Cross-layer design issues for quality of service provisioning in wireless networks

Leandros Tassiulas
University of Thessaly, Volos, Greece

Cross layer design of wireless systems

Network layer engineers cannot completely ignore physical layer subtleties in the design of the higher layers

Segregated layers design approach to telecommunication systems

- Physical layer engineers
  - deliver a bit-pipe with certain rate and bit error rate
- Higher layer designers
  - connect the bit-pipes
  - Manage the information flow to deliver the service

Bit-pipe abstraction inadequate for multi-user wireless systems

Wireless channel common resource for a multiplicity of nodes in the same locality

- Point to point bit-pipe abstraction (i.e. through TDMA or FDMA) inadequate for high performance wireless(broadband)
- Doesn't exploit multi-user diversity

Volatile, time-varying Mobile wireless channel

- Erratic medium
- Availability highly volatile
- Unpredictably time varying
- Performance may benefit significantly if access layer decisions are channel state aware at the time scale of channel variation (i.e. milisecaonts for fast fading induced variations)

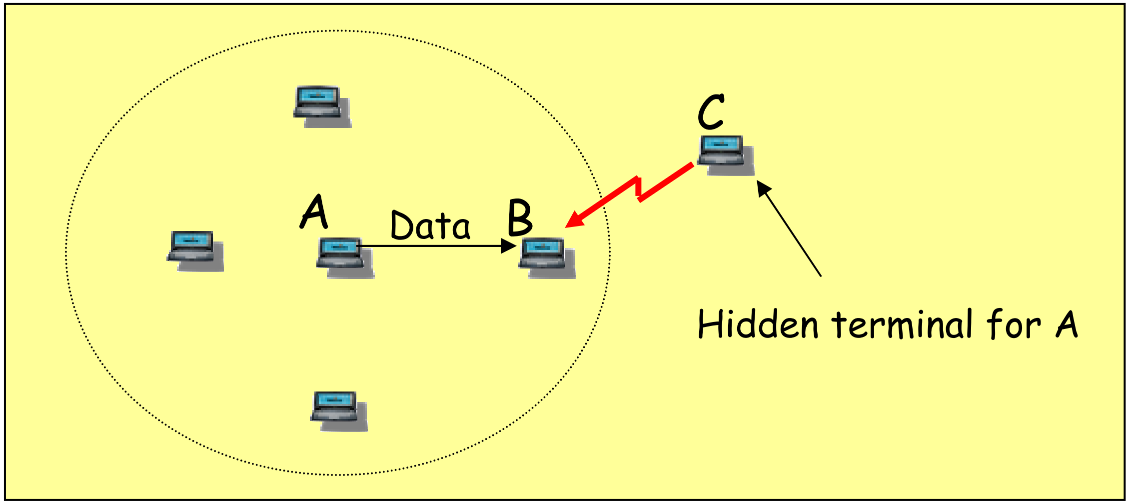
Sophisticated physical layer techniques may reach their full potential in performance improvement only in synergy with access layers

- Spatial processing techniques in multi antenna systems expected to provide an order of magnitude improvement in capacity
- Careful shaping of the radiation diagram suppresses interference in certain directions and improves SNR in other directions by concentrating signal energy
- Close synergy with access control to achieve desired improvement in:
  - node locationing
  - interaction with power control

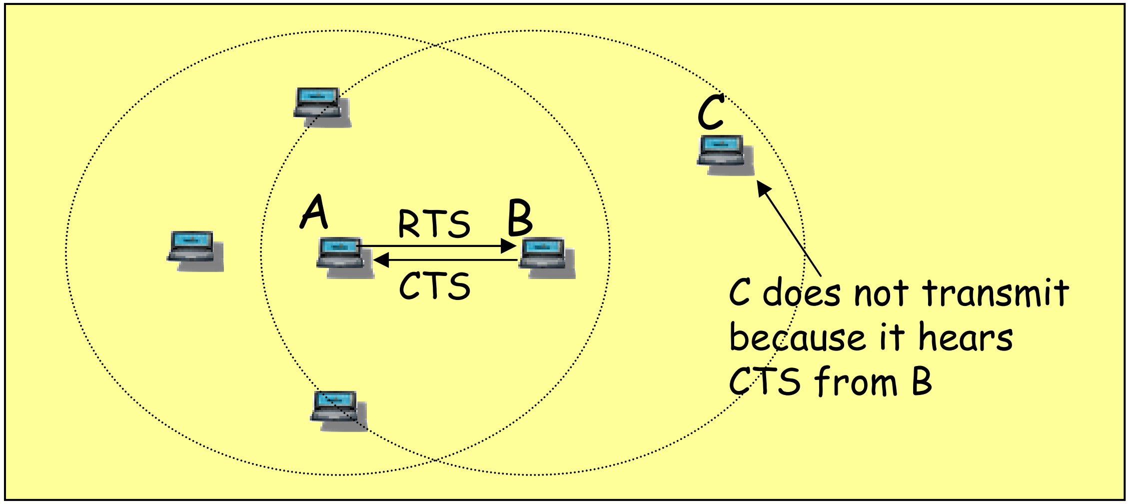
Signal propagation properties in the physical medium affect access control

- Medium sensing essential for random access protocols
- Straightforward in ethernet (CSMA/CD), signal propagation in coaxial cable very smooth
- Important challenge in WLAN (Hidden terminal problem) due to complicated nature of signal propagation in wireless

Carrier Sense alone inadequate in WLAN



RTS – CTS exchange to alleviate hidden terminal



- Directed transmissions in multiple antenna systems
- Medium sense ability diminished
- Hidden terminal numbers blow up
- Access protocol fails

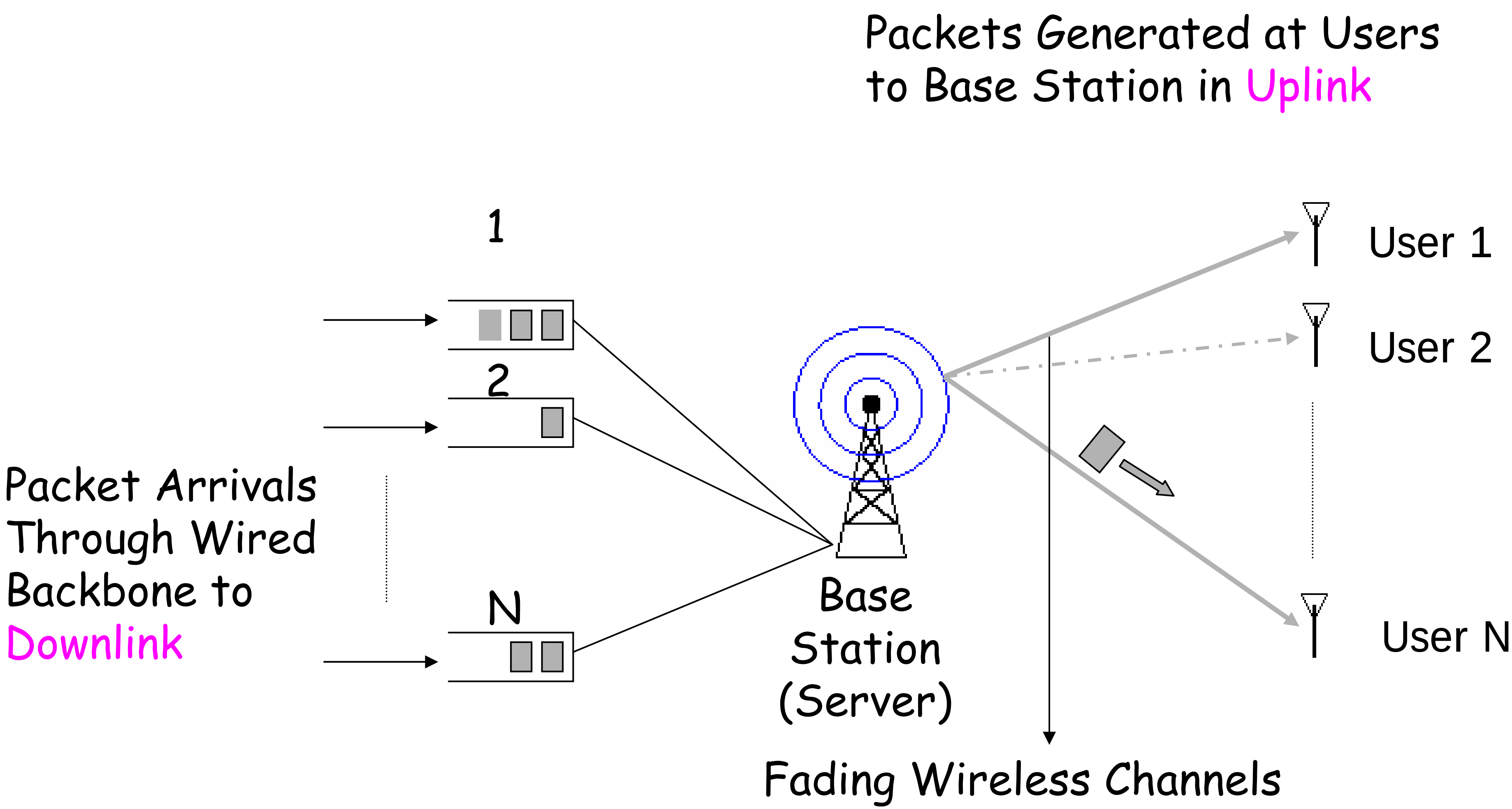
Tether-less nature of mobile wireless imposes new requirements: energy efficiency

- Energy conservation particularly important
- Issue to be addressed at multiple layers and modes of operation, physical, access, network and application

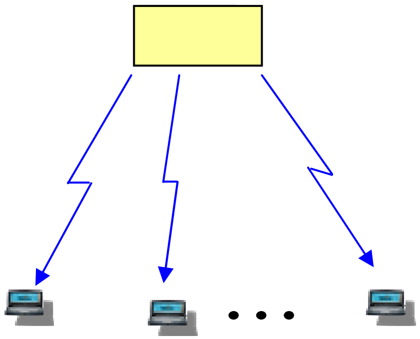
Cross-layer design issues

- Wireless channel common resource for a multiplicity of nodes in the same locality
- Volatile, time-varying Mobile wireless channel
- Sophisticated physical layer techniques may reach their full potential in performance improvement only in synergy with access layers
- Signal propagation properties in the physical medium affect access control
- Tether-less nature of mobile wireless imposes new requirements: energy efficiency

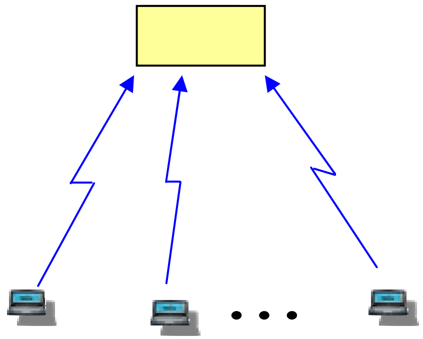
Fading multiuser channel



Information Theoretic Models



Broadcast channel



Multi-access channel

- Capacity studied by many, among others: Gallager, Knop,Humblet, Tse, Hanly

Information theoretic models miss temporal dimension of data traffic (all information available a-priori to the disposal of the encoders)

Recent information theoretic results for multiaccess channel

- Capacity characterization for fading channels with full channel information available to the encoder (Tse, Hanly 99)

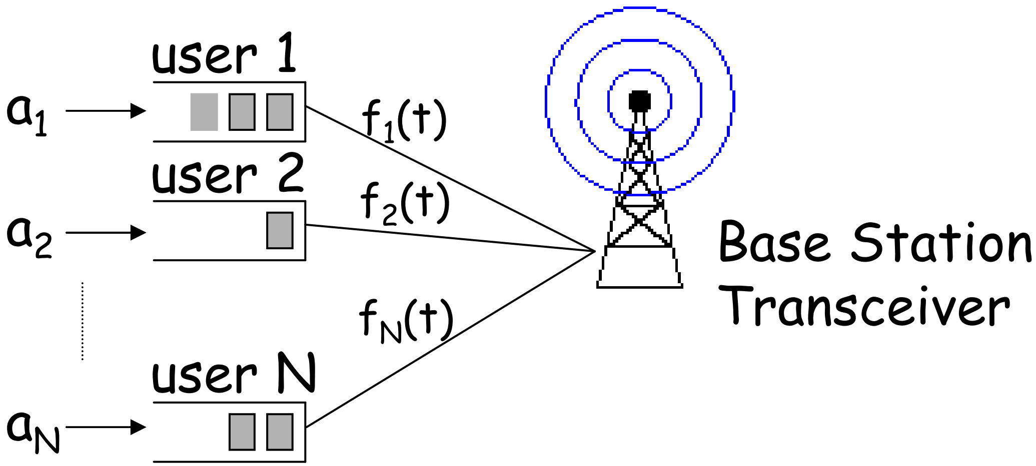
Symmetric fading, limitation on the average power transmission (Knopp, Humblet 95)

Capacity achieved by "Multiuser water filling" :

- When all channel states are sufficiently unfavorable no one transmits
- Otherwise only the user with best channel transmits

A (gross) model of a fading multi access system with teletraffic

(Tassiulas, Ephremides, 93)



- Fading state of channel n $f_n(t) = \begin{cases} 1 & \text{Channel available} \\ 0 & \text{Channel unavailable} \end{cases}$
- a_n: user n traffic generation rate
- Throughput capacity region

C = { (a_1, …, a_N) for which stable operation feasible for some appropriate access control policy}

Independent Bernoulli fading processes

$P[f_n(t) = 1] = P_n$

Capacity region characterization

$$\sum_{n \in S} a_n \leq 1 - \prod_{n \in S} (1 - P_n) \quad \forall S \in \{1, \ldots, N\}$$

Maximum throughput access policy

"Among the users with good channel enable the one with largest backlog"

Adaptive, no need for channel statistics

Refined model of a fading multi access system with teletraffic

(Tassiulas 97)

Multi access channel state f(t)
(arbitrary Hidden Markov process)

Access allocation decision a(t)
(may include e.g. user encoding order in successive decoding

Rate function (R1(t), …, RN(t)) = R(f(t), a(t))

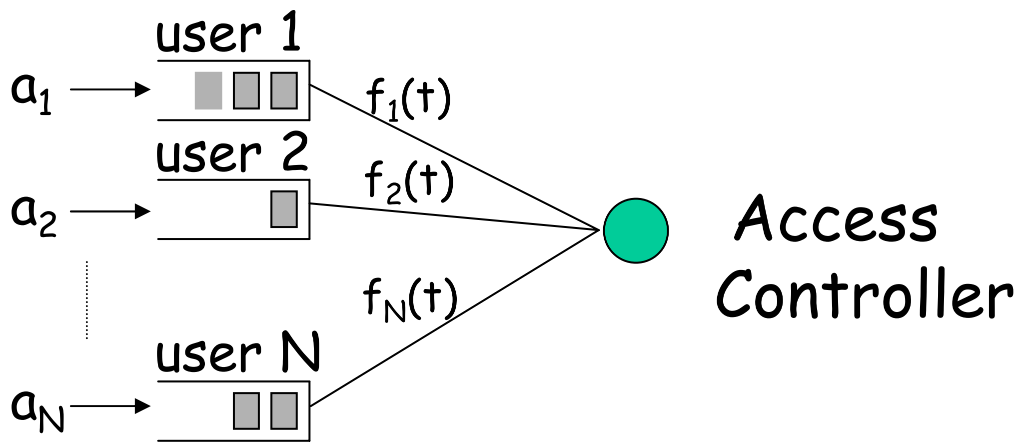
Throughput region:
Convex combination of polytopes like those in "gross" model

Maximum throughput policy
Select a(t) to maximize Σb(t)* R(f(t), a(t))

Recent work by
Yeh 02, Neely-Modiano 01-03, Tong 02
throughput results in models that incorporate
power control and limited energy constraints

Providing some notion of Quality of Service beyond maximum throughput

(Tsimbonis, Georgiadis, Tassiulas 03)



- Rn: long term average throughput rate of user n

Linear Quantitative Quality of Service Objective

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

Decreasing weights reflect decreasing user priorities from 1 to N

Find Access Control to maximize

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

Why linear objective ?

Crucial step for achieving more general objectives like sums of convex functions that approximate fair allocation (maximum or proportional)

Optimal Policy in two extreme cases

Case A

- $(a_1, a_2, \ldots, a_N)$ belongs to feasible throughput region
- The LCQ policy achieves $r_n = a_n$, all n, therefore optimal

Case B

- $(a_1, a_2, \ldots, a_N)$ too big, all users operate in saturated (backlogged) mode
- Strict priority does the job
- Among the users with available channel ($f_n(t) = 1$) activate the one with largest weight
- Backlog independent

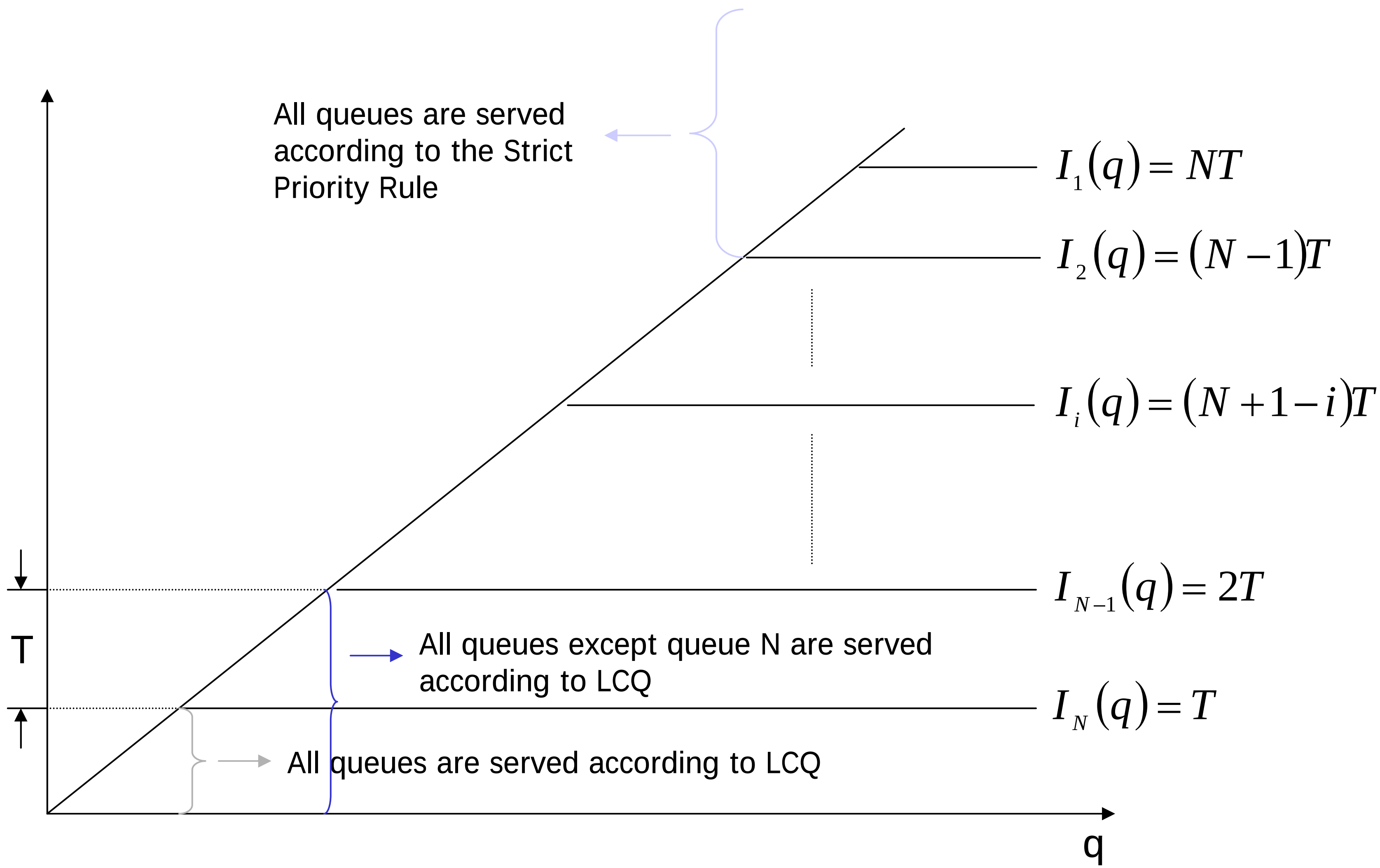
Optimal policy in general case

Index In(t) = I(bn(t)) associated with user n where bn(t) backlog of user n and

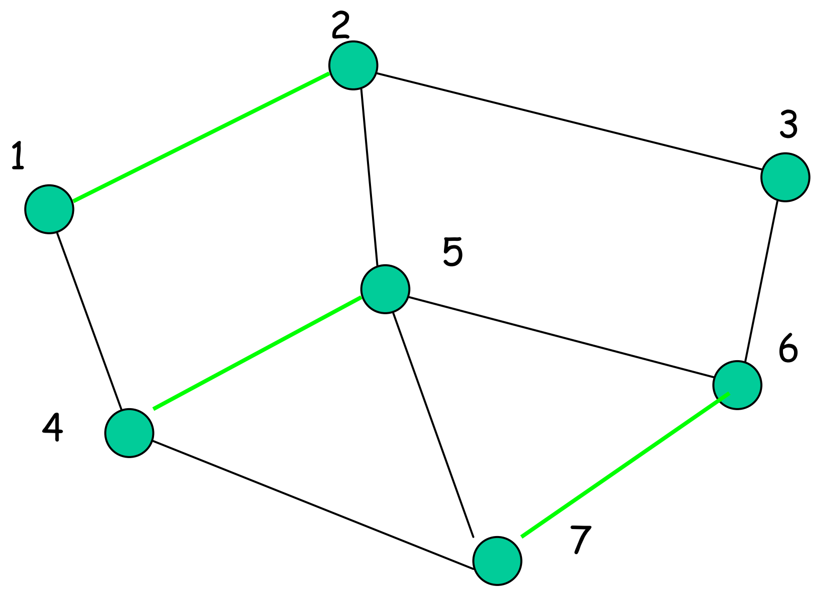
$$I(b) = \begin{cases} b, & \text{if } b \leq (N+1-i)T \\ (N+1-i)T, & \text{if } b > (N+1-i)T \end{cases}$$

Among the users with available channel select at each slot the one with largest index

Operating Diagram



Sharing the locally common channel in different neighborhoods of a multihop ad-hoc network



- Simultaneous transmissions of several links may result in conflicts
- Transmission conflict conditions depend on: signaling (spread spectrum, narrowband,..), number of wireless transceivers per node, directivity of transmissions,..

Access Controller dictates transmissions at each time

either explicitly through a conflict free link activation schedule

or implicitly i.e successful transmissions in a distributed random access scheme

Sets of non-conflicting links that may transmit Simultaneously, transmission sets

Transmission scheduling policy designates the transmission set at each time

$$I(t), \quad t=1,2,\dots$$

Tele-traffic Throughput Capacity

Feasible rate vector is any rate vector that is realized by some scheduling policy

Capacity region C of a multihop wireless network is the collection of all feasible rate vectors

The capacity region is equal to the convex hall of A

Access Control to maximize throughput

- The traffic generated at the different nodes is such that the arrival rate vector belongs to the capacity region
- The network controller observes the backlogs and schedules the transmissions
- The scheduling policy that selects for scheduling the maximum weighted transmission set at each time where the weights are the current backlogs guarantees stable operation
- In combination with a hop-by-hop routing algorithm maximum throughput for end-to-end

(Tassiulas Ephremides 92, IEEE Trans on Automatic Control)

QoS provisioning beyond throughput ?

QoS provisioning in wireline systems

Link schedulers for QoS provisioning: RR,FQ,GPS, PGPS, etc..

Network level fair session rate allocation

End-to-end fair rate allocation through fair queueing at the link level and hop-by-hop flow control

Challenges in wireless

- Neighboring links cannot be scheduled independently due to local interdependencies in transmission.
- When a link is scheduled it is as if it receives service from both its end nodes.
- Each node is viewed as an independent server that allocates service to the links emenating from him.
- A link can be served only if its two end nodes are synchronized to provide their service at the same time.
- Challenge: Scheduling at a node should be done in coordination with its neighbors.

QoS provisioning beyond throughput

Operate the network with a scheduling policy that provides a feasible rate vector that satisfies certain minimum rate requirements and maximum rate constraints and furthermore it is maxmin fair

Maxmin Fairness

A rate vector is *maxmin fair* if subject to feasibility, one **can not increase** the rate of a flow, **without decreasing** the rate of another flow having **equal or lesser** rate

Dynamic scheduling for maxmin fairness in multihop ad-hoc (Sarkar, Tassiulas 02)

Our Set-up

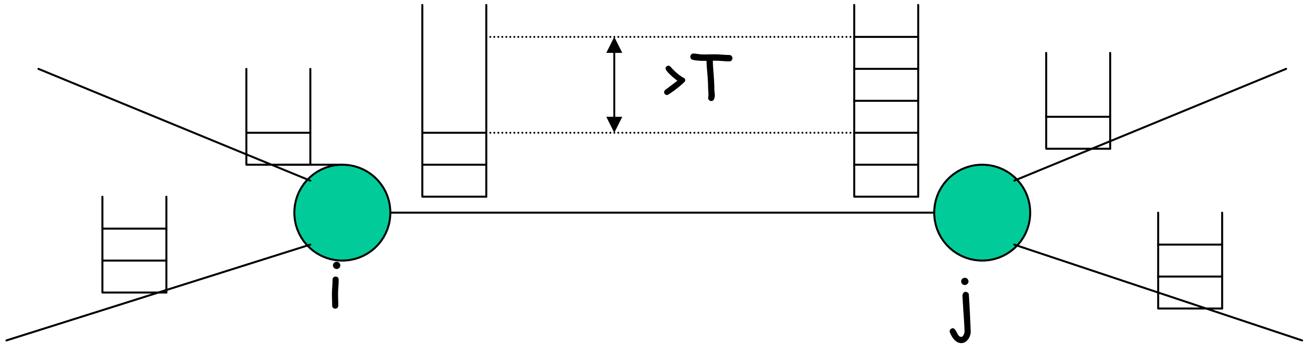
- There is a single flow per link in one of the two directions
- The transmission conflict constraints are that links sharing the same node cannot be active simultaneously
- Bipartite topology graph
- Saturated packet buffers

Scheduling for rate guarantees in wireless

- Each node allocates service tokens to the links emenating from him in a "round-robin-like" fashion.
- Each link maintains two service token buckets, one for each end node, where it stores the tokens received by the corresponding end node.
- The "service credit" of the link equals to the minimum of the two service token buckets.
- The collection of non-conflicting links with maximum service credit is selected for service at each slot.
- Whenever a link is served one token is deducted from each one of its token buckets.

Service token allocation: *saturated system*

- Assume that each link has an infinite packet supply
- A link (i,j) is eligible to receive a service token at slot t from node i if the size of the token bucket i of the link does not exceed the size of the token bucket j of the link by more than T service tokens.
- Each node i allocates the service tokens in a round-robin fashion, considering at each slot only the elligible links at that slot



Link elligibility is node dependent, i.e. a link may be elligible for service by one of its nodes (i) and ineligible by the other (j)

Maxmin-fairness in saturated system

$R(t)$ the vector of tokens allocated to each link in $[0,t)$

There is a threshold T such that

$$R(t)/t \longrightarrow R_0 \quad \text{as } t \longrightarrow \infty$$

where R_0 is maxmin vector in the region of achievable rate vectors

The token buffer lengths are bounded

System with arrivals

The packets in link (i,j) are generated according to an arrival process with rate aij.

The service token allocation is done as in the saturated system with the difference that if a link packet buffer is empty then the link is ineligible for service.

Link scheduling relies on service credits and not on queue lengths

The service rate vector of the links converges to the maxmin feasible service rate vector. Furthermore the links for which the arrival rate equals the service rate the packet length process is stable.

Other Issues

- Multiple flows per link: easily extensible, each flow its own buckets
- Other topologies and wireless constraints: extensions possible with multiple credit buckets per link, one for each constraint that affects the link.
- Minimal control information exchange and only between one hop away neighbors
- Distributed versus centralized: in the current algorithm the maximum matching computation is the only centralized part.

Multihop flows

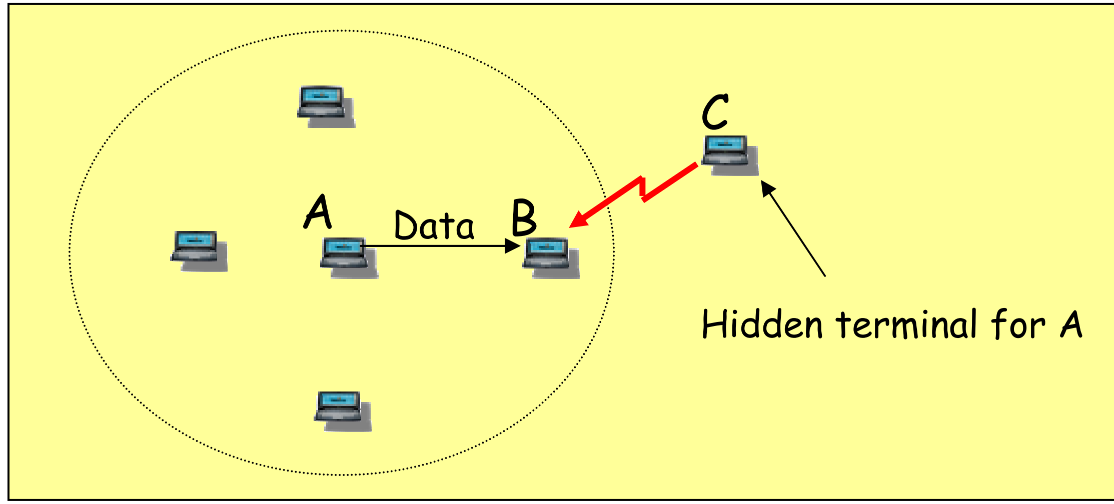
A combination of the token based scheduling algorithm with a back pressure hop-by-hop algorithm achieve end-to-end fair rate allocation.

(Sarkar,Tassiulas 03)

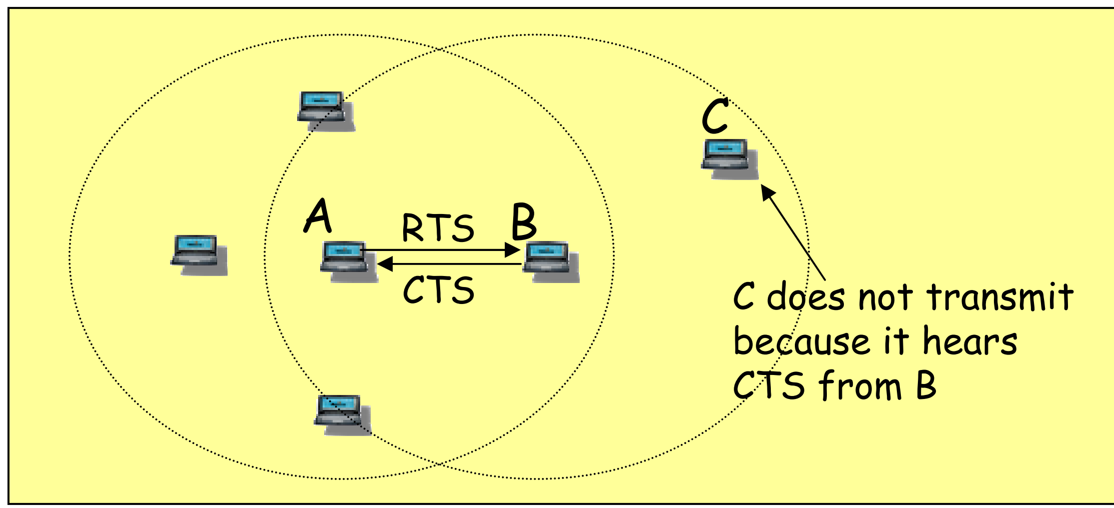
Modify 802.11 MAC for directional transmission systems

- Access Protocol designer needs to deal effectively with several new challenges in a system with directed transmissions
  - Hidden terminal explosion
  - Deafness
  - Node locating

Carrier Sense alone inadequate in WLAN

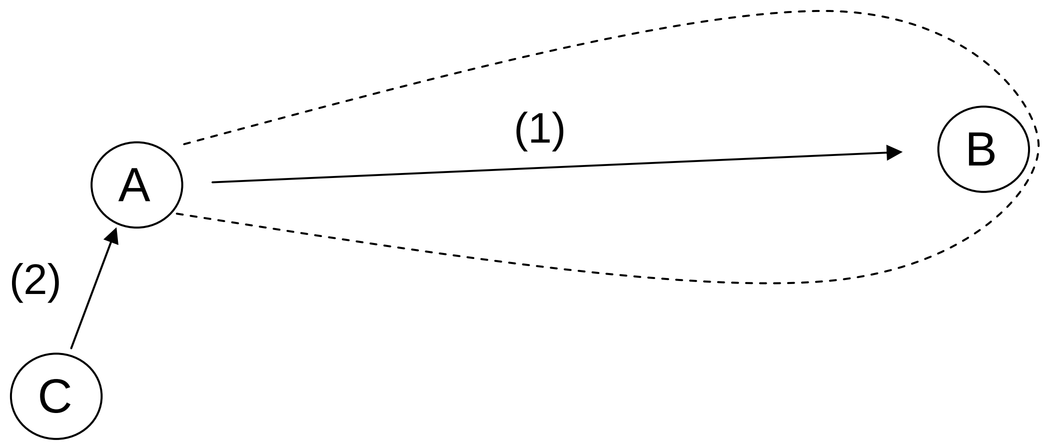


RTS – CTS exchange to alleviate hidden terminal



- Directed transmissions in multiple antenna systems
- Medium sense ability diminished
- Hidden terminal numbers blow up
- Access protocol fails

Deafness

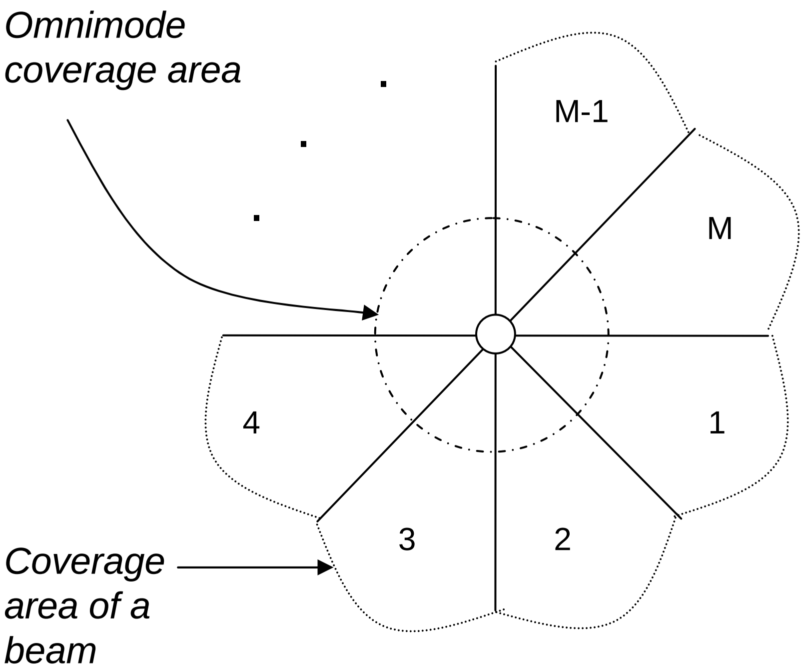


- Transmission C -> A does not collide with ongoing A -> B but fails since has no means to realize that A is busy transmitting to B

A MAC protocol for full exploitation of Directional Antennas in Ad-hoc Wireless Networks

Korakis, Jakllari, Tassiulas 03

- ✓ Circular Directional RTS

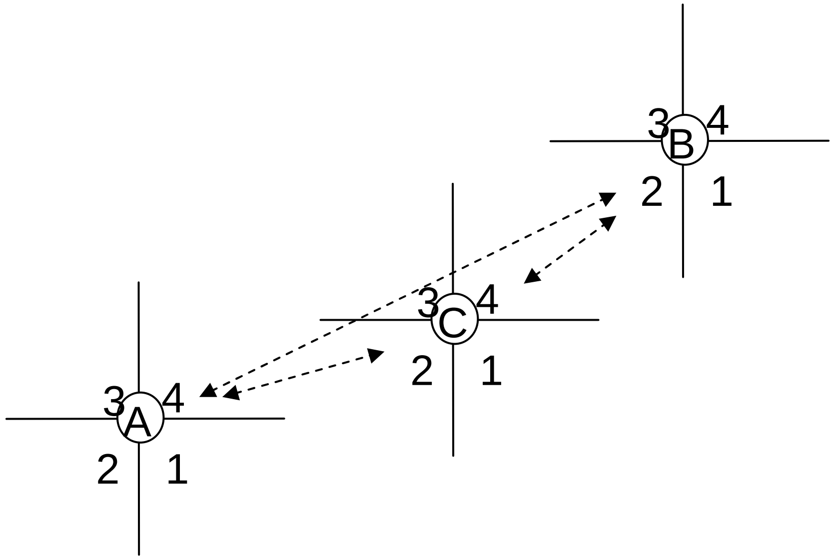


- ✓ Directional CTS
- ✓ Directional Data frame
- ✓ Directional Ack

Every station maintains a Location Table

For node C:

Me	*My Beam*	*Neighbor*	*Neighbor's Beam*
C	2	A	4
C	4	B	2



✓ Every frame contains information about the beam by which it is transmitted

✓ In the reception of a frame the receiver updates the corresponding record

✓ RTS contains the record of the Location Table, that corresponds to the transmission.

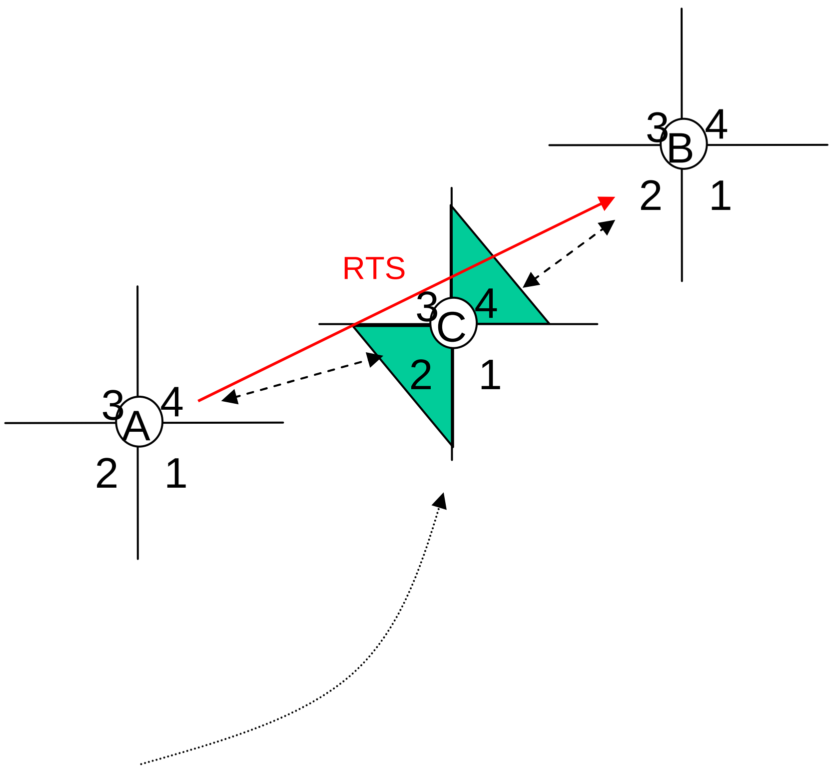
The use of the Location Table leads to an efficient update of D-NAV

Added information into RTS:

Me	My Beam	Neighbor	Neighbor's Beam
A	4	B	2

Location Table for node C:

Me	My Beam	Neighbor	Neighbor's Beam
C	2	A	4
C	4	B	2



✓ C defers transmission towards beams C2 and C4

Cross-layer design issues

- Wireless channel common resource for a multiplicity of nodes in the same locality
- Volatile, time-varying Mobile wireless channel
- Sophisticated physical layer techniques may reach their full potential in performance improvement only in synergy with access layers
- Signal propagation properties in the physical medium affect access control
- Tether-less nature of mobile wireless imposes new requirements: energy efficiency