A Unified Theory of Random Access

Lin Dai

Department of Electrical Engineering
City University of Hong Kong

lindai@cityu.edu.hk

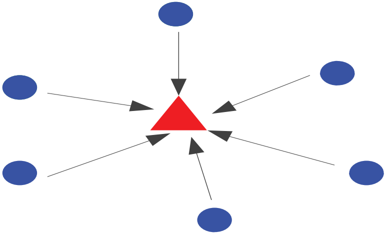
June 2025

Outline

- Random Access: A Huge Gap between Theory and Practice
- A Unified Theory of Random Access: From the Queueing-Theoretical Perspective
- Queue-Stability Region and Transmission Control
- Access Design for Next-Generation Communication Networks

Random Access: A Huge Gap between Theory and Practice

Multiple Access (MAC)



Multiple users transmit to a common receiver: How to share the resources?

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

Centralized Access

- Adopted in cellular systems since the first generation: Each Base Station (BS) allocates dedicated resources to users for data transmission.
- FDMA $\longrightarrow$ TDMA $\longrightarrow$ CDMA $\longrightarrow$ OFDMA
- Optimal resource allocation by BSs
- **Extensive signalling exchange between BSs and users**: Inefficient when the number of users is large and each with a small amount of data.

Random Access

- Adopted in both cellular and WiFi networks:
  - — Cellular in the licensed spectrum: signalling exchange
  - — Cellular in the unlicensed spectrum: data transmission
  - — WiFi: data transmission
- Performance degradation due to lack of coordination among users
- **Small overhead and scalable**: Increasing importance due to the paradigm shift from Human-to-Human (H2H) communications to Machine-to-Machine (M2M) communications.

M2M Communications

- M2M communications is expected to play a dominant role in the next-generation communication networks.
    - 80 billions machine-type devices to be connected to mobile networks by 2025
    - Wide applications in various domains such as smart grid, transportation, health care, manufacturing and monitoring
- Two out of three main services of 5G networks are for M2M communications: *massive Machine-Type Communications (mMTC) and Ultra-Reliable Low-Latency Communications (URLLC)*.
- Four out of six usage scenarios of 6G networks are related to M2M communications: *hyper reliable and low-latency communications, massive communications, ubiquitous connectivity, integrated sensing and communications, integrated AI and communications*.

Features of M2M Communications

- A typical scenario of conventional H2H communications: A relatively small number of users each with a large amount of data to transmit
- M2M Communications:

  — massive number of devices

  — short packet payload

  — diverse and more stringent Quality-of-Service (QoS) requirements

- *Challenge: How to properly design random access schemes for M2M communications?*

Design of Random Access Networks: Three Key Questions

For each node:

- When to start a transmission?
- When to end the transmission?
- What if the transmission fails?

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]

- Examples:

  — *Aloha-based*: Random access process of cellular networks (1G-5G) in the licensed spectrum

  — *CSMA-based*: Distributed Coordination Function (DCF) of WiFi networks, 5G New Radio Unlicensed (NR-U) in the unlicensed spectrum

Question 2: When to End the Transmission?

- Stop when the packet transmission is completed.
- Any "smarter" solution? **Stop when the transmission is deemed a failure**.
    - With full-duplex (i.e., able to receive signals during transmissions): Stop when other on-going transmissions are sensed.
    - With half-duplex (i.e., unable to receive signals during transmissions): Send a short request to reserve the channel first before the data packet transmission. **Connection (Grant)-based Access**
- Examples:

— *Connection-based*: 4-step Random Access-Small Data Transmission (RA-SDT) in 5G networks in the licensed spectrum, Request-To-Send/Clear-To-Send (RTS/CTS) access mechanism of DCF in WiFi networks

— *Connection-free*: 2-step RA-SDT in 5G networks in the licensed spectrum, basic access mechanism of DCF in WiFi networks

Question 3: What if the Transmission Fails?

- Inherent transmission failures caused by incoordination among transmitters.
- The definition of transmission failure depends on what type of receivers is adopted. Various assumptions on the receiver have been made, which can be broadly divided into three categories.
    - *Collision Model*: 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 Model*: 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**, e.g., Successive Interference Cancellation (SIC).

Question 3: What if the Transmission Fails?

- Resolving transmission failures: **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.
- How to set the transmission probability?
  - **Adjust the transmission probability according to the number of transmission failures** i that the packet has experienced, i.e., $q_i = q_0 \cdot \mathcal{Q}(i)$, where $\mathcal{Q}(i)$ is an arbitrary monotonic non-increasing function of the number of transmission failures i, $i = 0, 1, \ldots$.
  - *Binary Exponential Backoff (BEB)*: $\mathcal{Q}(i) = 2^{-i}$.
    Adopted in DCF of WiFi networks and 5G NR-U.

    Constant Backoff: $\mathcal{Q}(i) = 1$.
    Adopted in the random access process of cellular systems in the licensed spectrum.

Design Degrees of Freedom of Random Access Networks

- Sensing-free (Aloha) or Sensing-based (CSMA)
- Connection-free or Connection-based
- Backoff: Constant, Exponential, ...
- ...

Performance Metrics of Random Access Networks

- **Network Throughput**: the average number of successfully decoded packets of the network per time slot.
- **Network Sum Rate**: the average number of successfully decoded information bits of the network per time slot.
- **Delay** of a packet
  - Access delay (service time): the time interval from the instant that it becomes the Head-of-Line (HOL) packet to its successful transmission.
  - Queueing delay (waiting time + service time): the time interval from the packets arrival to its successful transmission.
- **Stability**: A network is stable if all the nodes' queues are stable.
  - Queue-stability: A queue is defined as queue-stable if the steady-state distribution of its queue length exists.
  - Throughput-stability: A queue is defined as throughput-stable if its throughput is equal to the input rate.
- ......

Challenges for Random Access Design

- The network performance is sensitive to the setting of access parameters.
- When the number of users/devices increases, the network performance may drastically degrade due to a significant increase of transmission failures if the access parameters are not properly selected.
- To support **massive access** and **stringent QoS requirements** of M2M communications, the random access schemes need to be carefully designed, with the parameters optimally tuned. — *But how?*

Lack of a Unified Theory of Random Access

- Despite a long history and wide applications, the theory of random access is much less developed than centralized access, which has been the focus of the MAC theory.
- Multiuser information theory:
  - Assume a continuous **bit stream** for each user.
  - Determine the maximum information transmission/encoding rate of each user for reliable communications.
  - Require **coordination** among users' transmissions.
- For random access:
  - Users decide when/how to transmit in a **distributed** manner.
  - Each user has a buffer to store the incoming **packets**.
  - In addition to the maximum information transmission rate, performance metrics related to **data queues**, e.g., stability and delay, are also of central interest.

Fundamental Questions to be Answered

- **How to stabilize the network?**
  - How to determine the **stability region of input rates**, within which and only within which the network can be stabilized?
  - For given input rates within the stability region, how to properly choose the **transmission probabilities of nodes** to stabilize the network?
- **How to optimize the performance?**
  - How to maximize the network **throughput** by optimizing the transmission probabilities of nodes?
  - How to minimize the mean queueing **delay** by optimizing the transmission probabilities of nodes?
  - ......
- **What are the key factors and how would they affect the optimal performance?**
  - What are the effects of **sensing** on the optimal throughput/delay performance?
  - What are the effects of **backoff** functions on the optimal throughput/delay performance?
  - ......

Random Access from a Coding Perspective

- Aloha with advanced coding/decoding design
  - Coded Slotted Aloha (CRDSA [Casini,De Gaudenzi&del Rio Herrero'2007], IRSA [Liva'2011][Munari'2021], Frameless Aloha [Stefanovic,Popovski&Vukobratovic'2012], ...)

    — Assumptions: Time slots are grouped into frames. Each user transmits coded segments in multiple slots of each frame with a certain probability.

    — Objective: Maximize the network throughput.
  - Unsourced Multiple Access [Polyanskiy'2017], [Liva&Polyanskiy'2024]

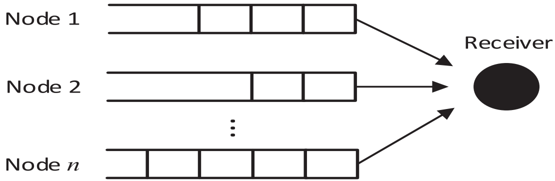
    — Assumptions: K_a out of K ($K \gg K_a$) users are actively transmitting and the receiver is unable to associate messages to users.

    — Objective: Minimize the per-user probability of error

- No consideration of data queues

Random Access from a Queueing Perspective

- Multiple access as a multi-queue-single-server system:

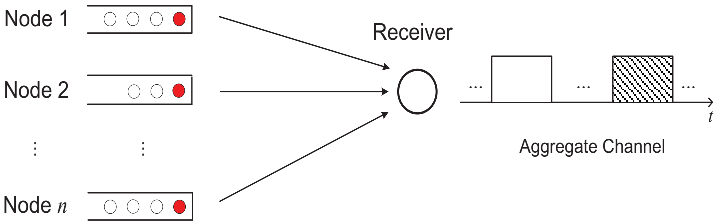


- Nodes have their own data arrivals and uncoordinated transmissions, but coupled service processes.

- How to characterize the coupled service processes?

Modeling of Random Access Networks

- Analytical models are usually customized for specific random access schemes to tackle specific problems.
- Modeling approaches can be broadly divided into two categories: **Channel-centric** and **Node-centric**, where the former focuses on modeling **the aggregate channel/traffic**, and the latter focuses on modeling **nodes' behavior**.



Channel-Centric Modeling

- Modeling focus: Aggregate traffic
- Representative models:
  - [Abramson'1970], [Kleinrock&Tobagi'1975]
    - Assume the aggregate traffic follows the Poisson distribution.
  - [Carleial&Hellman'1975], [Kleinrock&Lam'1975]
    - Model the dynamic change of the aggregate traffic
- Widely adopted in performance analysis of random access procedure in cellular networks.
- Capture the essence of contention among nodes and simplify the throughput analysis.
- **Ignore nodes' queues**.

Node-Centric Modeling

- Modeling focus: Nodes' queues
- Representative models:
  - [Tsybakov&Mikhailov'1979]
    - Model the queue lengths of n nodes as an n-dimensional Markov chain.
    - **Tractable only for two-node Aloha**.
  - [Rao&Ephremides'1988], [Szpankowski'1994] [Luo&Ephremides'1999], [Dimitriou&Pappas'2018]
    - Develop a hypothetical dominant system, where a node would send dummy packets when its queue is empty.
    - **Tractable only for two-node Aloha**.
    - **Approximations and bounds** for $n \geq 3$.

Node-Centric Modeling

- Modeling focus: Nodes' backoff process
- Representative models:
  - [Bianchi'2000]
    - Widely adopted in performance analysis of WiFi (**Symmetric CSMA+BEB**) networks.
    - Model the **backoff** behavior of each single node by a two-dimensional Markov chain.
    - Accurate characterization of network throughput and mean access delay of packets when all the nodes' queues are assumed to be saturated.
    - Nodes' states are only partially captured by the two-dimensional Markov chain – Difficult to be extended to the **unsaturated** condition.

Unsolved Issues

- How to characterize the coupled service processes when the queues are not all-saturated?
  - How to determine the stability region of input rates, only within which the network can be stabilized?
  - For given input rates within the stability region, how to tune the access parameters of nodes to stabilize the network?
  - How to optimize the delay performance?
  - ......
- **A unified framework based on which various performance metrics can be analyzed and optimized.**

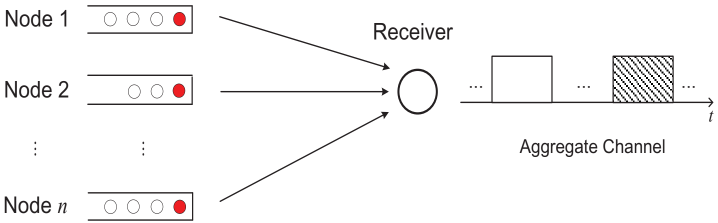
Unsolved Issues

- How to characterize the effects of key factors on the optimal performance?
  - To sense or not to sense: When is sensing beneficial?
  - Connection-based or connection-free: When is establishing a connection beneficial?
  - Constant backoff or exponential backoff: What backoff function is the best?
  - ......
- **A unified framework based on which various design features can be incorporated.**

A Unified Theory of Random Access:
From the Queueing-Theoretical Perspective

Toward a Unified Analytical Framework for Random Access

- For modeling of a multi-queue-single-server system, the main challenge lies in characterization of the **coupled** service processes of queues.



- For each node's queue, the service process is determined by the aggregate activities of *all* **Head-Of-Line (HOL)** packets and how many of them can be successfully received. The former depends on the access/transmission strategy, and the latter depends on the receiver design.

Toward a Unified Analytical Framework for Random Access

- Key to characterization of coupled service processes of nodes' queues [Dai'12] [Dai'13] [Dai'22] [Dai'24]:
  - **Modeling HOL packets' behavior**: Discrete-time Markov renewal process
  - **Capturing the coupling of service processes**: Fixed-point equations of steady-state probabilities of successful transmission of HOL packets $\mathbf{p}$

L. Dai, "A theoretical framework for random access: Effects of carrier sensing on stability," *IEEE Trans. Commun.*, vol. 73, no. 1, pp. 275–291, Jan. 2025.

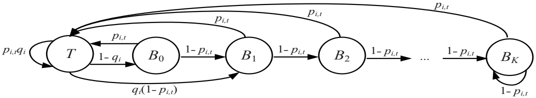
L. Dai, "A theoretical framework for random access: Stability regions and transmission control," *IEEE/ACM Trans. Networking*, vol. 30, no. 5, pp. 2173–2200, Oct. 2022.

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

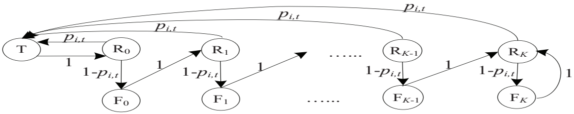
L. Dai, "Stability and delay analysis of buffered Aloha networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2707–2719, Aug. 2012.

Modeling of HOL Packets

- HOL packets' behavior can be modeled as a discrete-time Markov renewal process $(\mathbf{X}_i, \mathbf{V}_i) = \{(X_{i,j}, V_{i,j}), j = 0, 1, \ldots\}$, $i \in \{1, 2, \ldots, n\}$:
  - The embedded Markov chain $\mathbf{X} = \{X_j\}$ **without Sensing**:



  - The embedded Markov chain $\mathbf{X} = \{X_j\}$ **with Sensing**:



  - With sensing, the sensing states need to be distinguished from the transmission states.
  - The holding time of each state depends on sensing, backoff scheme, and protocol overhead.
  - $p_{i,t}$: probability of success given that the HOL packet of Node i is transmitted at t.

Characterization of Steady-State Probabilities of Success of HOL Packets $\mathbf{p}$

- Two basic assumptions:
  - **State-independence Assumption**: $\lim_{t\to\infty} p_{i,t} = p_i$, $i \in \mathcal{N}$.
  - **Queue-independence Assumption**: $\beta_{ij} = 1$ for all $i \neq j$, $i, j \in \mathcal{N}$, where

    $$\beta_{ij} = \frac{\Pr\{\text{Node } j\text{'s queue is empty} \mid \text{Node } i\text{'s queue is busy}\}}{\Pr\{\text{Node } j\text{'s queue is empty}\}}$$

    indicates the dependency between queues of Node i and Node j.

- Based on the **state-independence assumption**, the steady-state probability of successful transmission of HOL packets of Node $i \in \mathcal{N}$, p_i, can be written as

$$p_i = \sum_{\mathcal{S}_i \subseteq \mathcal{N}\setminus\{i\}} \Pr\{\text{Success of Node } i \mid \text{Node } i \text{ requests and the set of transmitters with concurrent transmissions is } \mathcal{S}_i\} \cdot \Pr\{\text{Set of transmitters with concurrent transmissions is } \mathcal{S}_i \mid \text{Node } i \text{ requests}\}. \quad (1)$$

Characterization of Steady-State Probabilities of Success of HOL Packets $\mathbf{p}$

- Based on the **queue-independence assumption**, (1) can be written as

$$p_i = \sum_{\mathcal{S}_i \subseteq \mathcal{N} \setminus \{i\}} r_{i|\mathcal{S}_i} \cdot \prod_{j \in \mathcal{S}_i} \omega_j \cdot \prod_{j \in \mathcal{N} \setminus (\mathcal{S}_i \bigcup \{i\})} (1 - \omega_j), \quad i \in \mathcal{N} \qquad (2)$$

  - I_i: Boolean indicator to indicate whether Node i transmits
  - ω_i: Transmission probability of Node i
    - ω_i depends on whether Node i's queue is **unsaturated/saturated** and whether sensing is available.
  - $r_{i|\mathcal{S}_i}$: Conditional success probability of Node i for given set $\mathcal{S}_i$ of transmitters with concurrent transmissions
    - $r_{i|\mathcal{S}_i}$ depends on specific **receiver and channel models**.

A Unified Analytical Framework of Random Access

- Markov renewal process of HOL packets:
  - Various design features of random access (e.g., sensing-based or sensing-free, connection-based or connection-free, backoff, retry limit, ...) can be incorporated.
- Fixed-point equations of steady-state probabilities of successful transmission of HOL packets $\mathbf{p}$:
  - Various receiver and channel models (e.g., collision receiver, capture receiver, SIC receiver, AWGN, Rayleigh fading, ...) can be incorporated.
- Markov renewal process of HOL packets + Fixed-point equations of steady-state probabilities of successful transmission of HOL packets $\mathbf{p}$:
  - Service process of each node's queue can be characterized, based on which various performance metrics can be analyzed and optimized.

Based on the Proposed Unified Analytical Framework

- **Fundamental Limits**
  - Maximum network throughput
  - Minimum mean access/queueing delay
  - Maximum network sum rate
  - Stability region of input rates
- **Insights to Network Design**
  - Optimal tuning of backoff parameters (transmission probability, backoff window size, ...) based on the long-term traffic input rates of nodes
  - Effects of key factors (sensing, backoff function, connection establishment, network size, receiver design, ...) on limiting performance and performance tradeoffs
  - Applications to practical networks

A Glimpse of Our Work

<table>
<tr><th rowspan="2"></th><th rowspan="2">Aloha</th><th rowspan="2">CSMA</th><th rowspan="2">WiFi Networks</th><th colspan="2">4G/5G Networks</th></tr>
<tr><th>Licensed Bands</th><th>Unlicensed Bands</th></tr>
<tr><td rowspan="2">Network Throughput Optimization</td><td>[Dai’12]
[Gao-Dai’19]</td><td>[Dai’13], [Sun-Dai’16]
[Gao-Dai’19]</td><td rowspan="2">[Dai-Sun’13], [Gao-Sun-Dai’13], [Gao-Dai’13], [Gao-Sun-Dai’14], [Sun-Dai’15], [Sun-Dai’16], [Gao-Dai-Hei’17]</td><td rowspan="2">[Zhan-Dai’18]
[Zhan-Dai’19]</td><td rowspan="2">[Sun-Dai’20]</td></tr>
<tr><td colspan="2">[Gao-Fang-Song-Dai’23]</td></tr>
<tr><td rowspan="2">Delay Optimization</td><td>[Dai’12]
[Li-Zhan-Dai’21]
[Zhao-Dai’24]
[Wang-Dai-Sun’24]</td><td>[Dai’13], [Sun-Dai’16]</td><td rowspan="2">[Dai-Sun’13], [Sun-Dai’15], [Sun-Dai’16]</td><td rowspan="2">[Zhan-Dai’19]
[Li-Zhan-Dai’21]
[Zhao-Dai’24]
[Wang-Dai-Sun’24]</td><td rowspan="2"></td></tr>
<tr><td colspan="2">[Zhao-Dai’25]</td></tr>
<tr><td>Network Sum Rate Optimization</td><td>[Li-Dai’16]
[Li-Dai’18]</td><td>[Sun-Dai’17]
[Sun-Dai’19]</td><td>[Sun-Dai’17], [Gao-Sun-Dai’19]</td><td></td><td></td></tr>
<tr><td>Stability Region</td><td>[Dai’22]
[Yang-Dai’23]</td><td>[Dai’25]</td><td></td><td></td><td></td></tr>
</table>

Queue-Stability Region and Transmission Control

What is Network Stability?

- Stability can be defined in many ways.
- **Queue-stability**: A queue is defined as stable if the steady-state distribution of its queue length $Q(t)$ exists:

$$\lim_{t\to\infty} \Pr\{Q(t) < x\} = F(x) \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1,$$

  which holds when the input rate is smaller than the service rate for stationary arrival and service processes, that is, when the queue is unsaturated.
- **Network stability**: A network is queue-stable if all the nodes' queues are queue-stable.

Stability Region of Input Rates

- For given access and receiver design, there exists a **stability region of input rates** of nodes, only within which the network can stay stable by properly setting the access parameters of nodes.

- Queues can be stabilized only when the *input rates* are within the *stability region*.
  (Reliable communication is achievable only when the *transmission rates* of users are within the *capacity region*.)

- How to characterize the stability region of input rates?

Transmission Control

- For given input rates within the stability region, the access parameters, e.g., transmission probability, still need to be carefully set to achieve network stability.
  (For given transmission rates of users within the capacity region, proper coding schemes need to be adopted to achieve reliable communications.)

- For given input rates within the stability region, how to properly set access parameters for achieving stability?

Let Us Start from Connection-Free Aloha

- Assume that n nodes transmit to a common receiver, each with an infinite buffer to store the incoming data packets. Nodes have independent stationary arrival processes, and packets are served on a first-come-first-served basis. Packets are removed from buffers only after being successfully transmitted.
- In each time slot, Node i transmits its HOL packet with probability of $q_{i,k} = q_i \cdot \mathcal{Q}(k)$, $k = 0, 1, \ldots, K$, $i \in \mathcal{N} = \{1, 2, \ldots, n\}$, where
  - $0 < q_i \leq 1$ is the initial transmission probability of Node i;
  - k is determined by the number of transmission failures experienced by the HOL packet until it reaches the cutoff phase K;
  - $\mathcal{Q}(k)$ is the backoff function, which is usually chosen as a monotonic non-increasing function of k.
- Assume the collision receiver. That is, a packet transmission is successful only when there are no concurrent transmissions.

Let Us Start from Connection-Free Aloha

- $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]$: Input rate vector of n nodes;
  $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]$: Service rate vector of n nodes;
  $\boldsymbol{q} = [q_1, \ldots, q_n]$: Initial transmission probability vector of n nodes.
- For Node i's queue: If $\lambda_i < \mu_i$, then the queue is unsaturated, i.e., with a non-zero probability of being empty;. Otherwise, it is saturated.
- When $\boldsymbol{\lambda} < \boldsymbol{\mu}$, all the nodes' queues are unsaturated – the network is queue-stable.

Let Us Start from Connection-Free Aloha

- Queue-stability region $S_Q(\boldsymbol{\lambda})$:

$$S_Q(\boldsymbol{\lambda}) \triangleq \bigcup_{\boldsymbol{q}} \{\boldsymbol{\lambda} : \boldsymbol{\lambda} < \boldsymbol{\mu}\}.$$

  There exists some initial transmission probability vector to achieve stability if and only if $\boldsymbol{\lambda} \in S_Q(\boldsymbol{\lambda})$.

- All-unsaturated region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$:

$$S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda}) \triangleq \{\boldsymbol{q} : \boldsymbol{\lambda} < \boldsymbol{\mu}\}.$$

  - For $\boldsymbol{\lambda} \in S_Q(\boldsymbol{\lambda})$, stability can be achieved with $\boldsymbol{q} \in S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$.
  - The all-unsaturated region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$ can also be defined with regard to the input rate vector $\boldsymbol{\lambda}$.

The 2-Node Case

- Queue-stability region [Rao&Ephremides'1988]:

$$S_Q(\boldsymbol{\lambda}) = \{\boldsymbol{\lambda} : \sqrt{\lambda_1} + \sqrt{\lambda_2} < 1\}.$$

- All-unsaturated region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$ in [Tsybakov&Mikhailov'1979] as:

$$\begin{aligned} &\lambda_1 \geq q_1(1-q_2), \quad \text{and} \quad \lambda_2 < (1-q_1)(1-\tfrac{\lambda_1}{q_1}), \\ \text{or} \quad &\lambda_1 < q_1(1-q_2), \quad \text{and} \quad \lambda_2 < q_2(1-\tfrac{\lambda_1}{1-q_2}), \end{aligned}$$

and in [Rao&Ephremides'1988] as

$$\begin{aligned} &\lambda_1 < q_1(1-\tfrac{\lambda_2}{1-q_1}), \quad \text{and} \quad \lambda_2 < q_2(1-q_1), \\ \text{or} \quad &\lambda_1 < q_1(1-q_2), \quad \text{and} \quad \lambda_2 < q_2(1-\tfrac{\lambda_1}{1-q_2}). \end{aligned}$$

They are equivalent.

What if $n > 2$?

For characterization of the queue-stability region:

- Approximations/Bounds were developed.
- Why is it so challenging to characterize the exact stability region?
  - Design of the *dominant* system becomes extremely complicated even for $n = 3$.
  - To obtain $S_Q(\boldsymbol{\lambda})$ through $S_Q(\boldsymbol{\lambda}) = \bigcup_{\boldsymbol{q}} S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$: $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$ is even more difficult to characterize!

What if $n > 2$?

For transmission control (tuning of initial transmission probability of each node to stabilize the network):

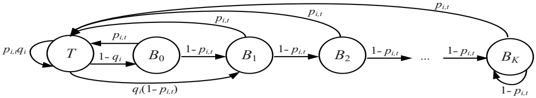
- Adaptive tuning of transmission probability based on a drift analysis [Carleial&Hellman'1975], [Kleinrock&Lam'1975]......:
  - Symmetric case where each node has identical transmission probability and is equipped with a one-packet buffer.
  - Stability defined from a control perspective, which is different from the queue-stability.
  - Tuning based on the number of backlogged nodes in the network, which requires fast tracking of time-varying network state.
- To achieve queue-stability, the initial transmission probability of each node only need to be set based on the input rates: $\boldsymbol{q} \in S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$. But how to characterize $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$?

L. Dai, "A theoretical framework for random access: Stability regions and transmission control," *IEEE/ACM Trans. Networking*, vol. 30, no. 5, pp. 2173-2200, Oct. 2022.

What if $n > 2$?

The proposed unified analytical framework of random access:

- Markov renewal process of each node's HOL packets with Aloha: $(\mathbf{X}_i, \mathbf{V}_i) = \{(X_{i,j}, V_{i,j}), j = 0, 1, \ldots\}, i \in \{1, 2, \ldots, n\}$



- Steady-state probability of successful transmission of each node's HOL packets based on

  1) State-independence assumption: $\lim_{t\to\infty} p_{i,t} = p_i,\ i \in \mathcal{N}$

  2) Queue-independence assumption:

  $\beta_{ij} = \dfrac{\Pr\{\text{Node } j\text{'s queue is empty} \mid \text{Node } i\text{'s queue is busy}\}}{\Pr\{\text{Node } j\text{'s queue is empty}\}} = 1$ for all $i \neq j$, $i, j \in \mathcal{N}$

With Aloha and the collision receiver, (2) can be written as:

$$p_i = \prod_{j \in \mathcal{N}^{\mathcal{U}} \setminus \{i\}} \left(1 - \frac{\lambda_j}{p_j}\right) \cdot \prod_{j \in \mathcal{N}^{\mathcal{S}} \setminus \{i\}} \left(1 - \frac{\mu_j}{p_j}\right), \tag{3}$$

$i \in \mathcal{N}$. Instead of jointly solving n fixed-point equations, let us convert them into a single one by introducing

$$c = \prod_{j \in \mathcal{N}^{\mathcal{U}}} \left(1 - \frac{\lambda_j}{p_j}\right) \cdot \prod_{j \in \mathcal{N}^{\mathcal{S}}} \left(1 - \frac{\mu_j}{p_j}\right),$$

which can be written as

$$c = \prod_{i=1}^{n} \frac{c}{p_i} = \prod_{i=1}^{n} \frac{c}{v_i^{-1}(c)},$$

where

$$c = v_i(p_i) = \begin{cases} p_i \cdot \left(1 - \frac{\lambda_i}{p_i}\right) & \theta_i = U \\ p_i \cdot \left(1 - \frac{\mu_i}{p_i}\right) & \theta_i = S. \end{cases}$$

In the All-Unsaturated Condition

To achieve queue-stability, all the nodes' queues should be unsaturated, i.e., $\mathcal{N}^{\mathcal{U}} = \mathcal{N}$. (3) can be written as

$$p_i = \prod_{j \in \mathcal{N} \setminus \{i\}} \left(1 - \frac{\lambda_j}{p_j}\right), \tag{4}$$

$i \in \mathcal{N}$, which can be transformed into a single fixed-point equation

$$c = \prod_{i=1}^{n} \frac{c}{c + \lambda_i}. \tag{5}$$

(5) has the following properties.

Lemma

(5) *has zero or two positive real roots if* $\sum_{i=1}^{n} \lambda_i < 1$. *Otherwise* (5) *has zero positive real roots.*

In the All-Unsaturated Condition

Let $f(c) = \frac{c^n}{\prod_{j \in \mathcal{N}}(c+\lambda_j)}$. Lemma 1 indicates that $f(c)$ has two positive real fixed points. Denote the two positive fixed points as $0 < c_S < c_L < 1$. Lemma 2 shows that only the larger one is an attracting fixed point.

Lemma

c_L is an attracting fixed point and c_S is a repelling fixed point.

Consider the fixed-point iteration $c_{t+1} = f(c_t)$, $t = 0, 1, \ldots$. Lemma 3 presents a sufficient and necessary condition for the convergence of c_t to c_L.

Lemma

A sufficient and necessary condition for the convergence of c_t to c_L is $\min_t c_t > c_S$.

In the All-Unsaturated Condition

- Lemma 1 and Lemma 2 suggest that the n fixed-point equations of $\boldsymbol{p}$ given in (4) have two roots $\boldsymbol{p}_L = (p_{1,L}, p_{2,L}, \ldots, p_{n,L})$ and $\boldsymbol{p}_S = (p_{1,S}, p_{2,S}, \ldots, p_{n,S})$, where

$$p_{i,L} = \lambda_i + c_L, \qquad p_{i,S} = \lambda_i + c_S,$$

  $i \in \mathcal{N}$. Only $\boldsymbol{p}_L$ is the steady-state point in the all-unsaturated condition. $\boldsymbol{p}_S$ is a transient equilibrium point. We refer to $\boldsymbol{p}_L$ as the all-unsaturated steady-state point.

- Existence of $\boldsymbol{p}_L$ and $\boldsymbol{p}_S$ requires the existence of two positive real roots c_L and c_S of (5).
  - The sufficient and necessary condition of the **existence** of c_L and c_S indeed dictates the queue-stability region $S_Q(\boldsymbol{\lambda})$.
  - The sufficient and necessary condition of the **convergence** to c_L is the key to characterization of the all-unsaturated region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$.

Queue-Stability Region $S_Q(\boldsymbol{\lambda})$

- $n = 2$: For the two-node case, (5) becomes

$$c^2 + (\lambda_1 + \lambda_2 - 1)c + \lambda_1\lambda_2 = 0,$$

which has two positive real roots if and only if

$$(1 - \lambda_1 - \lambda_2)^2 - 4\lambda_1\lambda_2 > 0.$$

We have

$$S_Q^{n=2}(\boldsymbol{\lambda}) = \{\boldsymbol{\lambda} : \sqrt{\lambda_1} + \sqrt{\lambda_2} < 1\}.$$

Queue-Stability Region $S_Q(\boldsymbol{\lambda})$

- $n = 3$: For the three-node case, (5) becomes

$$c^3 + (\lambda_1 + \lambda_2 + \lambda_3 - 1)c^2 + (\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3)c + \lambda_1\lambda_2\lambda_3 = 0,$$

which has three real roots (i.e., two positive and one negative) if and only if

$$(R_1^3 - \tfrac{9}{2}R_1R_2 + \tfrac{27}{2}R_3)^2 < (R_1^2 - 3R_2)^3,$$

where

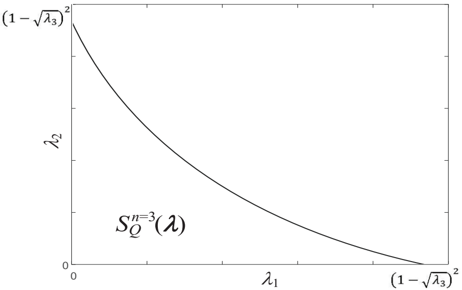
$$R_1 = \sum_{i\in\mathcal{N}} \lambda_i - 1, \quad R_2 = \sum_{i,j\in\mathcal{N}, i\neq j} \lambda_i\lambda_j,$$
$$R_3 = \sum_{i,j,k\in\mathcal{N}, i\neq j\neq k} \lambda_i\lambda_j\lambda_k.$$

We have

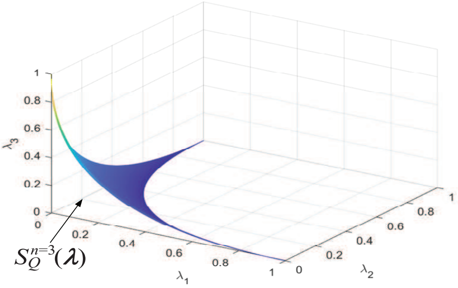
$$S_Q^{n=3}(\boldsymbol{\lambda}) = \{\boldsymbol{\lambda} : 4R_1^3R_3 - R_1^2R_2^2 + 4R_2^3 - 18R_1R_2R_3 + 27R_3^2 < 0\}.$$

Queue-Stability Region $S_Q(\boldsymbol{\lambda})$

- $n = 3$: $S_Q^{n=3}(\boldsymbol{\lambda}) = \{\boldsymbol{\lambda} : 4R_1^3 R_3 - R_1^2 R_2^2 + 4R_2^3 - 18 R_1 R_2 R_3 + 27 R_3^2 < 0\}$



(a)



(b)

Figure: Illustration of queue-stability region $S_Q^{n=3}(\boldsymbol{\lambda})$ for $n = 3$.

Queue-Stability Region $S_Q(\boldsymbol{\lambda})$

- $n \geq 4$: When $n = 4$, the condition for existence of roots becomes excessively complicated. It cannot be expressed using radicals for $n \geq 5$.

  Considering that $\prod_{i=1}^{r} \lambda_i$ becomes insignificant for $r \geq 4$. With $\prod_{i=1}^{r} \lambda_i \approx 0$ for $r \geq 4$, (5) reduces to a cubic equation:

  $$c^3 + R_1 c^2 + R_2 c + R_3 = 0.$$

  We have

  $$S_Q^{n\geq 4}(\boldsymbol{\lambda}) \approx \{\boldsymbol{\lambda} : 4R_1^3 R_3 - R_1^2 R_2^2 + 4R_2^3 - 18 R_1 R_2 R_3 + 27 R_3^2 < 0\}.$$

  Accuracy can be verified by simulations.

Queue-Stability Region $S_Q(\boldsymbol{\lambda})$

- Symmetric case: $\lambda_i = \lambda$, $q_i = q$, $i \in \mathcal{N}$.

  (4) becomes
  $$p = \left(1 - \frac{\lambda}{p}\right)^{n-1},$$
  and (5) becomes
  $$c^{n-1} = (c+\lambda)^n. \tag{6}$$
  (6) has two positive real roots $c_L > c_S$ if and only if $\lambda < \frac{(n-1)^{n-1}}{n^n}$.

  The queue-stability region is then
  $$S_Q(\lambda) = \left\{\lambda : \lambda < \frac{(n-1)^{n-1}}{n^n}\right\}. \tag{7}$$

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

- To achieve queue-stability, we need
  - convergence to the all-unsaturated steady-state point $\boldsymbol{p}_L$;
  - the input rate of each node's queue being smaller than its service rate at $\boldsymbol{p}_L$.
- Take the example of constant backoff, i.e., the backoff function $\mathcal{Q}(k) = 1$. In this case, the service rate $\boldsymbol{\mu} = \boldsymbol{p} \circ \boldsymbol{q}$.

  The all-unsaturated region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$ can be written as

$$S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda}) = \{\boldsymbol{q} : \boldsymbol{p} = \boldsymbol{p}_L,\ \boldsymbol{\lambda} < \boldsymbol{p}_L \circ \boldsymbol{q}\}.$$

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

Let us first check a few sufficient conditions for achieving queue-stability:

1)

$$\{\boldsymbol{q} : \boldsymbol{\lambda} < \boldsymbol{p}_A \circ \boldsymbol{q}\}. \tag{8}$$

$\boldsymbol{p}_A$ denote the all-saturated steady-state point. As $\boldsymbol{p}_A < \boldsymbol{p}_L$, if $\boldsymbol{\lambda} < \boldsymbol{p}_A \circ \boldsymbol{q}$, then $\boldsymbol{\lambda} < \boldsymbol{p}_L \circ \boldsymbol{q}$ must hold.

2)

$$\{\boldsymbol{q} : \boldsymbol{\lambda} < \boldsymbol{p}_L \circ \boldsymbol{q},\ \boldsymbol{p}_A > \boldsymbol{p}_S\}. \tag{9}$$

According to Lemma 3, a sufficient and necessary condition for convergence to c_L, or equivalently, $\boldsymbol{p}_L$, is $\min_t c_t > c_S$, that is,

$$\prod_{i=1}^{n}\left(1 - \frac{\lambda_i}{p_{i,t}}\right) > \prod_{i=1}^{n}\left(1 - \frac{\lambda_i}{p_{i,S}}\right), \quad \text{for all } t. \tag{10}$$

If $\min_t p_{i,t} > p_{i,S}$ for all $i \in \mathcal{N}$, then (10) holds.

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

3)

$$\left\{\boldsymbol{q} : \boldsymbol{\lambda} < \boldsymbol{p}_L \circ \boldsymbol{q}, \ \bigcup_{i \in \mathcal{N}} \left\{\boldsymbol{p}_A \setminus \{p_{i,A}\} > \boldsymbol{p}_S \setminus \{p_{i,S}\}\right\}\right\}. \tag{11}$$

According to the fixed-point equations of $\boldsymbol{p}$ in the all-unsaturated condition given in (4), we have $p_{i,t} = \prod_{j \in \mathcal{N} \setminus \{i\}} \left(1 - \frac{\lambda_j}{p_{j,t-1}}\right)$ for $i \in \mathcal{N}$. $p_{i,t} > p_{i,S}$ for all t as long as for all $j \neq i$, $\min_t p_{j,t} > p_{j,S}$.

- In the symmetric case, (8), (9) and (11) all reduce to the same form, which is a sufficient and necessary condition.
- In the general asymmetric case, neither (8) nor (9) is a necessary condition. (11) is a necessary condition only when $n = 2$.

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

- $n = 2$

Theorem

The all-unsaturated region of $\boldsymbol{q}$ *for* $n = 2$ *is*

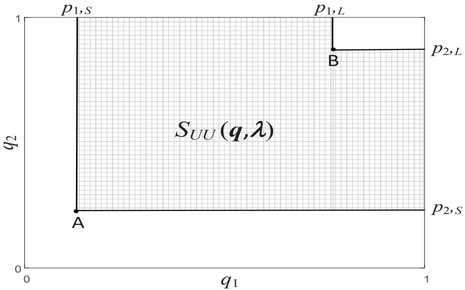
$$S_{UU}(\boldsymbol{q}, \boldsymbol{\lambda}) = \left\{ \boldsymbol{q} : \bigcup_{\substack{i,j \in \{1,2\} \\ i \neq j}} \left\{ \frac{\lambda_i}{p_{i,L}} < q_i < \frac{\lambda_i}{p_{i,S}},\ \frac{\lambda_j}{p_{j,L}} < q_j \leq 1 \right\} \right\}.$$

Note that with $n = 2$, $\frac{\lambda_i}{p_{i,L}} = 1 - p_{j,L} = p_{i,S}$ for $i \neq j$, $i, j \in \{1, 2\}$. As a result, $S_{UU}(\boldsymbol{q}, \boldsymbol{\lambda})$ can also be written as

$$S_{UU}(\boldsymbol{q}, \boldsymbol{\lambda}) = \left\{ \boldsymbol{q} : \bigcup_{\substack{i,j \in \{1,2\} \\ i \neq j}} \left\{ p_{i,S} < q_i < p_{i,L},\ p_{j,S} < q_j \leq 1 \right\} \right\}.$$

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

- $n = 2$



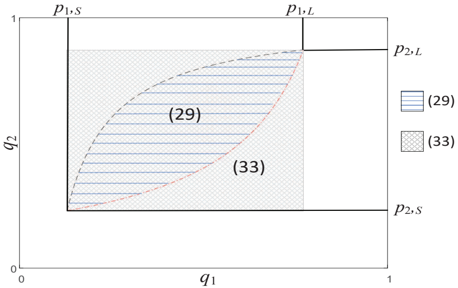


Figure: Illustration of (a) all-unsaturated region $S_{UU}(\boldsymbol{q}, \boldsymbol{\lambda})$ and (b) sufficient but not necessary conditions for queue-stability (8) and (9) for $n = 2$.

- The all-unsaturated region $S_{UU}(\boldsymbol{q},\boldsymbol{\lambda})$ can also be written as a region of input rates $\boldsymbol{\lambda}$ for given transmission probabilities $\boldsymbol{q}$:

$$S_{UU}(\boldsymbol{q},\boldsymbol{\lambda}) = \begin{cases} \left\{\boldsymbol{\lambda} : \bigcap_{i,j\in\{1,2\},i\neq j}\left\{\lambda_i < q_i\left(1-\frac{\lambda_j}{1-q_i}\right)\right\}\right\} & q_1+q_2<1 \\ \left\{\boldsymbol{\lambda} : \bigcup_{i,j\in\{1,2\},i\neq j}\left\{\lambda_i < q_i\left(1-\frac{\lambda_j}{1-q_i}\right)\right\}\right\} & q_1+q_2\geq 1, \end{cases}$$

which is consistent with the results in [Tsybakov&Mikhailov'1979] and [Rao&Ephremides'1988] despite different expressions.

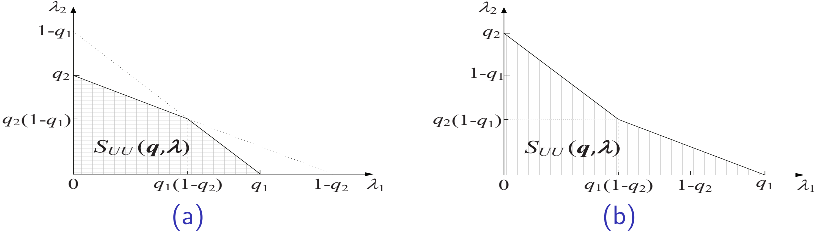


Figure: (a) $q_1+q_2<1$. (b) $q_1+q_2>1$.

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

- With $n > 2$, (11) is no longer a necessary condition. Yet given that (11) provides a sufficient and necessary condition for $n = 2$, we can develop $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$ in a recursive manner.
- For $n = 3$, for instance, we can first determine the all-unsaturated region of two nodes' transmission probabilities q_j and q_k as functions of q_i, $j \neq k \neq i$, given that Node i's queue is busy, and then obtain a sufficient and necessary condition of q_i for (10).

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

- $n = 3$

Theorem

The all-unsaturated region of $\boldsymbol{q}$ for $n = 3$ is

$$S_{UUU}(\boldsymbol{q}, \boldsymbol{\lambda}) = \Bigg\{ \boldsymbol{q} : \bigcup_{i \in \{1,2,3\}} \Bigg\{ \frac{\lambda_i}{p_{i,L}} < q_i < \frac{\lambda_i}{p_{i,S}}, \bigcup_{\substack{j,k \in \{1,2,3\} \setminus \{i\} \\ j \neq k}} \Bigg\{ \frac{\lambda_j}{p_{j,L}} < q_j < \frac{\lambda_j}{p_{jk,S}}, \ \frac{\lambda_k}{p_{k,L}} < q_k \leq 1 \Bigg\} \Bigg\} \Bigg\}.$$

where

$$p_{jk,S} = \frac{1}{2} \left(1 - q_i + \lambda_j - \lambda_k - \sqrt{(1 - q_i + \lambda_j - \lambda_k)^2 - 4\lambda_j (1 - q_i)} \right).$$

- $n = 3$

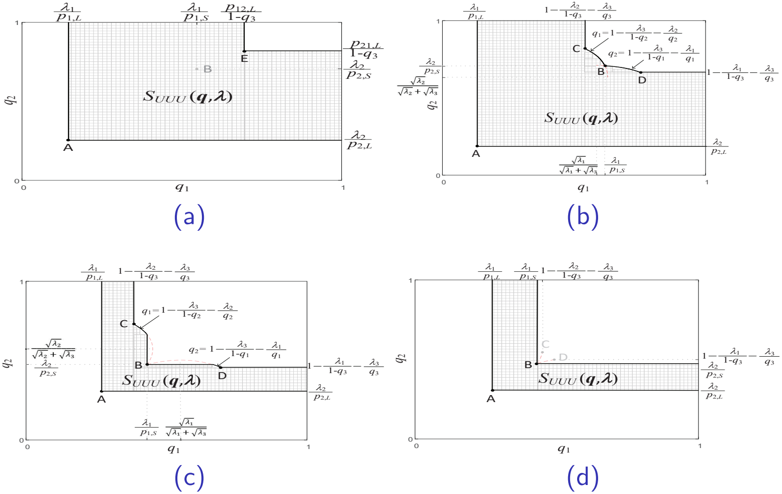


Figure: (a) $\frac{\lambda_3}{p_{3,L}} < q_3 < \frac{\lambda_3}{p_{3,S}}$. (b)(c)(d) $\frac{\lambda_3}{p_{3,S}} \leq q_3 \leq 1$.

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

- $n = 3$

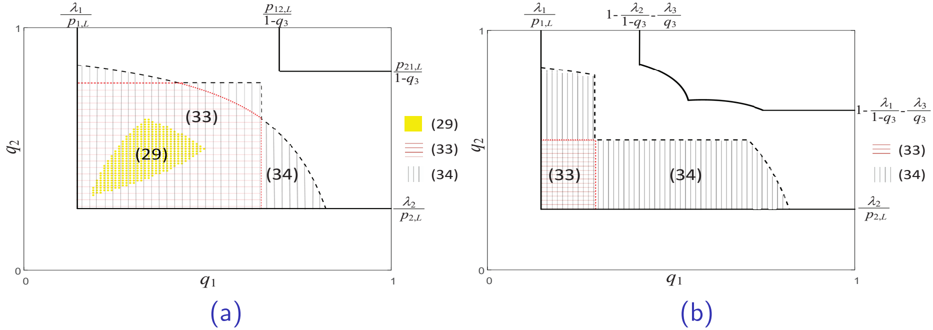


Figure: Illustration of sufficient but not necessary conditions for queue-stability (8), (9) and (11) for $n = 3$. (a) $\frac{\lambda_3}{p_{3,L}} < q_3 < \frac{\lambda_3}{p_{3,S}}$. (b) $\frac{\lambda_3}{p_{3,S}} \leq q_3 \leq 1$.

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

- By denoting $\bar{i} = \mathcal{N} \setminus \{i\}$, $\boldsymbol{q}_{\bar{i}} = \boldsymbol{q} \setminus \{q_i\}$ and $\boldsymbol{\lambda}_{\bar{i}} = \boldsymbol{\lambda} \setminus \{\lambda_i\}$, the all-unsaturated region $S_{UUU}(\boldsymbol{q}, \boldsymbol{\lambda})$ can be written as

$$S_{UUU}(\boldsymbol{q}, \boldsymbol{\lambda}) = \left\{ \boldsymbol{q} : \bigcup_{i \in \{1,2,3\}} \left\{ \frac{\lambda_i}{p_{i,L}} < q_i < \frac{\lambda_i}{p_{i,S}}, S^{\bar{i}}_{UU}(\boldsymbol{q}_{\bar{i}}, \boldsymbol{\lambda}_{\bar{i}}) \right\} \right\},$$

  where
  $S^{\bar{i}}_{UU}(\boldsymbol{q}_{\bar{i}}, \boldsymbol{\lambda}_{\bar{i}}) = \left\{ \boldsymbol{q}_{\bar{i}} : \bigcup_{j,k \in \bar{i}, j \neq k} \left\{ \frac{\lambda_j}{p_{j,L}} < q_j < \frac{\lambda_j}{p_{jk,S}}, \ \frac{\lambda_k}{p_{k,L}} < q_k \leq 1 \right\} \right\}$
  denotes the all-unsaturated region of $\boldsymbol{q}_{\bar{i}}$ for given transmission probability q_i of Node i.

- In general, for $n \geq 3$, the all-unsaturated region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$ can be characterized based on $S^{\bar{i}}_{U^{n-1}}(\boldsymbol{q}_{\bar{i}}, \boldsymbol{\lambda}_{\bar{i}})$ as

$$S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda}) = \left\{ \boldsymbol{q} : \bigcup_{i \in \mathcal{N}} \left\{ \frac{\lambda_i}{p_{i,L}} < q_i < \frac{\lambda_i}{p_{i,S}}, \ S^{\bar{i}}_{U^{n-1}}(\boldsymbol{q}_{\bar{i}}, \boldsymbol{\lambda}_{\bar{i}}) \right\} \right\}.$$

All-Unsaturated Region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$

- Symmetric case: $\lambda_i = \lambda$, $q_i = q$, $i \in \mathcal{N}$.

  In the symmetric case, the nodes are either all unsaturated at p_L or all saturated at p_A. The convergence to p_L requires that for any time t, $p_t > p_S$. As $\min_t p_t = p_A = (1-q)^{n-1}$ and $p_S = \left(1 - \frac{\lambda}{p_S}\right)^{n-1}$, we have $q < \frac{\lambda}{p_S}$. Combining with $\lambda < p_L \cdot q$, the all-unsaturated region $S_{U^n}(q, \lambda)$ can be written as

$$S_{U^n}(q, \lambda) = \left\{ q : \frac{\lambda}{p_L} < q < \frac{\lambda}{p_S} \right\},$$

  which can also be written as

$$S_{U^n}(q, \lambda) = \left\{ \lambda : \lambda < q(1-q)^{n-1} \right\}.$$

Complete Operating Regions of $\boldsymbol{q}$

- Let θ_i denote whether Node i's queue is saturated or unsaturated, i.e., $\theta_i \in \{U, S\}$.
- For any given network state $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]$, we can define the corresponding region $S_{\boldsymbol{\theta}}(\boldsymbol{q}, \boldsymbol{\lambda})$ of $\boldsymbol{q}$, within which $\lambda_i < \mu_i$ if $\theta_i = U$, and $\lambda_i \geq \mu_i$ if $\theta_i = S$, $i \in \mathcal{N}$.
- For an n-node network, there are totally 2^n network states. The corresponding 2^n regions of $\boldsymbol{q}$ provide a complete picture of how the network performance varies with different transmission probabilities of nodes.
- Note that in the symmetric case, as all the nodes have the same queue status, there are only two network states, all-unsaturated or all-saturated.

- $n = 2$: Four network states UU, US, SU and SS.

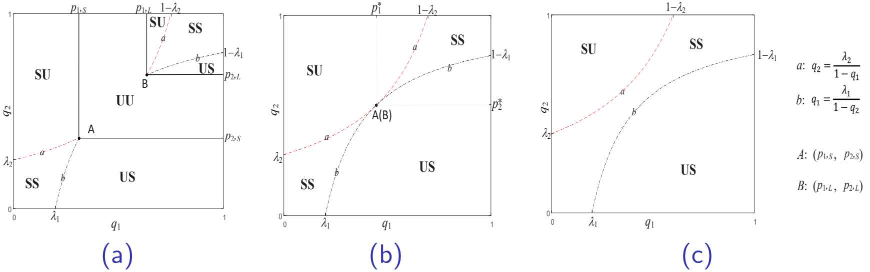
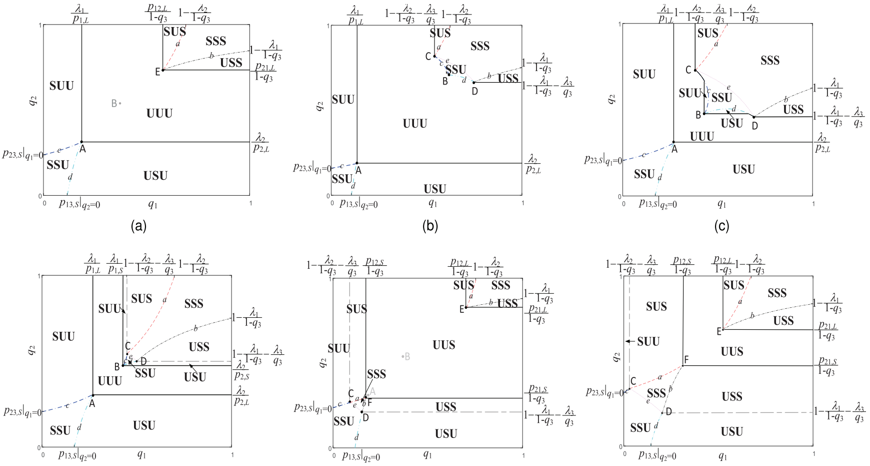


Figure: Illustration of the complete regions of $\boldsymbol{q}$ for $n = 2$. (a) $\boldsymbol{\lambda} \in S_Q^{n=2}(\boldsymbol{\lambda})$. (b) $\boldsymbol{\lambda} \in S_Q^{*,n=2}(\boldsymbol{\lambda})$. (c) $\boldsymbol{\lambda} \notin S_T^{n=2} = S_Q^{n=2}(\boldsymbol{\lambda}) \bigcup S_Q^{*,n=2}(\boldsymbol{\lambda})$.

$S_Q^*(\boldsymbol{\lambda})$ is the boundary of $S_Q(\boldsymbol{\lambda})$, i.e., $S_Q^*(\boldsymbol{\lambda}) = \bigcup_{\boldsymbol{q}} \{\boldsymbol{\lambda} : \boldsymbol{\lambda} = \boldsymbol{\mu}\}$.

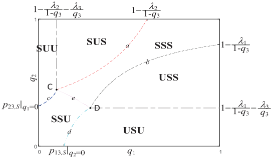
Complete Operating Regions of $\boldsymbol{q}$

- $n = 3$: Eight network states *UUU*, *UUS*, *USU*, *SUU*, *USS*, *SUS*, *SSU* and *SSS*.



(a) (b) (c) (d) (e) (f)

- $n = 3$: Eight network states *UUU*, *UUS*, *USU*, *SUU*, *USS*, *SUS*, *SSU* and *SSS*.



(g)

a: $q_2 = \frac{\lambda_2}{(1-q_1)(1-q_3)}$

b: $q_1 = \frac{\lambda_1}{(1-q_2)(1-q_3)}$

c: $q_1 = 1 - \frac{\lambda_3}{1-q_2} - \frac{\lambda_2}{q_2}$

d: $q_2 = 1 - \frac{\lambda_3}{1-q_1} - \frac{\lambda_1}{q_1}$

e: $q_3 = \frac{\lambda_3}{(1-q_1)(1-q_2)}$

A: $\left(\frac{\lambda_1}{p_{1,L}}, \frac{\lambda_2}{p_{2,L}}\right)$

B: $\left(\frac{\lambda_1}{p_{1,S}}, \frac{\lambda_2}{p_{2,S}}\right)$

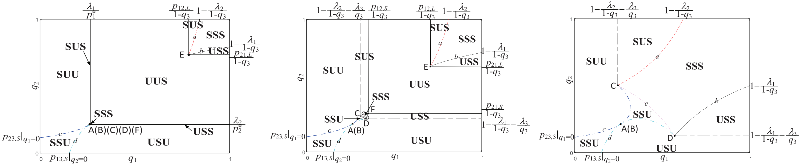
C: $\left(1 - \frac{\lambda_2}{1-q_3} - \frac{\lambda_3}{q_3}, \frac{\lambda_2}{\lambda_2 + \frac{\lambda_3(1-q_3)}{q_3}}\right)$

D: $\left(\frac{\lambda_1}{\lambda_1 + \frac{\lambda_3(1-q_3)}{q_3}}, 1 - \frac{\lambda_1}{1-q_3} - \frac{\lambda_3}{q_3}\right)$

E: $\left(\frac{p_{12,L}}{1-q_3}, \frac{p_{21,L}}{1-q_3}\right)$

F: $\left(\frac{p_{12,S}}{1-q_3}, \frac{p_{21,S}}{1-q_3}\right)$

Fig. 10. Illustration of the complete regions of $\boldsymbol{q}$ for $n = 3$ with $\boldsymbol{\lambda} \in S_Q^{n=3}(\boldsymbol{\lambda})$ or $\boldsymbol{\lambda} \notin S_T^{n=3}(\boldsymbol{\lambda})$. (a) $\boldsymbol{\lambda} \in S_Q^{n=3}(\boldsymbol{\lambda})$ and $\frac{\lambda_3}{p_{3,L}} < q_3 < \frac{\lambda_3}{p_{3,S}}$. (b)(c)(d) $\boldsymbol{\lambda} \in S_Q^{n=3}(\boldsymbol{\lambda})$ and $\frac{\lambda_3}{p_{3,S}} \leq q_3 \leq 1$. (e) $\boldsymbol{\lambda} \in S_Q^{n=3}(\boldsymbol{\lambda})$ and $q_3 \leq \frac{\lambda_3}{p_{3,L}}$. (f) $\boldsymbol{\lambda} \notin S_T^{n=3}(\boldsymbol{\lambda})$ with $q_3 \leq 1 - (\sqrt{\lambda_1} + \sqrt{\lambda_2})^2$. (g) $\boldsymbol{\lambda} \notin S_T^{n=3}(\boldsymbol{\lambda})$ with $q_3 > 1 - (\sqrt{\lambda_1} + \sqrt{\lambda_2})^2$.



(a) (b) (c)

Fig. 11. Illustration of the complete regions of $\boldsymbol{q}$ for $n = 3$ with $\boldsymbol{\lambda} \in S_Q^{*,n=3}(\boldsymbol{\lambda})$. (a) $q_3 = q_3^*$. (b) $q_3^* < q_3 \leq 1 - (\sqrt{\lambda_1} + \sqrt{\lambda_2})^2$. (c) $q_3 > 1 - (\sqrt{\lambda_1} + \sqrt{\lambda_2})^2$.

- Symmetric case: As the network has only two states, all-unsaturated and all-saturated, the complete operating regions of transmission probability q only consist of the all-unsaturated region $S_{U^n}(q,\lambda)$ and the all-saturated region $S_{S^n}(q,\lambda) = \overline{S_{U^n}(q,\lambda)}$.

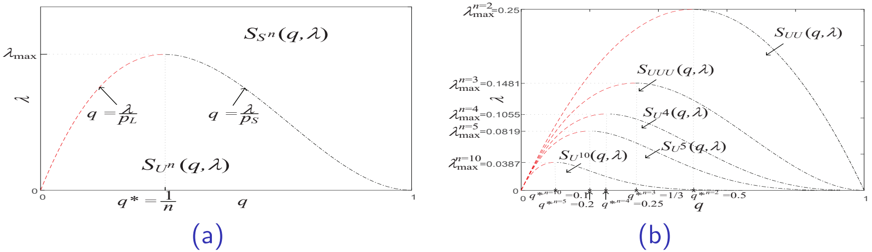


Figure: Illustration of (a) complete operating regions of q and (b) all-unsaturated region of q under various values of n.

Complete Operating Regions of $\boldsymbol{q}$

- Different from the queue-stability region $S_Q(\boldsymbol{\lambda})$ that is independent of the backoff function $\mathcal{Q}(k)$, the complete operating regions of $\boldsymbol{q}$ are crucially determined by the backoff function.
- Example: All-unsaturated region of q in the symmetric case with Exponential Backoff $\mathcal{Q}(k) = r^k$ for $r \leq 1$.

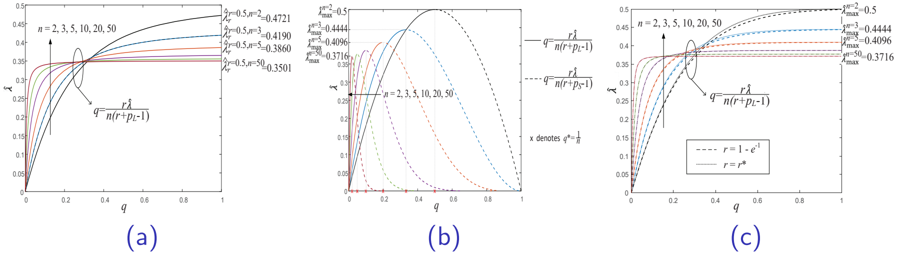


Figure: All-unsaturated region of q with Exponential Backoff $S_{U^n}^{EB}(q,\lambda)$. (a) $r = 0.5$. (b) $r = 1$. (c) $r = r^*$ or $1 - e^{-1}$.

Remark: When does the Queue-Independence Assumption introduce Errors?

- Queue-independence assumption:
  $\beta_{ij} = \frac{\Pr\{\text{Node } j\text{'s queue is empty} \mid \text{Node } i\text{'s queue is busy}\}}{\Pr\{\text{Node } j\text{'s queue is empty}\}} = 1$ for all $i \neq j$, $i, j \in \mathcal{N}$.

- $\beta_{ij} = 1$ when events of Node j's queue being empty and Node i's queue being busy are independent, which holds true when either Node i or Node j is saturated, or both are saturated.

- The queue-independence assumption holds true when there is no more than one node unsaturated, i.e., at least $n - 1$ out of n nodes are saturated.

Remark: When does the Queue-Independence Assumption introduce Errors?

- When the input rates $\boldsymbol{\lambda}$ fall on the boundary of queue-stability region, i.e., $\boldsymbol{\lambda} \in S_Q^*(\boldsymbol{\lambda}) = \bigcup_{\boldsymbol{q}} \{\boldsymbol{\lambda} : \boldsymbol{\lambda} = \boldsymbol{\mu}\}$:
  - The two positive real roots of (5) become equal and reduced to a single root $c_L = c_S = c^*$, leading to $\boldsymbol{p}_L = \boldsymbol{p}_S = \boldsymbol{p}^* = (\lambda_1 + c^*, \lambda_2 + c^*, \ldots, \lambda_n + c^*)$.
  - The all-unsaturated region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$ becomes an empty set, and its boundary reduces to

$$S_{U^n}^*(\boldsymbol{q}) = \left\{ \boldsymbol{q} : \bigcup_{i \in \mathcal{N}} \left\{ q_i^* \leq q_i \leq 1, \bigcap_{j \in \mathcal{N} \setminus \{i\}} \left\{ q_j = q_j^* \right\} \right\} \right\}, \qquad (12)$$

    where $\boldsymbol{q}^* = \left(\frac{\lambda_1}{p_1^*}, \frac{\lambda_2}{p_2^*}, \ldots, \frac{\lambda_n}{p_n^*} \right)$.
  - With $\boldsymbol{q} \in S_{U^n}^*(\boldsymbol{q})$, at least $n-1$ nodes are saturated.

Remark: When does the Queue-Independence Assumption introduce Errors?

- **The queue-stability region $S_Q(\boldsymbol{\lambda})$ derived based on the queue-independence assumption is accurate.**
  - As the input rates $\boldsymbol{\lambda}$ increase and approach the boundary $S_Q^*(\boldsymbol{\lambda})$, the all-unsaturated region $S_{U^n}(\boldsymbol{q}, \boldsymbol{\lambda})$ shrinks and approaches the boundary $S_{U^n}^*(\boldsymbol{q})$, on which at least $n-1$ nodes are saturated.
  - The queue-independence assumption holds true when $\boldsymbol{\lambda}$ falls on the boundary $S_Q^*(\boldsymbol{\lambda})$.
- For the operating regions of $\boldsymbol{q}$, errors could be introduced by the queue-independence assumption for boundaries shared by two network states both of which have more than one node unsaturated.
  - For network states with more than one node unsaturated, the network steady-state point $\boldsymbol{p}$ derived based on the queue-independence assumption is an approximation,

Remark: When does the Queue-Independence Assumption introduce Errors?

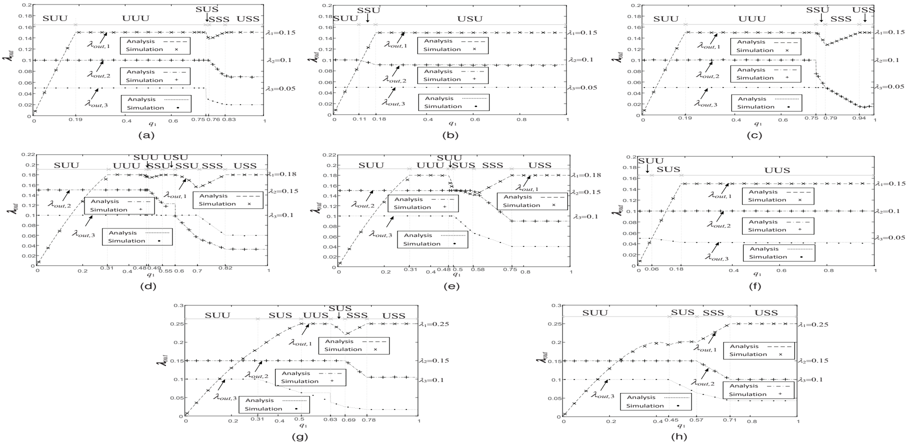


Fig. 15. Node throughput $\boldsymbol{\lambda_{out}}$ versus transmission probability q_1 for $n = 3$. (a)-(c) $\boldsymbol{\lambda} = (0.15, 0.1, 0.05)$. (a) $q_3 = 0.4$ and $q_2 = 0.7$. (b) $q_3 = 0.4$ and $q_2 = 0.12$. (c) $q_3 = 0.6$ and $q_2 = 0.6$. (d)-(e) $\boldsymbol{\lambda} = (0.18, 0.15, 0.1)$. (d) $q_3 = 0.6$ and $q_2 = 0.45$. (e) $q_3 = 0.4$ and $q_2 = 0.6$. (f) $\boldsymbol{\lambda} = (0.15, 0.1, 0.05)$. $q_3 = 0.06$ and $q_2 = 0.5$. (g)-(h) $\boldsymbol{\lambda} = (0.25, 0.15, 0.1)$. (g) $q_3 = 0.2$ and $q_2 = 0.6$. (h) $q_3 = 0.3$ and $q_2 = 0.5$.

Remark: When does the Queue-Independence Assumption introduce Errors?

- With a higher input rate or larger transmission probability, a node has more chances to transmit, and thus has stronger coupling with others.
- As the number of nodes increases, the transmission probability of each node should be reduced accordingly to stay in the unsaturated condition, leading to less dependency of nodes' queues.

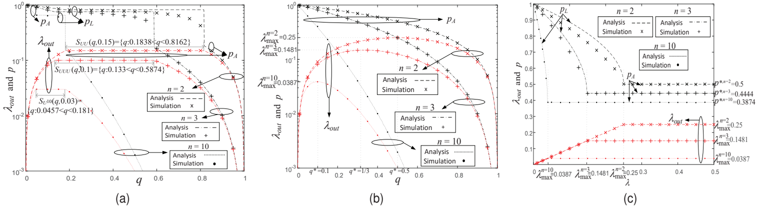


Fig. 19. Network steady-state point p and node throughput λ_{out} versus (a)(b) transmission probability q and (c) input rate λ. (a) $\lambda = \frac{0.3}{n}$ and (b) $\lambda = \lambda_{\max}$. (c) $q = q^* = \frac{1}{n}$.

- Intuitively, with more nodes in the network, the effect of each individual node on others is diminished.

Remark: About the Dominant Approach

- For $n = 2$, recall that the all-unsaturated region $S_{UU}(\boldsymbol{q}, \boldsymbol{\lambda})$ was also derived in [Rao&Ephremides'1988] by introducing a hypothetical dominant system where one of the queues sends "dummy" packets when the queue is empty.
- Essentially it can be regarded as the original network in states SU and US, where one of the queues is saturated. As $S_{UU}(\boldsymbol{q}, \boldsymbol{\lambda})$ shares the boundary with $S_{SU}(\boldsymbol{q}, \boldsymbol{\lambda})$ and $S_{US}(\boldsymbol{q}, \boldsymbol{\lambda})$, it can be obtained based on the shared boundary.
- For the network state SU, the service rates of Node 1 and Node 2 are given by $\mu_1 = q_1\left(1 - \frac{\lambda_2}{1-q_1}\right)$ and $\mu_2 = q_2(1 - q_1)$, respectively. The boundary shared with $S_{SU}(\boldsymbol{q}, \boldsymbol{\lambda})$ can then be written as $\lambda_1 < q_1\left(1 - \frac{\lambda_2}{1-q_1}\right)$ and $\lambda_2 < q_2(1 - q_1)$. Similarly, the boundary shared with $S_{US}(\boldsymbol{q}, \boldsymbol{\lambda})$ can be written as $\lambda_2 < q_2\left(1 - \frac{\lambda_1}{1-q_2}\right)$ and $\lambda_1 < q_1(1 - q_2)$. $S_{UU}(\boldsymbol{q}, \boldsymbol{\lambda})$ can then be obtained by combining both, which, despite a different form, is indeed consistent with Theorem 1.

Remark: About the Dominant Approach

- The success largely relies on how the dominant system is set up, i.e., what network states to look at, which was found to be extremely difficult for $n>2$.
- For $n=3$, $S_{UUU}(\boldsymbol{q},\boldsymbol{\lambda})$ was constructed in [Szpankowski'1994] based on a dominant system where only **one node** sends dummy packets, which correspond to states UUS, USU and SUU.
- However, the boundary of $S_{UUU}(\boldsymbol{q},\boldsymbol{\lambda})$ is not shared with that of $S_{UUS}(\boldsymbol{q},\boldsymbol{\lambda})$, $S_{USU}(\boldsymbol{q},\boldsymbol{\lambda})$, and $S_{SUU}(\boldsymbol{q},\boldsymbol{\lambda})$. Instead, $S_{UUU}(\boldsymbol{q},\boldsymbol{\lambda})$ shares the boundary with different regions in different scenarios.
- In addition to $S_{SUU}(\boldsymbol{q},\boldsymbol{\lambda})$ and $S_{USU}(\boldsymbol{q},\boldsymbol{\lambda})$, $S_{UUU}(\boldsymbol{q},\boldsymbol{\lambda})$ may also share the boundary with $S_{SSU}(\boldsymbol{q},\boldsymbol{\lambda})$, $S_{SUS}(\boldsymbol{q},\boldsymbol{\lambda})$ and $S_{USS}(\boldsymbol{q},\boldsymbol{\lambda})$, where two out of three nodes are saturated.

Extension: from Aloha to CSMA

- With sensing:
  - Each node senses the channel before it transmits.
  - Due to the propagation delay, a node cannot sense a transmission immediately after it occurs. Therefore, it should continuously sense for a while to ensure that it senses the channel states properly.
- For slotted CSMA, the length of a mini-slot is determined by the sensing time, which is the maximum propagation delay.
  - All the nodes are synchronized and can start a transmission only at the beginning of a mini-slot.
  - With the sensing time of one mini-slot, if a node senses the channel idle at mini-slot t, it would transmit with a certain probability at mini-slot $t+1$.

Extension: from Aloha to CSMA

- Let a denote the sensing time normalized by the data packet transmission time, which is also the mini-slot length.
- Each data packet transmission lasts for one slot, or equivalently, $\frac{1}{a}$ mini-slots.
- Each failed transmission lasts for $0 \leq x \leq \frac{1}{a}$ mini-slots, or $0 \leq xa \leq 1$ slot.
  - Connection-free (example: the basic access mechanism of DCF of WiFi): $xa \approx 1$
  - Connection-based (example: the RTS/CTS access mechanism of DCF of WiFi): $xa \ll 1$

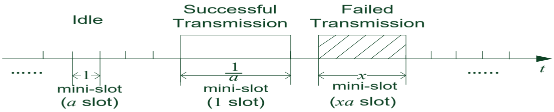


Figure: Illustration of the aggregate channel of CSMA.

Extension: from Aloha to CSMA

- In the literature, the normalized sensing time a and the failed transmission time x are often assumed to be zero.
- The normalized sensing time a is determined by both the sensing time and the data packet transmission time.
  - The sensing time, i.e., the maximum propagation delay, is always non-zero and could be large in some applications.
  - With the rapid increasing data transmission rate and rising popularity of short-packet transmissions, the data packet transmission time has been considerably reduced.
- The failed transmission time x could be large when nodes operate in the half-duplex mode and with connection-free access. Even with connection-based access, it takes time to request and establish a connection with the receiver.

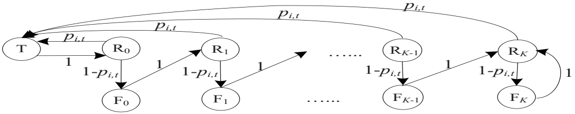
Extension: from Aloha to CSMA

- In the literature, Aloha is often regarded as a special case of CSMA when $a = x = 1$.
- With $a = x = 1$, Aloha outperforms CSMA as sensing has its cost: It takes time for a node to hear others' transmissions.
- When is sensing beneficial?
- What are the effects of the normalized sensing time a and the failed transmission time x on the stability performance?

Extension: from Aloha to CSMA

The proposed unified analytical framework of random access:

- Markov renewal process of each node's HOL packets with CSMA: $(\mathbf{X}_i, \mathbf{V}_i) = \{(X_{i,j}, V_{i,j}), j = 0, 1, \ldots\}, i \in \{1, 2, \ldots, n\}$



- Steady-state probability of successful transmission of each node's HOL packets based on

  1) State-independence assumption: $\lim_{t\to\infty} p_{i,t} = p_i,\ i \in \mathcal{N}$

  2) Queue-independence assumption:

  $\beta_{ij} = \dfrac{\Pr\{\text{Node } j\text{'s queue is empty} \mid \text{Node } i\text{'s queue is busy}\}}{\Pr\{\text{Node } j\text{'s queue is empty}\}} = 1$ for all $i \neq j$, $i, j \in \mathcal{N}$

L. Dai, "A theoretical framework for random access: Effects of carrier sensing on stability," *IEEE Trans. Commun.*, vol. 73, no. 1, pp. 275–291, Jan. 2025.

Fixed-point Equations of $\boldsymbol{p} = [p_1, p_2, \ldots, p_n]$

With CSMA and the collision receiver, (2) can be written as:

$$p_i = \prod_{j \in \mathcal{N}^{\mathcal{U}} \setminus \{i\}} \left(1 - \frac{a\lambda_j}{\alpha p_j}\right) \prod_{j \in \mathcal{N}^{\mathcal{S}} \setminus \{i\}} \left(1 - \frac{a\mu_j}{\alpha p_j}\right), \tag{13}$$

$i \in \mathcal{N}$, where α denotes the steady-state probability of the channel being idle:

$$\alpha = \frac{1}{x + 1 + \left(\frac{1}{a} - x\right)\left(\sum_{i=1}^{n} p_i - n\left(\prod_{i=1}^{n} p_i\right)^{\frac{1}{n-1}}\right) - x\left(\prod_{i=1}^{n} p_i\right)^{\frac{1}{n-1}}}.$$

Again let us convert n fixed-point equations into a single one by introducing

$$c = \prod_{j \in \mathcal{N}^{\mathcal{U}}} \left(1 - \frac{a\lambda_j}{\alpha p_j}\right) \cdot \prod_{j \in \mathcal{N}^{\mathcal{S}}} \left(1 - \frac{a\mu_j}{\alpha p_j}\right).$$

In the All-Unsaturated Condition

To achieve queue-stability, all the nodes' queues should be unsaturated, i.e., $\mathcal{N}^{\mathcal{U}} = \mathcal{N}$. (13) can be written as

$$p_i = \prod_{j \in \mathcal{N} \setminus \{i\}} \left(1 - \frac{a\lambda_j}{\alpha p_j}\right), \tag{14}$$

$i \in \mathcal{N}$, which can also be transformed into a single fixed-point equation

$$\tilde{c} = \prod_{i=1}^{n} \frac{\tilde{c}}{\tilde{c} + \tilde{\lambda}_i}, \tag{15}$$

where $\tilde{c} = c \cdot \prod_{i=1}^{n}(1 - z\lambda_i)$, $\tilde{\lambda}_i = \frac{x+1}{x} z\lambda_i \cdot \prod_{j \in \mathcal{N} \setminus \{i\}}(1 - z\lambda_j)$, and $z = \frac{xa}{1-(1-xa)\sum_{j=1}^{n} \lambda_j}$.

Queue-Stability Region $S_Q(\boldsymbol{\lambda})$

- (15) has the same expression as (5).
- The queue-stability region with CSMA is determined by the existence of positive real roots of the fixed-point equation (15).
- The queue-stability region with CSMA can be written as

$$S_Q^{CSMA}(\boldsymbol{\lambda}) = S_Q^{Aloha}(\tilde{\boldsymbol{\lambda}}).$$

- Example: Based on the queue-stability region of 2-node Aloha $S_Q^{n=2}(\boldsymbol{\lambda}) = \{\boldsymbol{\lambda} : \sqrt{\lambda_1} + \sqrt{\lambda_2} < 1\}$, we immediately have

$$S_Q^{n=2}(\boldsymbol{\lambda}) = \Bigg\{\boldsymbol{\lambda} : \sqrt{\lambda_1(1-(1-xa)\lambda_1-\lambda_2)}+ \sqrt{\lambda_2(1-(1-xa)\lambda_2-\lambda_1)} \le \frac{1-(1-xa)(\lambda_1+\lambda_2)}{\sqrt{a(1+x)}}\Bigg\}.$$

Queue-Stability Region $S_Q(\boldsymbol{\lambda})$ of 2-Node CSMA

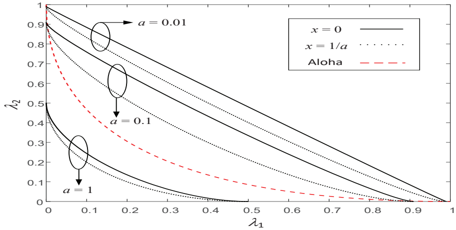


Figure: Illustration of queue-stability region $S_Q^{n=2}(\boldsymbol{\lambda})$ of 2-node CSMA.

- With CSMA, the queue-stability region can be improved by reducing the normalized sensing time a or the failed transmission time x.
- As $a \to 0$, the queue-stability region of CSMA approaches that of TDMA, i.e., $S_T^{n=2}(\boldsymbol{\lambda}) \to \{\boldsymbol{\lambda} : \lambda_1 + \lambda_2 < 1\}$.
- With a large a, the queue-stability region with CSMA could be smaller than that with Aloha.

Complete Operating Regions of $\boldsymbol{q}$ of 2-Node CSMA with Constant Backoff

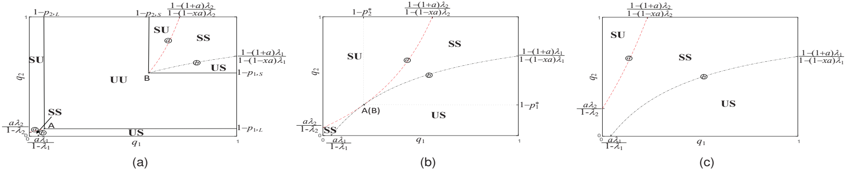


Fig. 5. Illustration of the complete operating regions of $\boldsymbol{q}$ for 2-node CSMA. (a) $\boldsymbol{\lambda} \in S_Q^{n=2}(\boldsymbol{\lambda})$. (b) $\boldsymbol{\lambda} \in S_Q^{*,n=2}(\boldsymbol{\lambda})$. (c) $\boldsymbol{\lambda} \notin S_T^{n=2}(\boldsymbol{\lambda})$. ⓐ: $q_2 = \frac{\lambda_2(a+q_1)}{1-q_1+\lambda_2((2-xa)q_1-1)}$. ⓑ: $q_1 = \frac{\lambda_1(a+q_2)}{1-q_2+\lambda_1((2-xa)q_2-1)}$. A: $(1-p_{2,L}, 1-p_{1,L})$. B: $(1-p_{2,S}, 1-p_{1,S})$.

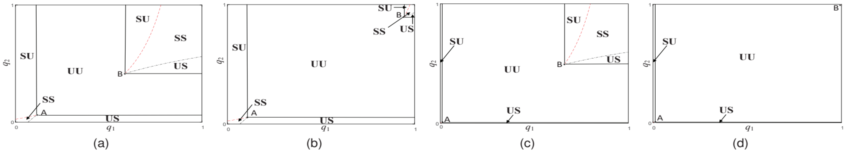


Fig. 6. Complete operating regions of $\boldsymbol{q}$ for 2-node CSMA with $\boldsymbol{\lambda} = (0.4, 0.2)$. (a) $a = 0.1$, $x = \frac{1}{a}$. (b) $a = 0.1$, $x = 0$. (c) $a = 0.01$, $x = \frac{1}{a}$. (d) $a = 0.01$, $x = 0$.

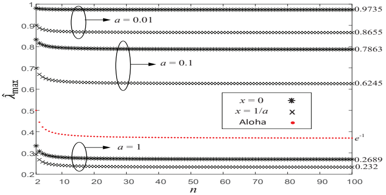
- The all-unsaturated region $S_{UU}(\boldsymbol{q}, \boldsymbol{\lambda})$ can be significantly enlarged by reducing the normalized sensing time a or failed transmission time x.

- Queue-stability region:
  - With Aloha:

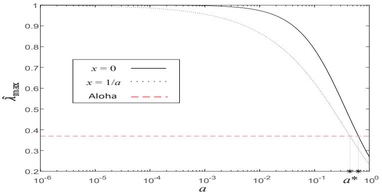
$$S_Q^{sym}(\lambda) = \left\{ \lambda : \lambda < \frac{(n-1)^{n-1}}{n^n} \right\}.$$

  - With CSMA:

$$S_Q^{sym}(\lambda) = \left\{ \lambda : \frac{\lambda}{1-(1-xa)n\lambda} \left(1 - \frac{xa\lambda}{1-(1-xa)n\lambda}\right)^{n-1} < \frac{(n-1)^{n-1}}{a(1+x)n^n} \right\}.$$



(a)



(b)

Figure: Maximum aggregate throughput $\hat{\lambda}_{\max}$ for symmetric n-node CSMA. (a) $\hat{\lambda}_{\max}$ versus number of nodes n. $a \in \{0.01, 0.1, 1\}$. (b) $\hat{\lambda}_{\max}$ versus normalized sensing time a. $n = 100$.

Symmetric n-Node CSMA

- As the network has only two states, all-unsaturated and all-saturated, the complete operating regions of transmission probability q only consist of the all-unsaturated region $S_{U^n}(q,\lambda)$ and the all-saturated region $S_{S^n}(q,\lambda) = \overline{S_{U^n}(q,\lambda)}$.
  - With Aloha:

$$S_{U^n}(q,\lambda) = \left\{q : \frac{\lambda}{g(p_L)} < q < \frac{\lambda}{g(p_S)}\right\} \bigcap \{q : 0 < q \leq 1\}.$$

  - With CSMA:

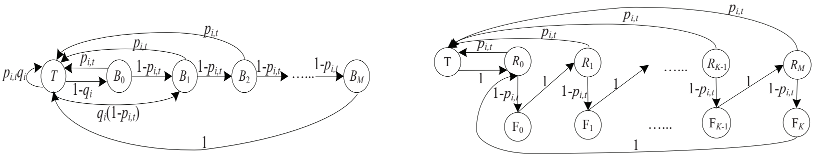
$$S_{U^n}(q,\lambda) = \left\{q : \frac{a\lambda}{\alpha(p_L)g(p_L)} < q < \frac{a\lambda}{\alpha(p_S)g(p_S)}\right\} \bigcap \{q : 0 < q \leq 1\}.$$

  where $g(p) = \frac{1}{\sum_{k=0}^{K-1} \frac{(1-p)^k}{\mathcal{Q}(k)} + \frac{(1-p)^K}{p\mathcal{Q}(K)}}$.

- The complete operating regions of transmission probability q depend on the backoff function $\mathcal{Q}(k)$.

Extension to Finite Retry Limit

- In practice, packets are often dropped after a few failed attempts.
- With a finite retry limit M, a HOL packet is discarded if the number of transmission failures reaches M.
- To incorporate a finite retry limit, the state transition process of HOL packets need to be revised. For a State-B_{M-1} (R_{M-1}) HOL packet in the Aloha (CSMA) case, instead of shifting to State B_M (R_M) when its transmission fails, with probability 1 it returns to the initial state T.



(a) (b)

Figure: (a) Aloha (b) CSMA

Extension to A General Network Topology

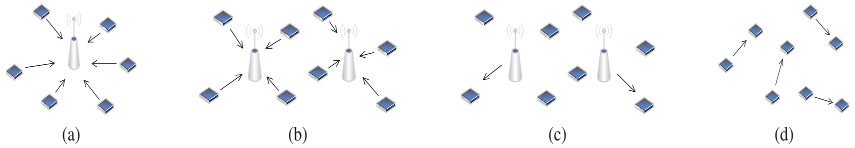


Figure: (a) Uplink of a single cell network (MAC); (b) Uplink of a multi-cell network; (c) Downlink of a multi-cell network; (d) An ad-hoc network.

- In practice, there could be multiple transmitter-receiver pairs co-existing.
- Assume K transmitters and L receivers. For uplink of single-cell (MAC), $K > L = 1$. For uplink of multi-cell, $K > L > 1$. For downlink of multi-cell or ad-hoc, $K = L > 1$.

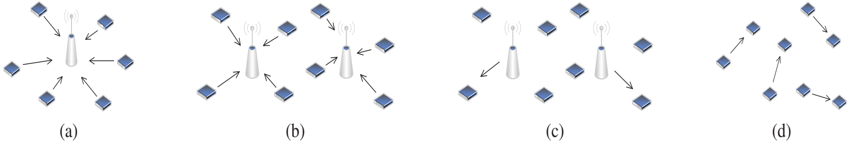


Figure: (a) Uplink of a single cell network (MAC); (b) Uplink of a multi-cell network; (c) Downlink of a multi-cell network; (d) An ad-hoc network.

- Let $\boldsymbol{p} = [p_1, p_2, \ldots, p_K]$ denote the steady-state probabilities of success transmission of HOL packets of K transmitters. K fixed-point equations of $\boldsymbol{p}$ can be written as

$$p_i = \sum_{\mathcal{S}_i \subseteq \mathcal{K} \backslash \{i\}} r_{i|\mathcal{S}_i} \cdot \prod_{j \in \mathcal{S}_i} \omega_j \cdot \prod_{j \in \mathcal{K} \backslash (\mathcal{S}_i \bigcup \{i\})} (1 - \omega_j), \quad i \in \mathcal{K}$$

where $\mathcal{S}_i$ denotes the set of transmitters that have concurrent transmissions with Transmitter i, $i \in \mathcal{K}$.

Extension to Other Receiver Models

Fixed-point equations of $\boldsymbol{p}$:

$$p_i = \sum_{\mathcal{S}_i \subseteq \mathcal{K} \setminus \{i\}} r_{i|\mathcal{S}_i} \cdot \prod_{j \in \mathcal{S}_i} \omega_j \cdot \prod_{j \in \mathcal{K} \setminus (\mathcal{S}_i \bigcup \{i\})} (1 - \omega_j), \quad i \in \mathcal{K}$$

- $I_i \in \{0, 1\}$: Boolean indicator to indicate whether Node i transmits
- ω_i: Transmission probability of Node i
  - With sensing-free Aloha: $\omega_i = \min\left(\frac{\lambda_i}{p_i}, \frac{\mu_i}{p_i}\right)$
  - With sensing-based CSMA: $\omega_i = \min\left(\frac{a\lambda_i}{\alpha p_i}, \frac{a\mu_i}{\alpha p_i}\right)$
- $r_{i|\mathcal{S}_i}$: Conditional success probability of Node i for given set $\mathcal{S}_i$ of transmitters with concurrent transmissions, which depends on **the channel and receiver model**.

Extension to Other Receiver Models

- $r_{i|\mathcal{S}_i}$: Conditional success probability of Node i for given set $\mathcal{S}_i$ of transmitters with concurrent transmissions, which depends on **the channel and receiver model**. For $k = |\mathcal{S}_i|$:
  - Collision receiver with ideal channels: $r_{i|0} = 1$ and $r_{i|k} = 0$ for $k > 0$.
  - Multi-Packet Reception (MPR): $r_{i|k} > 0$ for $k > 0$.

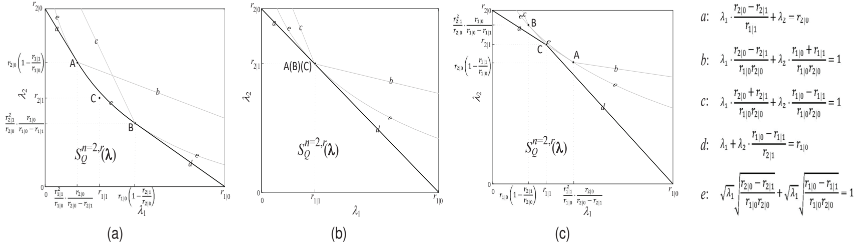


Fig. 24. Illustration of queue-stability region $S_Q^{n=2,r}(\boldsymbol{\lambda})$ with $r_{i|1} > 0$, $i \in \{1, 2\}$. (a) $0 < \frac{r_{1|1}}{r_{1|0}} + \frac{r_{2|1}}{r_{2|0}} < 1$. (b) $\frac{r_{1|1}}{r_{1|0}} + \frac{r_{2|1}}{r_{2|0}} = 1$. (c) $\frac{r_{1|1}}{r_{1|0}} + \frac{r_{2|1}}{r_{2|0}} > 1$.

Extension to Other Receiver Models

- $r_{i|\mathcal{S}_i}$: Conditional success probability of Node i for given set $\mathcal{S}_i$ of transmitters with concurrent transmissions, which depends on **the channel and receiver model**. For $k = |\mathcal{S}_i|$:
  - Collision receiver with ideal channels: $r_{i|0} = 1$ and $r_{i|k} = 0$ for $k > 0$.
  - Multi-Packet Reception (MPR): $r_{i|k} > 0$ for $k > 0$.
  - Capture receiver with symmetric AWGN channels (received SNR ρ and SINR threshold γ): $r_k{=}1$ for $k{\leq}\left\lfloor\frac{1}{\gamma}{-}\frac{1}{\rho}\right\rfloor$, and $r_k{=}0$ for $k{>}\left\lfloor\frac{1}{\gamma}{-}\frac{1}{\rho}\right\rfloor$.
  - Capture receiver with symmetric Rayleigh fading channels: $r_{i|k} = \frac{\exp\left\{-\frac{\gamma}{\rho}\right\}}{(\gamma+1)^k}$.
  - SIC receiver with symmetric Rayleigh fading channels [Li&Dai'2018]

Application to Cellular Networks and WiFi Networks

- Random access procedure from 1G to 4G: Connection-based Aloha
- Random Access-Based Small Data Transmission (RA-SDT) in 5G:
  - 4-step RA-SDT: Connection-based Aloha
  - 2-step RA-SDT: Connection-free Aloha
- DCF in WiFi:
  - Basic access mechanism: Connection-free CSMA
  - RTS/CTS mechanism: Connection-based CSMA
  - In DCF, σ, T_S and T_C are usually used to denote the mini-slot length, successful transmission time and failed transmission time, respectively, all in the unit of seconds (or mini-seconds). Using our notations: $a = \frac{\sigma}{T_S}$ and $x = \frac{T_C}{\sigma}$.
- In pratice, window-based backoff is adopted, that is, each node has a backoff window with size $W_{i,k} = W_i \cdot \tilde{\mathcal{Q}}(k)$, where W_i is the initial window size of Node i, $i \in \mathcal{N}$, and $\tilde{\mathcal{Q}}(k)$ is a monotonic non-decreasing function of the number of transmission failures k. To apply the analysis to window-based backoff, replace $q_i \cdot \mathcal{Q}(k)$ with $\frac{2}{1+W_i \cdot \tilde{\mathcal{Q}}(k)}$.

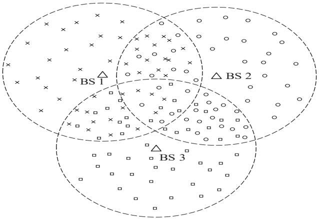
Access Design for Next-Generation Communication Networks

In the Future...

To support *more* devices with *higher* QoS requirements:

- More BSs/APs
- More “Intelligent” Access Design

More BSs/APs



- More transmitter-receiver pairs, more difficult to stabilize them.
- **Zero gain (and even worse performance)** if improperly designed.
- Inter-cell interference should be taken account of when optimizing the access design – **That requires information exchange among BSs/APs!**

Y. Yang and L. Dai, "Achieving stability of Aloha networks with multiple transmitter-receiver pairs," to appear in *IEEE Trans. Commun.*

Y. Yang and L. Dai, "Stability region and transmission control of multi-cell Aloha networks," *IEEE Trans. Commun.*, vol. 71, no. 9, pp. 5348-5364, Sep. 2023.

More "Intelligent" Access Design

- **Learning-based access design**: Each node independently determines when to access based on its own observations/measurements and past experience.
  - Supposed to outperform the traditional random access schemes by applying more efficient learning algorithms
  - **Abundant improvement in network throughput demonstrated**: Without sensing, the network throughput with the simple collision receiver can far exceed the well known limit of e^{-1}.
    - **Q-learning-based Framed Slotted Aloha** [Chu&etc'2015]
    - **Deep Q-Network (DQN)-based Multichannel Slotted Aloha** [Naparstek&Cohen'2019], [Zhong&etc'2019], [Sohaib,Jeong&Jeon'2022], [Jadoon&etc'2024], ...
    - **Multi-Armed Bandit (MAB)-based Slotted Aloha** [Peng&Dai'2024]: *Achieve the maximum throughput of* 1

N. Peng and L. Dai, "Multi-Armed-Bandit-based Framed Slotted Aloha for throughput optimization," *IEEE Commun. Lett.*, vol. 28, no. 42, pp. 847-851, Apr. 2024.

More "Intelligent" Access Design

- Most are direct applications of existing reinforcement learning algorithms.
  – Due to the lack of analytical framework, effects of key learning parameters such as the learning rate cannot be fully understood.
  – Design based on intuition and experience: More like an art than a science?

- *How can the proposed theory of random access guide the learning-based access design?*
  – Identify the learned access strategy.
  – Feed the learned access strategy into the analytical framework of random access for performance optimization.

L. Dai, "Harnessing the channel-capture phenomenon in slotted Aloha for achieving the optimal tradeoff between throughput and short-term fairness," to appear in *IEEE Trans. Commun.*

Acknowledgement

Collaborative work with my students:

Current PhD Students:

- Xinran Zhao, Yunshan Yang, Huaqiang Zhang, Xinlong Wang, Nian Peng, Liang Du, Mingyuan Ding

Former PhD Students:

- Xinghua Sun (Sun Yat-sen University)
- Yayu Gao (Huazhong University of Science and Technology)
- Zhiyang Liu (Nankai University)
- Junyuan Wang (Tongji University)
- Yitong Li (Zhengzhou University)
- Wen Zhan (Sun Yat-sen University)
- Yue Zhang (Shantou University)

Thank You!

You may find more information here:
`http://www.ee.cityu.edu.hk/~lindai/`

If you have any questions, please do not hesitate to contact me:
lindai@cityu.edu.hk