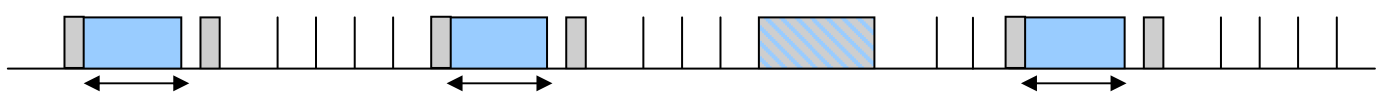
4. Performance Analysis of IEEE 802.11 DCF

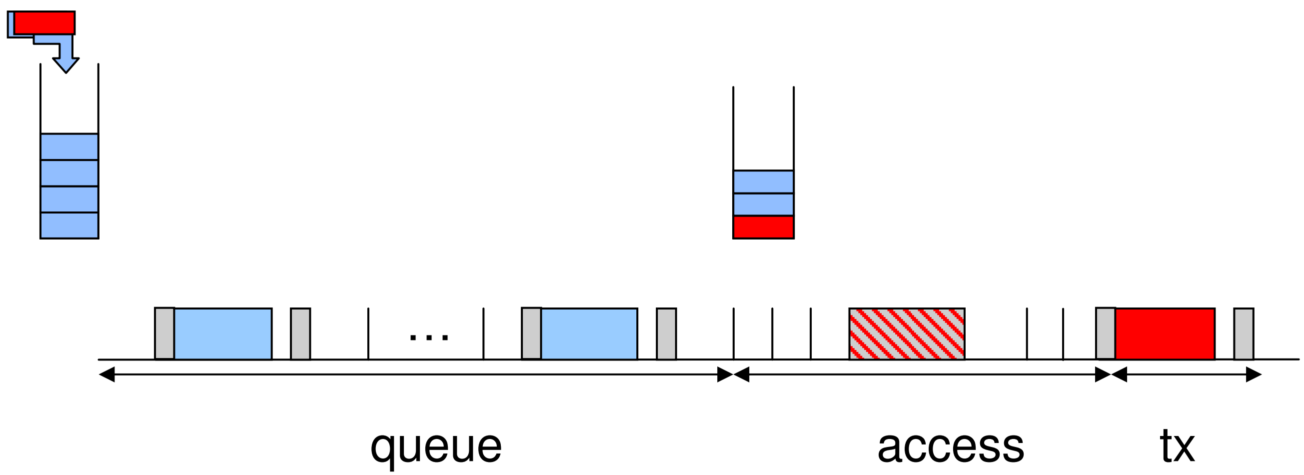
Throughput

- ➔ **Represents the system efficiency (different definitions).**
- ➔ **Overheads: backoff, monitoring times, collisions, headers, ACK/RTS/CTS.**
  - ⇨ Payload bits transmitted in average per each second
  - ⇨ Fraction of the channel time used for payload transmissions (normalized throughput)

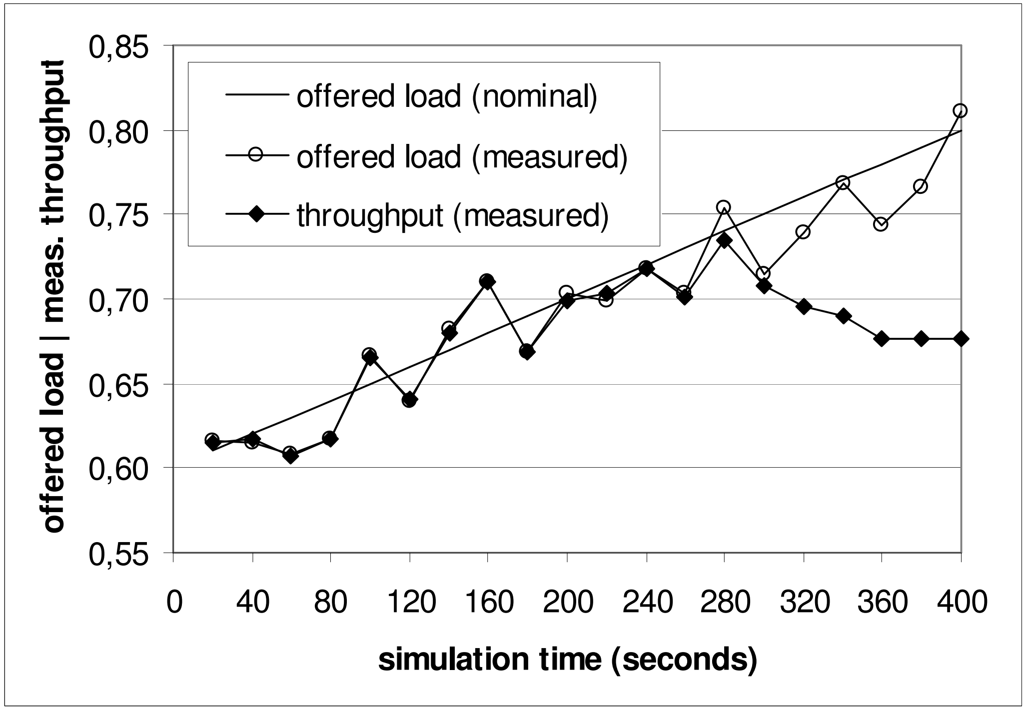


Delay

- ➔ **Time required for a packet to reach the destination after it leaves the source.**
- ➔ **Three different components: queue delay, service access delay, transmission delay.**



Saturation Analysis



Fixed number of stations, varying arrival rate

Arrival rate higher than maximum throughput -> transmission queues build up until saturation (always full)

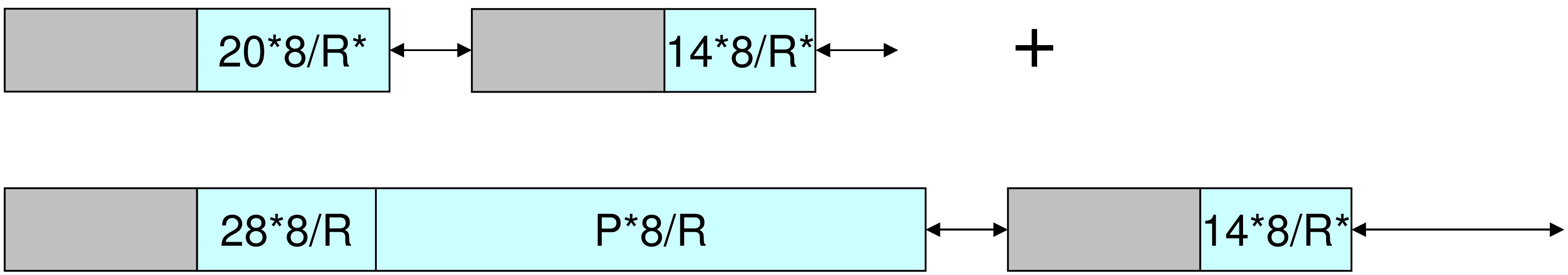
“Saturation throughput”: limit reached by the system throughput as the offered load increases

DCF Overheads

Frame Transmission Time

➔ **Let R be the data rate and R* the control rate.**

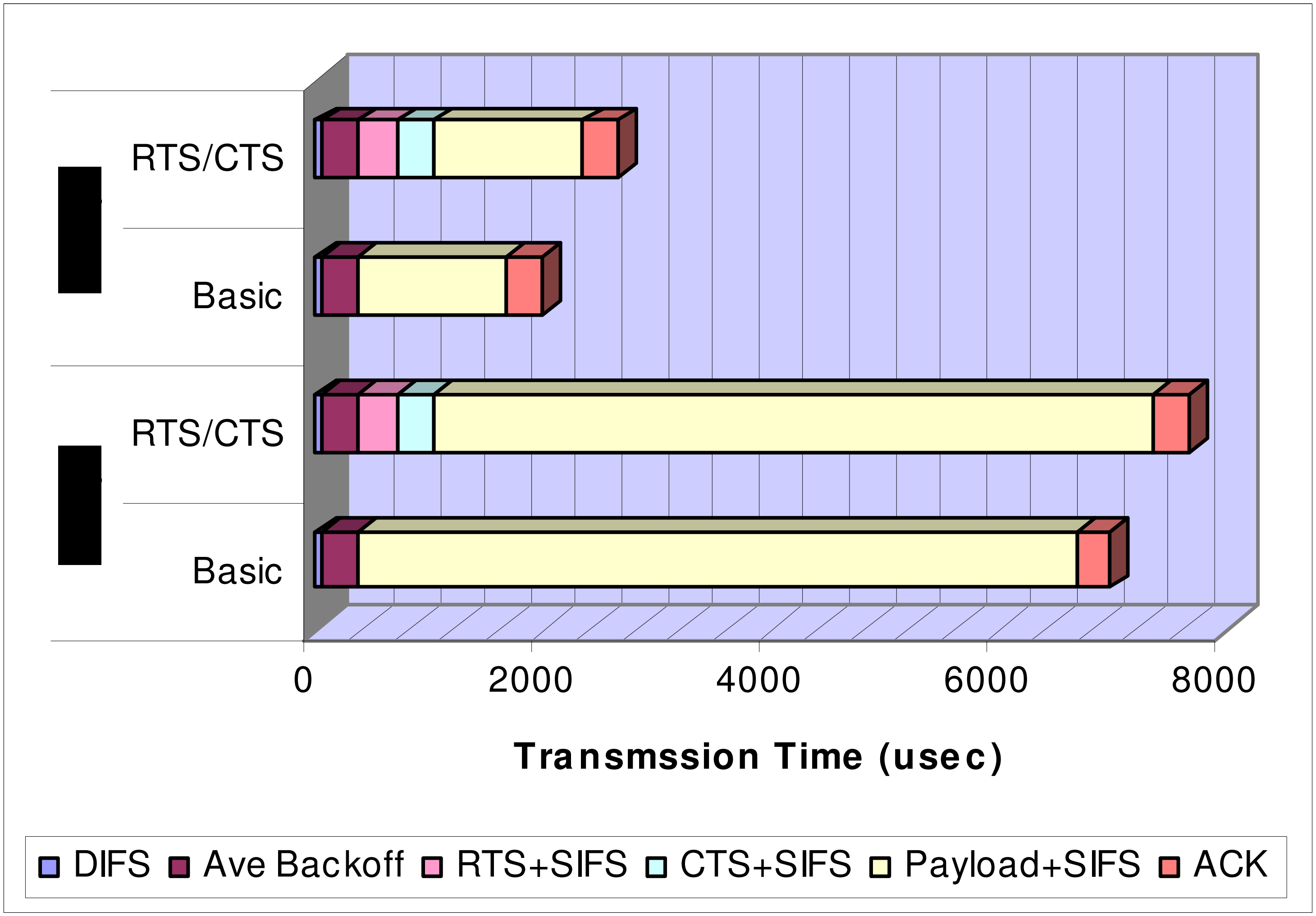
➔ **Let P be the MSDU size [byte].**



$$T_{FRAME} = T_{MPDU} + SIFS + T_{ACK} + DIFS$$

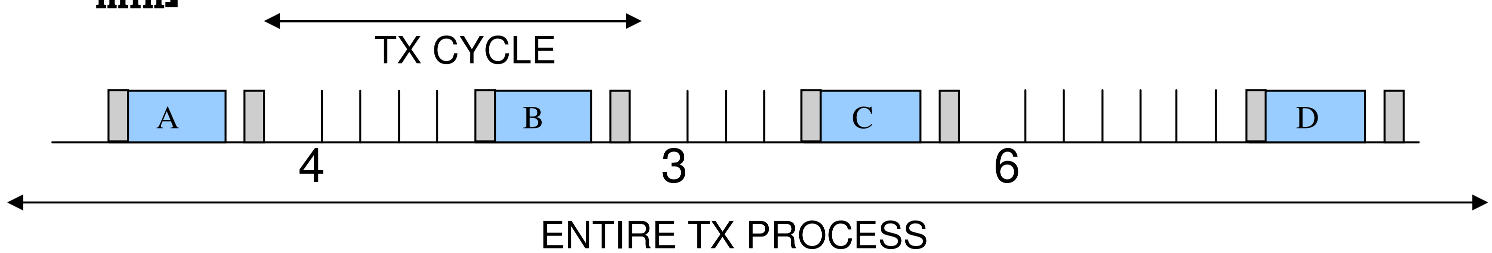
$$T_{FRAME} = T_{RTS} + SIFS + T_{CTS} + SIFS + T_{MPDU} + SIFS + T_{ACK} + DIFS$$

Overheads @ different rates
(P=1500 bytes)



Protocol Overhead

- ➔ Suppose to have just a single station, with a never empty queue
- ➔ Each transmission is originated after a backoff counter expiration
- ➔ Since no collision is possible, and no channel error is considered, each backoff is extracted in the range [0, CW_{min}]

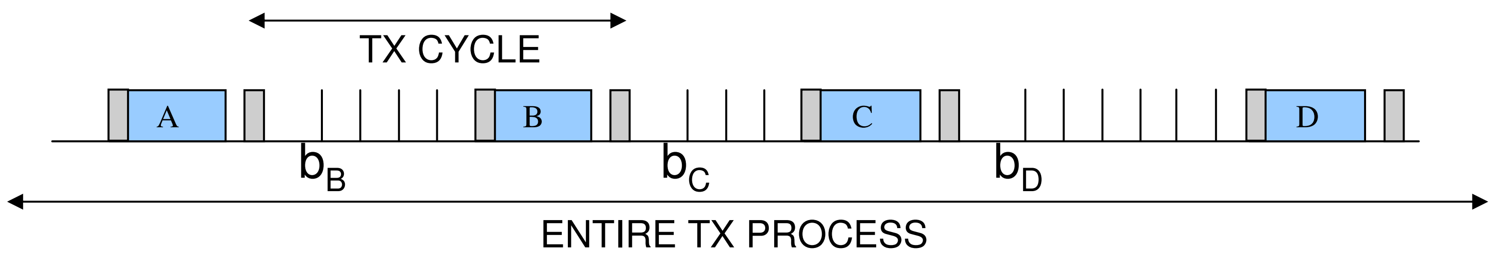


- ➔ Different transmission cycles on the channel, composed of:

  1) frame transmission time, which depends on the MSDU size;

  2) random delay time, which depends on the backoff extraction.

Max Throughput Computation



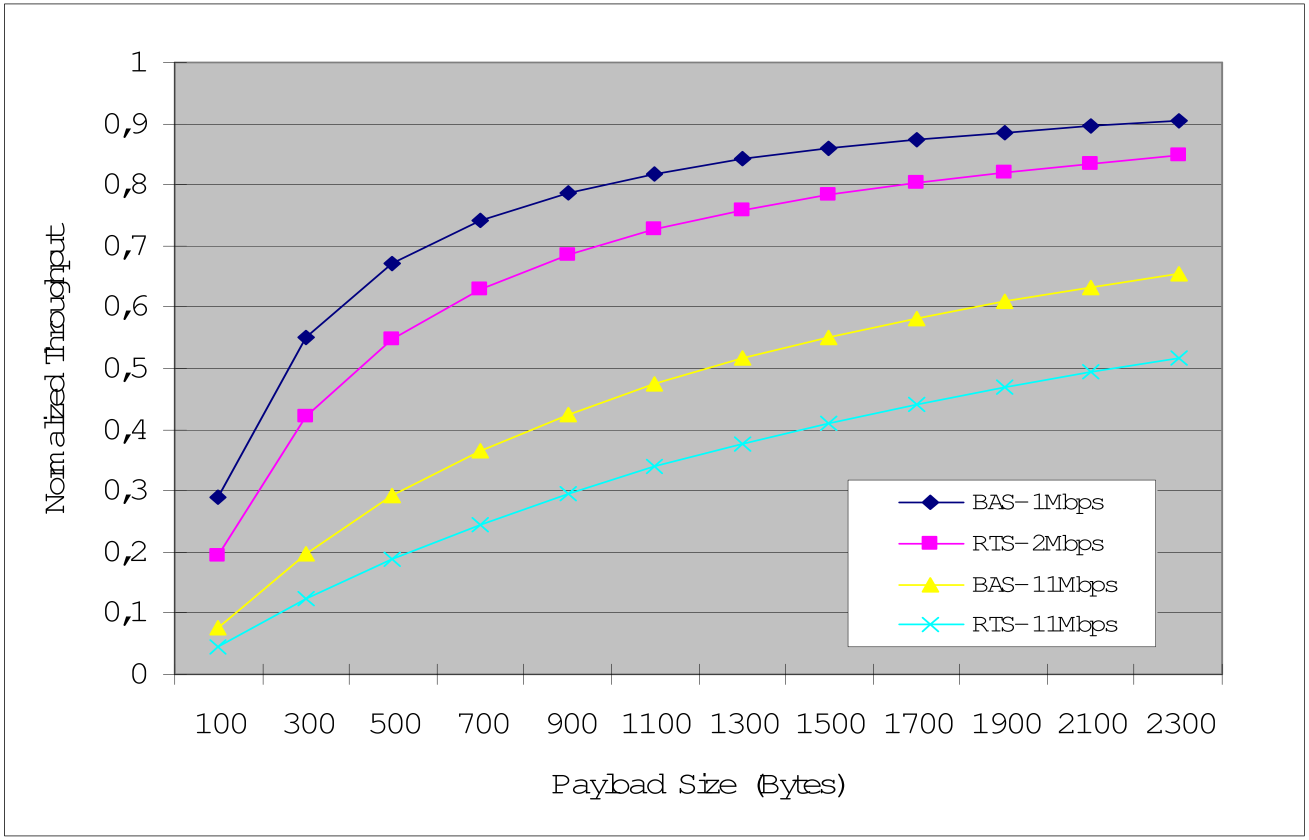
➔ **From the throughput definition:** $S = \frac{\sum P_i}{\sum \left(T_{FRAME_i} + b_i \right)}$

➔ **After each transmission, backoff counters are regenerated.**
From Renewal Theory: $S = \frac{E[P]}{E[T_{FRAME}] + E[b]}$

➔ **In the case of fixed packet size, given CW$_{\min}$:**

$$S = \frac{P}{T_{FRAME} + \sigma CW_{\min} / 2}$$

Max Throughput (normalized)



Saturation Throughput Analysis

802.11 DCF Bianchi's model approach

Step 1: Discrete-Time model of backoff for <u>tagged</u> STA

- ⇨ background STAs "summarize" into a unique collision probability value *p*

Step 2: find transmission prob. t

- ⇨ Result: t versus *p* non-lin function

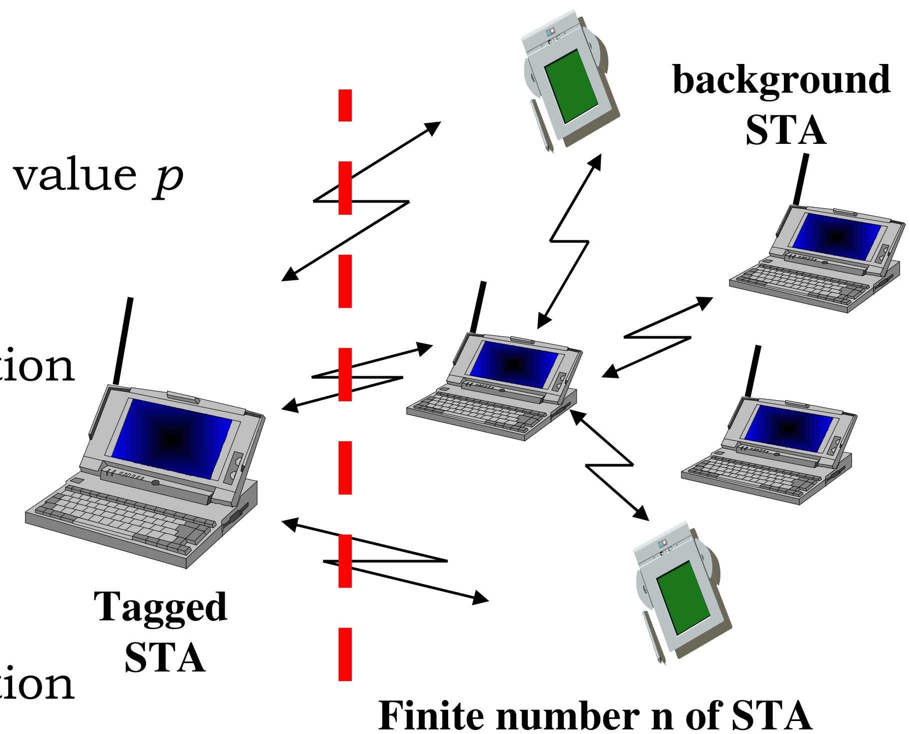
Step 3: assume background STAs behave as tagged STA, i.e. transmit with probability t

- ⇨ Result: *p* versus t non-lin function

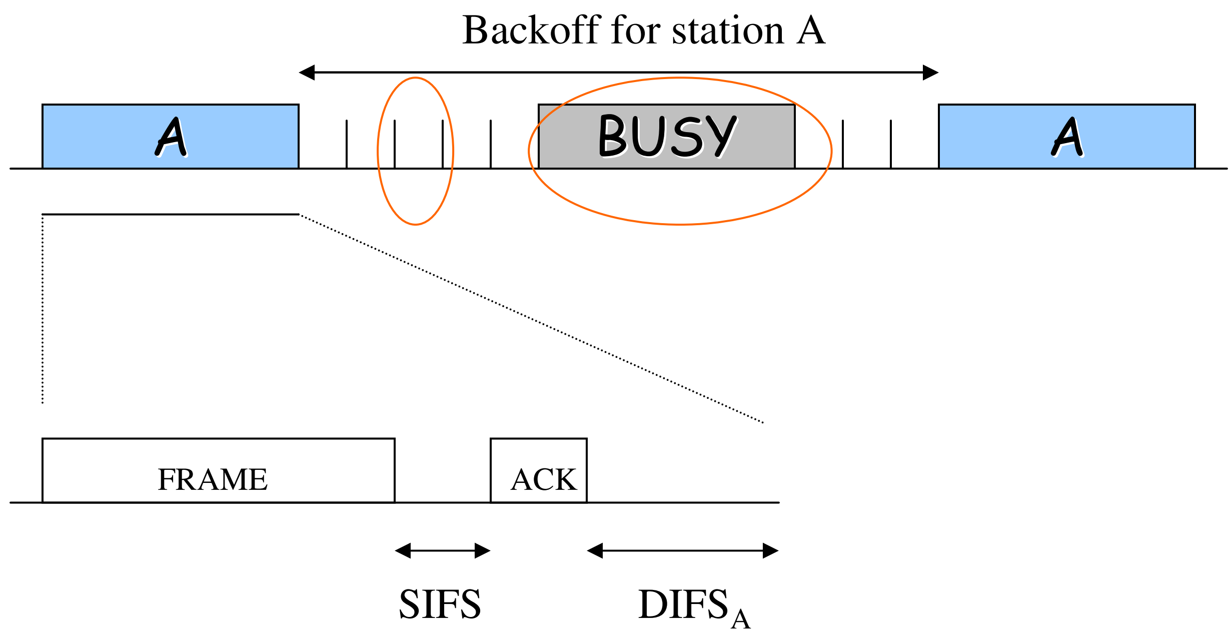
Step 4: solve non linear 2eqs system

Step 5: find performance figures

- ⇨ Throughput, Delay



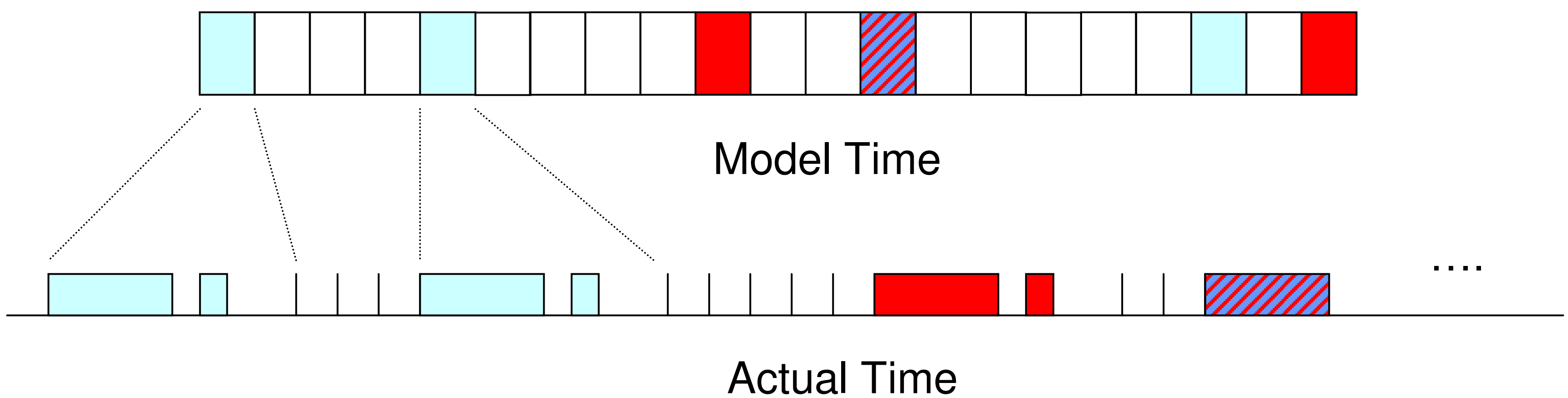
Discrete Not Uniform Time scale



Slot time = Time interval between two consecutive backoff time counter decrements:

- idle backoff slot
- busy time + DIFS + 1 backoff slot

DCF as τ-persistent CSMA



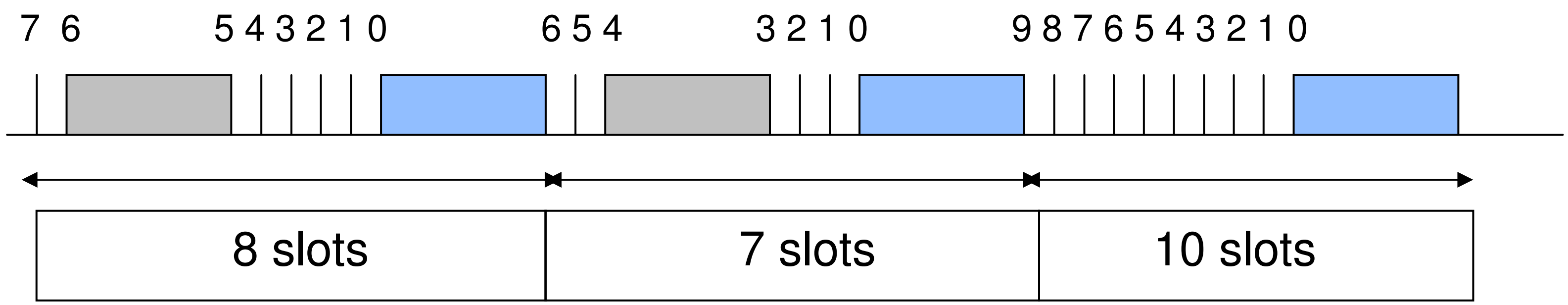
-In each system slot, each station accesses with probability τ (and does not access with probability 1-τ).

-Each system slot can assume 3 different sizes: idle slot, successful slot, collision slot.

-The key assumption is that τ is fixed slot by slot (and then also the collision probability p).

Channel Access Probability τ

Given the time scale, 1 transmission every backoff + 1 slot



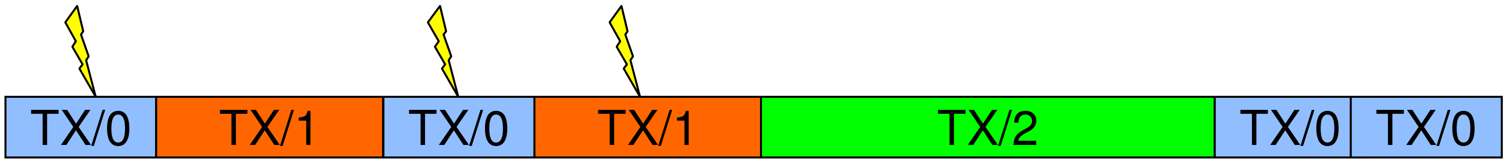
Note that the access cycle length in discrete time slot is not related to the actual channel time (e.g., the last cycle is the shortest cycle according the channel time, but the longest one according to the model time)

Whenever $W_0=W_1=..=W_R=W$, from renewal theory: $\tau = 1 / (W/2+1) = 1/ (E[W] + 1)$

In general, the access time depends on the *backoff stage,* i.e. on the number of consecutive collisions..

N.B. Access cycle = transmission cycle for a given station

τ=f(E[Access Cycle Length])



The average length of the access cycle conditioned by the fact that **i** collisions have been experienced is: $W_i/2 + 1$.

To evaluate the average access cycle length we have to compute the probability of the conditioning event Pr(s=i/TX) (i.e. the probability to be in backoff stage **i**, given that we are transmitting).

In our example, TX/0 occurs with probability 4/7, TX/1 with probability 2/7 and TX/2 with probability 1/7-> $E[W] = 4/7 \cdot W_0 + 2/7 \cdot W_1 + 1/7 \cdot W_2$.

$$\tau = \frac{1}{\sum_{i=0}^{R} \frac{W_i}{2} \Pr(s = i/TX) + 1}$$

Formal derivation

$$P(TX)P(s=i/TX)=P(s=i)P(TX/s=i)$$

$$P(s=i)=P(TX)\frac{P(s=i/TX)}{P(TX/s=i)}$$

$$\sum_{i=o}^{R}P(s=i)=1\rightarrow\tau=P(TX)=\frac{1}{\sum_{i=0}^{R}\frac{P(s=i/TX)}{P(TX/s=i)}}$$

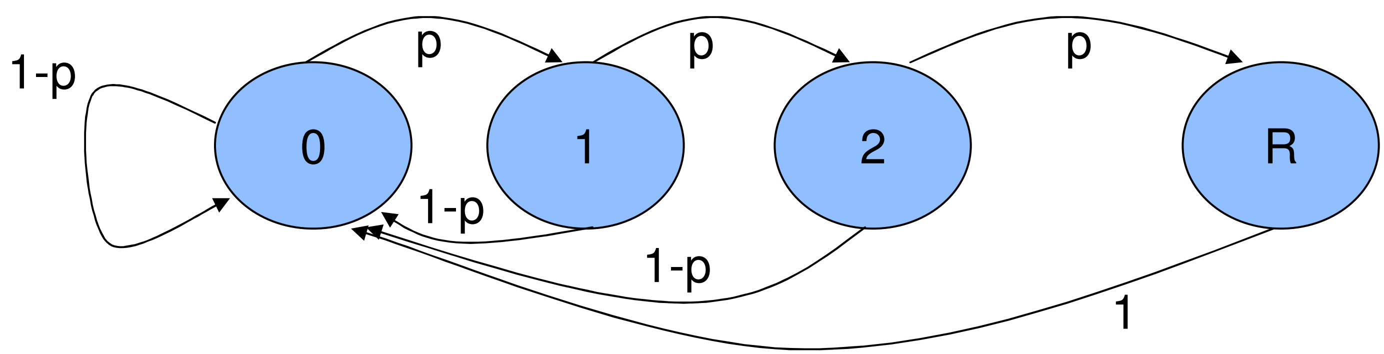
Backoff stage Probablity

Suppose to know the collision probability p:

We are in stage **i** if we were in stage **i-1** in the previous attempt and we experienced a collision: Pr(s=i/TX)=Pr(s=i-1/TX) p;

After a success or after R collisions, we come back to stage 0. Then, this probability has a geometric distribution:

$$P(s=i\,|\,TX)=\frac{(1-p)p^{i}}{1-p^{R+1}} \quad i\in(0,1,\ldots R)$$

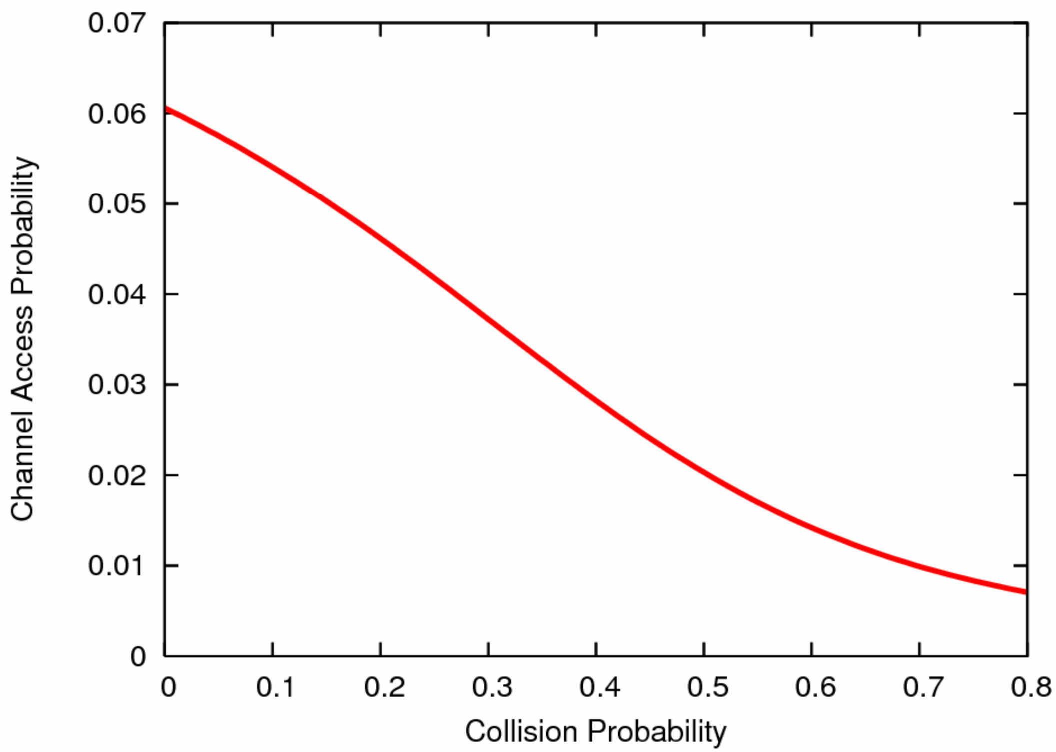


Channel Access Probability τ (2)

We can finally express the channel access probability as a function of the collision probability p and of the average backoff values $W_i/2$:

$$\tau = \frac{1}{\frac{1-p}{1-p^{R+1}} \sum_{i=0}^{R} \frac{W_i}{2} p^i + 1}$$

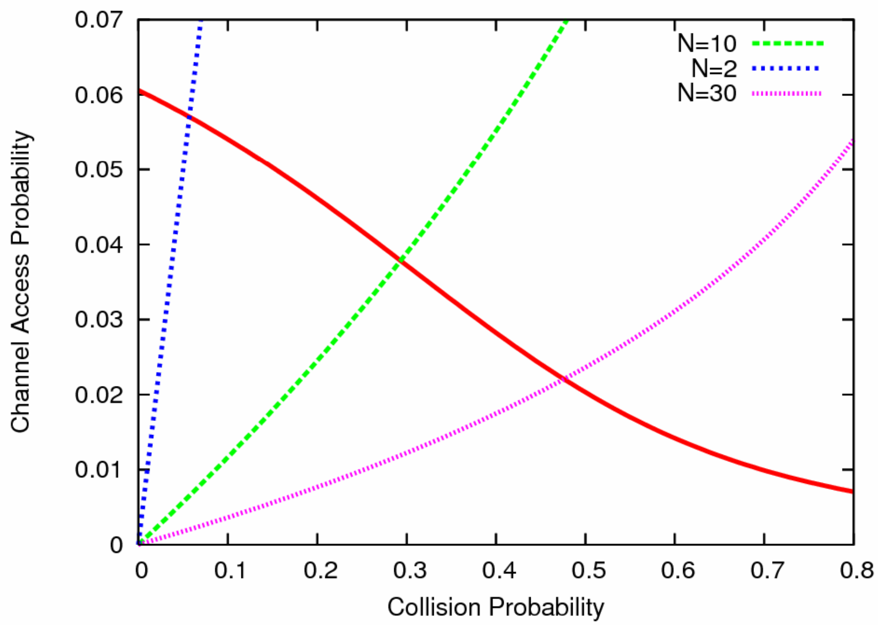
Note: τ does not depend on the backoff value distribution, but only on the average value!!!



How much is p???

The conditional collision probability p, i.e. the probability to experience a collision in a given slot, given that the tagged station is transmitting, is the probability that at least one of the other N-1 stations is accessing the channel.

If we assume that all the stations have the same behavior, and then access the channel with probability τ, it is easily expressed as: $p=1-(1-\tau)^{N-1}$



From τ&p to Throughput Performance

$$S = \frac{P_s E[P]}{P_s E[T_{SUCC}] + P_c E[T_{COLL}] + P_{idle}\sigma} = \frac{P_s E[P]}{E[slot]}$$

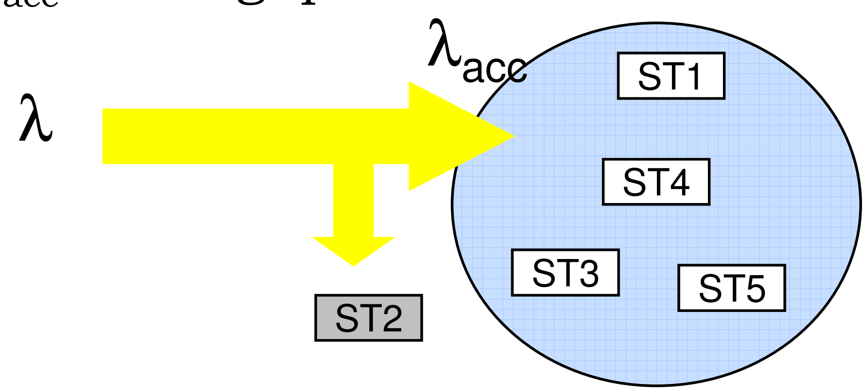
P_s: Pr success in a contention slot = $N\tau(1-\tau)^{N-1}$=N $\tau(1-p)$

$P_{idle} = 1- (1- \tau)^N$

$P_c = 1-P_S-P_{idle}$

Delay Computation

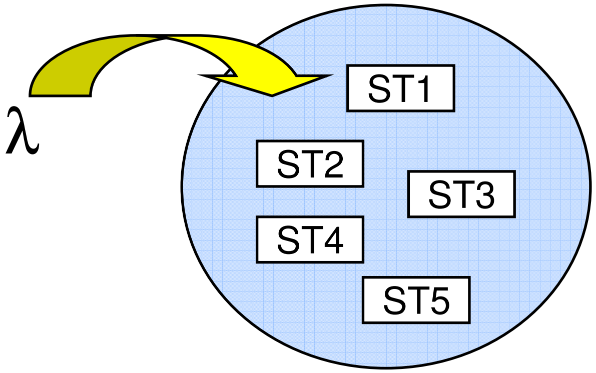
- ➔ **Very easy, via Little's Result!**
  - ⇨ Clients: Contending packets which will be ultimately delivered
  - ⇨ Server: DCF protocol
  - ⇨ Delay: system permanence time $\mathbf{D = E[N] / \lambda_{acc}}$
- ➔ **In static scenarios and saturation:**
  - ⇨ A new client is accepted if the packet is transmitted before the retry limit expiration
  - ⇨ Arrival rate λ: e new packet arrives to the system :
    - →When a packet is successfully transmitted
    - →When a packet is dropped because of a retry limit expiration
  - ⇨ Accepted traffic λ_{acc}=throughput!

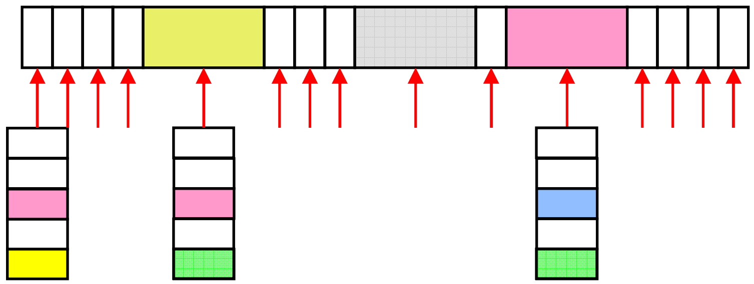


Retry Limit = ∞

➔λ = **Throughput [pk/s]**

$$D = \frac{N}{\lambda} = \frac{N}{S / E[P]} = \frac{N}{P_S} E[slot]$$

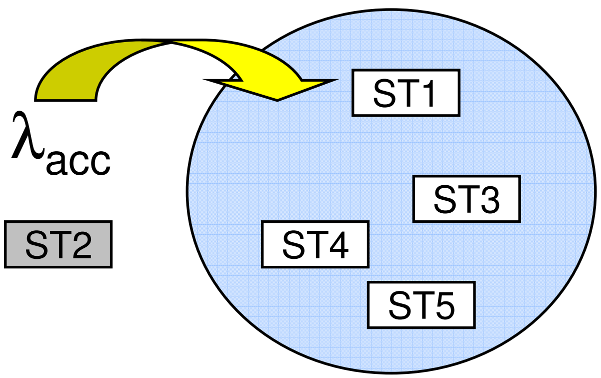


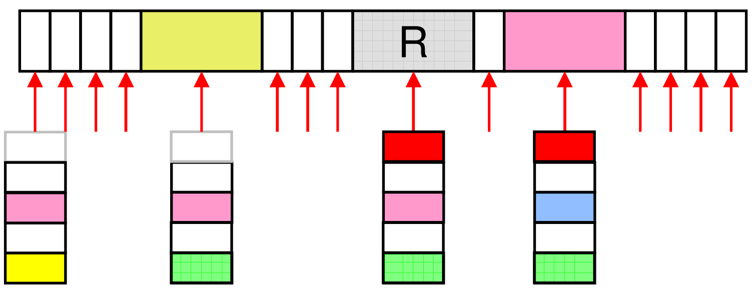


Retry Limit = R

➔ **$E[N] = N(1-p_{lost})$**

$$D = \frac{E[N]}{\lambda_{acc}} = \frac{N(1-p_{lost})}{S/E[P]}$$





Packet loss probability

➔ **In a generic contention slot, with access probability τ, the packet loss probability depends on the number of already suffered collisions:**

$$p_{lost} = \sum_{i=0}^{R} P(lost / s = i) P(s = i)$$

$$P(lost / s = i) = p^{R+1-i}$$

$$P(s = i) = P(TX) \frac{P(s = i / TX)}{P(TX / s = i)} = \tau \frac{(1-p)p^{i}}{1-p^{R+1}} (1 + W_i / 2)$$

How to take into account the 2-way or 4-way access mode?

➔ **By simply defining opportunely frame transmission times and collision times. Assuming fixed MPDU size:**

BASIC ACCESS:

$$T_{FRAME} = T_{MPDU} + SIFS + T_{ACK} + DIFS$$

$$T_{COLL} = T_{MPDU} + DIFS$$

RTS/CTS:

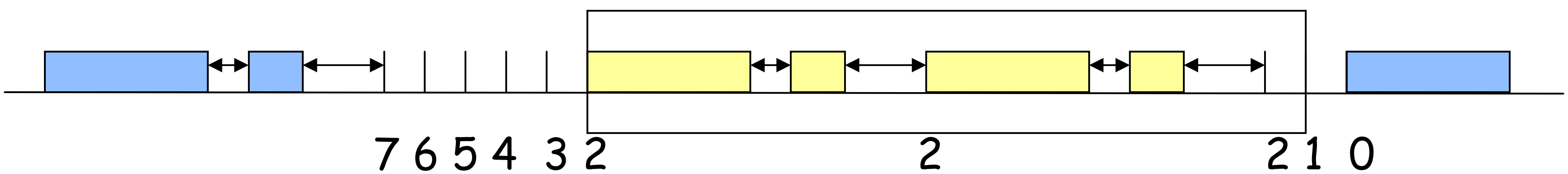
$$T_{FRAME} = T_{RTS} + SIFS + T_{CTS} + SIFS\ T_{MPDU} + SIFS + T_{ACK} + DIFS$$

$$T_{COLL} = T_{RTS} + DIFS$$

Note: The channel access probability and the collision probability do not depend on the employed access mode!

A last refinement...

- **We defined a system slot as the time interval between two successive backoff decrements.**
- **This allows to conclude that the number of system slots among two successive transmissions by the same stations is equal to the backoff value.**
  - Busy times due to successful transmissions can include more than 1 data frame!!! This happens whenever the transmitting stations extracts 0. The data frames which follow the first one are always successful.
- **Backoff expiration example:**



• The exact successful time expression is then $T_{FRAME}' = T_{FRAME}/(1-B_0)+\sigma$ (and $E[P]' = E[P]/(1-B_0)$)

(where B_0 represents the probability to extract 0 in stage 0 and $1/(1-B_0)$ is the average number of consecutive transmissions).

EDCA model

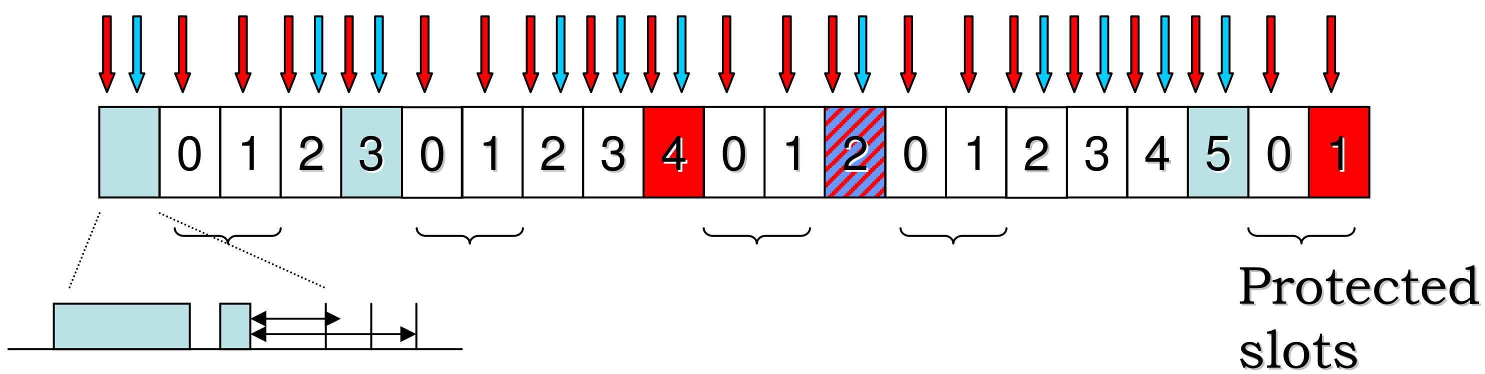
Is EDCA p-persistent?

-Proposal: differentiating τ_1 and τ_2 for different service classes

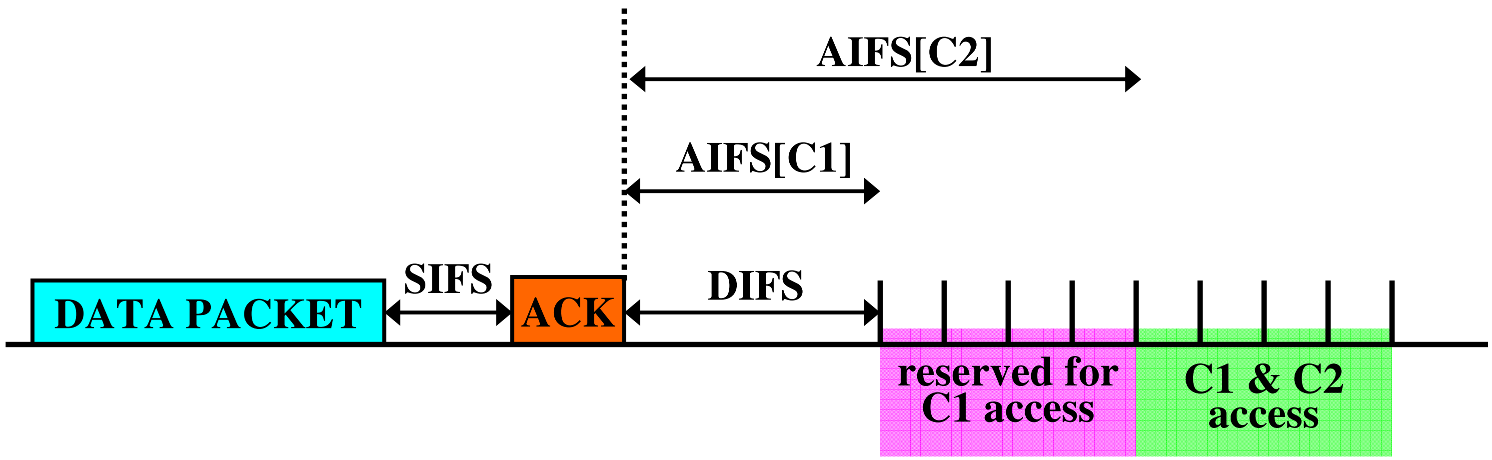
-Conceptually wrong: some slots can be accessed only by some stations and the collision probabilities p_1 and p_2 are not constant slot by slot

-Slots are numbered according to the time *n* elapsed by the last busy slot

-For each service class, we need the distribution $\tau(n)$ and $p(n)$



..wrong.. but easy to be corrected



- Seems easy: differentiate “C1 slots” from “C1+C2” slots
  - Split t_1 and find t_{11} and t_{12}
  - Use properly t_{11}& t_{12} in different slot ranges
- A bit more complex than backoff differentiation
  - Need to increase space-state of an additional dimension
- But appeared in the literature
  - Zhao et al, Globecom 02

Where’s the problem?

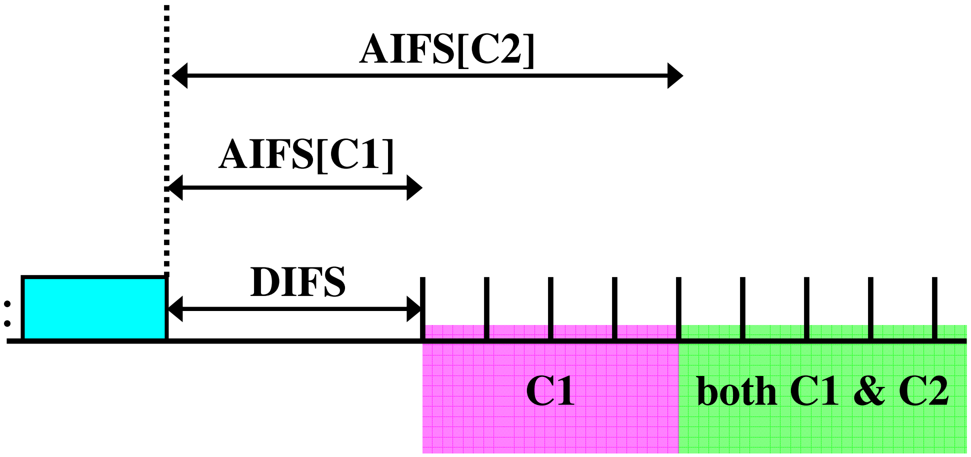
Modeling EDCA as p-persistent: failure

One C1 station; parameters:

- CWmin=CWmax=5;
  AIFS=DIFS

Several C2 station; parameters:

- AIFS=DIFS+4;
  CWmin=CWmax=32



<u>RESULTS (real):</u>

- C1 STA collision-free in protected slots
- All C2 STA always in collision in slot C1&C2!

<u>RESULTS (model)</u>

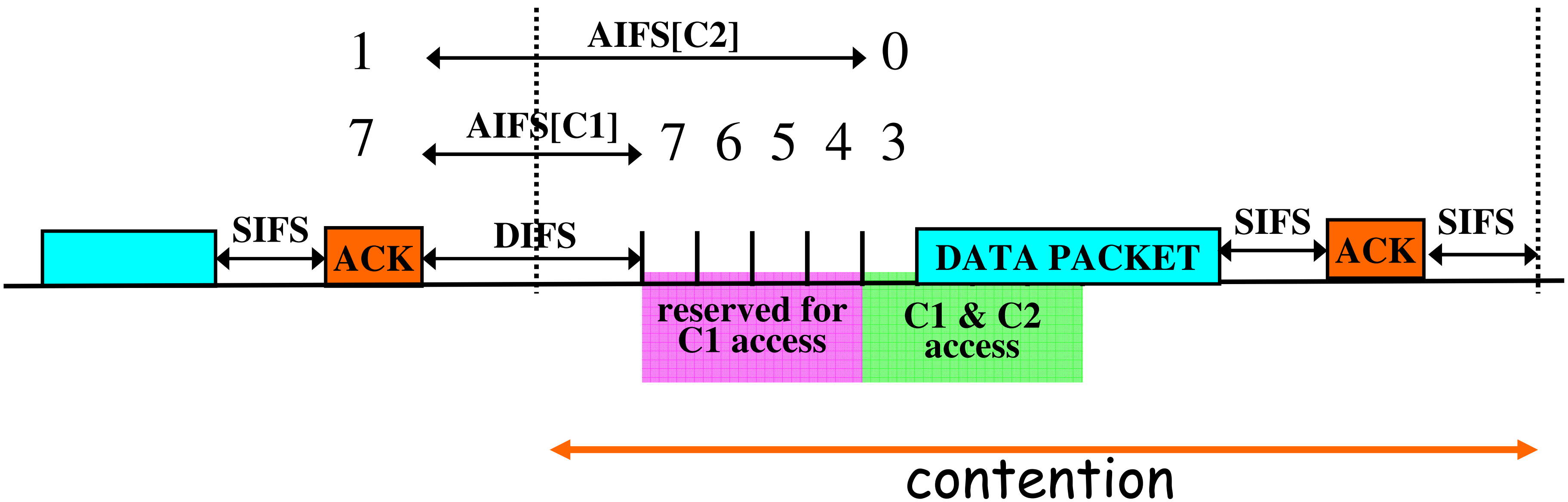
- τ_{11} as a constant in the first C1 slots
- Finite, NOT NULL, τ_2 (and throughput) value!

1) ***Numerical problems in several other (non pathological) conditions***
2) ***Problems also in standard DCF modeling (next slot after TX)***
3) ***To (convincingly) go further, need to completely rethink '98 model***

Beyond '98 model

- **'98 model - limiting factor:**
  - Based on the computation of the PER-SLOT transmission probability t
- **New approach:**
  - Channel accesses described in terms of consecutive contentions (no equivalent p-persistent processes)
  - Based on the computation of the transmission probability DISTRIBUTION $t(l)$, $p(l)$
  - … whatever this means…

Who wins the next contention?



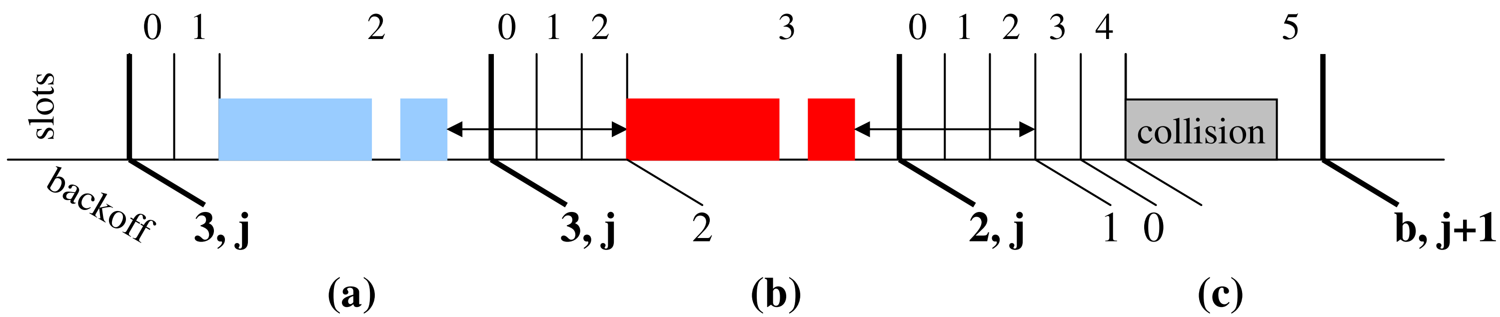
Given the IFS time = SIFS + δ_i * SlotTime, and the backoff value b_i, the winner is who minimizes $b_i + \delta_i$

- -Different AIFS modeled just in terms of different δi

Collision whenever the minimum is not unique

All the performance figures depend on the b_i distributions!

Backoff process evolution



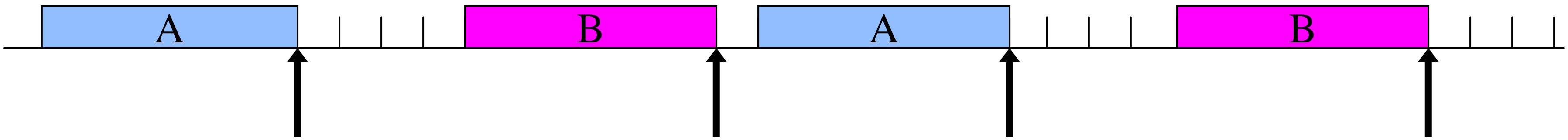
➔ **Contention: time interval until next channel activity + channel release (final SIFS)**

➔ **Backoff process: counter + stage**

➔ **At the end of each contention:**

- ⇨ (b, s) -> (b, s)
- ⇨ (b, s) -> (b-k, s)
- ⇨ (b, s) -> (b*, s*)

Model Specification



$B_k(b) = \sum_{j=0}^{m_k} \Pi_k(b, j)$ = backoff counter steady-state probability distribution

$T_k(b)$ = prob next TX by other stations after b slots

$$\Pi_k(i,j) = \Pi_k(i,j) \sum_{l=0}^{\delta_k - 1} T_k(l) +$$

$$\sum_{l=1}^{W_k^j - i} \Pi_k(l+i, j) T_k(l + \delta_k - 1) +$$

$$\frac{1}{W_k^j + 1} \sum_{l=0}^{W_k^{j-1}} \Pi_k(l, j-1) T_k(l + \delta_k)$$

$T_1(b)$ in turns can be expressed as a function of the cumulative distributions of the backoff probabilities $B_1(b)$: non linear system (details in the paper..)

Low level performance figures

System performance depends on the backoff counter comparison, and then on the marginal distribution $B_i(b)$

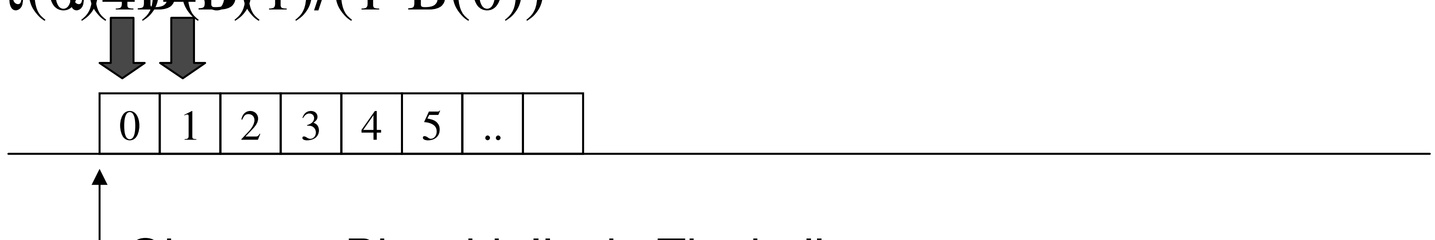
From $B_i(b)$, not only throughput and delay, but also access and collision probability slot by slot..

$t_k(l)$ = Pr{transmission in slot l / l idle slots elapsed} = $\dfrac{B_k(l-\delta_k)}{1-\sum_{j=0}^{l-\delta_k-1} B_k(j)}$

$p_k(l)$ = Pr{collision / transmission in slot l } = $\dfrac{T_k(l)}{1-\sum_{j=0}^{l-1} T_k(j)}$

$\tau(0)=B(0)$

$\tau(1)=B(1)/(1-B(0))$



Numerical Results

Custom made C++ simulator vs. Model results
Reference scenario: two contending classes
CWmin-based differentiation: p-persistent OK

HP : 8 – 256 LP: 64 – 256

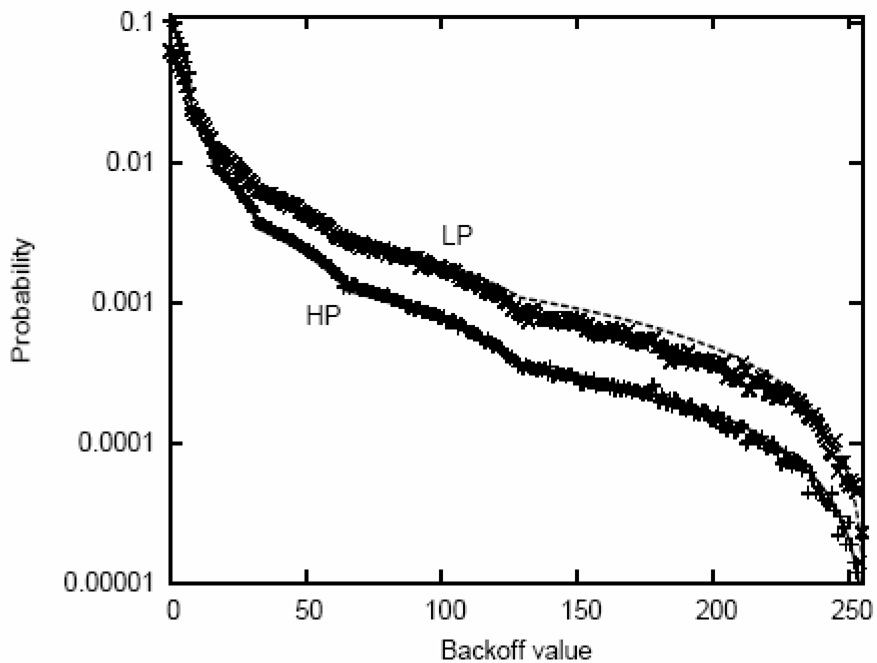
Class	n	Simulation	Analysis	P-Persistent
LP	5	0.090826	0.089161	0.089654
	10	0.104662	0.104200	0.104843
	15	0.112198	0.110334	0.110921
HP	5	0.599017	0.598843	0.601997
	10	0.500808	0.499121	0.501726
	15	0.438395	0.437382	0.439584

AIFS-based differentiation: p-persistent does not work

HP: $\delta = 2$ LP: $\delta = 5$

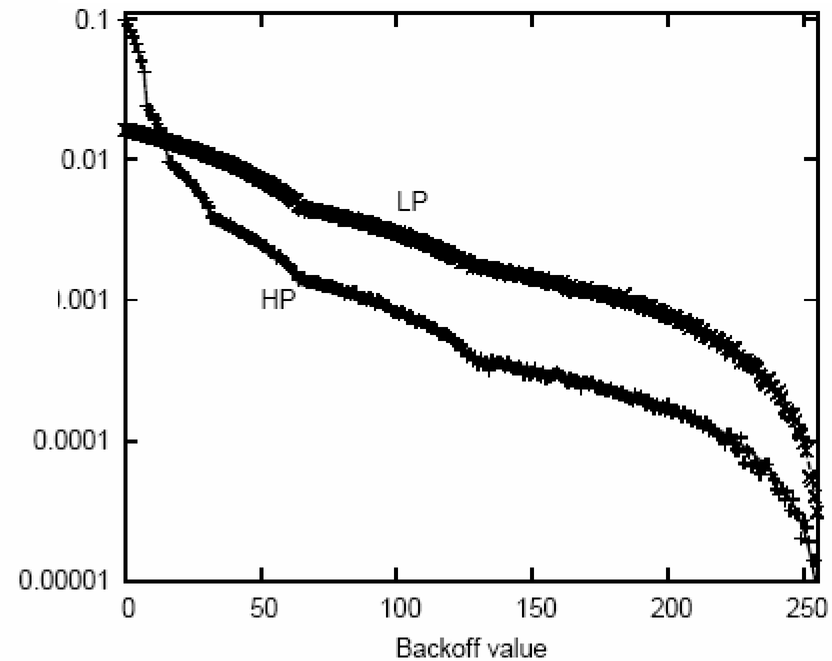
Class	n	Simulation	Analysis	P-Persistent
LP	5	0.040671	0.037502	0.058873
	10	0.021185	0.020012	0.033938
	15	0.013096	0.012422	0.022678
HP	5	0.658001	0.657933	0.636946
	10	0.601720	0.601074	0.588803
	15	0.563343	0.561148	0.552609

Model validation: whole bk distribution!

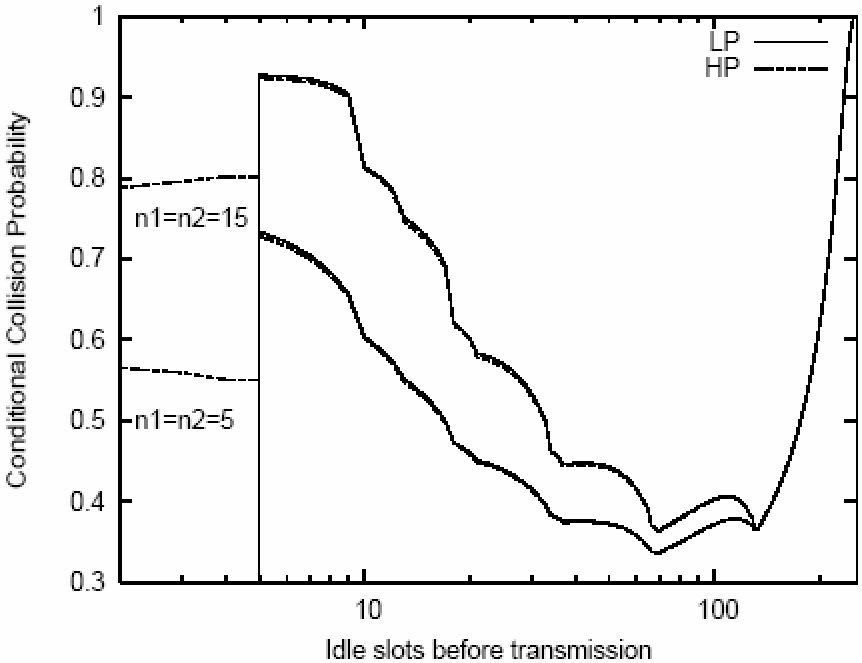


AIFS

CW_{min}



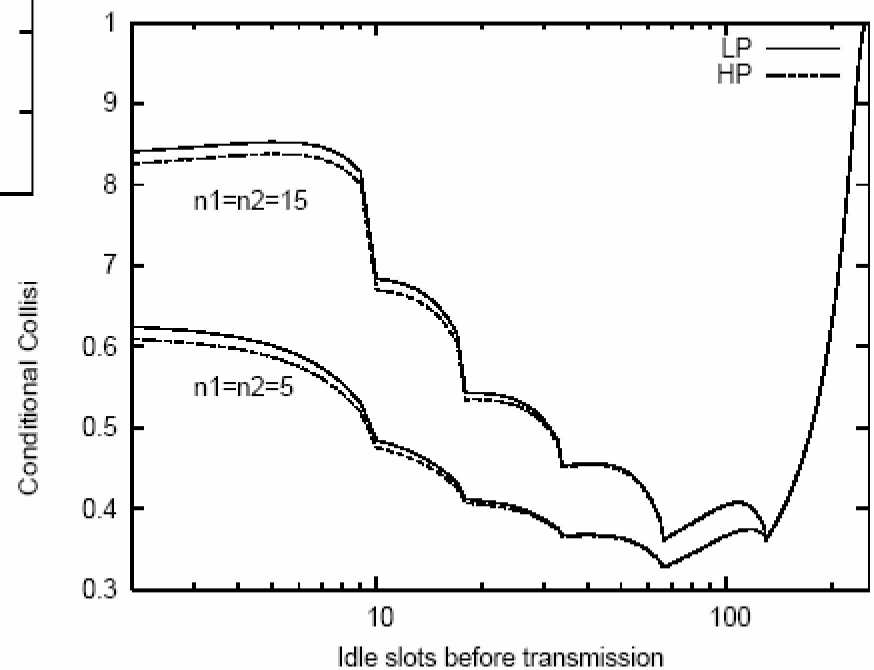
p-persistent Accuracy



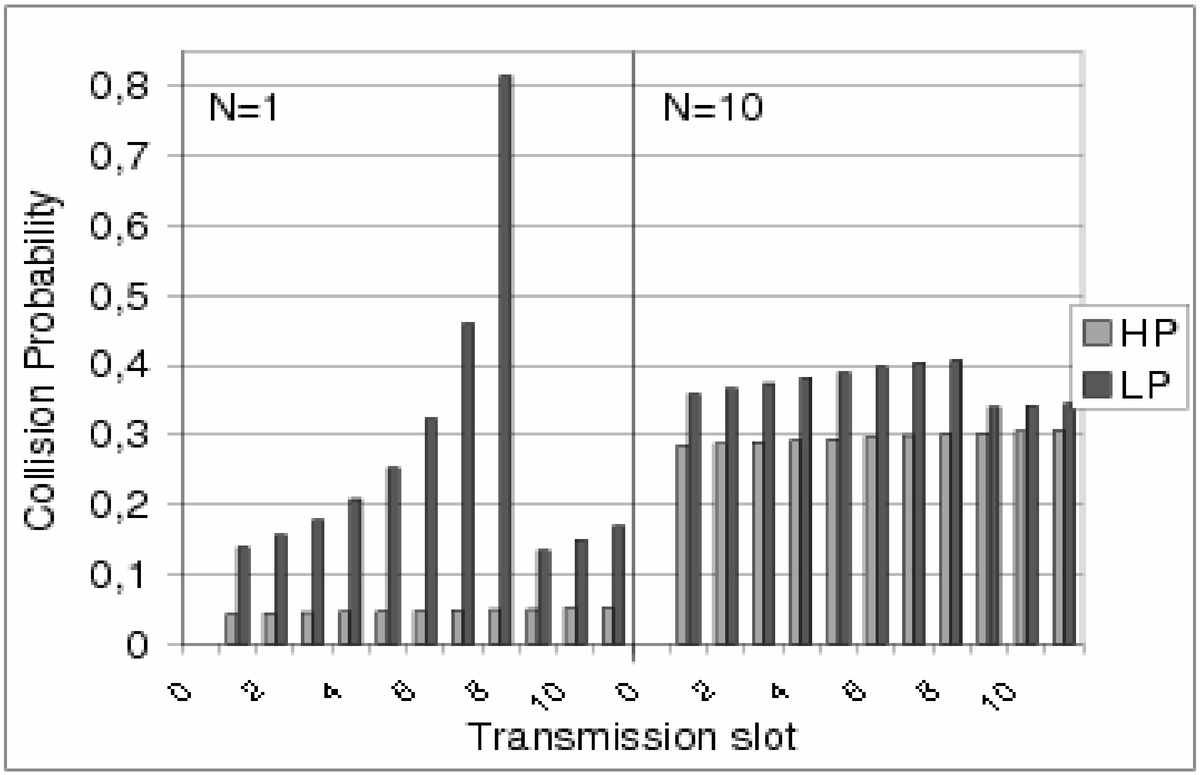
AIFS

From low level figures, we can verify some common simplificative assumptions!

CW_{min}



p-persistent Accuracy



1 HP station vs. N LP station

Conclusions

- Proposed new CSMA-CA model
  - Completely different approach
  - Not “just” an extension to ’98 model

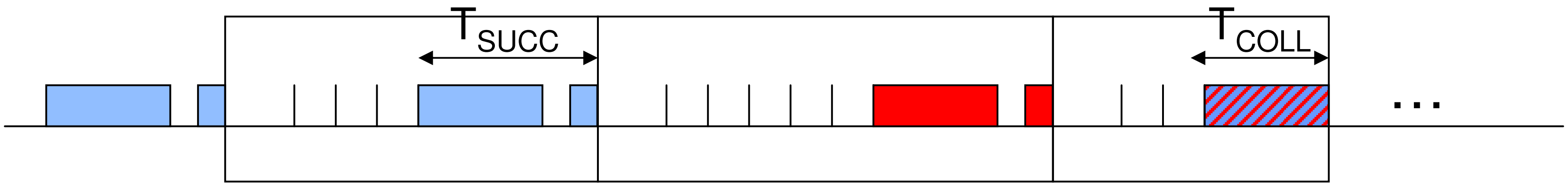
- Extremely accurate
  - Tightly matches distributions, not just mean values
  - Allows to verify all previous assumptions
    - p-persistent models for CWmin differentiation only

- EDCA evaluation

  Here just shown how to derive performance figures
  - Useful for MAC parameter tuning:
    - IFS for service differentiation
    - CWmin for throughput maximization

Performance Optimizations

What parameters are more critical?



$$S = \frac{P_s E[P]}{E[TXcycle]} = \frac{P_s E[P]}{P_s E[T_{SUCC}] + (1-P_s) E[T_{COLL}] + E[b]\sigma + DIFS}$$

-Throughput: Average payload bits in average transmission cycle

-How can we improve the thr? (short TXcycle and high P_S)

-Small backoff slots, small SIFS

-Small Collision Probability

-Small Collision Times

-Small Transmission Times (High data rates, Small Overheads)

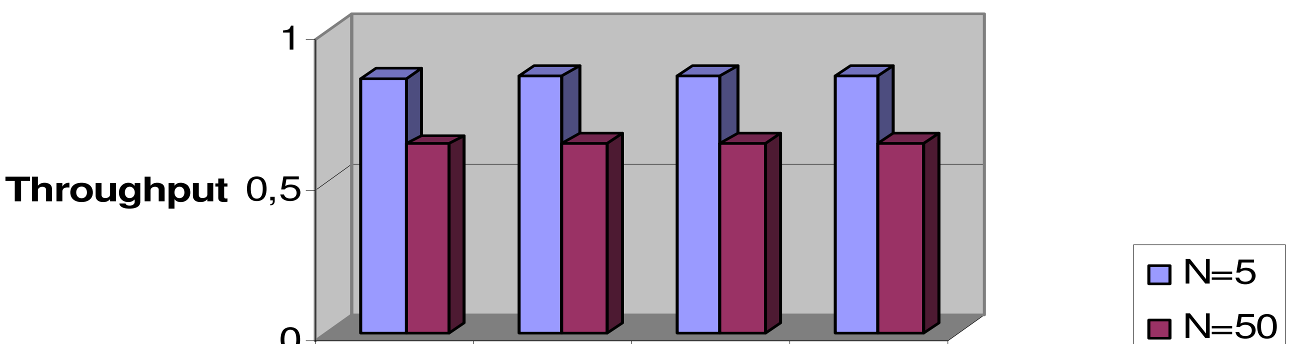
Is the bk SlotTime really critical?

-**SIFS = RxRFDelay + MACPrcDelay + RxTxTurnaroundTime**

-**SlotTime = CCAdel + RxTxTurnaroundTime + MACPrcDelay + RxRFDelay**

SIFS and SlotTime interval are not arbitrary, since they depend on critical PHY operations (basically, Carrier Sensing and Rx/Tx switch time) ->

Reducing SIFS and SlotTime is a very expensive operation (it requires complex hardware), but does not affect performance too much!!!



	(20/8)	(10/4)	(5/2)	(0/0)
N=5	0,8459	0,8501	0,8521	0,8541
N=50	0,6251	0,6267	0,6275	0,6283

SlotTime/SIFS

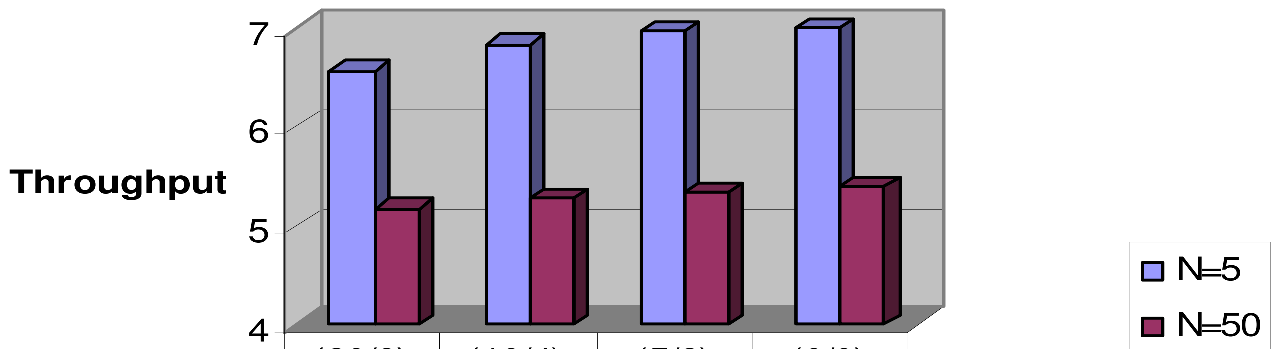
And for higher data rates?

-**SIFS = RxRFDelay + MACPrcDelay + RxTxTurnaroundTime**

-**SlotTime = CCAdel + RxTxTurnaroundTime + MACPrcDelay + RxRFDelay**

As the data rate increases, inter-frame inactivity times due to backoff expiration are more and more significant, since they can be comparable to the data transmission time -> Performance slightly improve with small bk slots, especially in low load conditions.

N.B. In 802.11a SlotTime < SIFS!



	(20/8)	(10/4)	(5/2)	(0/0)
N=5	6,57	6,83	6,96	7
N=50	5,17	5,28	5,34	5,4

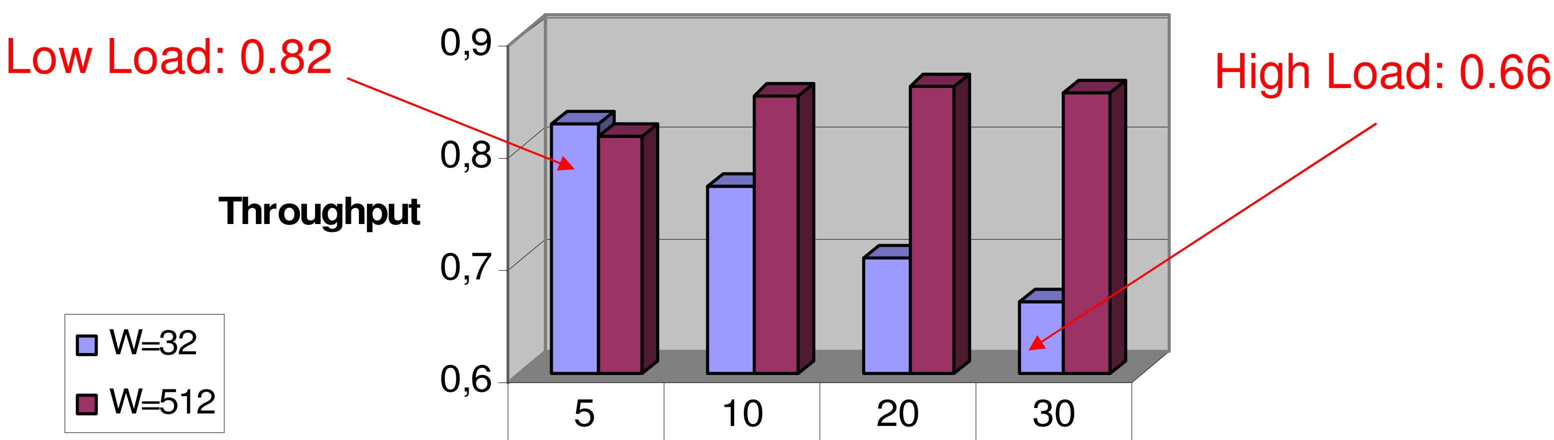
SlotTime/SIFS

Collision Probability Impact

-Unlike the SlotTime/SIFS case, collision probability is really critical for the overall system performance, since the lack of immediate collision detection causes large channel time wastes.

-The collision probability mainly depends on the CW_{min} value and on the number of contending users. For low CW and high loaded networks, is very likely than two or more stations select the same bk value and collide.

-Optimal CW_{min} = f(n) as a tradeoff among backoff time and collision pr minimization



	5	10	20	30
W=32	0,8229	0,7669	0,7034	0,6639
W=512	0,8111	0,8472	0,8557	0,8497

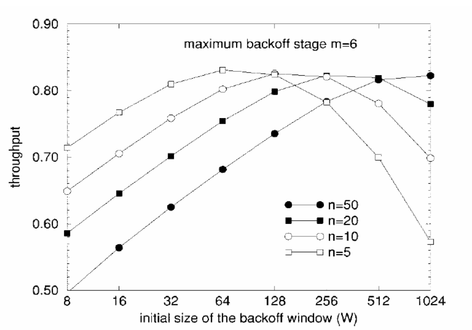
Number of Users

Collision Probability Impact (2)

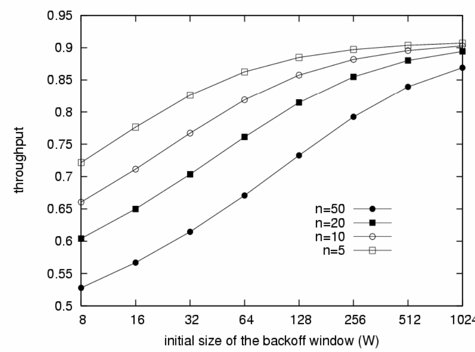
-Optimal CW_{min} setting should depend on the number of contending stations in the network.

-In absence of such a number, for low rates is better a big CW value, for high rates is better a small CW value

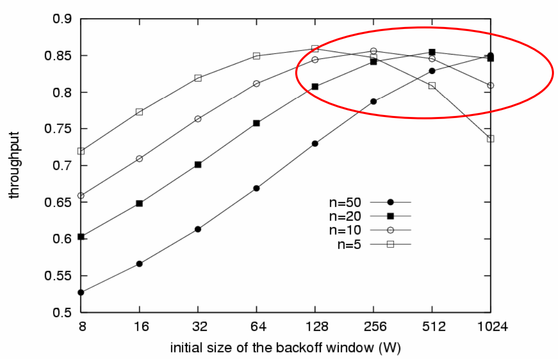
50μs



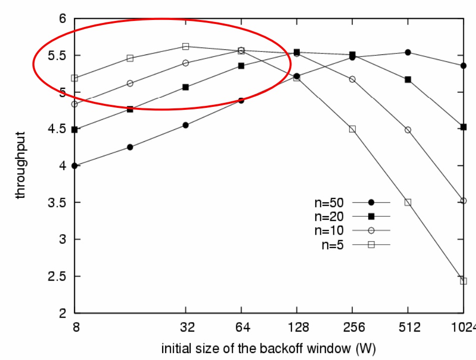
0μs



20μs



20μs@
11Mbps



And in actual networks?

- ➔ **We could use analytical models instead of simulation curves:** CWopt=f(load, AIFSN), **but..**
  - ⇨Load estimation
  - ⇨Function f evaluation
    - →f depends on the traffic sources, which need to be estimated and modeled !
    - →complex f expressions for not-saturated traffic sources
    - →no close expression for every traffic conditions..

How to compute CWopt?

- Our solution: not fixed CWmin, but adaptive corrections on the basis of the channel monitoring status, to force idle slots and collision equalization **[Gallagher]**

$\cancel{\text{CWmin} \overset{?}{=} \text{CWopt}}$ **$\text{CWmin}(t) = \text{CWmin}(t-1) + \Delta\text{CW}$**

- CW has an opposite effects on the two different events of channel wastes of the access protocol: BACKOFF, COLLISIONS
  - Large CW -> too long backoff expirations
  - Small CW -> too high collision probability
- Optimal CW as a tradeoff between these channel wastes
- different tuning algorithms are possible based on:

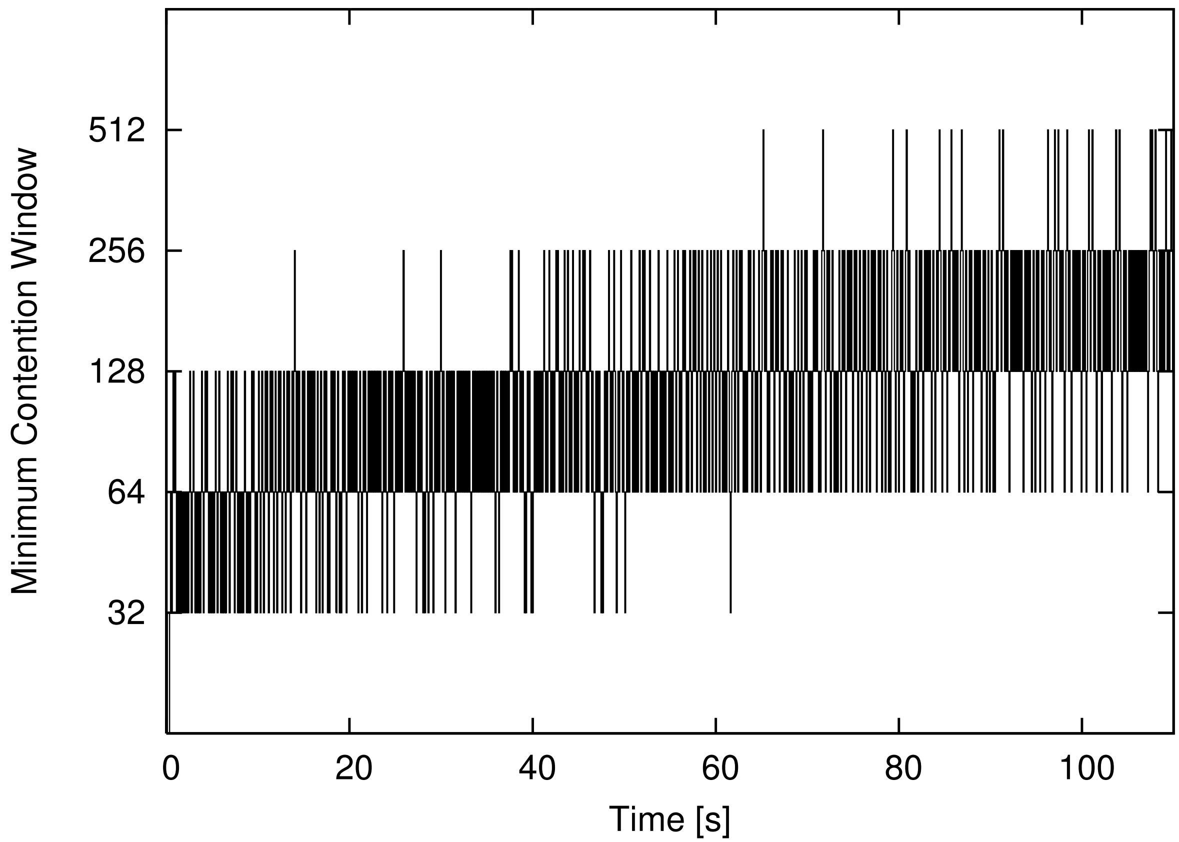
  If (COLLISIONS > BACKOFF) -> increase the CWmin

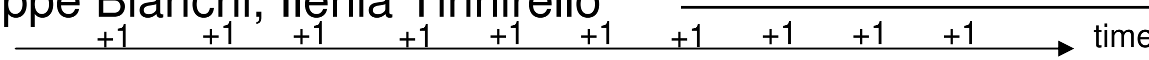
  If (BACKOFF > COLLISIONS) -> decrease the CWmin

No estimation of the system status, but simple channel monitoring of BACKOFF and COLLISIONS
Intrinsically suitable for dynamic network conditions

An example

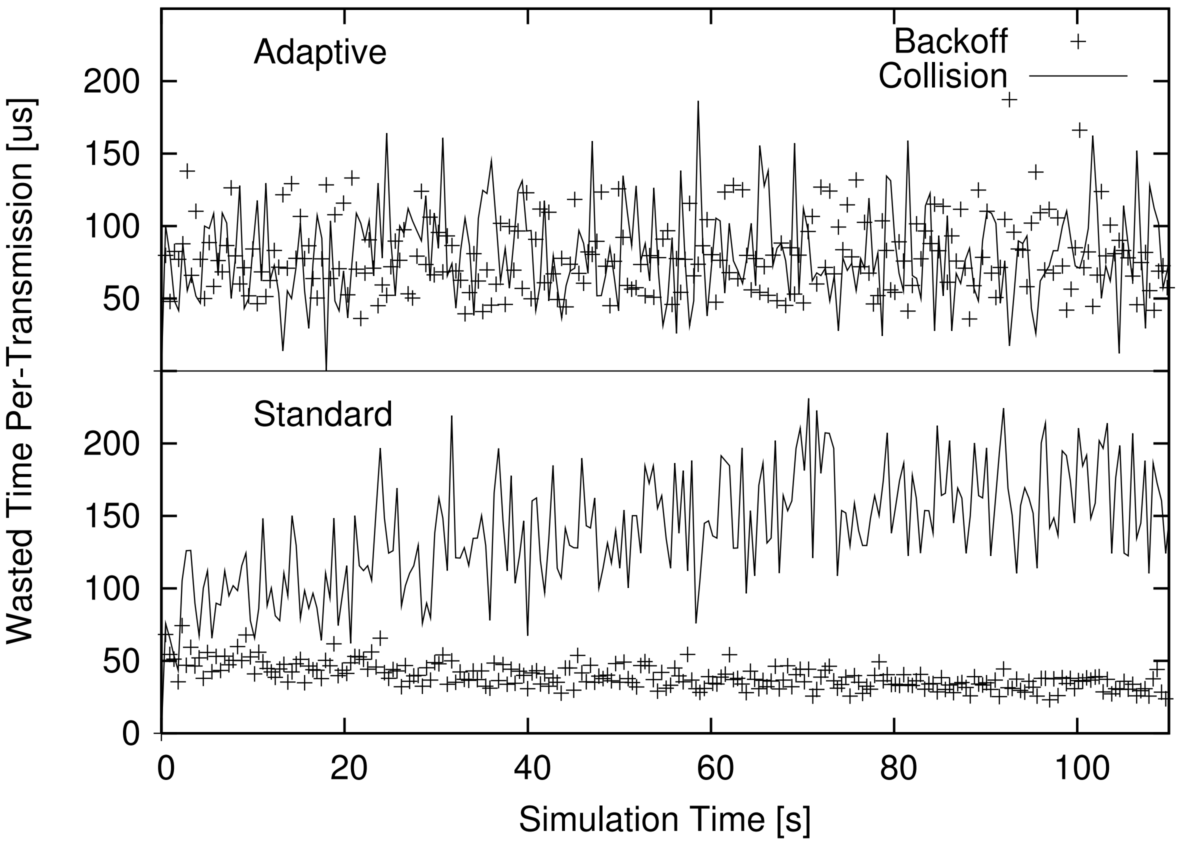
- ➔ **Assume that every 10 seconds a new data station joins the network..**
- ➔ **At each beacon: double the contention window at each beacon in which the collisions overcome the backoff times; half the contention window at each beacon in which the backoff overcomes the collision times.**

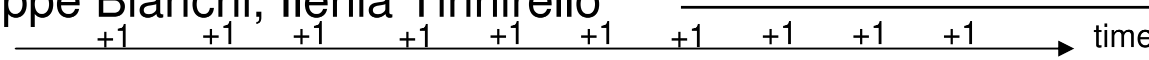




An example

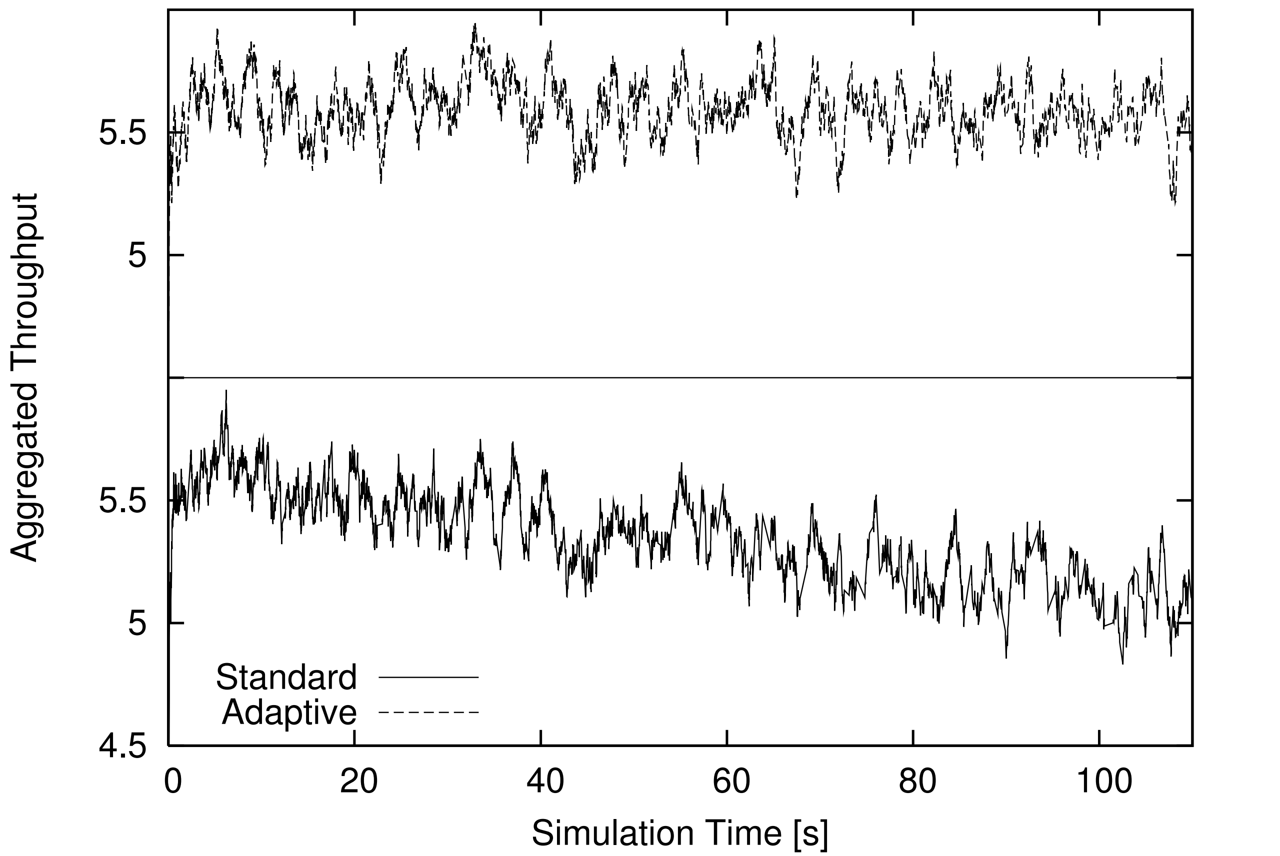
- Assume that every 10 seconds a new data station joins the network..
- At each beacon: double the contention window at each beacon in which the collisions overcome the backoff times; half the contention window at each beacon in which the backoff overcomes the collision times.

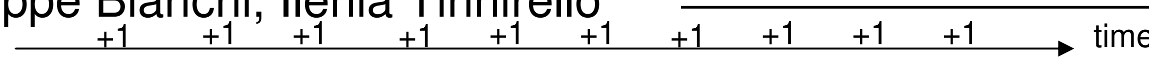




An example

- Assume that every 10 seconds a new data station joins the network..
- At each beacon: double the contention window at each beacon in which the collisions overcome the backoff times; half the contention window at each beacon in which the backoff overcomes the collision times.



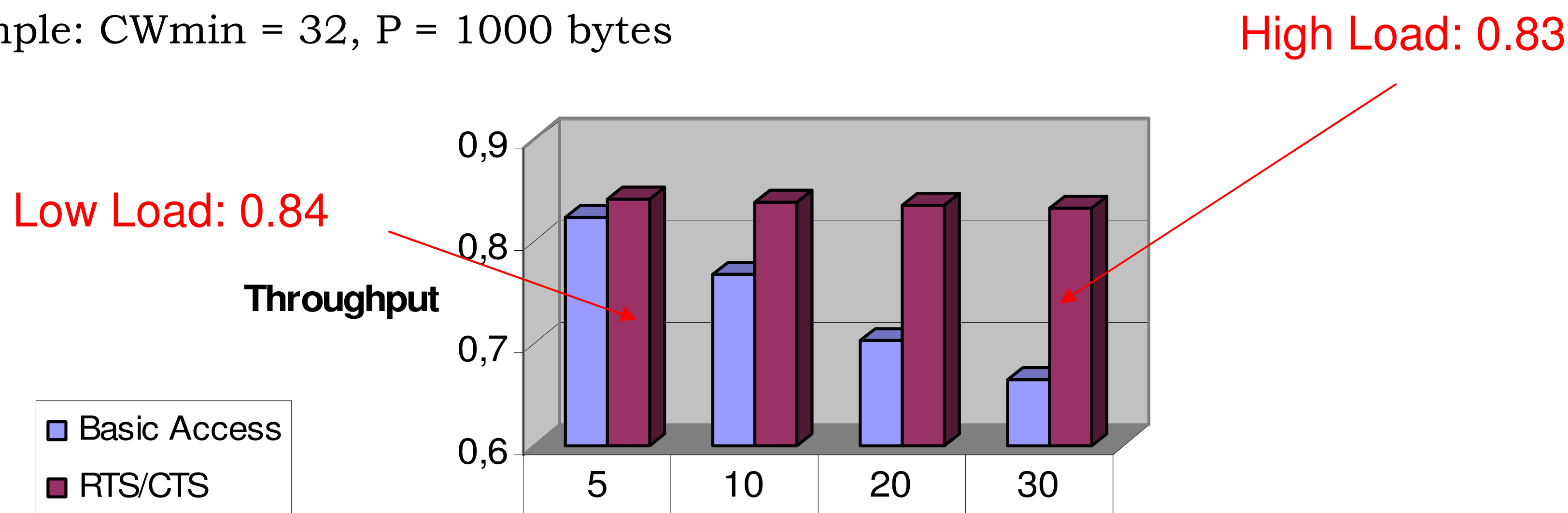


Collision Times Reduction

-For a given collision probability (i.e., number of users), how can we reduce the time wasted for collision detection??

-RTS/CTS mechanism, introduced for hidden terminals, limits the collision times to the short control RTS size (but introduces some overheads).

-Example: CWmin = 32, P = 1000 bytes



	5	10	20	30
Basic Access	0,8229	0,7669	0,7034	0,6639
RTS/CTS	0,8417	0,8385	0,8346	0,8318

Number of Users

Payload Size Effects

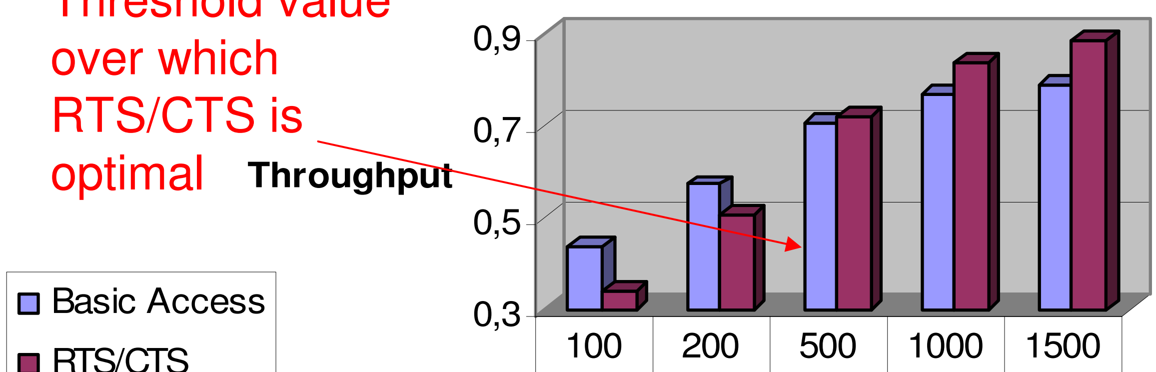
-As the payload sizes grows, there are two different beneficial effects:

-The fixed per-bit overhead (in terms of headers, RTS/CTS, and ACK) is reduced.

-The random per-bit overhead due to collisions is reduced too (for a given amount of data, a lower number of accesses are required).

-Example: 10 Stations

Threshold value over which RTS/CTS is optimal



	100	200	500	1000	1500
Basic Access	0,4361	0,5731	0,7073	0,7671	0,7892
RTS/CTS	0,3412	0,5095	0,7228	0,8385	0,8862

Payload Size [byte]

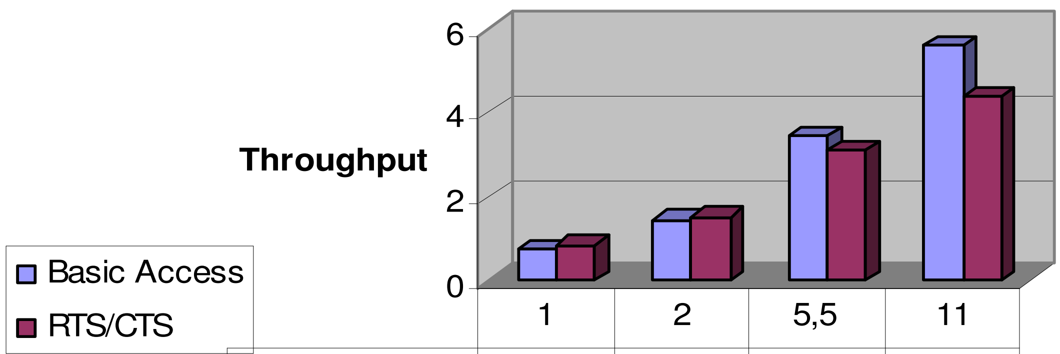
4-way access is opportune only for data frames which require a long transmission time (long frames?)

And for higher data rates?

-The throughput cannot grows proportionally, since some overheads are fixed (physical preamble, backoff times). Moreover, control frames are transmitted at lower rates.

-Channel occupancy times depend on both frame size and employed data rates

-Example: N=10, P=1000 bytes



	1	2	5,5	11
Basic Access	0,7671	1,466	3,44	5,589
RTS/CTS	0,8385	1,5348	3,099	4,374

Data rate [Mbps]

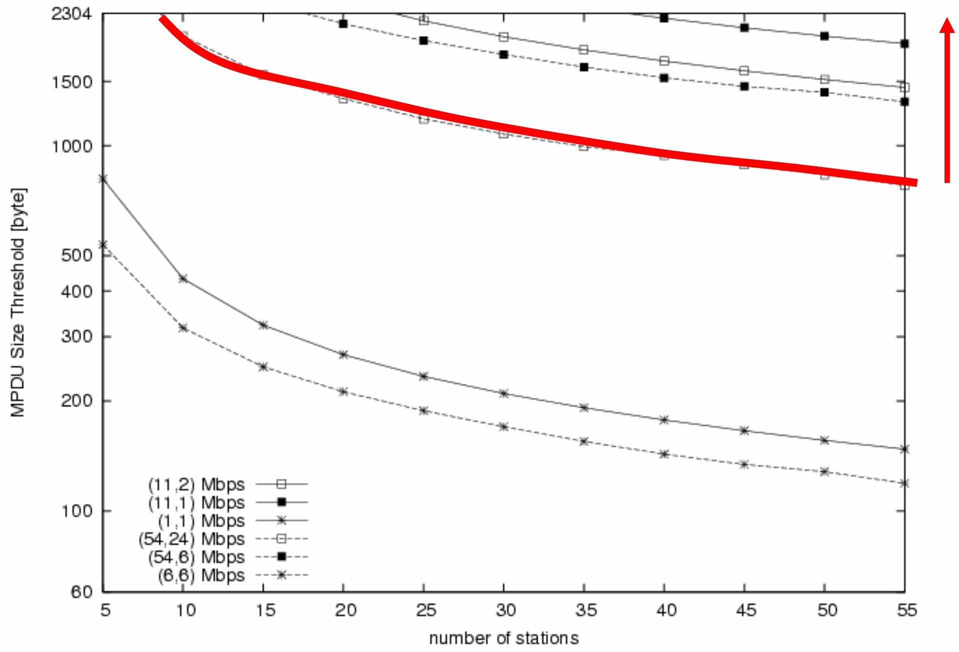
4-way access is opportune only for low data rates!! For high data rates, RTS time is comparable with the data frame transmission time.

RTS thresholds

RTS/CTS mechanism: on one side it introduces an higher overhead, but on the other side it reduces the collision times..

$S_{RTS} > S_{BASIC}$ whenever $E[TXcycle]_{RTS} < E[TXcycle]_{BASIC}$

For fixed MSDU size: $T_{MPDU} > P_s/(1-P_S)\ O_{RTS} + T_{RTS}$



Note: for high data rates in most cases is not the optimal access mode

Final Remarks

Due to the new available PHYs, some consolidated conclusions about DCF optimal settings need to be redefined.

- 4-way Access Mode is not opportune for High Data Rates
- SlotTime is more and more critical as the data rate increases
- CW_{min} values can be shortened

New channel utilization mechanisms can be performed on the basis of the emerging 802.11e MAC extensions:

- TXOP: transmission burst, in order to provide temporal fairness
- BACK: in order to further reduce High Rate overheads.

References:

- ➔ G. Bianchi, “Performance Analysis of the IEEE 802.11 Distributed Coordination Function,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, March 2000
- ➔ G. Bianchi, and I. Tinnirello, “Remarks on IEEE 802.11 DCF Performance Analysis”, to appear in *IEEE Communication Letters*
- ➔ F. Cali, M. Conti, and E. Gregori, "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit," IEEE/ACM Trans. Networking (TON), vol. 8, No. 6, Dec. 2000
- ➔ Giuseppe Bianchi and Sunghyun Choi, "IEEE 802.11 Ad-Hoc Networks," a chapter in *Wireless Ad hoc and Sensor Networks*, Ed. Kluwer, to be published in 2004.