

低延迟无线通信 性能极限和逼近方法

陈巍
清华大学

报告提纲

- 通信中的实时性需求
- 产生通信延时的主要因素
- 几类延时性能测度
- 通信延时的理论界、逼近方法和低复杂度策略

通信中的实时性需求

有实时性需求的典型场景



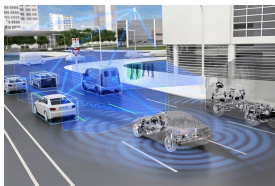
电力系统



工厂自动化



高频交易



自动驾驶

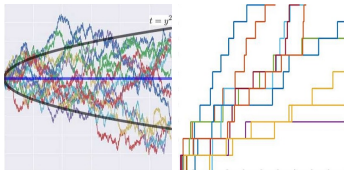


远程手术

为什么有实时性需求?

1、感兴趣的观测对象随时间变化

- 如：布朗运动，计数过程等



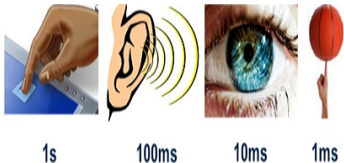
2、控制决策中代价随决策延时增加

- 如：白酒的罐装系统



3、人类感官系统的舒适性需求

- 几种感官的延时容忍情况



产生延时的主要因素

1. 传播延时——物理极限，不可改变
2. 编译码重传——由可靠性来决定，信息论极限
3. 衰落引发的延时——衰落导致速率降低，乃至中断，引起分组的堆积等待
 - 信道求逆 ($P \propto |h|^{-2}$) 可消除衰落延时，但会导致巨大的功率代价
4. 突发引起的延时——分组的突发到达导致分组的排队等待

几类延时性能测度（一）

1. 平均延时

$$\bar{D} = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K D_k$$

D_k —— 第 k 个分组的延时

由 Little 定理

$$\bar{D} = \lambda^{-1} \bar{Q}$$

\bar{Q} —— 平均队列长度，一般由 Markov 队列模型的稳态分析得到

$$\bar{Q} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T Q[t]$$

$Q[t]$ —— 第 t 时隙观测到的队列长度

几类延时性能测度（二）

2. 超时概率

$$\begin{aligned}\varepsilon_P &= \Pr\{D \geq D_{th}\} \\ &= \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \mathbb{1}_{\{D_k \geq D_{th}\}}\end{aligned}$$

D_{th} ——可容忍的最大延时

3. 队列超长概率

$$\begin{aligned}\varepsilon_L &= \Pr\{Q \geq Q_{th}\} \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{1}_{\{Q[t] \geq Q_{th}\}}\end{aligned}$$

几类延时性能测度（三）

- 平均延时与超长概率之间的关系

$$\Pr\{Q \geq Q_{th}\} \leq \frac{\bar{Q}}{Q_{th}} = \frac{\lambda \bar{D}}{Q_{th}} (\text{Markov 不等式})$$

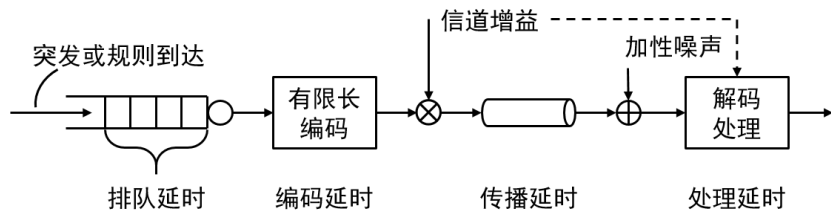
意义：优化平均延时，也是优化超长概率上界

- 超时概率与超长概率之间的关系

$$\text{对于 } G/D/1 \text{ 队列 } \Pr\{D \geq D_{th}\} = \Pr\left\{Q \geq \frac{D_{th}}{\mu}\right\}$$

$$\text{对于 } D/G/1 \text{ 队列 } \Pr\{D \geq D_{th}\} = \Pr\left\{Q \geq \frac{D_{th}}{\lambda}\right\}$$

通信的延时模型



通信的实时性涉及多个层的联合分析，设计与优化

[1] X. Zhao, W. Chen, J. Lee and N. B. Shroff, "Delay-Optimal and Energy-Efficient Communications With Markovian Arrivals," in *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1508-1523, Mar. 2020.

平均延时的理论极限 (一)

- 问题形式

$$\begin{array}{ll}\min & \bar{D} \\ \text{s.t.} & \bar{P} \leq P^{\text{ave}}.\end{array}$$

- 物理意义:

在给定的平均功率约束下, 最小化平均延时/队长

- 直观思路:

- ① 当前信道越好, 发送功率越高
- ② 当前队列越长, 发送功率越高

平均延时的理论极限（二）

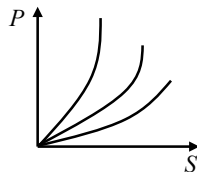
概率型跨层控制

$$P(q, h) \sim \begin{pmatrix} 0 & 1 & 2 & \dots, & p^{max} \\ \pi_{qh}^0 & \pi_{qh}^1 & \pi_{qh}^2 & \dots, & \pi_{qh}^{p^{max}} \end{pmatrix}$$

——发送功率是条件于 (q, h) 上的随机变量

$$S(P(q, h), h) = W \log \left(1 + \frac{|h|^2 P(q, h)}{n_0 W} \right)$$

$$Q_{k+1} = [Q_k + A_k - S_k]^+, \quad [x]^+ = \max\{0, x\}$$



该问题是一个受约束马氏决策过程

(*Constrained Markov Decision Process, CMDP*)

平均延时的理论极限 (三)

CMDP 的线性规划形式

$$\pi_{ijk} = \Pr \{Q = i, H = j, S = k\}.$$

$$\text{目标函数 } \min \sum_i i \times \left(\sum_j \sum_k \pi_{ijk} \right)$$

$$\text{约束条件 s.t. } \sum_j \sum_k P_{jk} \times \left(\sum_i \pi_{ijk} \right) \leq P^{\text{ave}} \rightarrow \text{功率约束}$$

$$\sum_k \pi_{lmk} = \sum_i \sum_j \sum_k \pi_{ijk} b_{ijklm}$$

$$\sum_i \sum_j \sum_k \pi_{ijk} = 1 \rightarrow \text{概率归一化}$$

$$\pi_{ijk} \geq 0 \quad \forall i, j, k. \rightarrow \text{概率非负}$$

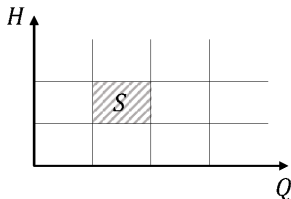
P_{jk} ——信道状态 j 下, 发送 k 所需的功率

b_{ijklm} ——转移概率 $\Pr \{Q = l, H = m | Q = i, H = j, S = k\}$, $\sum_l \sum_m b_{ijklm} = 1$

平均延时的理论极限 (四)

最优控制策略的门限结构：

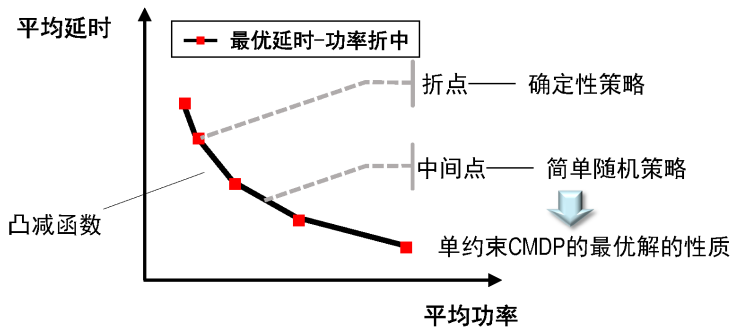
- 存在 Q (传输迫切度) 和 H (资源效率) 上的两组门限, 将状态空间 $Q \times H$ 分割成不相交的子集。
- 对于一个子集内的状态, 采用相同的物理层发送策略。
- 对于门限状态, 以一定概率选择相邻子集的发送策略。
- 传输功率和速率随队列长度和信道增益单调增加。



[2] M. Wang, J. Liu, W. Chen and A. Ephremides, "Joint Queue-Aware and Channel-Aware Delay Optimal Scheduling of Arbitrarily Bursty Traffic Over Multi-State Time-Varying Channels," in *IEEE Transactions on Communications*, vol. 67, no. 1, pp. 503-517, Jan. 2019.

平均延时的理论极限 (五)

功率—延时的分段线性关系:



[3] X. Chen, W. Chen, J. Lee and N. B. Shroff, "Delay-Optimal Buffer-Aware Scheduling With Adaptive Transmission," in *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2917-2930, Jul. 2017.

码率和信道增益的连续化 (一)

条件

- P, R 可以从 \mathbb{R} 中连续取值。
- H 为 $\mathbb{R}^+ \cup \{0\}$ 上的连续分布随机变量。

结论

- 最优策略为确定性策略，由函数 $P = P^*(Q, H)$ 给出。

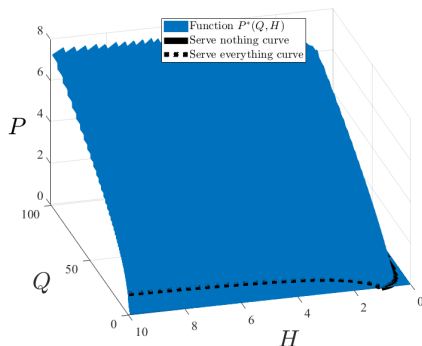
原因

- 门限结构导致简单随机策略的两个确定策略任意接近。

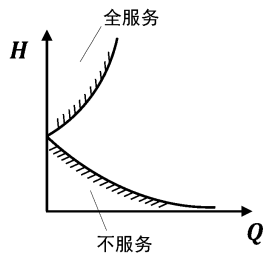
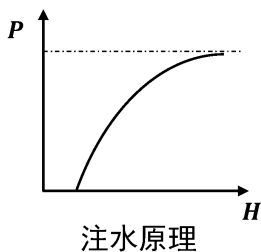
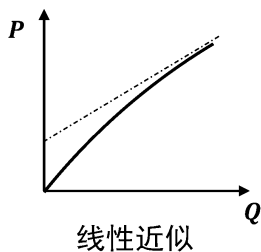
[4] J. Wu and W. Chen, "A Deterministic Scheduling Policy for Low-Latency Wireless Communication With Continuous Channel States," in *IEEE Transactions on Communications*, vol. 69, no. 10, pp. 6590-6603, Oct. 2021.

码率和信道增益的连续化 (二)

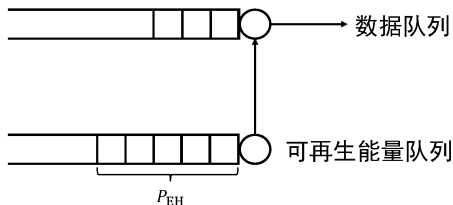
$$P^*(Q, H) = \left[(aQ + b) (1 - e^{-cQ}) - \frac{d}{H} \right]^+$$
$$[x]^+ = \max\{x, 0\}$$



码率和信道增益的连续化 (三)

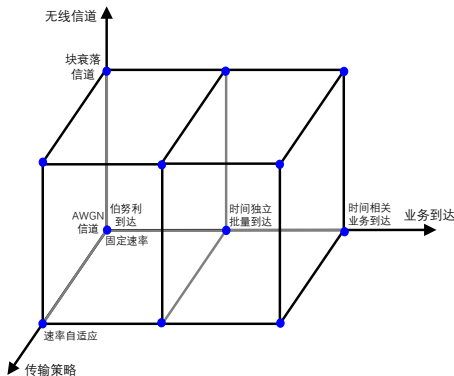


能量收割系统的低延时通信



- 最优策略：若能量 buffer（电池）无穷, 则采用一般系统的 $P^*(Q, H)$, 从中减去 P_{EH} 。
- 有限电池：若能量 buffer（电池）有限, 会导致无法存储的损失, 由大偏差方法可计算。

平均延时极限的逐点研究



[6] J. Liu, W. Chen and K. B. Letaief, "Delay Optimal Scheduling for ARQ-Aided Power-Constrained Packet Transmission Over Multi-State Fading Channels," in *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7123-7137, Nov. 2017.

[7] X. Chen, W. Chen, J. Lee and N. B. Shroff, "Delay-Optimal Buffer-Aware Scheduling With Adaptive Transmission," in *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2917-2930, Jul. 2017.

[8] M. Wang, J. Liu, W. Chen and A. Ephremides, "Joint Queue-Aware and Channel-Aware Delay Optimal Scheduling of Arbitrarily Bursty Traffic Over Multi-State Time-Varying Channels," in *IEEE Transactions on Communications*, vol. 67, no. 1, pp. 503-517, Jan. 2019.

[9] X. Zhao, W. Chen, J. Lee and N. B. Shroff, "Delay-Optimal and Energy-Efficient Communications With Markovian Arrivals," in *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1508-1523, Mar. 2020.

.....

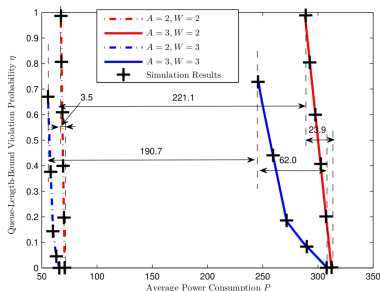
队列超长概率的理论极限 (一)

无（有）丢包情况下，CMDP 的线性规划形式

$$\begin{aligned} \min \quad & \overbrace{\sum_{i \geq Q_{th}} \sum_j \sum_k \pi_{ijk}}^{\text{无丢包情况}} \quad / \quad \overbrace{\sum_j \sum_k \pi_{Q_{th}jk}}^{\text{有丢包情况}} \\ \text{s.t.} \quad & \sum_j \sum_k P_{jk} \times \left(\sum_i \pi_{ijk} \right) \leq P^{\text{ave}} \\ & \sum_k \pi_{lmk} = \sum_i \sum_j \sum_k \pi_{ijk} b_{ijklm} \\ & \sum_i \sum_j \sum_k \pi_{ijk} = 1 \\ & \pi_{ijk} \geq 0 \end{aligned}$$

队列超长概率的理论极限 (二)

- 若 P^{ave} 可确保 $\Pr\{Q \geq Q_{\text{th}}\} = 0$ ，则令功率消耗最小的策略同样具有门限结构。
- 若此时，码率或信道增益可连续变化，最优策略为确定性策略。



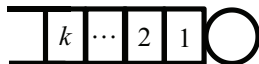
超长概率 -- 功率折中曲线

[10] S. Hu and W. Chen, "Minimizing the Queue-Length-Bound Violation Probability for URLLC: A Cross Layer Approach," in *IEEE Global Communications Conference*, pp. 1–6, Dec. 2020.

超时概率的理论极限 (一)

高维 CMDP 问题

- 状态 $\vec{D} = (\text{队列位置1的延时}, \text{队列位置2的延时}, \dots, \text{队列位置}k\text{的延时})$



- 代价：平均丢包数

$$= \Pr \left\{ \vec{D} = \vec{d} \right\} \Pr \{ H = j \} \times \{ \text{状态} \vec{d} \times j \text{ 下无法} \\ \text{服务且超时的包数} \}$$

超时概率的理论极限 (二)

高维 CMDP 问题的线性规划形式

- $\pi_{\vec{d},j,k} = \Pr \left\{ \vec{D} = \vec{d}, H = j, s = k \right\},$
- $\eta_{\vec{d},k}$: \vec{d} 状态下, 若发送 k 个 bit, 则需要丢弃的 bit 数。

$$\min \sum_{\vec{d}} \pi_{\vec{d},j,k} \eta_{\vec{d},k},$$

$$\text{s.t.} \quad \sum_k \sum_j p_{j,k} \left(\sum_{\vec{d}} \pi_{\vec{d},j,k} \right) \leq P^{ave},$$

$$\sum_k \pi_{\vec{d}',j',k} = \sum_{\vec{d}} \sum_j \sum_k \pi_{\vec{d},j,k} b_{\vec{d},j,k,\vec{d}',j'},$$

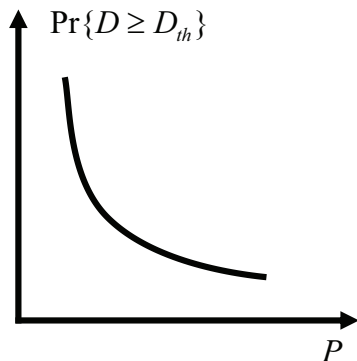
$$\sum_{\vec{d}} \sum_j \sum_k \pi_{\vec{d},j,k} = 1,$$

$$\pi_{\vec{d},j,k} \geq 0.$$

超时概率的理论极限 (三)

针对上述单约束的 CMDP 问题

- 仍可证明最优策略在 \vec{D} 和 H 上具有门限结构



低复杂度的低延时通信策略

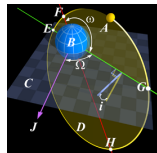
Lyapunov 方法

- 控制理论中提升稳定性的方法



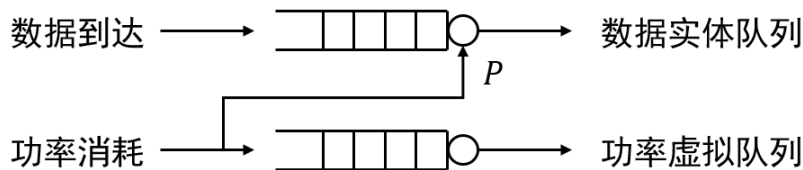
应用于排队系统法

- 当队列濒临于发散时，不发散的
系统就相对具有低延时的优势



[12] Z. Xie and W. Chen, "Pilot-Efficient Scheduling for Large-Scale Antenna Aided Massive Machine-Type Communications: A Cross-Layer Approach," in *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4262-4276, Jul. 2020.

基于 Lyapunov-Drift 的虚队列方法 (一)



队列演化过程

$$Q[t+1] = (Q[t] - s[t])^+ + a[t]$$

$$Q_p[t+1] = (Q_p[t] - P_{av})^+ + p[t]$$

Lyapunov 函数表征系统状态

$$L(Q[t], Q_p[t]) = \frac{1}{2} (Q^2[t] + Q_p^2[t])$$

基于 Lyapunov-Drift 的虚队列方法 (二)

含虚队列的 Lyapunov-Drift 上界

$$\begin{aligned}\Delta(Q[t], Q_p[t]) &\leq B + Q[t] \lambda - Q_p[t] P_{av} \\ &\quad - \mathbb{E}\{Q[t] s[t] - Q_p[t] p[t] | Q[t], Q_p[t]\}\end{aligned}$$

- 通过控制含虚队列的 Lyapunov-Drift 上界使系统稳定

队列长度加权的优化问题

$$\begin{aligned}\max \quad & Q \cdot s - Q_p \cdot p \\ \text{s.t.} \quad & p \geq 0\end{aligned}$$

含加权系数的 Lyapunov 调度方法

重载情型下 Lyapunov-Drift 的表达式

$$\Delta(Q[t]) = \mathbb{E} \{ a^2[t] + s^2[t] - 2s[t]Q[t] + 2a[t](Q[t] - s[t]) | Q[t] \}$$

- 通过直接控制含加权系数的 Lyapunov-Drift 使系统稳定

含加权系数的优化问题

$$\min \quad s^2 - 2s(Q+a) + 2aQ + Vp$$

$$\text{s.t.} \quad p \geq 0$$

大偏差理论及其在低延时通信中的应用

大偏差理论

- 大偏差理论是概率论中的一种分析系统中极端事件发生概率渐进特性的理论。

应用于排队系统

- 大偏差理论应用于排队系统中，可以给出队长尾分布的渐进特性：

$$\lim_{Q_{th} \rightarrow \infty} \frac{\ln(\Pr\{Q > Q_{th}\})}{Q_{th}} = -\beta.$$

有效带宽 (Effective Bandwidth)



$a[t]$ 时隙 t 的到达量

μ : 固定服务速率

- $s[t]$ 有效带宽定义为

$$\alpha_a(\beta) = \lim_{T \rightarrow +\infty} \frac{1}{\beta T} \ln \mathbb{E} \left\{ e^{\beta \sum_{t=1}^T a[t]} \right\}$$

- 有效带宽: 满足 QoS 要求下需要的最小的服务速率

有效容量 (Effective Capacity)



λ : 固定到达速率

$s[t]$ 时隙 t 的服务量

- $s[t]$ 有效容量定义为

$$\alpha_s(\beta) = \lim_{T \rightarrow +\infty} -\frac{1}{\beta T} \ln \mathbb{E} \left\{ e^{-\beta \sum_{t=1}^T s[t]} \right\}$$

- 有效容量: 满足 QoS 要求下支持的最大的到达速率

有效容量和有效带宽结合

- 当到达和发送速率均为随机时，可以结合有效容量和有效带宽得到保持队列稳定的临界衰减指数系数 β^* :

$$\alpha_a(\beta^*) = \alpha_s(\beta^*).$$

- 此时的队列超长概率满足:

$$\Pr\{Q > Q_{\text{th}}\} \approx e^{-\beta^* Q_{\text{th}}}.$$

- 队列超时概率可以表示为:

$$\Pr\{D > D_{\text{th}}\} \approx e^{-\beta^* \alpha_a(\beta^*) D_{\text{th}}}.$$

低延时通信中的编码延时（一）

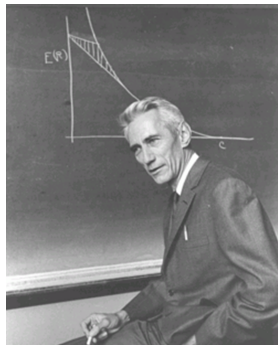
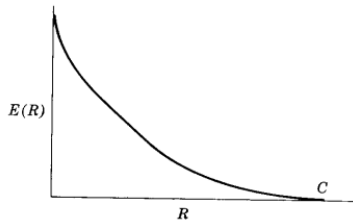
经典信息论中，关注码长

$$N \rightarrow +\infty$$

时的渐近性能。

由 $P_e \leq \exp[-NE_r(R)]$ 知,

若 $E_r(R) > 0$, 则 $N \rightarrow +\infty$ 时, $P_e \rightarrow 0$ 。



有限长编码延时下,可靠性和实时性之间构成一对矛盾。

低延时通信中的编码延时（二）

- 2010 年, Polyanskiy, Poor 和 Verdu 对有限长编码的性能进行了精细分析, 得到:

$$R = \ln(1 + |h|^2 \text{SNR}) - \sqrt{\frac{V|h|^2 \text{SNR}}{N}} Q^{-1}(\epsilon) + O\left(\frac{\ln N}{N}\right).$$

- 其中,

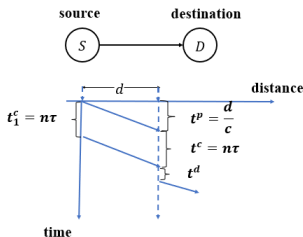
V — 信道色散系数

ϵ — 误码率

- 于是, 在可靠性约束下, 码长 (编码延时) 越小, 所需的信噪比 (功率) 越高。

延时，可靠性和通信距离的折中关系 (一)

- 有限长编码中的延时主要由编码延时、传输延时以及解码延时组成



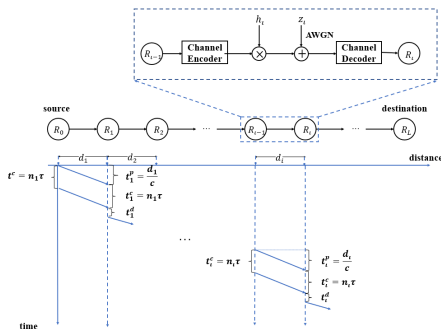
- 根据推导，点到点传输延时、可靠性和通信距离的折中关系为

$$\epsilon = Q\left(\left(\ln(1 + \gamma d^{-\alpha}) - \frac{k \ln 2}{n}\right) \sqrt{\frac{n}{V}}\right).$$

延时, 可靠性和通信距离的折中关系 (二)

- L 跳的通信链路传输延时、可靠性和通信距离的折中关系为

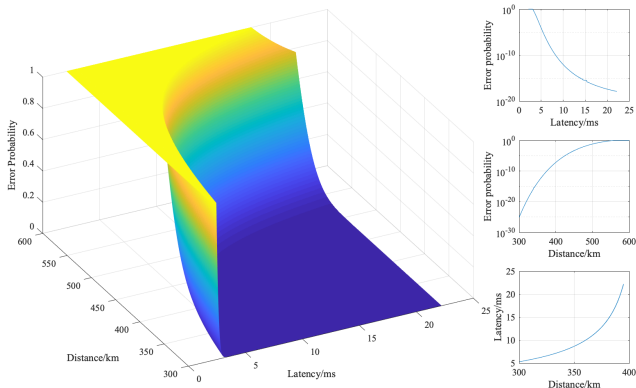
$$\epsilon = LQ \left(\left(\ln(1 + \bar{\gamma}d^{-\alpha}) - \frac{k \ln 2}{\bar{n}} \right) \sqrt{\frac{\bar{n}}{V}} \right).$$



[13] L. Huang, X. Zhao, W. Chen, and H. V. Poor, "Low-Latency Short-Packet Transmission over a Large Spatial Scale," *Entropy*, vol. 23, no. 7, p. 916,

延时, 可靠性和通信距离的折中关系 (二)

传输延时、可靠性和通信距离的折中关系曲面图



[13] L. Huang, X. Zhao, W. Chen, and H. V. Poor, "Low-Latency Short-Packet Transmission over a Large Spatial Scale," *Entropy*, vol. 23, no. 7, p. 916, Jul. 2021.

增益守恒方程 (一)

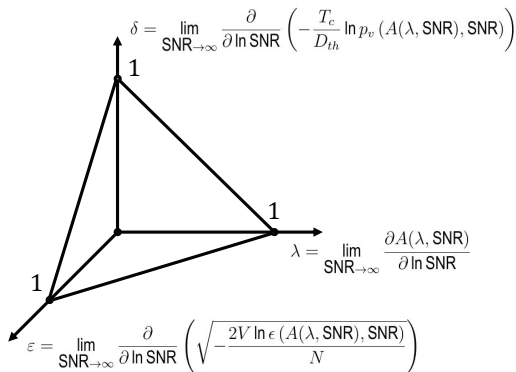
高信噪比下

$$\frac{\partial}{\partial \ln \text{SNR}} \left(\underbrace{A(\lambda, \text{SNR})}_{\text{服务率增益}} - \underbrace{\frac{T_c}{D_{th}} \ln p_v(A(\lambda, \text{SNR}), \text{SNR})}_{\text{实时性增益}} + \underbrace{\sqrt{-\frac{2V \ln \epsilon(A(\lambda, \text{SNR}), \text{SNR})}{N}}}_{\text{可靠性增益}} \right) = 1.$$

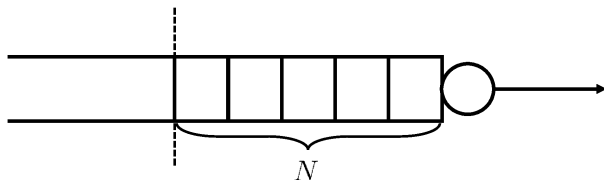
[14] Y. Wang, W. Chen, and H. V. Poor, "Ultra-Reliable and Low-Latency Wireless Communications in the High SNR Regime: A Cross-Layer Tradeoff," in *IEEE Transactions on Communications*, to appear.

增益守恒方程 (二)

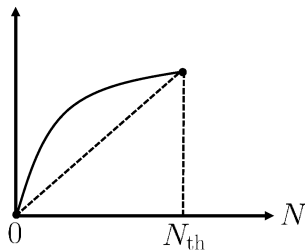
高信噪比下



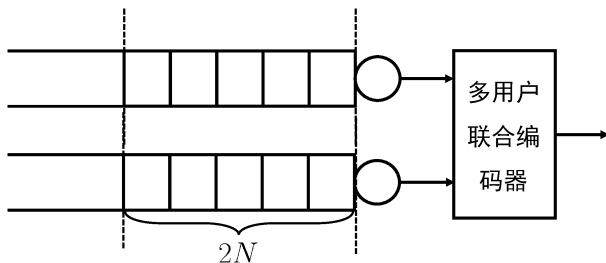
随机到达下的端点策略



攒够 N 个 bit 进行信道编码在随机到达下会有等待延时
最优策略为“端点策略”的随机组合



多用户联合编码



- 在相同的等待时间内，获得更长的码长
- 对于非同步的周期到达，这是一个非线性匹配问题

超长概率和误码率的折中关系

- 根据有限长编码公式，服务量 R 和误码率 ϵ 之间存在折中关系。
- 根据排队理论，超长概率 δ 和服务量 R 之间也存在折中关系。
- 因此，URLLC 的理论极限还可以由 $\epsilon - \delta$ 折中给出，即

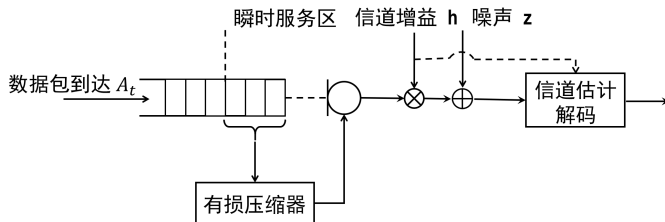
$$\ln \epsilon \approx -\frac{1}{2}(\psi - \Phi \ln \delta)^2.$$

ψ ——负载衡量系数；

Φ ——突发衡量系数。

[16] L. Li, W. Chen, and K. B. Letaief, "Short Packet Communications with Random Arrivals: An Effective Bandwidth Approach," in *IEEE Global Communications Conference (accepted)*, 2021.

失真容忍信源的低延时通信 (一)



失真容忍信源的低延时通信程序框图

降低延时的两个方法：

- ① 提升发射功率
- ② 增加压缩失真和压缩比

[17] S. Hu and W. Chen, "Joint Lossy Compression and Power Allocation in Low Latency Wireless Communications for IIoT: A Cross-Layer Approach," in *IEEE Transactions on Communications*, vol. 69, no. 8, pp. 5106-5120, Aug. 2021.

失真容忍信源的低延时通信 (二)

- 此时的动作空间

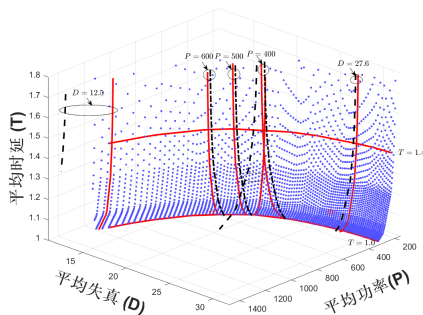
$$A = R(\text{发送速率}) \times C(\text{压缩比})$$

- 决策变量 $\pi_{ijkc} = \Pr\{Q = i, H = j, R = k, C = c\}$
- 于是，有线性规划问题

$$\begin{aligned} \min \quad & \sum_i i \times \left(\sum_j \sum_k \sum_c \pi_{ijkc} \right) \\ \text{s.t.} \quad & \sum_j \sum_k P_{jk} \times \left(\sum_i \sum_c \pi_{ijkc} \leq P^{\text{ave}} \right) \\ & \sum_c D_c \times \left(\sum_i \sum_j \sum_k \pi_{ijkc} \right) \leq D^{\text{ave}} \\ & \sum_k \sum_c \pi_{lmkc} = \sum_i \sum_j \sum_k \sum_c \pi_{ijkc} b_{ijkclm} \\ & \sum_i \sum_j \sum_k \sum_c \pi_{ijkc} = 1, \pi_{ijkc} \geq 0 \end{aligned}$$

失真容忍信源的低延时通信 (三)

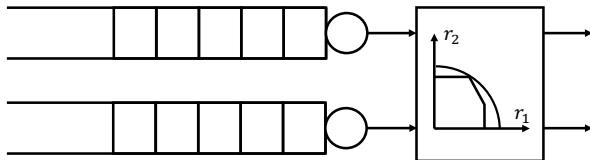
- 最优策略是由三个确定性策略随机组合得到
- 存在一个功率 - 失真 - 时延三者的最佳折中关系



功率 - 失真 - 时延折中三维图

多用户低时延通信 (一)

- 多用户共享无线资源，多了用户域分配的自由度



- $\vec{r} \in R(\vec{P}, \vec{h})$ -凸域
- MAC: $P_i \leq P_i^{max}$ BC: $\vec{I}^T \vec{P} \leq P^{max}$
- 可控变量: 功率 P_i 编译码顺序 θ_i

[18] X. Zhao and W. Chen, "Non-Orthogonal Multiple Access for Delay-Sensitive Communications: A Cross-Layer Approach," in *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5053-5068, Jul. 2019.

多用户低时延通信 (二)

CMDP 方法

- 动作空间 $A = \vec{r} \times \vec{\theta}$ (\vec{r} : 速率, $\vec{\theta}$: 译码顺序)
- 决策变量 $\pi_{\vec{i}, \vec{h}, \vec{r}, \vec{\theta}}$
- 线性规划形式: 该问题的生成一般依靠计算机算法

$$\begin{aligned} \min \quad & \sum_{\vec{i}} \mathbf{1}^T \vec{i} \times \sum_{\vec{h}} \sum_{\vec{r}} \sum_{\vec{\theta}} \pi_{\vec{i}, \vec{h}, \vec{r}, \vec{\theta}} \\ \text{s.t.} \quad & \sum_{\vec{h}} \sum_{\vec{r}} \sum_{\vec{\theta}} \mathbf{1}^T \vec{P}_{\vec{h}, \vec{r}, \vec{\theta}} \times \sum_{\vec{i}} \pi_{\vec{i}, \vec{h}, \vec{r}, \vec{\theta}} \leq P^{ave} \\ & \sum_{\vec{r}} \sum_{\vec{\theta}} \pi_{\vec{l}, \vec{m}, \vec{r}, \vec{\theta}} = \sum_{\vec{i}} \sum_{\vec{h}} \sum_{\vec{r}} \sum_{\vec{\theta}} \pi_{\vec{i}, \vec{h}, \vec{r}, \vec{\theta}} \times b_{\vec{i}, \vec{h}, \vec{r}, \vec{\theta}, \vec{l}, \vec{m}} \\ & \sum_{\vec{i}} \sum_{\vec{h}} \sum_{\vec{r}} \sum_{\vec{\theta}} \pi_{\vec{i}, \vec{h}, \vec{r}, \vec{\theta}} = 1 \\ & \pi_{\vec{i}, \vec{h}, \vec{r}, \vec{\theta}} \geq 0 \end{aligned}$$

多用户低时延通信 (三)

Lypunov 方法

- 无时隙间功率分配

$$\max \quad \vec{Q}^T \vec{r}$$

$$\text{s.t.} \quad \vec{r} \in R(\vec{p}, \vec{h})$$

- 有时隙间功率分配

$$\text{MAC} \quad \begin{cases} \max & \vec{Q}^T \vec{r} - \vec{Q}_P^T \vec{P} \\ \text{s.t.} & \vec{r} \in R(\vec{p}, \vec{h}) \end{cases}$$

$$\text{BC} \quad \begin{cases} \max & \vec{Q}^T \vec{r} - Q_P \mathbf{1}^T \vec{P} \\ \text{s.t.} & \vec{r} \in R(\vec{p}, \vec{h}) \end{cases}$$

[20] D. Han, W. Chen and Y. Fang, "Joint Channel and Queue Aware Scheduling for Latency Sensitive Mobile Edge Computing With Power Constraints," in *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 3938-3951, Jun. 2020.

[21] Y. Liu, J. Lee, and W. Chen, "Buffer-Aware Scheduling and Power Allocation for CoMP Transmission with Large-Scale Antennas," in *IEEE Transactions on Wireless Communications*, to appear.

多用户低时延通信（四）

大偏差分析方法

- 结合路径相关大偏差原理，超长概率由最可能发生超长的样本路径决定。
- 路径相关大偏差速率函数决定超长概率：

$$I_x^T(\vec{x}(\cdot)) = \inf_{\vec{x}(\cdot)} \left\{ \int_{-T}^0 H(\vec{\phi}(t) | \vec{p}) dt \right\}.$$

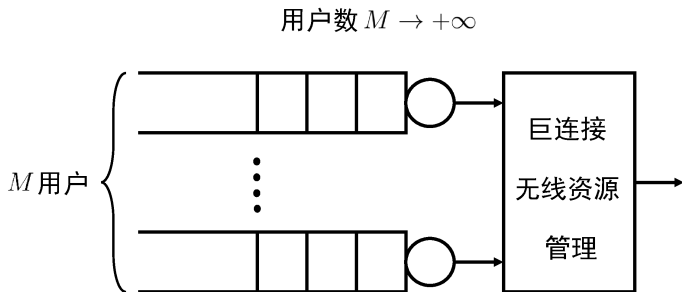
$H(x|y)$ —— 相对熵

\vec{p} —— 系统状态先验概率

$\vec{x}(\cdot)$ —— 发生超长的样本路径

$\vec{\phi}(\cdot)$ —— 样本路径 \vec{x} 对应的系统状态概率

平均场方法 (Mean Field Approach)

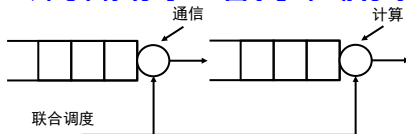


- ① 每个信道状态上，用户数 $M(h) \rightarrow +\infty$
- ② 组内 $M(h)$ 个用户，队列长度呈相同分布
- } 简化策略

[22] C. Li, W. Chen, and K. B. Letaief, "Mean-field approximation based scheduling for broadcast channels with massive receivers," in *IEEE Global Communications Conference (accepted)*, Dec. 2021.

通信-计算协同的低时延

存在通信队列和计算队列二者为串联关系



动作空间 $A = r_1 \times r_2$ (\vec{r} : 速率, $\vec{\theta}$: 译码顺序) } 2 维 CMDP

功率代价 $P = P_T(r_1, h) + P_C(r_2)$

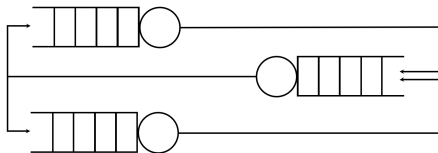
● 上述系统具有清晰的瓶颈分析

Scenario	Case 1: exponential	Case 2: power-law
Type 1	Transmission	Transmission ($N \leq 1$) and computation ($N \geq 2$)
Type 2	Transmission	Transmission ($N \leq 3$ and $3N < 5M$) and computation (in others)
Type 3	Transmission	Transmission ($N \leq 3$ and $3N < 5M$) and computation (in others)
Type 4	Transmission	Transmission ($N \leq 1$) and computation ($N \geq 2$)

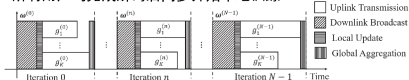
[23] D. Han, W. Chen and Y. Fang, "Joint Channel and Queue Aware Scheduling for Latency Sensitive Mobile Edge Computing With Power Constraints," in *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 3938-3951, Jun. 2020.

[24] D. Han, W. Chen, B. Bai and Y. Fang, "Offloading Optimization and Bottleneck Analysis for Mobile Cloud Computing," in *IEEE Transactions on Wireless Communications*, vol. 67, no. 9, pp. 6153-6167, Sept. 2019.

联邦学习系统的延时性能

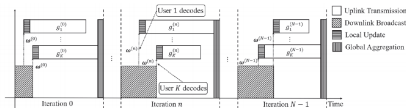


所有用户均完成解码后同步开始本地训练

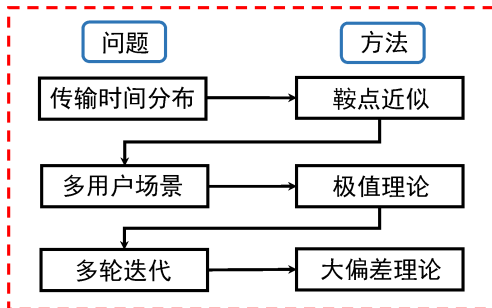


同步下行情况

利用rateless coding, 用户完成解码后立刻开始本地训练



异步下行情况



[25] L. Li, L. Yang, X. Guo, Y. Shi, H. Wang, W. Chen, and K. B. Letaief, "Delay Analysis of Wireless Federated Learning Based on Saddle Point Approximation and Large Deviation Theory," in *IEEE Journal on Selected Areas in Communications*, early access, Oct. 2021.

谢谢

Questions & Answers