Toward a Unified Theory of Random Access

Lin Dai

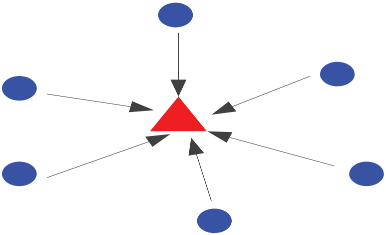
City University of Hong Kong

lindai@cityu.edu.hk

January 17, 2014

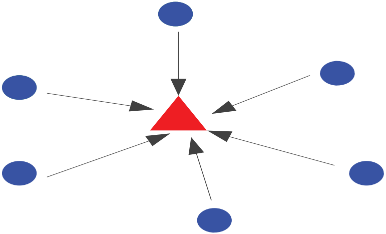
Overview

Multiple Access



Multiple nodes transmit to a common receiver: How to share the channel?

Multiple Access



Multiple nodes transmit to a common receiver: How to share the channel?

- Centralized Access: A central controller performs resource allocation/optimization.
- Random Access: Each node determines when/how to access in a distributed manner.

Random Access

Various assumptions on the receiver have been made, which can be broadly divided into three categories.

- *Collision*: When more than one node transmit their packets simultaneously, a collision occurs and none of them can be successfully decoded. A packet transmission is successful only if there are no concurrent transmissions.
- *Capture*: Each node's packet is decoded independently by treating others' as background noise. A packet can be successfully decoded as long as its received signal-to-interference-plus-noise ratio (SINR) is above a certain threshold.
- *Joint-decoding*: Multiple nodes' packets are jointly decoded.

Protocol Design: Three Key Questions

For each node:

- When to start a transmission?
- When to end a transmission?
- How to resolve collisions?

Question 1: When to Start a Transmission?

- Transmit if packets are awaiting in the queue.

  — Aloha [Abramson'1970]

- A more "polite" solution: Transmit if packets are awaiting in the queue **and the channel is sensed idle**.

  — Carrier Sense Multiple Access (CSMA) [Kleinrock&Tobagi'1975]

Question 2: When to End a Transmission?

- Stop when a packet transmission is completed.

- A "smarter" solution: Stop if a collision is sensed. Otherwise stop when a packet transmission is completed.

    — It requires full duplex.

Question 3: How to Resolve Collisions?

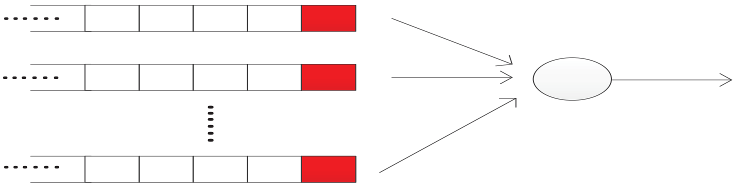
Backoff

- Probability-based: Retransmit with a certain probability at each time slot.
- Window-based: Choose a random value from a window and count down. Retransmit when the counter is zero.
- Example — Exponential Backoff [Metcalfe&Boggs'1976]: For each node, the transmission probability/window size at time slot t is exponentially decreased/increased according to the number of collisions the node has experienced by time slot t.

Performance Analysis: Three Key Questions

- How to model a random-access network?
- How to evaluate the performance of a random-access network?
- How to optimize the performance of a random-access network?

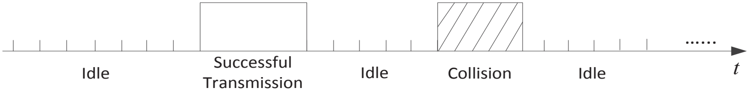
Question 1: How to Model a Random-access Network?



- A random-access network can be regarded as a multi-queue-single-server system.
- Numerous models have been proposed, which can be broadly divided into two categories: channel-centric and node-centric.

Question 1: How to Model a Random-access Network?

Channel-centric modeling: to characterize the aggregate channel.



— [Abramson'1970]: The aggregate traffic of Aloha at each time slot is modeled as a Poisson random variable with parameter G.

— [Kleinrock&Tobagi'1975]: The aggregate channel of CSMA is modeled as an alternating renewal process.

— [Hajek&Loon'1982], [Tan&Tsai'1996], [Yang&Yum'2003], [MacKenzie&O'Farrell'2010], [Wong,Yin&Lee'2013]...

Question 1: How to Model a Random-access Network?

Node-centric modeling: to characterize the behavior of each node.

— [Tsybakov'1979]: A two-node buffered Aloha network is modeled as a 2-dimensional random walk.

— [Rao&Ephremides'1988], [Anantharam'1991], [Szpankowski'1994]...: Generalization to an n-node buffered Aloha network leads to prohibitively high complexity.

— [Takagi&Kleinrock'1985], [Takin,Takahashi&Hasegawa'1988], [Wan&Sheikh'2000]...: Approximations are developed to simplify the analysis.

Question 2: How to Evaluate the Performance of a Random-access Network?

- Network Throughput: the time fraction that an effective output is produced.
- Delay
  1) Queueing delay (waiting time): the time interval from the packet's arrival to the instant that it becomes the head-of-line (HOL) packet;
  2) Access delay (service time): the time interval from the instant that it becomes the HOL packet to its successful transmission.
- Stability
  1) The network is stable if the network throughput is equal to the aggregate input rate.
  2) The network is stable if the mean access/queueing delay is finite.
  ...

Question 3: How to Optimize the Performance of a Random-access Network?

Observations:

- Performance of random-access networks crucially depends on backoff parameters.
- A random-access network may suffer from low throughput and large delay jitter if the backoff parameters are improperly selected.
- To stabilize the network, the transmission probability of each node should be adaptively adjusted.

Three Open Questions to be Answered

- Which modeling methodology, channel-centric or node-centric, is better?
- Is there a unified framework for throughput, delay and stability analysis?
- How to tune the backoff parameters to optimize the performance?

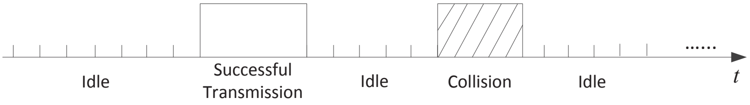
References

L. Dai, "Stability and delay analysis of buffered Aloha networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2707–2719, Aug. 2012.

L. Dai, "Toward a coherent theory of CSMA and Aloha," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3428–3444, Jul. 2013.

Modeling of Random-Access Networks

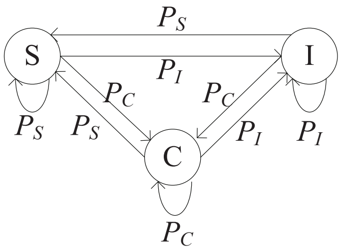
Modeling I: Channel-Centric



- The aggregate channel has three states: Idle (I), Successful Transmission (S) and Collision (C).
- Key Assumption: The aggregate traffic, i.e., the total number of attempts generated by both fresh and backlogged packets, of each time slot can be modeled a Poisson random variable with rate G.

State Characterization of Aggregate Channel

- The channel state transition can be modeled as a discrete-time Markov renewal process $(\mathbf{X}^c, \mathbf{V}^c) = \{(X_j^c, V_j^c), j = 0, 1, \ldots\}$.
- The embedded Markov chain $\mathbf{X}^c = \{X_j^c\}$:



Network Throughput

- The limiting state probabilities of the Markov renewal process $(\mathbf{X}^c, \mathbf{V}^c)$ are given by:

$$\tilde{\pi}_i = \frac{\pi_i \cdot \tau_i}{\sum_{j \in \{I,S,C\}} \pi_j \cdot \tau_j}, \quad i \in \{I, S, C\}.$$

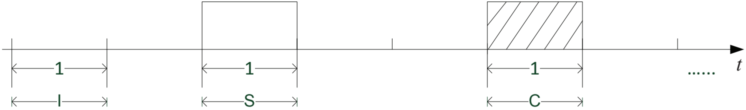
τ_i: holding time at State i, $i \in \{\mathrm{I,S,C}\}$.
π_i: limiting state probability of the embedded Markov chain at State i, $i \in \{\mathrm{I,S,C}\}$.

- Network Throughput: the time fraction that an effective output is produced.

$$\hat{\lambda}_{out} = \frac{1}{\tau_S} \cdot \tilde{\pi}_S = \frac{\pi_S}{\sum_{j \in \{I,S,C\}} \pi_j \cdot \tau_j}$$

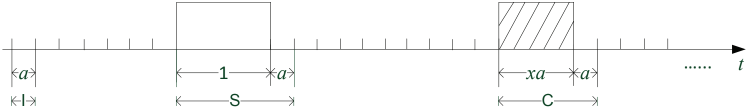
Assume that a packet transmission lasts for 1 time slot.

Network Throughput without Sensing



- $\tau_S = \tau_I = \tau_C = 1$ time slot
- $\pi_S = P_S = Ge^{-G}$, $\pi_I = P_I = e^{-G}$, $\pi_C = P_C = 1 - e^{-G} - Ge^{-G}$.
- Network throughput without sensing: $\hat{\lambda}_{out}^{Aloha} = Ge^{-G}$.
- Maximum network throughput: $\hat{\lambda}_{\max}^{Aloha} = e^{-1}$, achieved at $G = 1$.

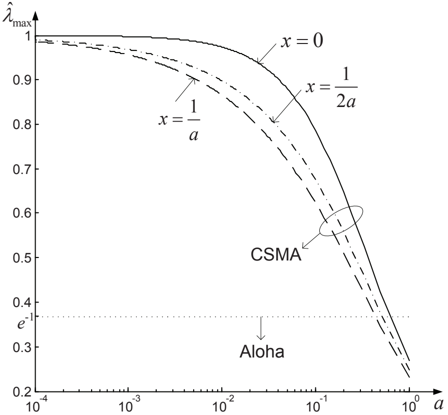
Network Throughput with Sensing



- a (mini-slot length): the ratio of the propagation delay required by each node for sensing the channel to the packet transmission time. x: how many mini-slots for each node to detect the collision and abort the ongoing transmission.
- $\tau_S = 1 + a$ time slot, $\tau_I = a$ time slot, $\tau_C = (x+1)a$ time slot; $\pi_S = P_S = aGe^{-aG}$, $\pi_I = P_I = e^{-aG}$, $\pi_C = P_C = 1 - e^{-aG} - aGe^{-aG}$.
- Network throughput: $\hat{\lambda}_{out}^{CSMA} = \frac{Ge^{-aG}}{x+1-xe^{-aG}+(1/a-x)aGe^{-aG}}$.
- Maximum network throughput: $\hat{\lambda}_{\max}^{CSMA} = \frac{-\mathbb{W}_0\left(-\frac{1}{e(1+1/x)}\right)}{xa-(1-xa)\mathbb{W}_0\left(-\frac{1}{e(1+1/x)}\right)}$,

achieved at $G = \frac{1}{a}\left(1 + \mathbb{W}_0\left(-\frac{1}{e(1+1/x)}\right)\right)$.

Maximum Network Throughput



- $\hat{\lambda}_{max}^{CSMA}$ increases as the normalized propagation delay a or the collision-detection time x decreases.
- $\hat{\lambda}_{max}^{CSMA} > \hat{\lambda}_{max}^{Aloha}$ if $a < e^{\frac{1}{e}} - 1 \approx 0.445$.

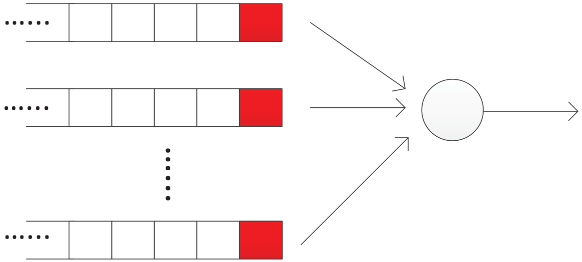
Summary of Channel-Centric Modeling

- Throughput analysis can be performed by modeling the state transition process of the aggregate channel.
- The maximum network throughput is independent of backoff parameters. It is solely determined by the holding time of the channel at State S (Successful Transmission), State I (Idle) and State C (Collision).

Questions:

1. What if the aggregate traffic cannot be approximated as a Poisson random variable?
2. How to achieve the maximum network throughput?
3. how to characterize the per-node performance such as delay?

Modeling II: Node-Centric

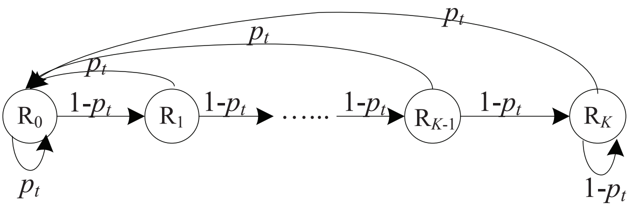


- An n-node buffered random-access network can be modeled as an n-queue-single-server system.
- Key Assumption: Each node's queue can be regarded as an independent queueing system with identically distributed service time if the number of nodes n is large.

State Characterization of Head-of-Line (HOL) Packet

- The state transition of each HOL packet can be modeled as a discrete-time Markov renewal process $(\mathbf{X}^h, \mathbf{V}^h) = \{(X_j^h, V_j^h), j = 0, 1, \dots\}$.
- The embedded Markov chain $\mathbf{X}^h = \{X_j^h\}$:

Without Sensing

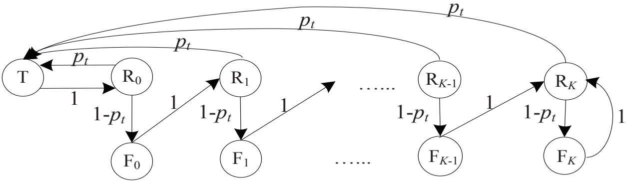


- K: cutoff phase. The HOL packet stays at State R_K if the number of collisions exceeds K.
- p_t: probability of successful transmission of HOL packets at time slot

State Characterization of Head-of-Line (HOL) Packet

- The state transition of each HOL packet can be modeled as a discrete-time Markov renewal process $(\mathbf{X}^h, \mathbf{V}^h) = \{(X_j^h, V_j^h), j = 0, 1, \ldots\}$.
- The embedded Markov chain $\mathbf{X}^h = \{X_j^h\}$:

With Sensing



- The states of $\{X_j^h\}$ are divided into three categories: 1) waiting to request (State R_i, $i = 0, \ldots, K$), 2) collision (State F_i, $i = 0, \ldots, K$) and 3) successful transmission (State T).
- p_t: probability of successful transmission of HOL packets at mini-slot

Dynamic Trajectory of p_t

- For each HOL packet, its transmission is successful if and only if all the other $n-1$ nodes are either idle with an empty queue, or, with a State-R_i HOL packet but not requesting any transmission.

Without Sensing

$$p_{t+1} = \left\{1 - \rho_t + \sum_{i=1}^{K} \rho_t \tilde{\pi}_{R_i,t}(1-q_i)\right\}^{n-1}.$$

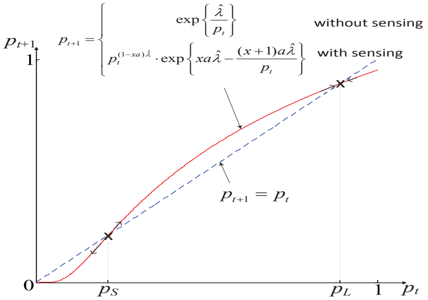
With Sensing

$$p_{t+1} = \left\{\frac{1 - \rho_t + \rho_t \cdot \sum_{i=0}^{K} \tilde{\pi}_{R_i,t} \cdot (1-q_i)}{1 - \rho_t + \rho_t \cdot \sum_{i=0}^{K} \tilde{\pi}_{R_i,t}}\right\}^{n-1}.$$

q_i: Transmission probability of a State-R_i HOL packet.
ρ_t: Offered load at time slot t.
$\pi_{i,t}$: Probability that the HOL packet stays at State R_i at time slot t.

- Approximation: $\rho_t \approx \rho(p_t)$ and $\tilde{\pi}_{R_i,t} \approx \tilde{\pi}_{R_i}(p_t)$, $i = 0, \ldots, K$.

Bi-stable Property



- If $p_t \geq p_S$ at any t and $\lim_{t\to\infty} \rho_t = \rho \leq 1$: $\lim_{t\to\infty} p_t \to p_L$.

  p_L and p_S are roots of the fixed-point equation:

$$p = \begin{cases} \exp\left\{\frac{\hat{\lambda}}{p}\right\} & \text{without sensing} \\ \exp\left\{\frac{xa\hat{\lambda}}{1-(1-xa)\hat{\lambda}}\right\} \cdot \exp\left\{-\frac{(x+1)a\hat{\lambda}}{1-(1-xa)\hat{\lambda}} \cdot \frac{1}{p}\right\} & \text{with sensing.} \end{cases}$$

Bi-stable Property

- Otherwise, the network becomes saturated and all the nodes are busy with non-empty queues. In this case, $\lim_{t\to\infty} p_t \to p_A$, where p_A is the single non-zero root of

$$p = \exp\left\{ -\frac{n}{\sum_{i=0}^{K-1} \frac{p(1-p)^i}{q_i} + \frac{(1-p)^K}{q_K}} \right\}$$

if $\{q_i\}$ is a monotonic decreasing sequence.

Desired Stable Point p_L and Undesired Stable Point p_A

- Desired Stable Point p_L is determined by the aggregate input rate $\hat{\lambda}$ and independent of the backoff parameters $\{q_i\}$.

- Undesired Stable Point p_A is determined by the backoff parameters $\{q_i\}$ and the number of nodes n.

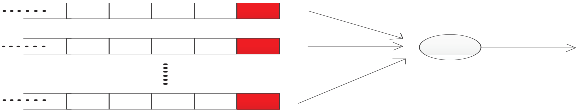
	Without Sensing	With Sensing
Desired Stable Point p_L	$p_L^{Aloha} = \exp\left\{\mathbb{W}_0(-\hat{\lambda})\right\}$	$p_L^{CSMA} = \exp\left\{\mathbb{W}_0\left(-\frac{(x+1)a\hat{\lambda}}{1-(1-xa)\hat{\lambda}}\cdot\exp\left\{-\frac{xa\hat{\lambda}}{1-(1-xa)\hat{\lambda}}\right\}\right)+\frac{xa\hat{\lambda}}{1-(1-xa)\hat{\lambda}}\right\}$
Undesired Stable Point p_A	$p_A = \exp\left\{-\frac{n}{\sum_{i=0}^{K-1}\frac{p_A(1-p_A)^i}{q_i}+\frac{(1-p_A)^K}{q_K}}\right\}$	

Summary of Node-Centric Modeling

- The key to node-centric modeling lies in the proper characterization of 1) the state transition process of each HOL packet, and 2) the steady-state probability of successful transmission of HOL packets.
- Both steady-state points are obtained as explicit functions of system parameters, based on which a unified analysis of throughput, delay and stability can be further performed.

A Unified Analysis of Throughput, Delay and Stability

Performance Evaluation of Random-access Networks



Multi-Queue-Single-Server System

- Network throughput: the time fraction that an effective output is produced.
- Access delay: the time interval from the instant that the packet becomes the HOL to its successful transmission.
- Stability: the network is stable if the network throughput is equal to the aggregate input rate.

Questions

- How to characterize the throughput and delay performance at the bi-stable points p_L and p_A?
- How to properly tune the backoff parameters to stabilize the network?
- How to properly tune the backoff parameters to optimize the throughput/delay performance?

Network Throughput at the Desired Stable Point p_L

- The desired stable point p_L is a function of the aggregate input rate $\hat{\lambda}$.
- For the network to operate at the desired stable point p_L, the offered load ρ of each queue should not exceed 1.
- Network throughput at p_L:

  Without Sensing: $\hat{\lambda}_{out,p=p_L}=\hat{\lambda}=-p_L \ln p_L$.

  With Sensing: $\hat{\lambda}_{out,p=p_L}=\hat{\lambda}=\frac{-p_L \ln p_L}{(x+1)a-(1-xa)p_L \ln p_L - xap_L}$.

- The maximum network throughput is achieved when the aggregate input rate

$$\hat{\lambda}=\hat{\lambda}_{\max}=\begin{cases} e^{-1} & \text{without sensing} \\ \dfrac{-\mathbb{W}_0\left(-\frac{1}{e(1+1/x)}\right)}{xa-(1-xa)\mathbb{W}_0\left(-\frac{1}{e(1+1/x)}\right)} & \text{with sensing.} \end{cases}$$

Network Throughput at the Undesired Stable Point p_A

- The undesired stable point p_A is a function of the backoff parameters $\{q_i\}$.
- The aggregate service rate at p_A:

  Without Sensing: $-p_A \ln p_A$.

  With Sensing: $\frac{-p_A \ln p_A}{(x+1)a-(1-xa)p_A \ln p_A - xap_A}$.
- Network throughput at p_A: $\hat{\lambda}_{out,p=p_A} = \hat{\lambda}$ if $p_S \leq p_A \leq p_L$. Otherwise, $\hat{\lambda}_{out,p=p_A} < \hat{\lambda}$.
- The maximum network throughput is $\hat{\lambda}_{\max}$, which is achieved when the backoff parameters $\{q_i\}$ are properly adjusted such that $p_A = p_A^*$, where $p_A^* = e^{-1}$ without sensing and $p_A^* = -\frac{1+x}{x}\mathbb{W}_0\left(-\frac{1}{e(1+1/x)}\right)$ with sensing.

Stability

The network is stable if $\hat{\lambda}_{out} = \hat{\lambda}$.

- The network is stable if it operates at the desired stable point p_L.
  Conditions for the network to operate at p_L: 1) the offered load of each queue $\rho \leq 1$; 2) $p_t \geq p_S$ for any t.
- The network may become unstable if it operates at the undesired stable point p_A.
  Conditions for the network to be stable at p_A: $p_S \leq p_A \leq p_L$.

How to properly tune the backoff parameters $\{q_i\}$ to stabilize the network?

Stable Region

- Let $q_i = q_0 \cdot \mathcal{Q}(i)$, where the backoff function $\mathcal{Q}(i)$ is an arbitrary monotonic decreasing function of i with $\mathcal{Q}(0) = 1$ and $\mathcal{Q}(i) < 1$, $i = 1, \ldots, K$.
- Define the complete stable region of the initial transmission probability q_0 as $S = \{q_0 | \hat{\lambda}_{out} = \hat{\lambda}\}$.

Absolute-stable Region

- Absolute-stable Region: $S_L = [q_l,\ q_u]$. The network is stabilized at the desired stable point p_L if $q_0 \in S_L$.

- Lower-bound q_l:

  Without Sensing: $q_l = \frac{-p_L \ln p_L}{n} \left(\sum_{i=0}^{K-1} \frac{(1-p_L)^i}{\mathcal{Q}(i)} + \frac{(1-p_L)^K}{\mathcal{Q}(K) p_L} \right)$.

  With Sensing: $q_l = \frac{-p_L \ln p_L}{n - \hat{\lambda} - xa\hat{\lambda}(1/p_L - 1)} \left(\sum_{i=0}^{K-1} \frac{(1-p_L)^i}{\mathcal{Q}(i)} + \frac{(1-p_L)^K}{\mathcal{Q}(K) p_L} \right)$.

  For large number of nodes n,
  $q_l \approx \frac{-p_L \ln p_L}{n} \left(\sum_{i=0}^{K-1} \frac{(1-p_L)^i}{\mathcal{Q}(i)} + \frac{(1-p_L)^K}{\mathcal{Q}(K) p_L} \right)$.

- Upper-bound $q_u = -\frac{1}{n} \ln p_S$.

Quasi-stable Region

- Quasi-stable Region: $S_A = \{q_0 \notin S_L | p_S \leq p_A \leq p_L\}$. The network is stabilized at the undesired stable point p_A if $q_0 \in S_A$.
- $S_A = S_L' \bigcap$ $\left[\frac{-p_L \ln p_L}{n}\left(\sum_{i=0}^{K-1} \frac{(1-p_L)^i}{\mathcal{Q}(i)} + \frac{(1-p_L)^K}{\mathcal{Q}(K)p_L}\right), \frac{-p_S \ln p_S}{n}\left(\sum_{i=0}^{K-1} \frac{(1-p_S)^i}{\mathcal{Q}(i)} + \frac{(1-p_S)^K}{\mathcal{Q}(K)p_S}\right)\right]$.

Complete Stable Region

The complete stable region $S = S_L \bigcup S_A$. It is determined by:

- Backoff function $\mathcal{Q}(i)$ and cutoff phase K.
- System parameters including the aggregate input rate $\hat{\lambda}$, the number of nodes n, the collision-detection time x and the normalized propagation delay a (i.e., with sensing).

Access Delay

- Let Y_i denote the holding time of a HOL packet in State R_i and D_i denote the time spent from the beginning of State R_i until the service completion, $i = 0, \ldots, K$.

- D_0 is the access delay of the HOL packet.

- Let $G_X(z)$ denote the probability generating function of X.

  Without Sensing: $G_{D_i}(z) = pG_{Y_i}(z) + (1-p)G_{D_{i+1}}(z)$, $i = 0, \ldots, K-1$, and $G_{D_K}(z) = G_{Y_K}(z)$.

  With Sensing: $G_{D_i}(z) = pzG_{Y_i}(z) + (1-p)z^{xa}G_{Y_i}(z)G_{D_{i+1}}(z)$, $i = 0, \ldots, K-1$, and
  $G_{D_K}(z) = pzG_{Y_K}(z) + (1-p)z^{xa}G_{Y_K}(z)G_{D_K}(z)$.

- The access delay performance is closely determined by which stable point the network operates at.

Mean Access Delay

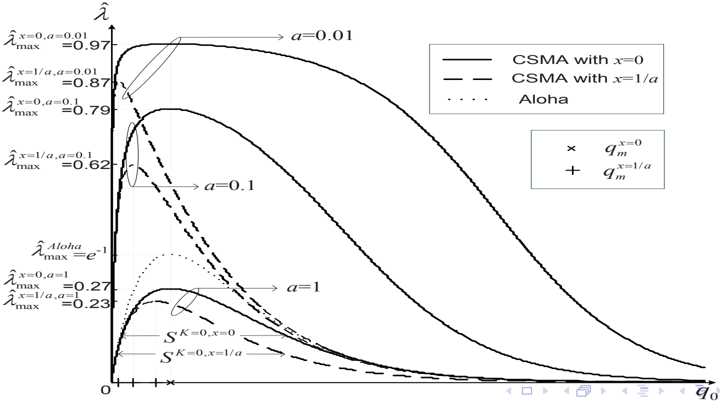
Without Sensing: $E[D_0] = \sum_{i=0}^{K-1} \frac{(1-p)^i}{q_i} + \frac{(1-p)^K}{q_K p}$.

With Sensing: $E[D_0] = 1 + xa \cdot \frac{1-p}{p} + \frac{a}{\alpha}\left(\sum_{i=0}^{K-1} \frac{(1-p)^i}{q_i} + \frac{(1-p)^K}{q_K p}\right)$.

- At the desired stable point p_L: $E[D_{0,p=p_L}]$ can be reduced by increasing the transmission probabilities $\{q_i\}$.
- At the undesired stable point p_A: $E[D_{0,p=p_A}] = \frac{n}{\hat{\lambda}_{out,p=p_A}} > \frac{n}{\hat{\lambda}} \geq E[D_{0,p=p_L}]$ if $p_A > p_L$, or $p_A < p_S$. $E[D_{0,p=p_A}]$ is minimized when the backoff parameters $\{q_i\}$ are properly adjusted such that $p_A = p_A^*$.

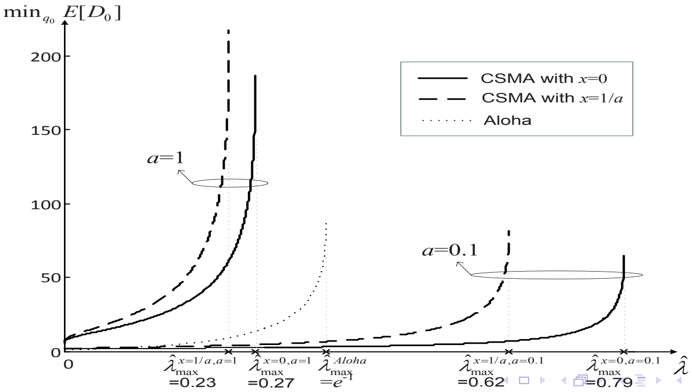
Example 1: p-persistent CSMA versus Aloha with Geometric Retransmission

- Cutoff phase $K = 0$. Assume that the holding time at State R_0, Y_0, is a geometric random variable with parameter q_0.
- Stable region of q_0:



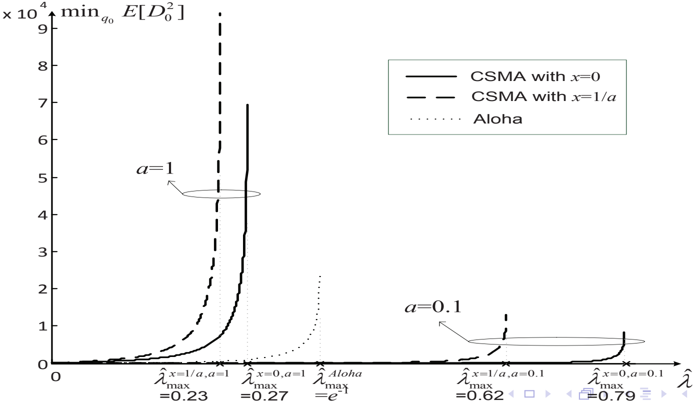
Example 1: p-persistent CSMA versus Aloha with Geometric Retransmission

- Cutoff phase $K = 0$. Assume that the holding time at State R_0, Y_0, is a geometric random variable with parameter q_0.
- Minimum mean access delay:



Example 1: p-persistent CSMA versus Aloha with Geometric Retransmission

- Cutoff phase $K = 0$. Assume that the holding time at State R_0, Y_0, is a geometric random variable with parameter q_0.
- Minimum second moment of access delay:



Remarks

- Better performance (i.e., larger stable region, higher maximum network throughput and lower minimum first and second moments of access delay) is achieved by CSMA when the sensing time (i.e., the normalized propagation delay a and the collision-detection time x) is sufficiently small.

- Backoff parameters should be properly set to optimize the network performance.

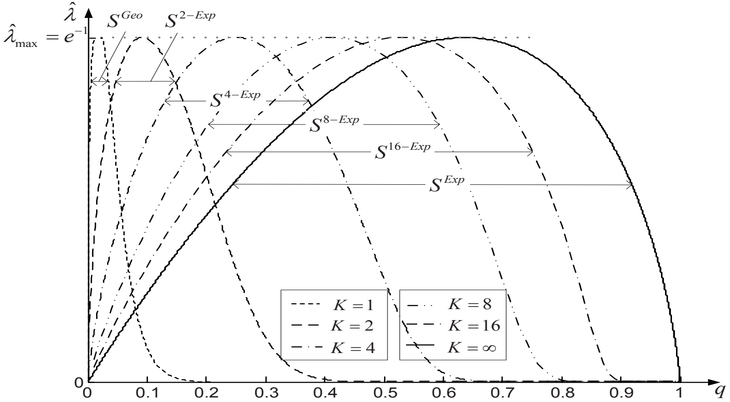
  For instance, to achieve the maximum network throughput, the initial transmission probability q_0 should be set as

$$q_m = \begin{cases} \frac{1}{n} & \text{without sensing} \\ \frac{1}{n}\left(1 + \mathbb{W}_0\left(-\frac{1}{e(1+1/x)}\right)\right) & \text{with sensing.} \end{cases}$$

  As the number of nodes n increases, the transmission probability of each node should be reduced!

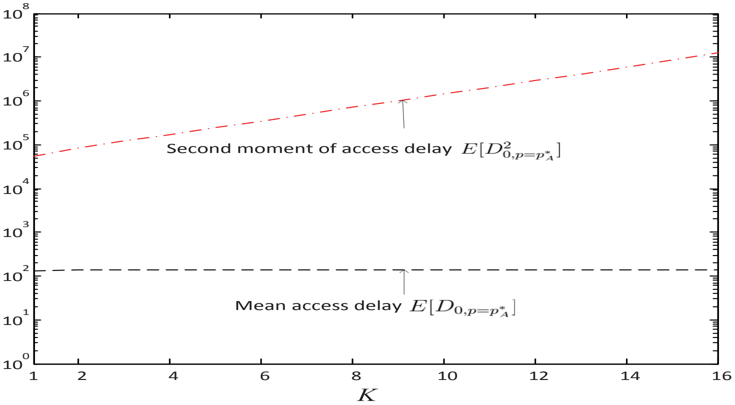
Example 2: Aloha with K-Exponential Backoff

- Assume that the holding time at State R_i, Y_i, is a geometric random variable with parameter q_i. $q_i = q^i$, $i = 0, \cdots, K$. $K = 0, \cdots, \infty$.
- Stable region of q:



Example 2: Aloha with K-Exponential Backoff

- Assume that the holding time at State R_i, Y_i, is a geometric random variable with parameter q_i. $q_i = q^i$, $i = 0, \cdots, K$. $K = 0, \cdots, \infty$.
- Access delay with $q = q_m$ at the undesired stable point p_A:



Remarks

As the cutoff phase K increases:

- HOL packets have more room to reduce their transmission probabilities to alleviate the contention $\rightarrow$ the network has a better capability of remaining stable.
- The difference of transmission probability between a fresh HOL packet and a deeply backlogged one is enlarged $\rightarrow$ a larger second moment of access delay (delay jitter) at the undesired stable point.

Proper backoff design is required to reach a balance between stability and delay performance.

Summary of Unified Analysis of Throughput, Delay and Stability

- The maximum network throughput is independent of backoff parameters and solely determined by the sensing capability of nodes.

  To achieve the maximum network throughput, nevertheless, the backoff parameters should be properly selected.

- Both delay and stability performance crucially depend on the backoff design.

Conclusion

- Three key elements for random-access protocol design: when to start a transmission, when to end a transmission, and how to resolve collisions.
- A unified analysis of throughput, delay and stability for Aloha and CSMA is presented.
- The analysis sheds important light on the practical network design and control, and serves as a crucial step toward a unified theory of random access.

Application to IEEE 802.11 DCF Networks

References

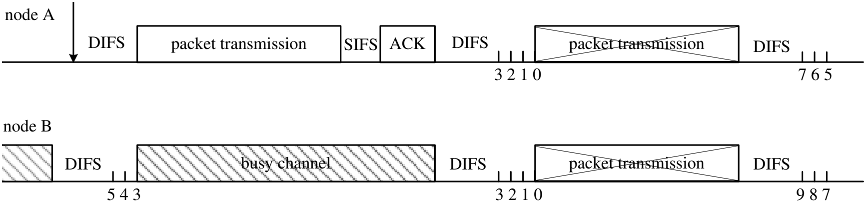
L. Dai and X. Sun, "A unified analysis of IEEE 802.11 DCF networks: stability, throughput and delay," *IEEE Trans. Mobile Computing*, vol. 12, no. 8, pp. 1558–1572, Aug. 2013.

IEEE 802.11 Distributed Coordination Function (DCF) [1]

- Carrier Sense Multiple Access (CSMA) with Binary Exponential Backoff (BEB).
- Two access mechanisms: basic access and request-to-send/clear-to-send (RTS/CTS) access.

[1] *IEEE Std. 802.11-2007, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, June 2007.

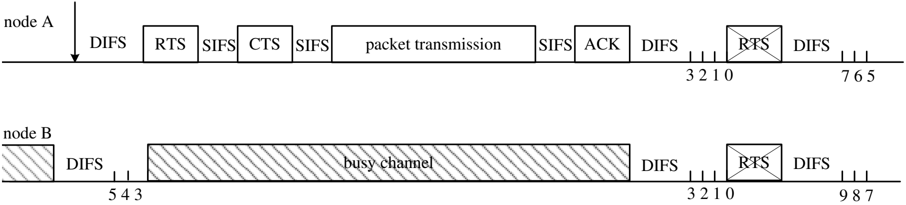
Operation of DCF: Basic Access



If a node has a packet to transmit:

- Sense the channel before the transmission.
- Transmit if the node senses the channel idle.
- Backoff if the node senses the channel busy.
  - Wait until the channel is idle for a duration of DIFS.
  - Choose a random backoff value from $\{0, ..., W_i - 1\}$.
  - Count down when idle, freeze when busy.
  - Transmit when the counter reaches zero.

Operation of DCF: RTS/CTS Access

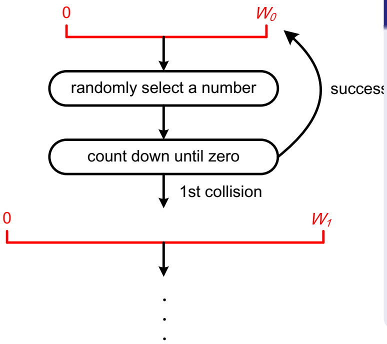


RTS/CTS access mechanism is firstly introduced to deal with the hidden-terminal problem.

- If a node has a packet to transmit, it sends an RTS frame to the destination first to indicate the transmission attempt.
- When the destination receives the RTS frame, it broadcasts a CTS frame to permit the transmission.
- The packet transmission starts after the successful exchange of the RTS and CTS frames.

Operation of DCF: Backoff Scheme

- A sequence of non-decreasing backoff window sizes $\{W_0, W_1, W_2, W_3, ..., W_i, ...\}$. i is the number of collisions the node has experienced.

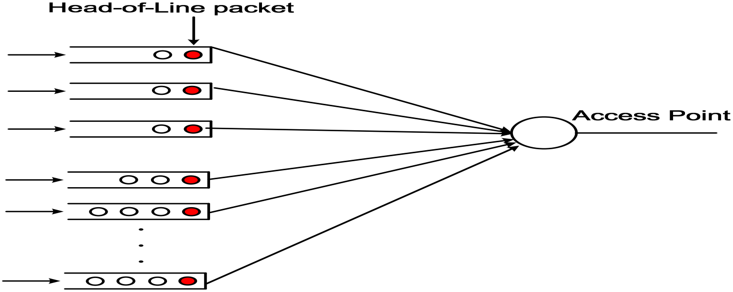


Binary Exponential Backoff

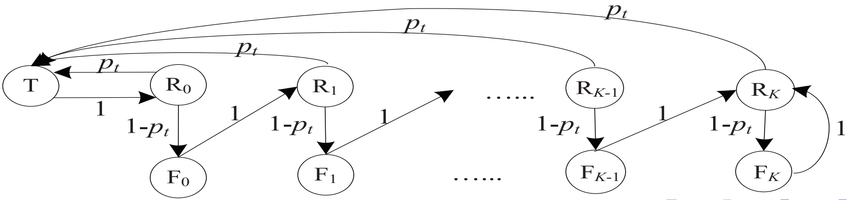
- Backoff window size is doubled after ever collision.
- W: initial backoff window size.
- K: the maximum backoff stage (cutoff phase).

$$W_i = \begin{cases} W \cdot 2^{\min\{i,K\}} & \text{if i-th collision} \\ W & \text{if success} \end{cases}$$

Modeling of IEEE 802.11 DCF Networks



- State characterization of Head-of-Line (HOL) packet



State Characterization of HOL Packets: State T

- State T: The HOL packet makes a successful transmission.
- Mean holding time τ_T at State T depends on the access mechanism.

- Basic Access:

$$\tau_T^{Basic} = \frac{\text{PHY+MAC+Payload+SIFS+ACK+DIFS}}{\text{slot time}}$$

- RTS/CTS Access:

$$\tau_T^{RTS} = \frac{\text{RTS+SIFS+CTS+SIFS+PHY+MAC+Payload+SIFS+ACK+DIFS}}{\text{slot time}}$$

State Characterization of HOL Packets: State F_i

- State F_i: The HOL packet experiences a collision.
- Mean holding time τ_F at State F_i depends on the access mechanism.
- Basic Access:

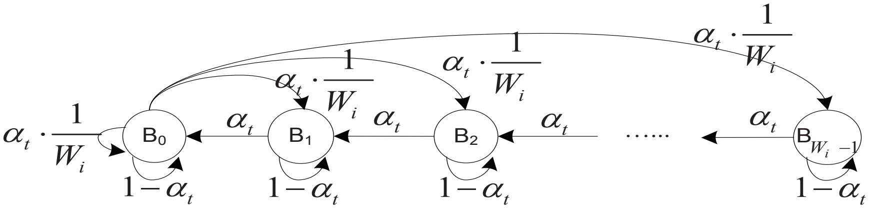
$$\tau_F^{Basic} = \frac{\text{PHY+MAC+Payload+DIFS}}{\text{slot time}}$$

- RTS/CTS Access:

$$\tau_F^{RTS} = \frac{\text{RTS+DIFS}}{\text{slot time}}$$

State Characterization of HOL Packets: State R_i

- State R_i: The HOL packet waits to request a transmission.
- Mean holding time τ_{R_i} at State R_i depends on the backoff window size W_i.



- $\tau_{R_i} = \frac{1}{2\alpha} \cdot (1 + W_i)$, where $\alpha = \lim_{t\to\infty} \alpha_t$ is the limiting probability of sensing the channel idle, which is given by $\alpha = \frac{1}{1+\tau_F - \tau_F p - (\tau_T - \tau_F) p \ln p}$.

Steady-state Operating Points

Desired stable point p_L

$$p_L = \exp\left\{\mathbb{W}_0\left(-\frac{\hat{\lambda}(1+\tau_F)/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}}\cdot\exp\left\{-\frac{\hat{\lambda}\tau_F/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}}\right\}\right)+\frac{\hat{\lambda}\tau_F/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}}\right\}.$$

- p_L is determined by the aggregate input rate $\hat{\lambda}$, and the holding time in successful transmission and collision states, τ_T and τ_F.

Undesired stable point p_A

- When the network becomes saturated, it shifts to the undesired stable point p_A, which is the root of the following fixed-point equation:

$$p = \exp\left\{-\frac{2n}{W\left(\frac{p}{2p-1}+\left(1-\frac{p}{2p-1}\right)(2(1-p))^K\right)}\right\}.$$

- p_A is determined by the network size n and the backoff parameters, i.e., initial backoff window size W and the cutoff phase K.

$\mathbb{W}_0(.)$ is the principal branch of the Lambert W function.

Maximum Network Throughput

The maximum throughput $\hat{\lambda}_{\max}$ of IEEE 802.11 DCF networks is given by

$$\hat{\lambda}_{\max} = \frac{-\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)}{\tau_F/\tau_T - (1-\tau_F/\tau_T)\,\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)}.$$

- $\hat{\lambda}_{\max}$ is determined by the holding time of HOL packets in the successful transmission and collision states, τ_T and τ_F.
- $\hat{\lambda}_{\max}$ varies under different access mechanisms.

Maximum Network Throughput: Basic Access vs RTS/CTS Access

Table : System Parameter Setting in the 802.11n standard

Packet payload	4096*8 bits
MAC header	288 bits
PHY header	136 bits
ACK	248 bits
RTS	288 bits
CTS	240 bits
Channel Bit Rate	54 Mbps
Slot Time	$9\mu s$
SIFS	$16\mu s$
DIFS	$34\mu s$

Basic Access

- $\tau_T^{Basic} = 74.4$ time slots and $\tau_F^{Basic} = 72.1$ time slots.
- $\hat{\lambda}_{max}^{Basic} = 0.85$.

RTS/CTS Access

- $\tau_T^{RTS} = 78.1$ time slots and $\tau_F^{RTS} = 4.4$ time slots.
- $\hat{\lambda}_{max}^{RTS} = 0.94$.

Maximum Network Throughput: Basic Access vs RTS/CTS Access

Other often-used definitions of network throughput:

- The fraction of time that the payload is transmitted:

$$\hat{\lambda}_{\text{max}} \cdot \frac{\text{Payload}}{\tau_T} = \begin{cases} 0.77 & \text{Basic Access} \\ 0.81 & \text{RTS/CTS Access.} \end{cases}$$

- The number of bits that are successfully transmitted in a second:

$$\hat{\lambda}_{\text{max}} \cdot \frac{\text{Payload}}{\tau_T} \cdot \text{Channel Bit Rate} = \begin{cases} 41.6 \text{ Mbps} & \text{Basic Access} \\ 43.8 \text{ Mbps} & \text{RTS/CTS Access.} \end{cases}$$

Optimal Initial Backoff Window Size to Achieve $\hat{\lambda}_{max}$

$$W_m = n \cdot \frac{-\frac{4(1+\tau_F)}{\tau_F}\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right) - 2}{\frac{1+\tau_F}{\tau_F}\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)\ln\left(-\frac{1+\tau_F}{\tau_F}\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)\right)}$$

- To achieve the maximum network throughput $\hat{\lambda}_{max}$, the initial backoff window size should be linearly increased with the network size n.
- The increasing rate is closely dependent on the holding time in collision state τ_F.
  - Basic access ($\tau_F^{Basic} = 72.1$ time slots): $W_m = 10.6n$.
  - RTS/CTS access ($\tau_F^{RTS} = 4.4$ time slots): $W_m = 1.3n$.

Mean Access Delay

At the desired stable point $p_L \approx 1$

$$E[D_{0,p=p_L}] = \tau_T + \frac{1+W}{2}.$$

At the undesired stable point p_A

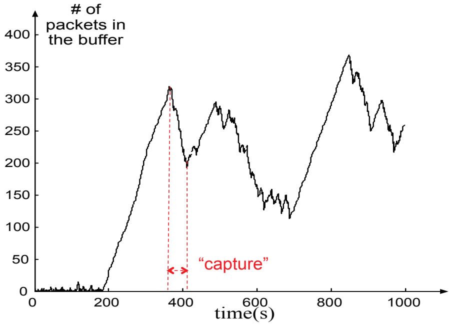
$$E[D_{0,p=p_A}] = \left(\tau_T + \tau_F \frac{1-p_A}{p_A} + (1+\tau_F(1-p_A) - (1-\tau_F)\, p_A \ln p_A)\right. \\ \left.\cdot \left(\frac{1}{2p_A} + \frac{W}{2}\left(\frac{1}{2p_A-1} + \left(\frac{1}{p_A} - \frac{1}{2p_A-1}\right)(2(1-p_A))^K\right)\right)\right).$$

- $E[D_{0,p=p_A}]$ is inversely proportional to the network throughput.
- $\min_W E[D_{0,p=p_A}] = \frac{n}{\hat{\lambda}_{\max}}$, which is achieved when $W = W_m$.

Second Moment of Access Delay

At the undesired stable point p_A

$$E[D^{2,K=\infty}_{0,p=p_A}] = \infty \text{ when } W \leq \frac{4n}{3\ln\frac{4}{3}} \approx 4.63n.$$



- Capture phenomenon: Some node captures the channel and produces a continuous stream of packets, while others have to wait for a long time.
    - Poor queueing delay performance
    - Short-term unfairness

Summary of IEEE 802.11 DCF Networks

- The maximum network throughput is solely determined by the holding time of HOL packets in the successful transmission and collision states, τ_T and τ_F, which varies with the access mechanism, i.e., basic access or RTS/CTS access.
- To achieve the maximum network throughput, the initial backoff window size W should be linearly increased with the network size n.
- With BEB, the second moment of access delay may become infinite if a small initial backoff window size is chosen in saturated conditions, causing poor queueing performance and serious short-term unfairness.

What can be done next?

- Backoff design: In general, the backoff scheme can be characterized by a sequence of backoff window sizes $\{W_i\}$. What is the "best" backoff scheme? [1]
- From homogeneous DCF to heterogeneous DCF: How to model the network if nodes in different groups have distinct input rates and backoff parameters? [2]
- From DCF to EDCA: Given the throughput differentiation requirement, how to properly tune the backoff parameters to optimize the network throughput? [3]

References

- X. Sun and L. Dai, "Backoff Design for IEEE 802.11 DCF Networks: Fundamental Tradeoff and Design Criterion," to appear in *IEEE Trans. Networking*.
- Y. Gao, X. Sun, and L. Dai, "Throughput Optimization of Heterogeneous IEEE 802.11 DCF Networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 398–411, Jan. 2013.
- Y. Gao, X. Sun, and L. Dai, "IEEE 802.11e EDCA Networks: Modeling, Differentiation and Optimization," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3863–3879, July 2014.

The End