddeh, Tiago Rochas Goncalves

Some references

- **On queuing models for URLLC:**
  - S. E. Elayoubi, N. Naddeh, S. Ben Jemaa and T. Chahed, “A Large Deviations Model for Latency Outage for URLLC,” in **Valuetools** 2022, November 2022.
  - A. Chagdali, S. E. Elayoubi, A. Masucci and A. Simonian, On the design and performance of scheduling policies exploiting spatial diversity for URLLC, **Computer Communications**, Elsevier, 2023.
  - M. Abdullah, S. E. Elayoubi, T. Chahed and A. Lisser, “Performance modeling and dimensioning of latency-critical traffic in 5G networks,” in **IEEE Globecom**, December 2023.
  - M. Abdullah, S. E. Elayoubi, T. Chahed and A. Lisser, “Radio Resource Allocation for Extreme URLLC Under Partial Knowledge of Arrival Distributions,” in **IEEE PIMRC**, September 2024.
  - M. Abdullah, S. E. Elayoubi and T. Chahed, “Efficient queue control policies for latency-critical traffic in mobile networks,” in **IEEE Transactions on Network and Service Management**, 2024.
  - S. E. Elayoubi, M. ElHassan and A. Masucci, Optimal multi-connectivity strategies for delay-sensitive industrial IoT traffic, **IEEE Globecom**, Cape Town, December 2024.

- **On haptic communications**
  - Y. Yeh, V. Varma and S. E. Elayoubi, “AoI-based switching control for safe haptic teleoperation over a wireless network ,” in **IEEE CDC**, December 2024.
  - M. Abdullah, S. E. Elayoubi and T. Chahed, “Network control for ensuring haptic service performance in 5G/6G networks,” Preprint available on HAL, 2024.
  - J. Mirande T. Chahed and S. E. Elayoubi, “Optimal resource allocation for the transport of multi-modal visual-haptic metaverse flows in 5G,” Preprint available on HAL, 2024.

- **On goal-oriented communications and joint control/communications**
  - T. Rochas, V. Varma and S. E. Elayoubi, “Relay-assisted platooning in wireless networks: a joint communication and control approach,” in **IEEE Transactions on Vehicular Technology**, 2023.
  - T. Rochas, R. Cunha, V. Varma and S. E. Elayoubi, “Fuel-efficient switching control for platooning systems with a deep reinforcement learning approach,” **IEEE Transactions on Intelligent Transportation Systems**, 2023.