Real-time communication

P. R. Kumar

With I-Hong Hou and V. Borkar

Dept. of Electrical and Computer Engineering, and
Coordinated Science Lab
University of Illinois, Urbana-Champaign

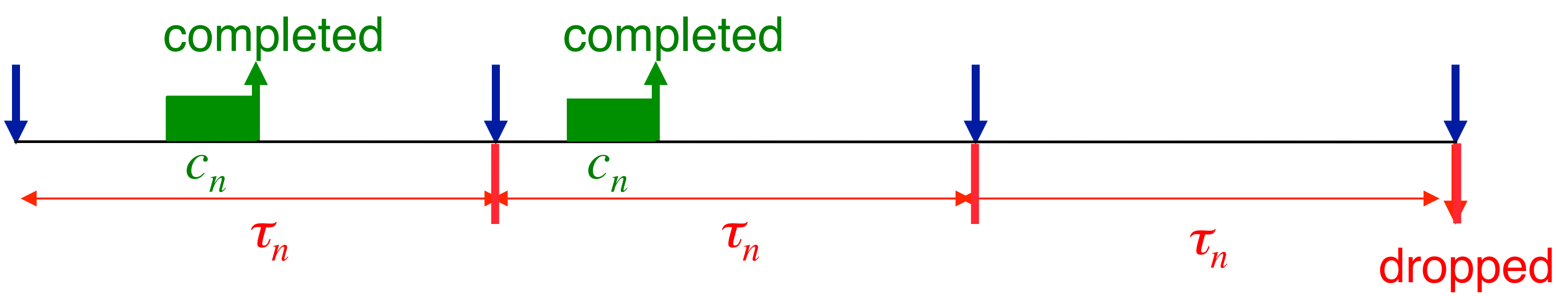
LCCC Workshop
Lund University
Feb 3-5, 2010

Email: prkumar@illinois.edu
Web: http://decision.csl.illinois.edu/~prkumar/

Delay guarantees for wireless

- Increasing use of wireless networks for serving traffic with delay constraints:
  - VoIP
  - Interactive Video
  - Networked Control
- Example
  - Average car has 70 microprocessors and kilometers of wiring
  - Replace with a Faraday cage and a base-station?
- Move from event-driven computing to event-cum-time-driven computing
  - Cyberphysical systems
    - » Vehicular networks, Medical plug-n-play
- How to support delay guarantees?

Backbone of Real-Time Scheduling: Liu-Layland (`73)



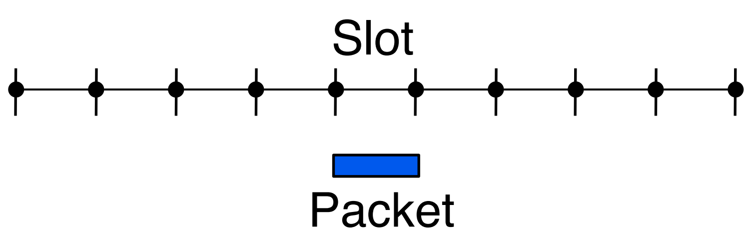
- N tasks
  - Jobs of Task n arrive with period τ_n
  - Deadline is end of period
  - Worst case execution time c_n
- Rate monotone scheduling: Priority to smallest period task
- All deadlines met if $\sum_{n=1}^{N} \frac{c_n}{\tau_n} \leq N(2^{1/N} - 1)$ $\quad (\to \ln 2 = 0.69$ as $N \to \infty)$
- If any priority policy can meet all deadlines, then this policy can

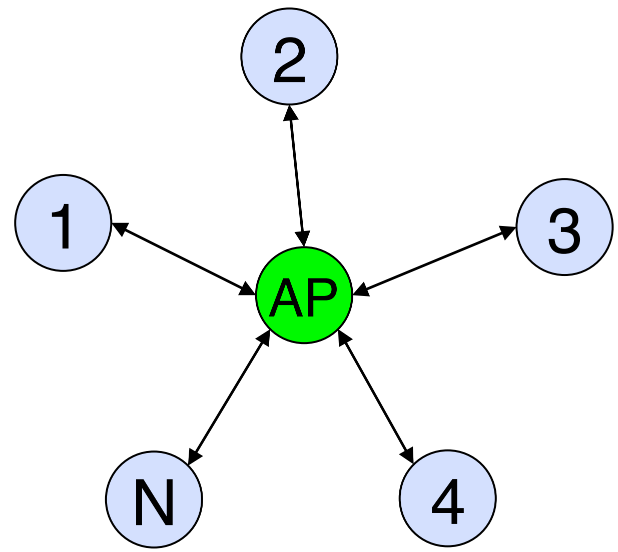
Real-time communication: Client-Server model

- A wireless system with an Access Point serving N clients

- Time is slotted

- One slot = One packet



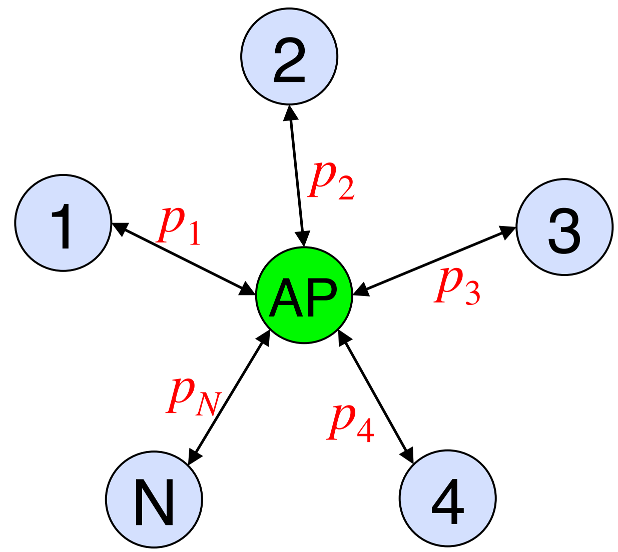


- AP indicates which client should transmit in each time slot

Model of unreliable channels

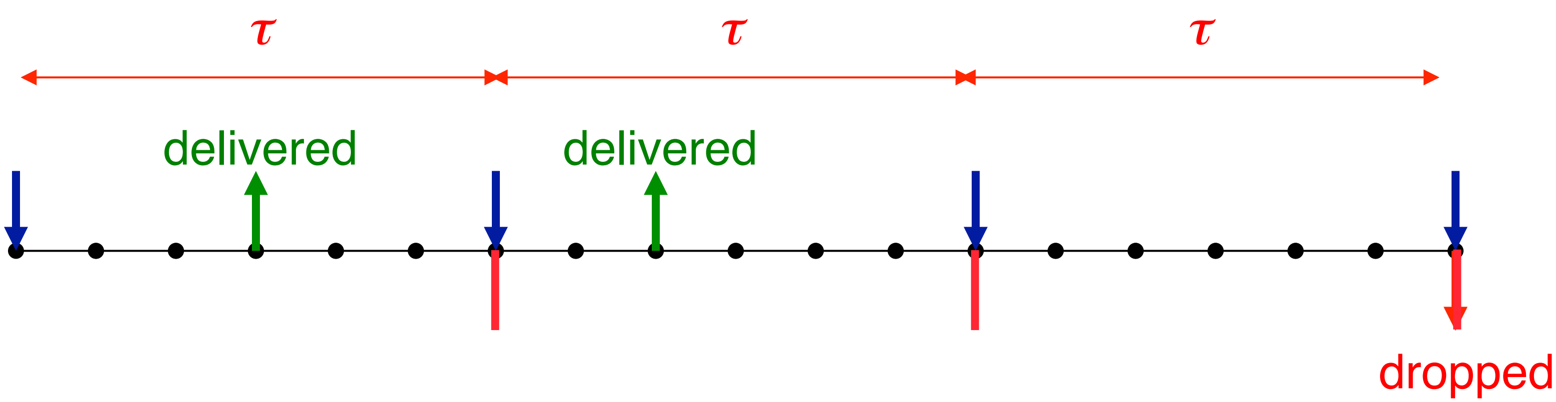
- Unreliable channels

- Packet transmission in each slot
  - Successful with probability p_n
  - Fails with probability $1\text{-}p_n$
  - So packet delivery time is a geometrically distributed random variable γ_n with mean $1/p_n$



- Non-homogeneous link qualities
  - $p_1, p_2, \ldots, p_N$ can be different

QoS model

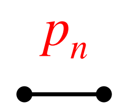


- Clients generate packets with fixed period τ
- Packets expire and are dropped if not delivered in the period
- Delay of successfully delivered packet is therefore at most τ
- Delivery ratio of Client n should be at least q_n

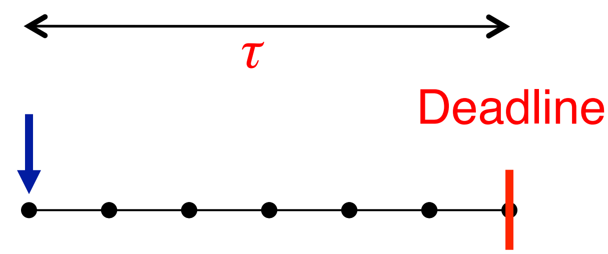
$$\liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} 1(\text{Packet delivered to Client } n \text{ in } t\text{-th period}) \geq q_n \quad a.s.$$

Multiple-time scale QoS requirements

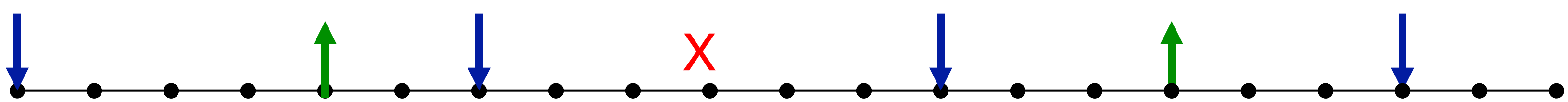
- Unreliable channels
  - Short time scale: Slots



- Arrivals and Deadlines
  - Medium time scale:
  - Period τ arrivals
  - Relative Deadline τ



- Delivery ratio requirements
  - Long time scale:



$$\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} 1(\text{Packet of client } n \text{ delivered in } t\text{-th period}) \geq q_n \ a.s.$$

Feasibility of a set of clients

Implied load

- Load due to Client n

$$w_n = \frac{q_n \cdot \frac{1}{p_n}}{\tau}$$

$$= \frac{E(\text{\# deliveries per period}) \cdot E(\text{\# slots per delivery})}{\text{\# of slots of per period}}$$

- The proportion of time slots needed by Client n is

$$w_n = \frac{q_n}{p_n \tau}$$

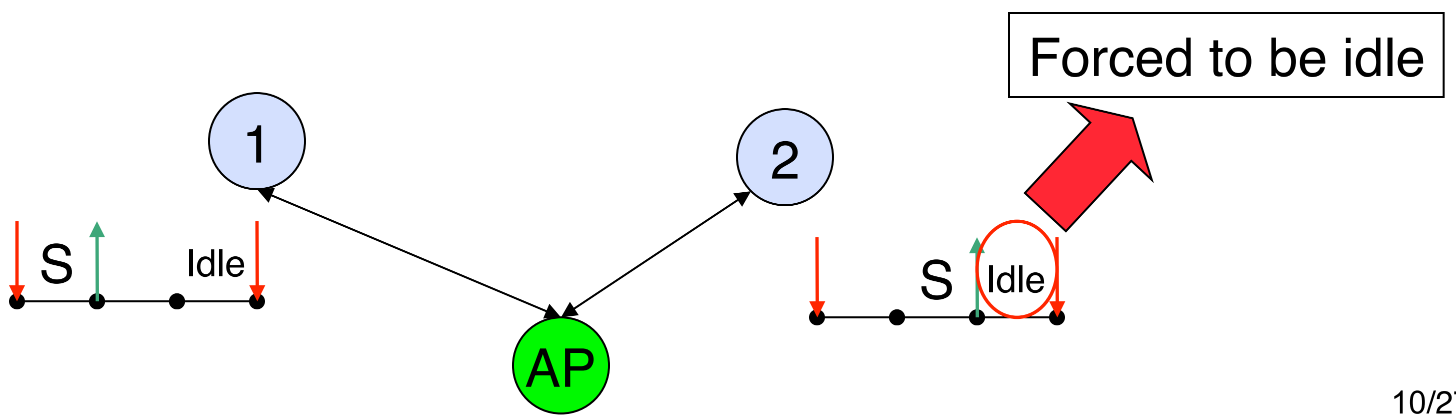
Necessary condition for feasibility of QoS requirements

- Necessary condition from classical queueing theory

$$\sum_{n=1}^{N} w_n \leq 1$$

- But not sufficient

- Reason: Unavoidable idle time
  - No queueing: At most one packet



Stronger necessary condition

- Let $I(1, 2,\ldots, N)$:= Unavoidable idle time after serving $\{1, 2,\ldots, N\}$

$$I(1,2,\ldots,N) = \frac{1}{\tau} E\left[\left(\tau - \sum_{n=1}^{N} \gamma_n\right)^+\right] \quad \text{where } \gamma_n \sim \text{Geom}(p_n)$$

- Stronger necessary condition

$$\sum_{n=1}^{N} w_n + I(1,2,\ldots,N) \le 1$$

- Sufficient?

- Still not sufficient!

Counterexample

- Two clients: Period $\tau = 3$

- Client 1
  - $p_1 = 0.5$
  - $q_1 = 0.876$
  - $w_1 + I_1 = 3.002/3 > 1$ ✗

$$w_1 = \frac{q_1}{p_1 \tau} = \frac{1.752}{3}$$

$$I_1 = \frac{\left(2p_1 + (1-p_1)p_1\right)}{3} = \frac{1.25}{3}$$

- Client 2
  - $p_2 = 0.5$
  - $q_2 = 0.45$
  - $w_2 + I_2 = 2.15/3 < 1$ ✓

$$w_2 = \frac{q_2}{p_2 \tau} = \frac{0.9}{3}$$

$$I_2 = \frac{1.25}{3}$$

- Clients {1,2}
  - $w_1 + w_2 + I_{\{1,2\}} = 2.902/3 < 1$ ✓

$$w_{\{1,2\}} = w_1 + w_2 = \frac{2.652}{3}$$

$$I_{\{1,2\}} = \frac{p_1 p_2}{3} = \frac{0.25}{3}$$

Even stronger necessary condition

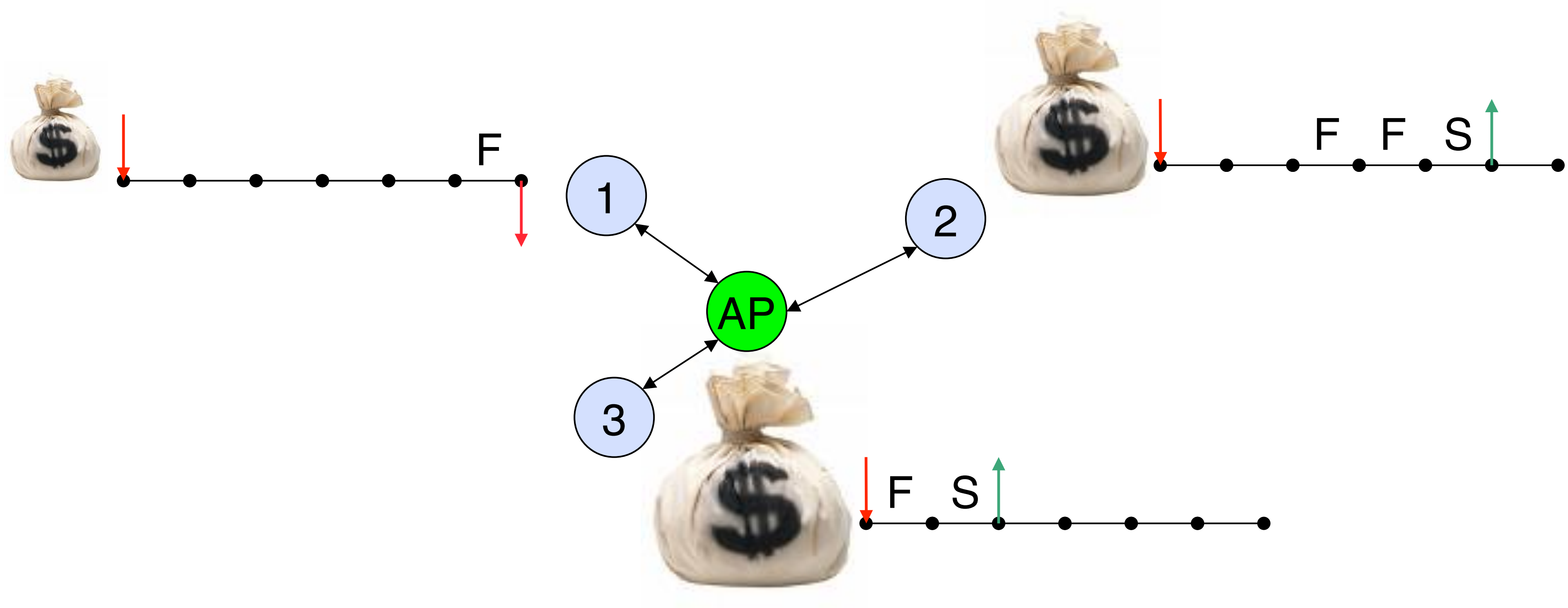
- Every *subset* of clients $S \subseteq \{1,2,\ldots,N\}$ should also be feasible
- Let $I(S) := \frac{1}{\tau} E\left[\left(\tau - \sum_{n \in S} \gamma_n\right)^+\right]$ = Idle time if only serving S
- Stronger necessary condition: $\sum_{n \in S} w_n + I(S) \leq 1, \quad \forall S \subseteq \{1,2,\ldots,N\}$

$$\underbrace{\nearrow \text{ with } S}_{} \quad \underbrace{\searrow \text{ with } S}_{}$$

- Not enough to just evaluate for the whole set $\{1, 2, \ldots, N\}$
- **Theorem (Hou, Borkar & K '09)**
  Condition is necessary and sufficient for a set of clients to be feasible

Scheduling policy

Debt-based scheduling policies



- Compute "debt" owed to each client at beginning of period
- A client with higher debt gets a higher priority on that period

Two definitions of debt

- The **time debt** of Client n

$$= (w_n - \text{Actual proportion of transmission slots given to Client } n)$$

- The **weighted delivery debt** of Client n

$$= \frac{q_n - \text{Actual delivery ratio of Client } n}{p_n}$$

- **Theorem (Hou, Borkar & K '09)**
  Both largest debt first policies fulfill every set of clients that can be fulfilled

Utility maximization framework and solution

Utility maximization

- Client n has a utility function $U_n(q_n)$
  - U_n positive, str incr, str concave, $U_n(0)$ = right limit …
- Maximize the total utility

- SYSTEM

Max $\sum_n U_n(q_n)$

s.t. $\sum_{n\in S} \frac{q_n}{\tau p_n} \leq 1 - I_S, \forall S$

over $q_n \geq 0$

Solving SYSTEM directly is difficult

Clients may have different utility functions U_n

2^N feasibility constraints

Two sub-problems

Considers own utility function

Client n

$$\underset{0\le\rho_n\le\psi_n}{Max}\; U_n\left(\frac{\rho_n}{\psi_n}\right)-\rho_n$$

Price ψ_n

Payment ρ_n

Considers feasibility

Access Point

$$\underset{\{q_n\}\text{ feasible}}{Max}\sum_{n=1}^{N}\rho_n \log q_n$$

Achieved by
Weighted Transmission Time Policy

Give priority to lowest $u_n(t)/\rho_n$

$u_n(t)$ = Number of slots in $[0, t]$ given to Client n

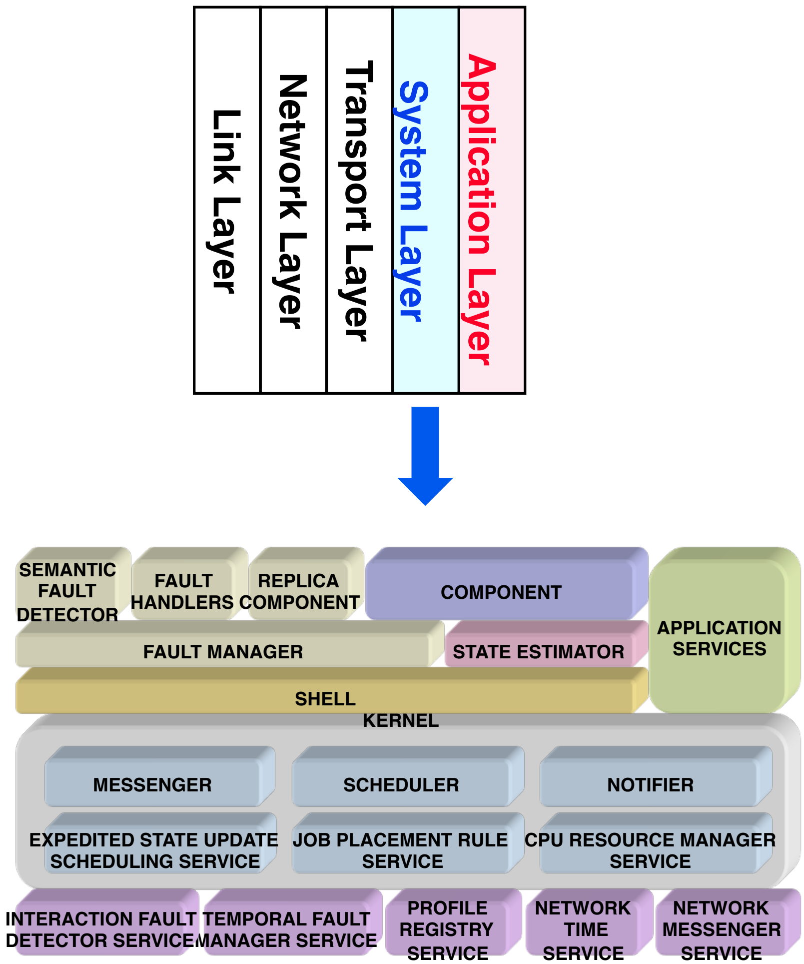
Is weighted max-min fair
And weighted proportionally fair

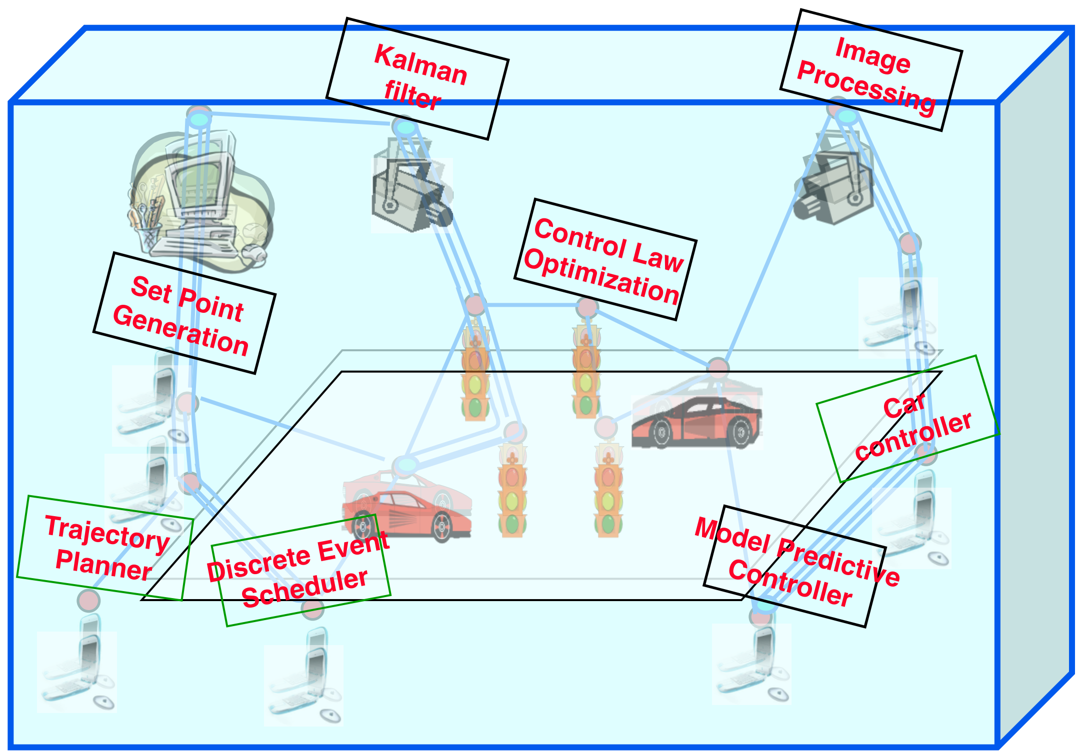
- Nobody needs to know the channel reliability p_n

(Hou & K '09)

Networked control

Real-time middleware for control



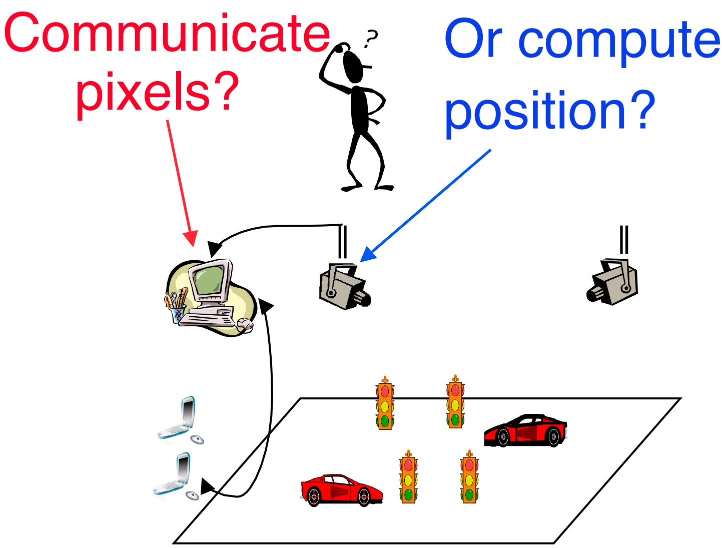


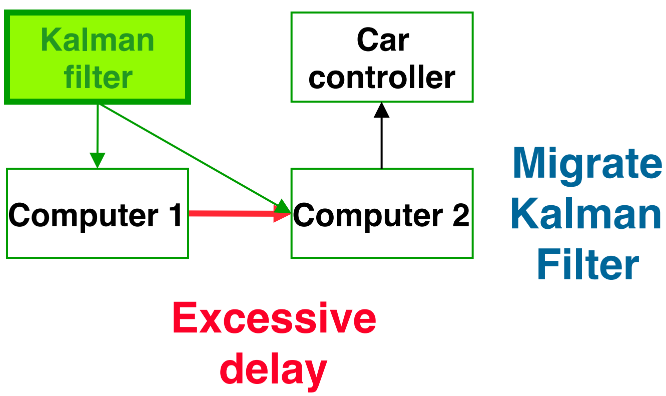
(Baliga, Graham, & K '05
And
Kim & K '08)

Collision avoidance (Schuetz, Robinson & K ‘05)

http://decision.csl.uiuc.edu/~testbed/videos/CollisionAvoidance.mpg

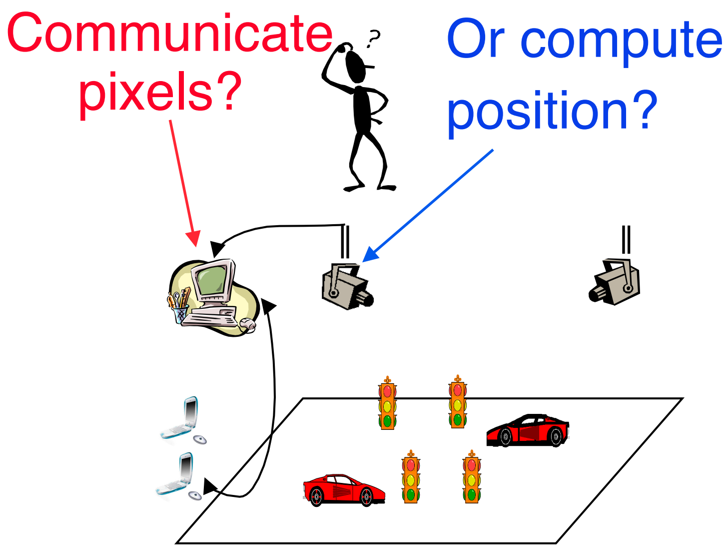
Example of capabilities: Component migration

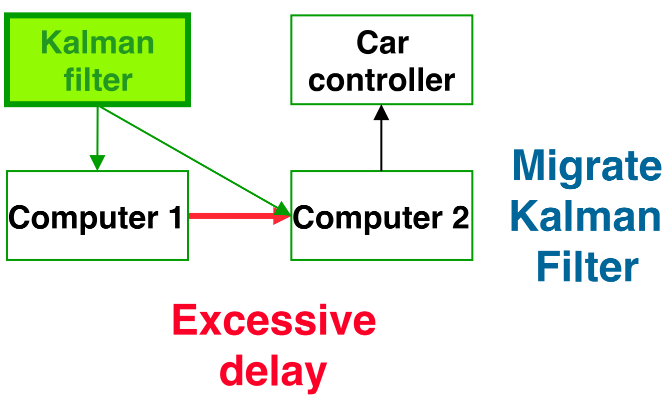




(Baliga, Graham & K ‘04)

Example of capabilities: Component migration





Real-time middleware

(Kim & K '09)

Thank you

References

- I-Hong Hou, V. Borkar and P. R. Kumar, “A Theory of QoS for Wireless.” *Proceedings of Infocom 2009*, April 19-25, 2009, Rio de Janeiro, Brazil.

- I-Hong Hou and P. R. Kumar, “Admission Control and Scheduling for QoS Guarantees for Variable- Bit-Rate Applications on Wireless Channels.” *Proceedings of MobiHoc 2009*, pp. 175–184. New Orleans, May 18-21, 2009.

- I-Hong Hou and P. R. Kumar, “Utility Maximization for Delay Constrained QoS in Wireless.” (http://arxiv.org/abs/0908.0362v1). July 31, 2009. To appear in Infocom 2010.

- I-Hong Hou and P. R. Kumar, “Scheduling Heterogeneous Real-Time Traffic over Fading Wireless Channels” (http://arxiv.org/abs/0908.0587v1). July 31, 2009. To appear in Infocom 2010.

Thank you