AI Based D2D Communication Networks

Choong Seon Hong

Professor, Department of Computer Science and Engineering,
Kyung Hee University

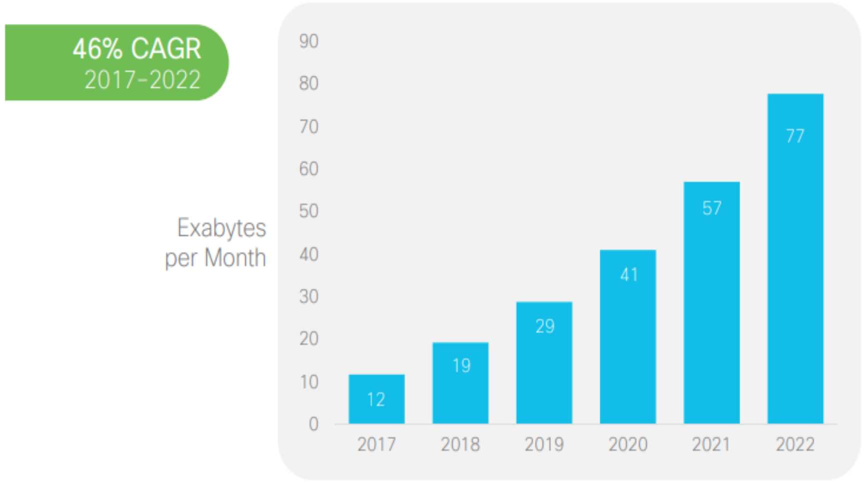
KYUNG HEE UNIVERSITY

NETWORKING INTELLIGENCE LAB

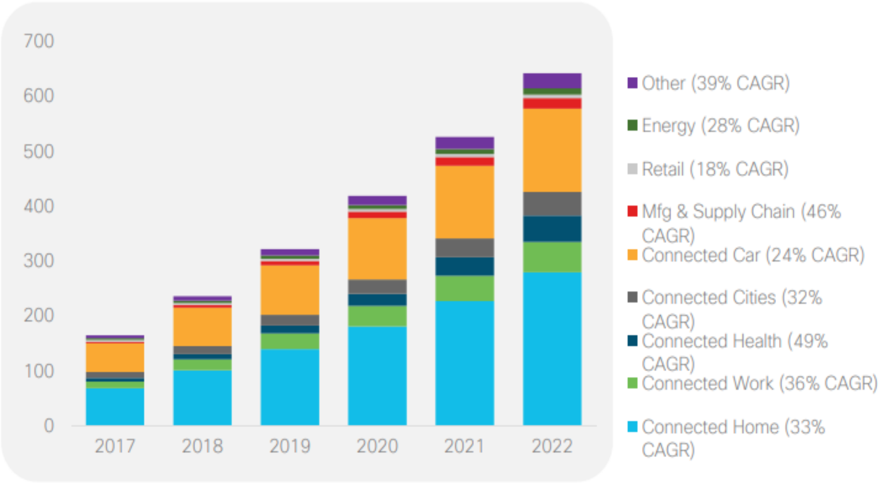
Outline

- Introduction
- D2D Usage Scenarios
- Why D2D Communications?
- D2D Communication Procedure
- Resource Management For D2D Communication
  - Problem Formulation
  - Solution Approaches:
    - Game Theory
    - Successive Convex Approximation
    - Deep Reinforcement Learning
- Challenges And Ongoing Research

- Mobile Internet and Smart Devices
  - Bandwidth and data traffic boost (Cisco)
    - Data traffic increases ~2 times/per year
    - Wireless network cannot support that!
  - Information aggregates to hotspot and local area
    - 70% in office and hotspot, over 90% in future
    - Hotspot QoS cannot be guaranteed!

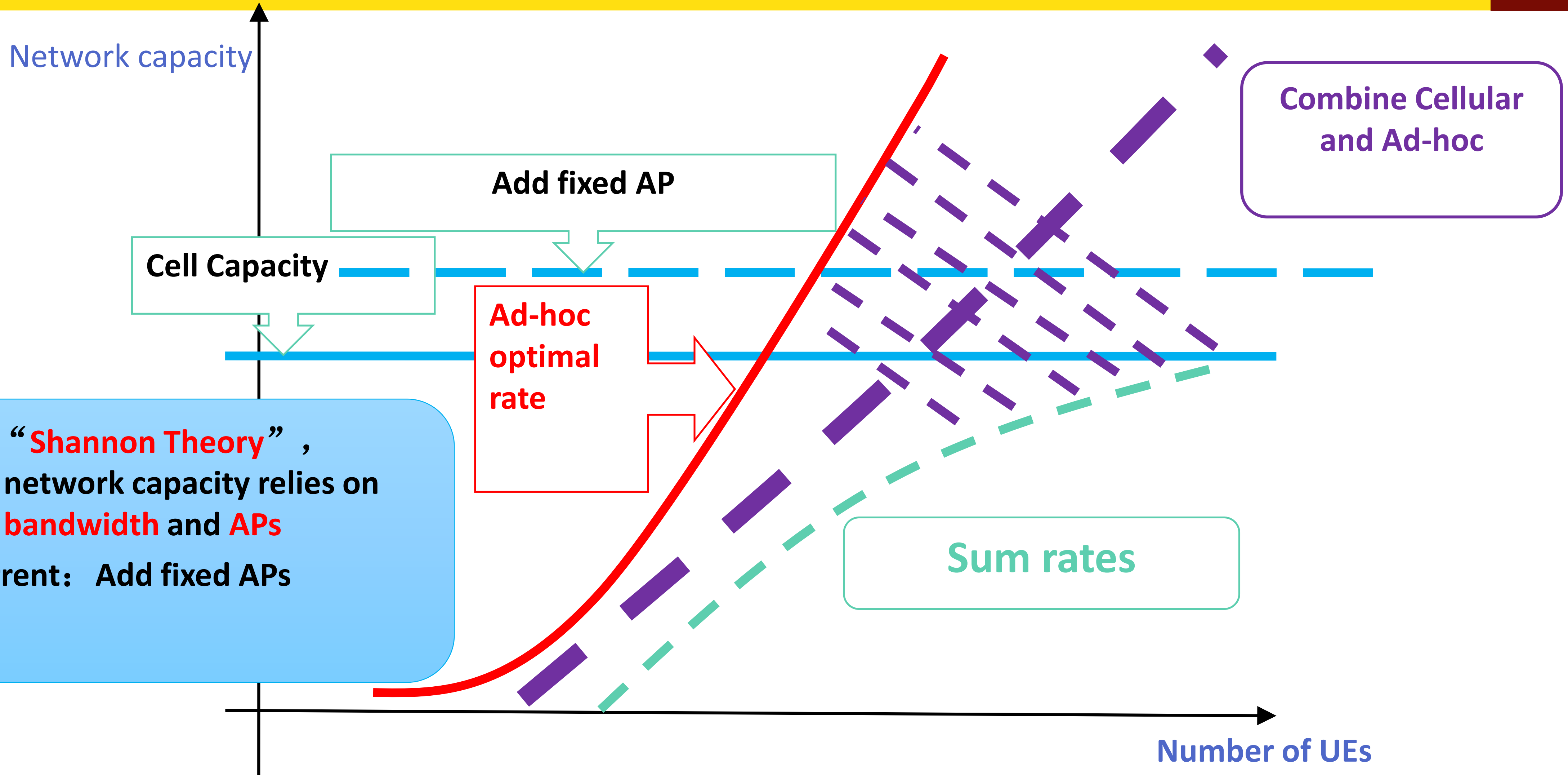


Source: Cisco VNI Global Mobile Data Traffic Forecast, 2017-2022



Source: Cisco VNI Global Mobile Data Traffic Forecast, 2017–2022

Introduction



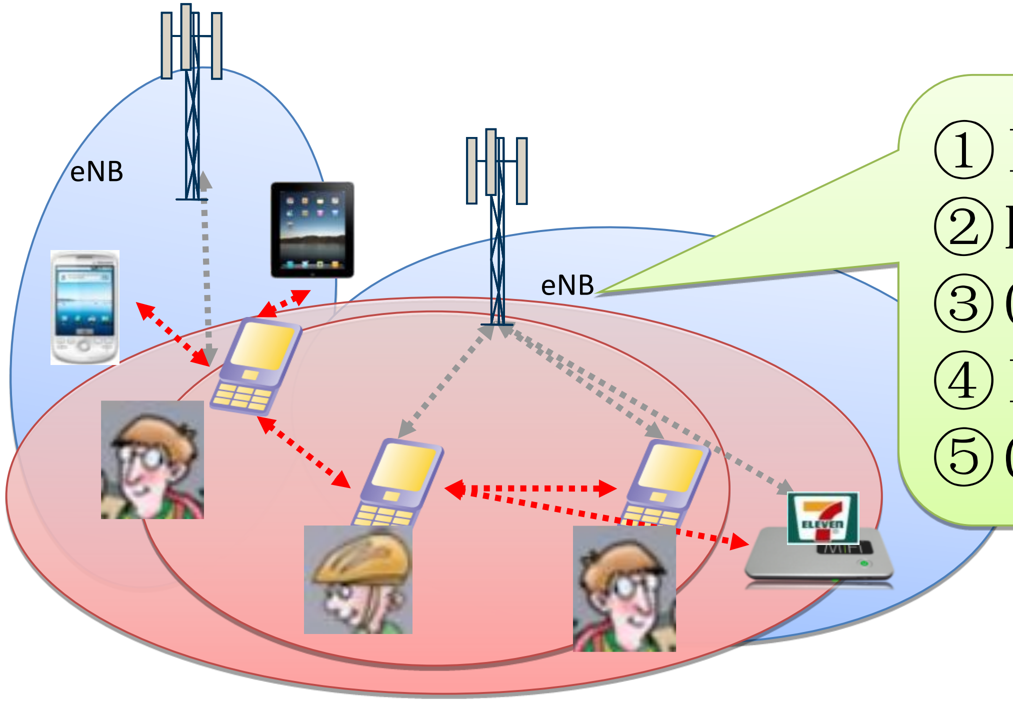
By "**Shannon Theory**", network capacity relies on **bandwidth** and **APs**

Current: Add fixed APs

P. Gupta and P. Kumar, "**The capacity of wireless networks**," *IEEE Transactions on Information Theory,* vol. 46, no. 2, pp. 388-404, Mar. 2000.

Introduction

- Definition of Device-to-Device (D2D) Communications
  - D2D communications commonly refer to the technologies that enable devices to communicate directly without an infrastructure of access points or base stations

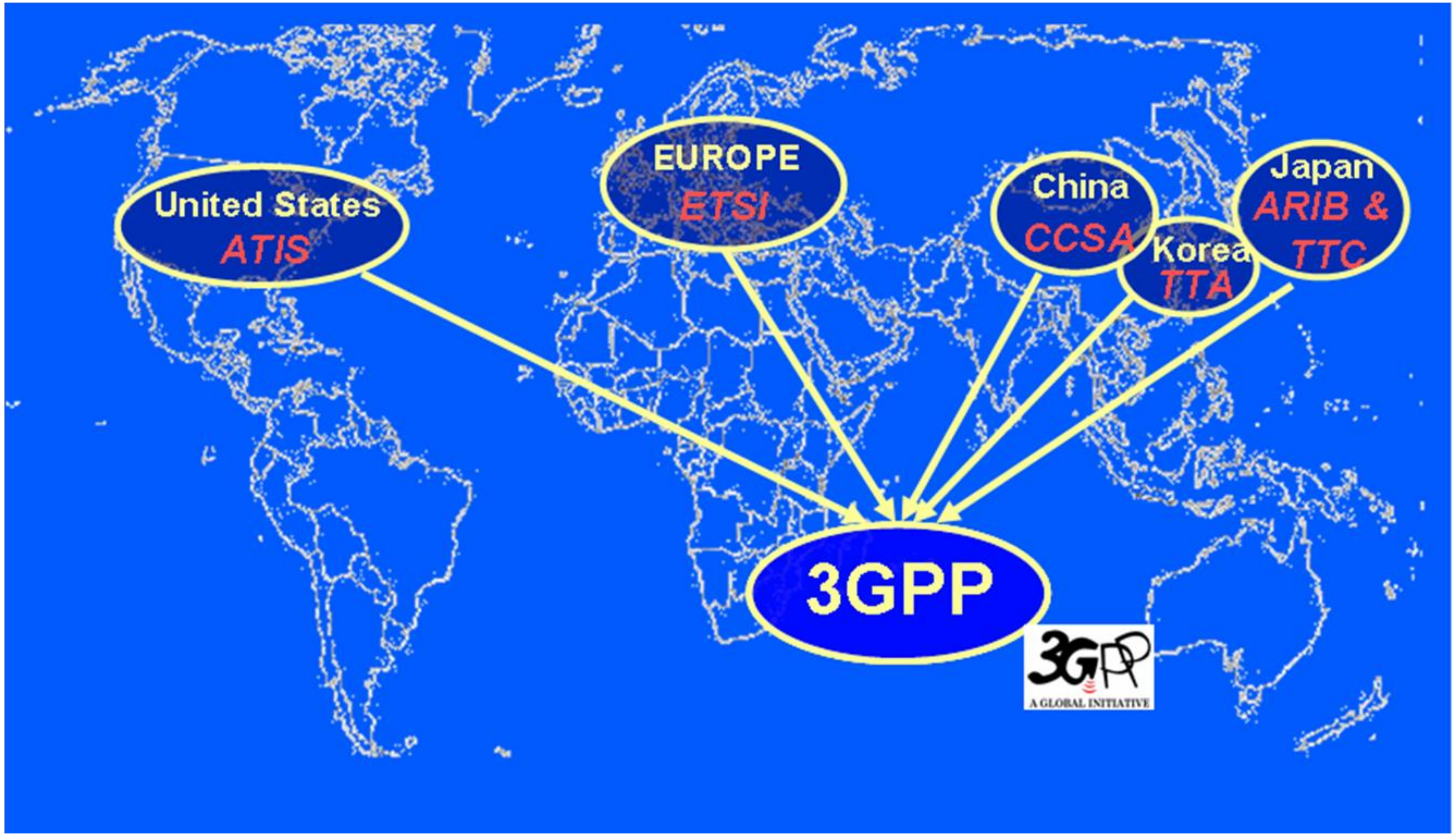


① Increase network capacity
② Extend coverage
③ Offload data
④ Improve energy efficiency
⑤ Create new applications

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication.** Berlin: Springer, 2014.

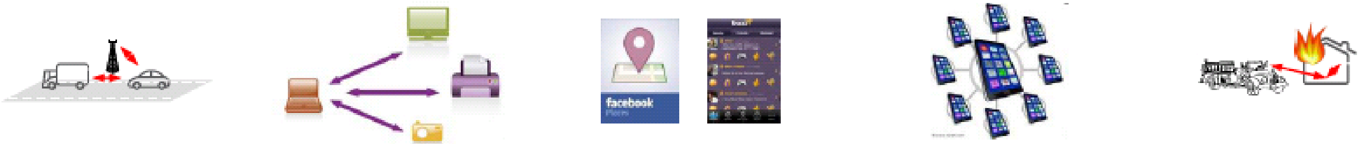
Standardization in a Global Partnership

- 3GPP: The 3rd Generation Partnership Project



3GPP R12 on D2D*

- Drastically different requirements



Traffic Safety | Local data transfer | Proximity based social networking | Proximity-enabled communication | NSPS

- The LTE platform would have the advantage over others, such as Wi-Fi and Bluetooth that operate device to device protocols, because they use license exempt spectrum.
- Proximity-based applications and services represent a recent and enormous social-technological trend
  - These applications and these services are based on the awareness that two devices or two users are close to each other
  - Awareness of proximity carries value, and generates demand for an exchange of traffic between them
- Direct D2D communication is also essential for public safety services
  - e.g. in case of lack of network infrastructure in disaster situations
- 3GPP has imitated work on enhancing the LTE-EPC platform to support these capabilities

KYUNG HEE UNIVERSITY

*3GPP R12 V0.2.0 page:49
URL: https://www.3gpp.org/ftp/Information/WORK_PLAN/Description_Releases/

NSPS : national security and public safety
LTE-EPC : Long Term Evolution-Evolved Packet Core

NETWORKING INTELLIGENCE LAB

3GPP R12 on D2D*

- Motivation of D2D Standard Issue(SI) and Works Items(WI) in 3GPP and other standardizations
- Development within 3GPP
  - SA1: - TR22.803: Feasibility study for Proximity Services (ProSe)
    - TS22.278: Service requirements for the Evolved Packet System (EPS)
  - SA2 WI TR22.969 : Feasibility study on enhancements of Public Warning System (PWS)
  - RAN SI: Study on LTE Device to Device Proximity Services
    - Objectives: evaluate LTE device-to-device proximity services
    - **Pro**ximity **Se**rvices (ProSe) Discovery over E-UTRA;
    - Evaluation of discovery and communication performance
      - Discovery research on carrier, timing, resource allocation, discovery signal/channel
      - Communication research on synchronization, timing, resource allocation, control/share channel
    - ProSe Communication over E-UTRA;
    - EPC support of ProSe Communication over WLAN

E-UTRA: Evolved Universal Terrestrial Radio Access
SA : System Aspect
TR: Technical Report
TS : Technical Specification

*3GPP R12 V0.2.0 page:49
URL: https://www.3gpp.org/ftp/Information/WORK_PLAN/Description_Releases/

Deployment Roadmap of D2D Communication

Cellular unaware D2D

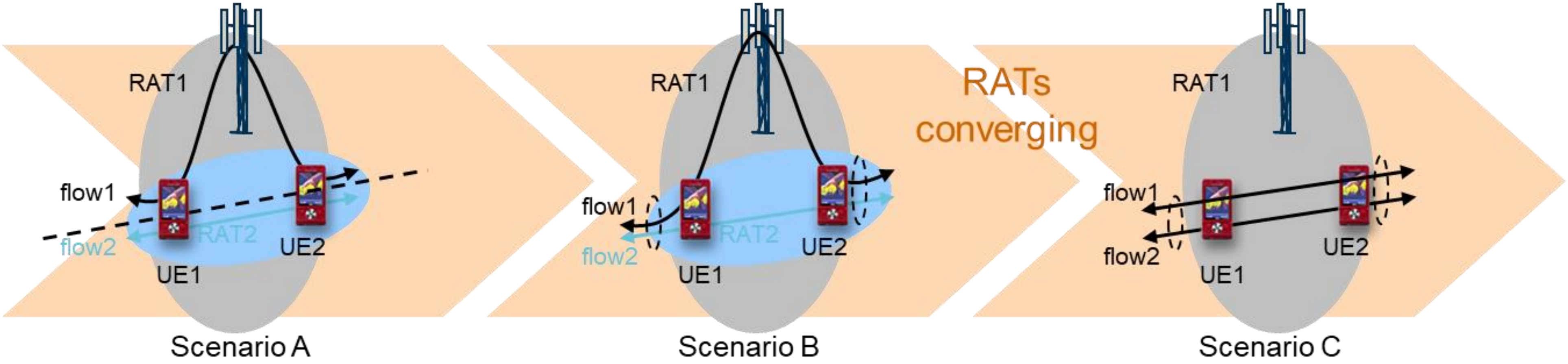
- Cellular network is not aware of D2D
- 2 RATs, e.g. 3G + Wifi
- No cooperation between cellular and D2D

Cellular aware D2D

- Cellular network is aware of D2D
- 2 RATs, e.g. LTE + Wifi
- Kind of cooperation between cellular and D2D

Cellular controlled D2D

- Cellular network fully controls D2D
- A single RAT, e.g. LTE-A
- D2D is a part of cellular communication



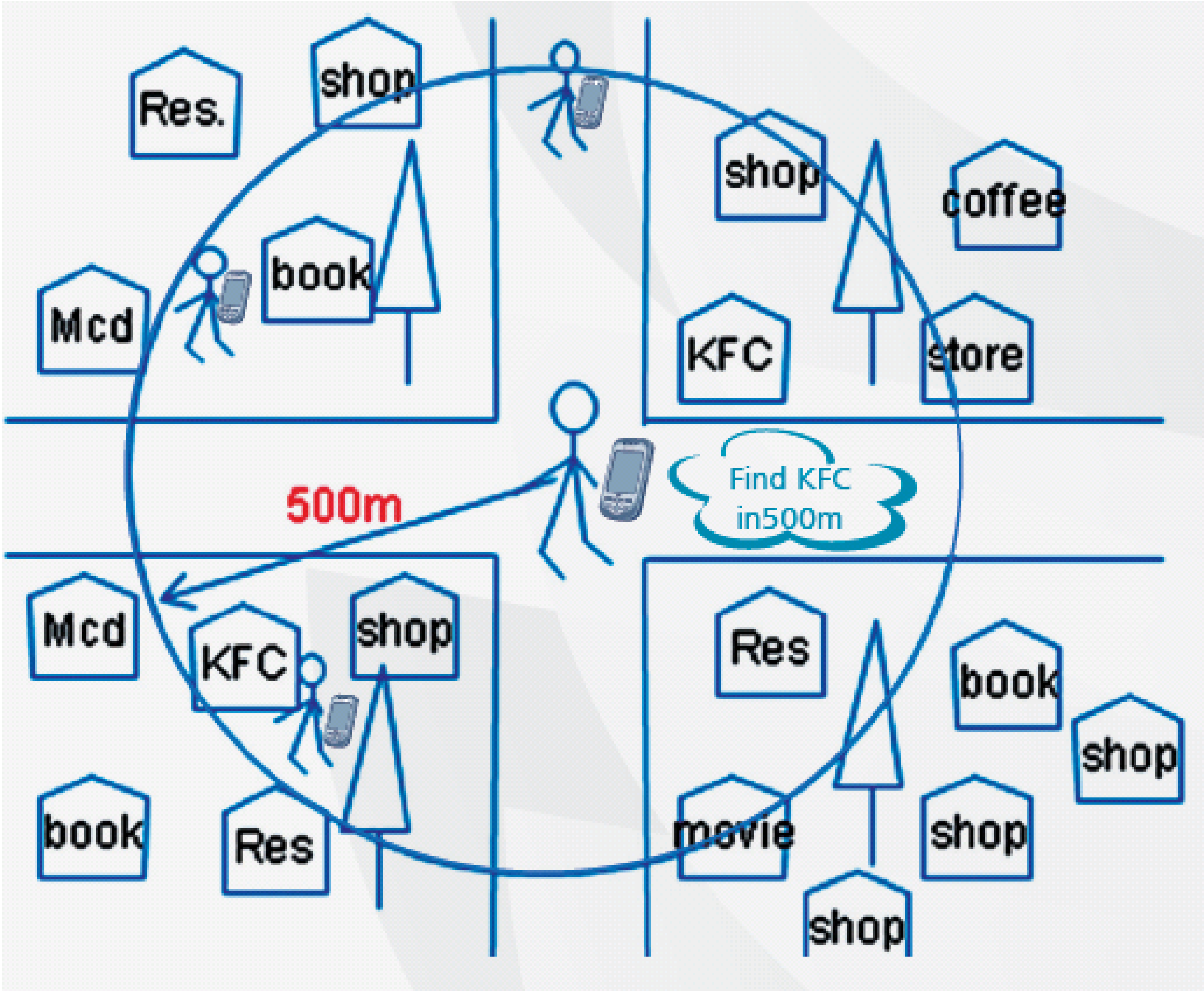
D2D Benefits	Scenario A	Scenario B	Scenario C
Traffic offload	★★★	★★★★	★★★★★
Unified & Simplified comm.	★★★	★★★★	★★★★★
User experience improvement	★	★★★	★★★★★
Cellular capacity enhancement	★	★	★★★★★

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication.** Berlin: Springer, 2014.

- **D2D Usage Scenarios**
  - Local Advertisement
  - Location Enhancement
  - Distance Based Applications
  - Disaster and Public Safety

Local Advertisement

1. The shops will automatically distribute the advertisement to the passage nearby
2. Applications in users' terminal discover the advertisers automatically and receive the information from them, including introduction, menus, coupons, etc.

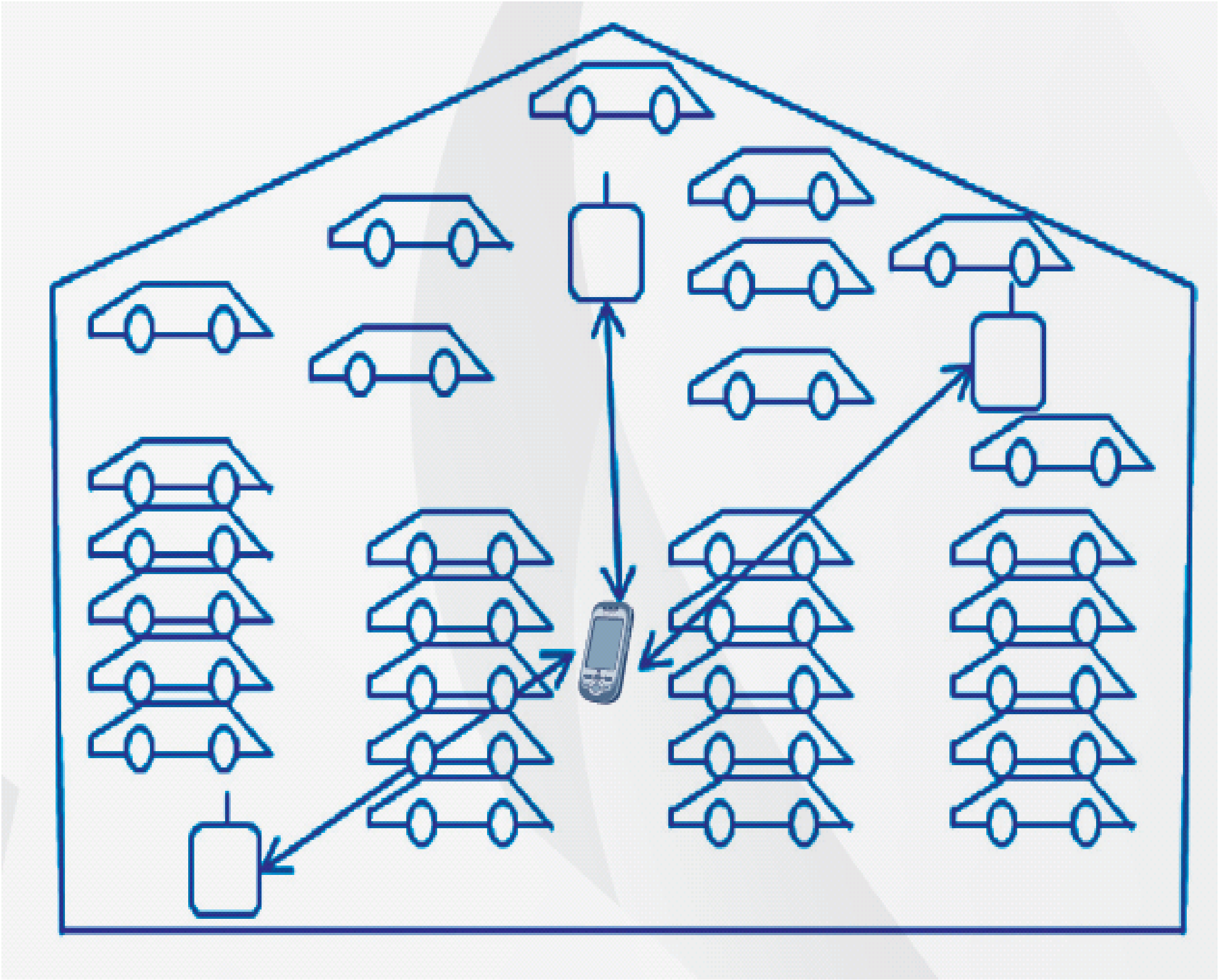


Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication**. Berlin: Springer, 2014.

Location Enhancement

- The D2D terminals receive the real-time parking space information that helps finding one's parking space easily

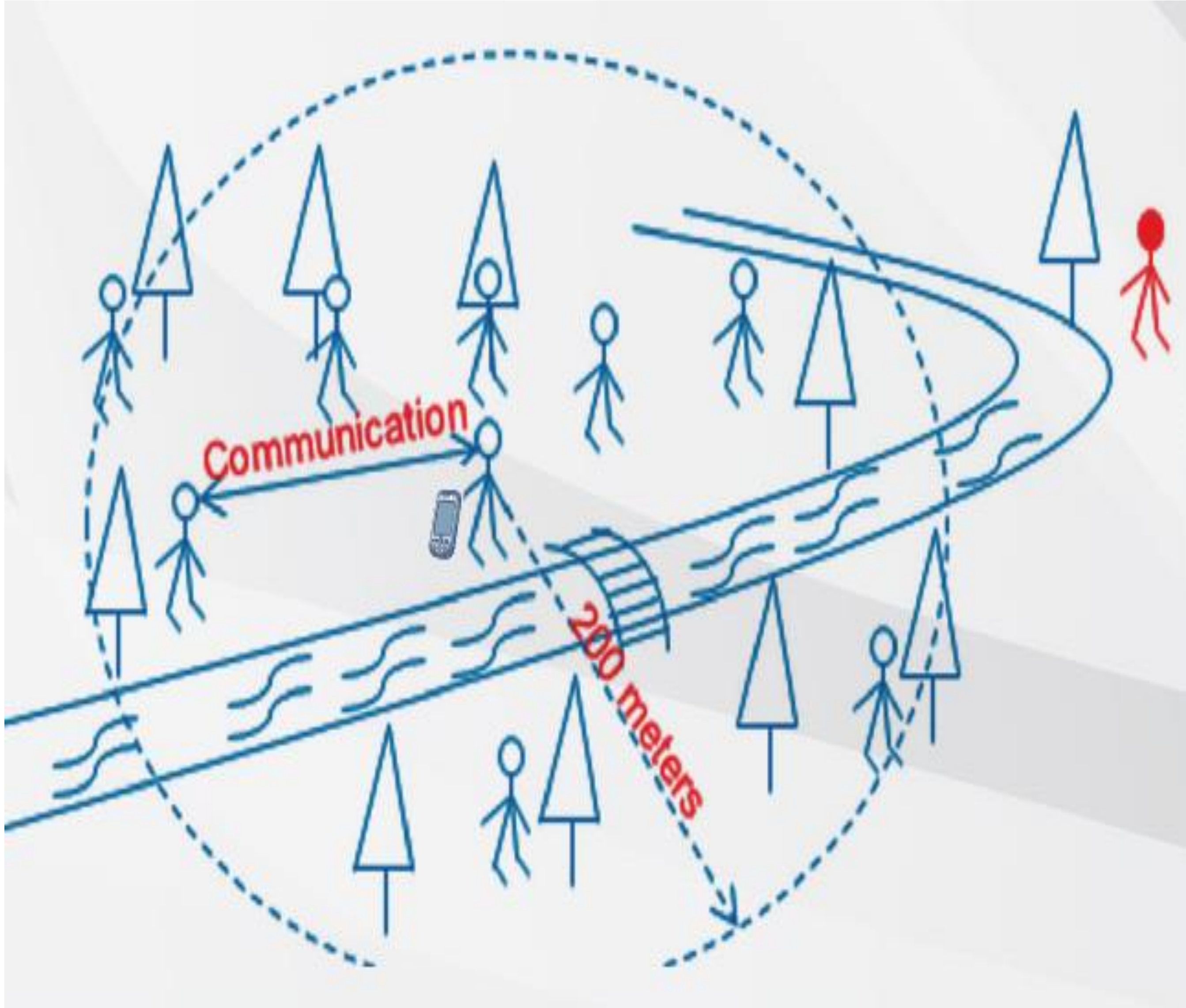
- It can provide more information than a GPS based application by D2D.



Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication**. Berlin: Springer, 2014.

Distance Based Applications

- Members of a team or group can obtain the sphere of activities for each other by D2D distance monitoring when touring, keeping a safe movement range to prevent occurring accident.



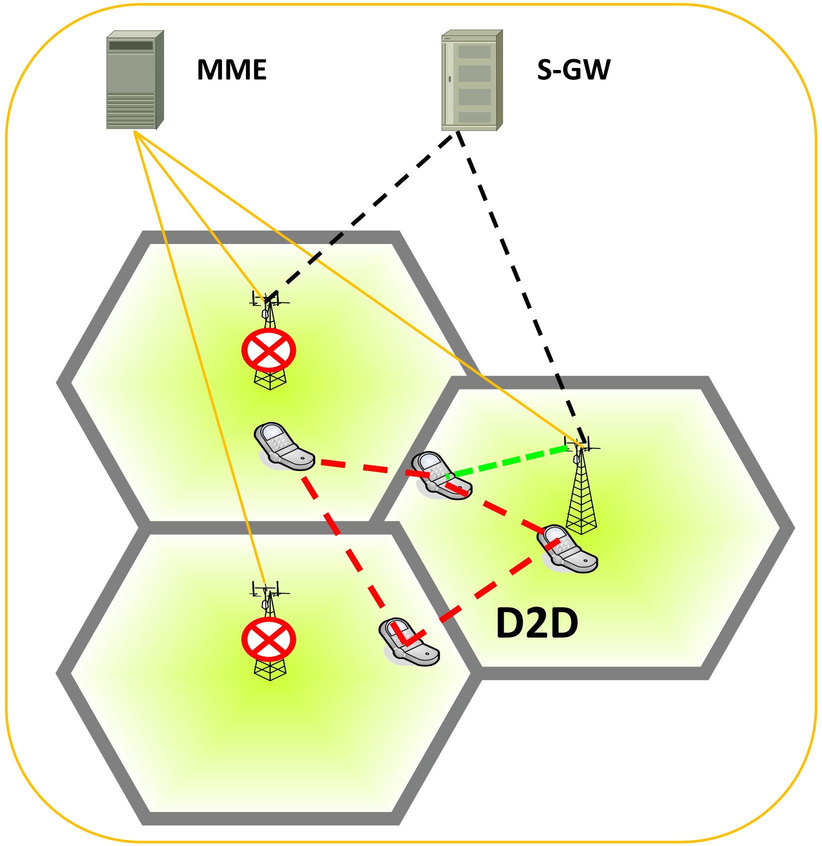
KYUNG HEE UNIVERSITY

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication**. Berlin: Springer, 2014.

NETWORKING INTELLIGENCE LAB

Disaster and Public Safety

- In case of disastrous situation, where the fixed infrastructures, such as BSs, are in failure, mobiles not in the coverage can possibly reach the BS with the aid of mobiles in the coverage area.
- This is similar to multi-hop relaying networks.



Emergency Communications

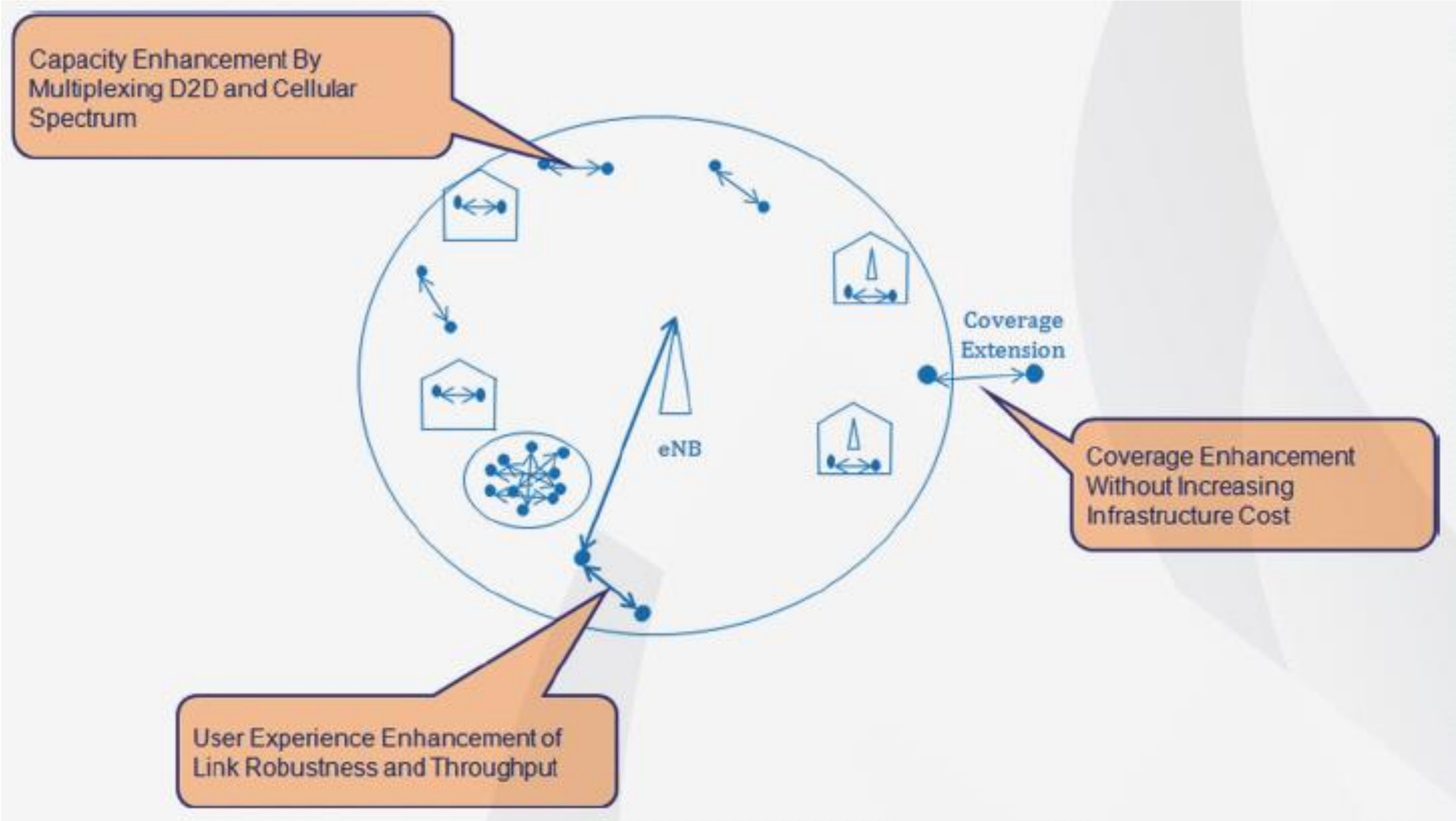
① Mobile Ad-hoc networks

② Active BS is **the** final destination

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication.** Berlin: Springer, 2014.

Enhance Network Capability (Offloading)

- D2D applications can provide coverage enhancement without increasing infrastructure cost, capacity enhancement by multiplexing D2D and cellular spectrum and user experience enhancement of link robustness and



Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication.** Berlin: Springer, 2014.

Why D2D Communication?

Why D2D Communication?

- Advantages of D2D Communications
  - ✓Proximity gain
  - ✓Hop gain
  - ✓Reuse gain

- Potential Benefits
  - ✓Higher date rate /capacity
  - ✓Lower latency
  - ✓Higher spectrum-, energy-, and cost-efficiency
  - ✓Better robustness

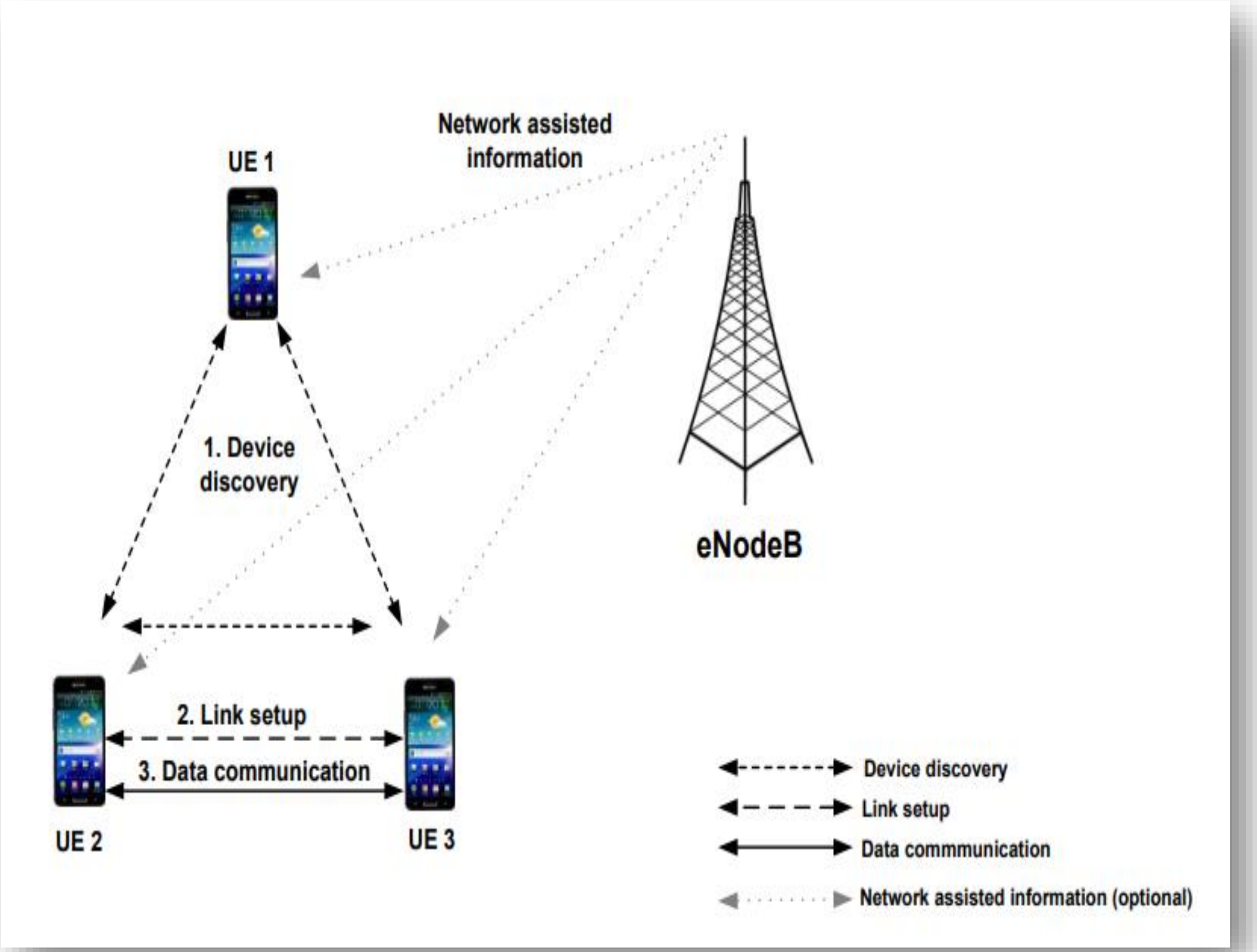
Why D2D Communication?

Feature name	D2D(4G)	D2D (5G)	Wi-fi Direct	Bluetooth
Standardization	3GPP R12*	3GPP R15**	IEE 802.11	Bluetooth SIG
Frequency band	LTE-A	mmWave (30GHz~300GHz)	2.4GHz – 5GHz	2.4GHz
Max transmission distance	Up to 500 m	10m ~ 500m	32 m	10-100 m
Max data rate	1Gb/s	Up to 10GB/s	54Mb/s	25Mb/s
Device discovery	Base station assisted or device assisted	Base station assisted or device assisted	Require user defined settings for access points	Require manual pairing
Uniformity of service provision	Yes	Yes	No	No
Application	Public safety, Local advertising, Cellular relay	Virtual Reality(VR), Augmented Reality (AR), Mixed Reality (MR)	Content sharing, Group gaming, Device connection	Object exchange, Peripherals connection

*3GPP R12
**3GPP TR 21.915

D2D Communication Procedure

- Device discovery
  - Detecting the presence of other devices in the neighborhood
- Link setup
  - Establishing links between interested devices
- Data communication
  - Transmitting or receiving data via established links



Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication**. Berlin: Springer, 2014.

Cellular Controlled D2D communication

BS controlled D2D Link setup:

1. A request of communication is initiated by one UE pair.

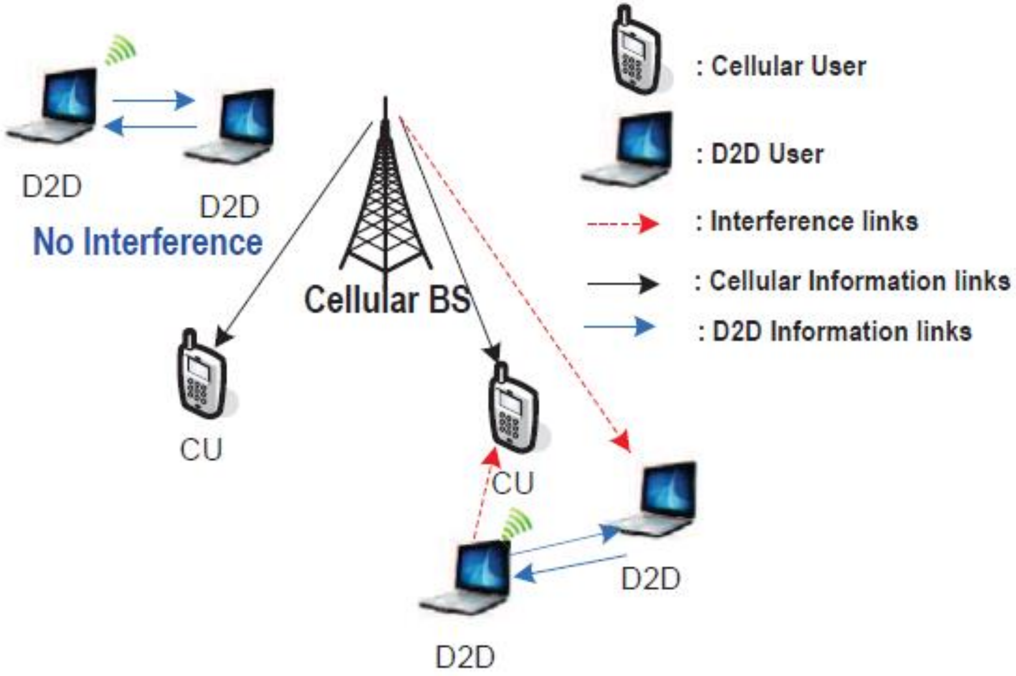
2. The system detects traffic originating from and destined to the UE in the same subnet.

3. If the traffic fulfills a certain criterion (e.g., data rate), the BS considers the traffic as the potential D2D traffic.

4. The BS checks if D2D communication offers higher throughput by calculating the surrounding interference.

5. If both UEs are D2D capable and D2D communication offers higher throughput, the BS may set up a D2D bearer.

6. Even if the D2D connection setup is successful, the BS still maintain detecting if UE should be back to the cellular communication mode.



Resource Management For D2D Communication

- Research Challenges
  - Access Methods
  - Mode Selection
  - Spectrum Sharing
  - Interference Management
- Problem Formulation
- Solution Approaches
  - Game Theory
  - Successive Convex Approximation(SCA)
  - Deep Reinforcement Learning

Access Methods

- The D2D networks can be configured in three ways to allow or restrict their usage by certain users:
  - Network Controlled: In this scheme, the base station and the core network control the communication signaling setup and the there after resource allocation for both cellular and D2D users.
  - Self-organized: In this scheme, D2D users themselves realizes the communications in a self-organizing way by finding the empty spectrum hole.
  - Network Assisted : The D2D users operates in a self-organized way, and exchanges with cellular system limited controlling information for resource management. And the cellular network can obtain the status of D2D communications for better control purposes.
    - This scheme comprises the merits of first two D2D users access methods.

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication**. Berlin: Springer, 2014.

Mode Selection

- Mode selection criteria
  - Cellular Mode: The mobile works in a traditional cellular way relaying data by a BS.
  - D2D Mode: The mobiles exchanges information directly.
  - Mode Adaptation: The mobiles can select the right mode for communication according to the predefined criterions.
- Typically, all the UEs are implemented with two modes, i.e. cellular mode, and D2D mode, and can **adaptively utilize** the proper way for transmission.

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication**. Berlin: Springer, 2014.

Spectrum Sharing

- Spectrum sharing as an **overlay**:
  - The D2D users occupies the vacant cellular spectrum for communication.
  - This approach that completely eliminates cross-layer interference is to divide the licensed spectrum into two parts (orthogonal channel assignment).
  - This way, a fraction of the subchannels would be used by the cellular users while another fraction would be used by the D2D networks.
  - Although optimal from a cross-layer interference standpoint, this approach is inefficient in terms of spectrum reuse
- Spectrum sharing as an **underlay**:
  - This scheme allows multiple D2D users to work as an underlay with cellular users, and thus to improve the spectrum efficiency.
  - Therefore, co-channel assignment of the cellular and D2D users seems more efficient and profitable for operators, although far more intricate from the technical point of view.

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication**. Berlin: Springer, 2014.

Interference Analysis

- D2D networks provides coverage at the customer premises, but radiates toward neighboring mobile users, introducing interference
- Considering that these networks define two separate layers, interference can be classified as follows:
  - Cross-layer: This refers to situations in which the aggressor (e.g., a D2D user) and the victim (e.g., a cellular user) of interference belong to different network layers.
  - Co-layer: In this case the aggressor (e.g., a D2D user) and the victim (e.g., a neighboring D2D user) of interference belong to the same network layer.
- Combating the interference
  - Transmitter: Frequency, time, space, and power allocations
  - Receiver: Signal processing for interference cancelation

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication.** Berlin: Springer, 2014.

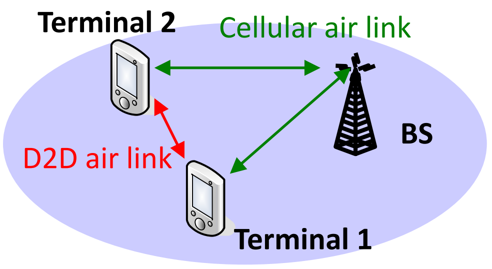
- Power control can be performed by two approaches:
  - Self-organized power control:
    - The D2D users make power changes in a self-organized way according to a predefined SINR threshold in order to meet the QoS, and meanwhile not affect the cellular users.
  - Network managed power control:
    - Both cellular and D2D users adaptively adjust the transmit power according to the SINR report.
    - Typically, the D2D users can control the transmit power at first, and then the cellular users make change afterward.
    - This iterative process terminates until all the users meet their SINR requirements.
- The first method is not going to change the behaviors of cellular users, such that D2D users are invisibly treated.
- The second method allows all the users to adjust the transmit power, but requires a certain amount of signaling overhead

SINR : Signal to Interference Noise Ratio

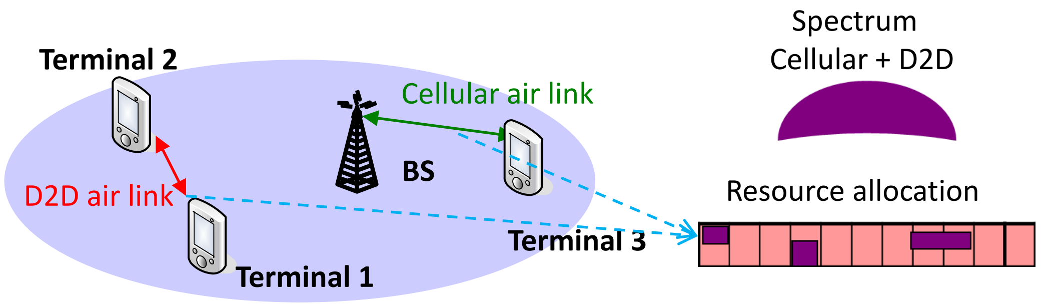
Threshold Based Power Control

Motivation & Problem

- Same frequency-time resources could be shared by cellular & D2D links to enhance the system capacity
- **Power control** is necessary to mitigate the interference & saving energy
  - Limit the amount of interference to cellular users
  - Achieve a reliable level of performance for D2D users



Cellular & D2D

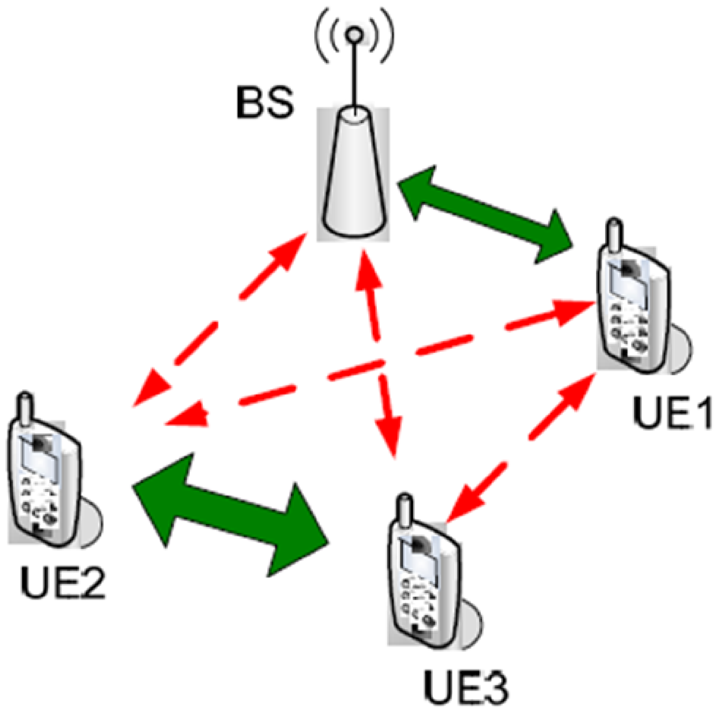


Resource reuse between D2D & cellular links

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication.** Berlin: Springer, 2014.

Signal and Interference Model

Resource allocation	Interference
Uplink (UL)	CEL— From: D2D user. To: BS D2D— From: cellular user. To: D2D user
Downlink (DL)	CEL— From: D2D user. To: cellular user D2D— From: BS. To: D2D user



D2D comm. reusing cellular resources. Solid arrows indicate the wanted signals and dashed arrows the interference.

Xu, Chen, Lingyang Song, and Zhu Han. **Resource management for device-to-device underlay communication.** Berlin: Springer, 2014.

Problem Formulation

- Only D2D pair:

Transmit power

Channel gain

$$\boxed{\gamma_k^r} = \frac{\boxed{P_k^r}\boxed{g_k^r}}{\boxed{\sigma^2}}$$

SINR

Additive White Gaussian Noise

- Cellular user and single D2D pair:

$$\gamma_k^r = \frac{P_k^r g_k^r}{\boxed{P_{k_0}^r g_{k_0,k}^r} + \sigma^2}$$

Interference from Cellular User

- Cellular user and multiple D2D pairs:

$$\gamma_k^r = \frac{P_k^r g_k^r}{P_{k_0}^r g_{k_0,k}^r + \boxed{\sum_{i \in \Omega_r, i \neq k} P_i^r g_{i,k}^r} + \sigma^2}$$

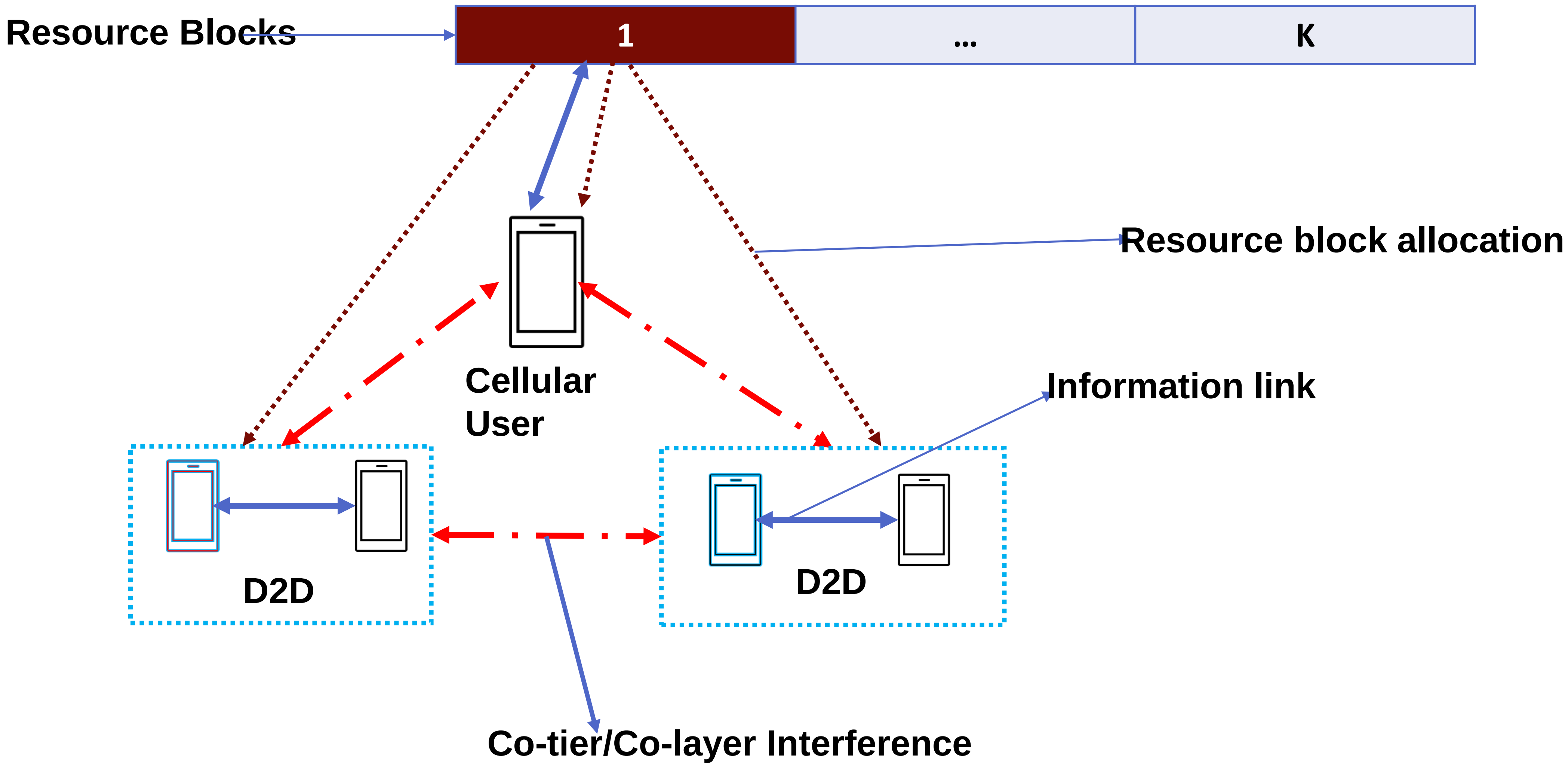
Co-layer/Co-tier interference

- Achievable rate of D2D user:

$$R_k^r = W^r \log\left(1 + \gamma_k^r\right)$$

S. M. A. Kazmi, **CS Hong** *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

Co-tier /Co-layer Interference



A Matching Game Approach

System Model for Mode Selection

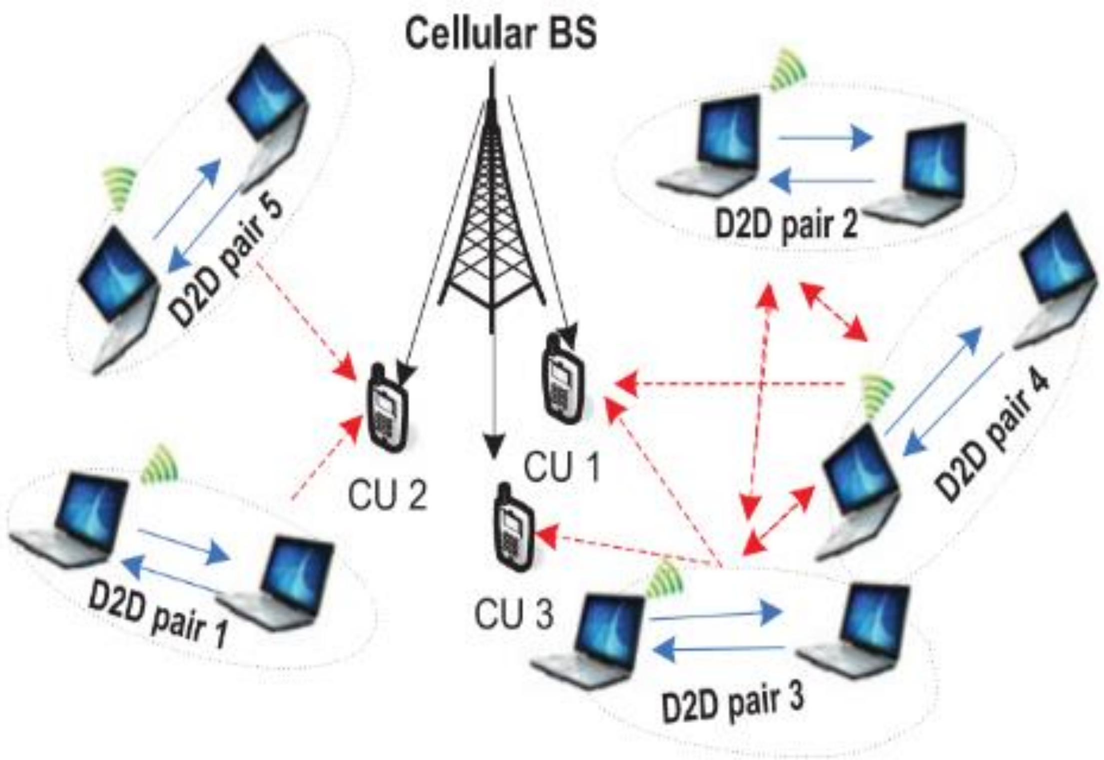


Figure: A downlink D2D communication system. The solid line shows the information links while the dashed line shows the interference links.

S. M. A. Kazmi, CS Hong *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

System Model for Mode Selection

- Consider the downlink of a cellular network consisting of a single BS and a set $\mathcal{K}$ of K D2D pairs located under its coverage
- We use the index k_0 to indicate the base station(BS)
- We let set $\mathcal{C}$ be the set of C cellular users(CUEs)
- The CUEs and D2D pairs use the same set $\mathcal{R}$ of R orthogonal resource blocks (RBs)
- For any given RB $r \in \mathcal{R}$, a predefined interference threshold $I_r^{\max}$ must be maintained for protecting the cellular users
- This system model is focused on a dense communication environment in which the density of the users is higher than the number of connections that a given BS can support

S. M. A. Kazmi, CS Hong *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

System Model for Mode Selection

- The D2D pairs at each time slot need to determine which RB is feasible in order to maximize the utility of the system while protecting the cellular users
- For RB allocation, we introduce the binary variables x_k^r

$$x_k^r = \begin{cases} 1, & \text{If D2D pair } k \text{ is assigned RB } r, \\ 0, & \text{Otherwise.} \end{cases} \tag{1}$$

- The received signal to interference noise ratio (SINR) pertaining to the transmission of the D2D pair k over RB r with transmit power $P_k^{r\,(*)}$ is:

SINR

$$\boxed{\gamma_k^r} = \frac{x_k^r P_k^r g_k^r}{\boxed{P_{k_0}^r g_{k_0,k}^r} + \boxed{\sum_{i \in \Omega_r, i \neq k} x_i^r P_i^r g_{i,k}^r} + \boxed{\sigma^2}}, \tag{2}$$

Co-tier interference

White noise

Co-tier interference

[(*)] We assume that all transmitters (BS and D2D pairs) transmit using a fixed power within the RB duration

S. M. A. Kazmi, CS Hong *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

System Model for Mode Selection

- Similarly, the SINR of cellular user c over RB r is given as:

Signal to Interference noise ratio(SINR)

Transmit power

Channel gain

White noise

Interference from D2D users

$$\gamma_c^r = \frac{x_{k_0}^r P_{k_0}^r g_{k_0,c}^r}{\sum_{i\in\Omega_r} x_i^r P_i^r g_{i,c}^r + \sigma^2}, \quad (7)$$

- In the D2D decision model, each D2D pair acts based on its achieved utility
- The action here represents the D2D decision to use a given mode or not
- We assume that each D2D pair selfishly and rationally acts in a way that maximizes its utility
- Each D2D pair has knowledge of its own utility functions

S. M. A. Kazmi, CS Hong *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

- Therefore, each D2D pair only acts to maximize its own utility
- A decision variable α_k is used to indicate if D2D pair k will follow a specific mode, as follows:

$$\alpha_k = \begin{cases} 1, & \text{if D2D pair k uses the mode,} \\ 0, & \text{Otherwise.} \end{cases} \tag{8}$$

- This D2D decision model assists the BS in the mode selection process
- For mode selection, we consider two modes that can be selected for RB allocation for the D2D pairs

S. M. A. Kazmi, CS Hong *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

Mode Selection Definition

Partial-reuse mode

Only one D2D pair can be allocated to an RB currently in use by a cellular user, only if the interference is below a pre-defined threshold. By using this mode, there exists no co-tier interference (i.e., between D2D pairs). This mode is suitable for scenarios in which the number of D2D pairs is limited compared to the RBs or the D2D pairs are in close proximity with each other

Full-reuse mode

A group of D2D pairs can share an RB only if the interference produced by this group is below the predefined threshold for protecting the cellular tier. This mode is preferred in the scenario where there exist a large number of D2D pairs compared to RBs

S. M. A. Kazmi, **CS Hong** *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

- A binary variable y is defined to represent the two modes, controlled by the BS:

$$y = \begin{cases} 1, & \text{partial-reuse mode,} \\ 0, & \text{full-reuse mode.} \end{cases} \tag{9}$$

- In this model, the BS does not choose a mode for individual D2D pairs based on their channel conditions and buffer status
- The BS chooses a mode depending upon the utility achieved by the network
- Our goal is to maximize a utility function that captures the sum rate of the D2D pairs by selecting the optimal mode for communication, admitting the best D2D pairs, and properly reusing the RBs already occupied by the cellular tier:

$$U(x, \alpha, y) = \sum_{k \in \mathcal{K}} \sum_{r \in \mathcal{R}} \left[y \alpha_k R_k^r + (1 - y) \alpha_k \tilde{R}_k^r \right] .^{(*)} \tag{10}$$

(*) R_k^r and $\tilde{R}_k^r$ k represents the rates achieved using partial and full-reuse modes, respectively.

S. M. A. Kazmi, **CS Hong** *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

Problem Formulation for Mode Selection

$$\max_{x,y,\alpha} \; U(x,y,\alpha)$$

Maximize network utility function

$$\text{s.t.} \sum_{r \in \mathcal{R}} x_k^r \leq 1, \forall k \in K,$$

Resource block allocations constraint
Each cluster is allocated at most one resource block

Set of D2D users in resource block r

$$y\alpha_k I_k^k + \sum_{k=1}^{|\Omega_r|} (1-y)\alpha_k I_k^r \leq I_{max}^r,$$

Maximum interference of D2D users to cellular users must be below the predefined threshold

$$x_k^r \in \{0,1\}, \forall k \in \mathcal{K}, \forall r \in \mathcal{R},$$

Clustering variable

$$\alpha_k \in \{0,1\}, \forall k \in \mathcal{K},$$

Control variables whether D2D uses the mode or not

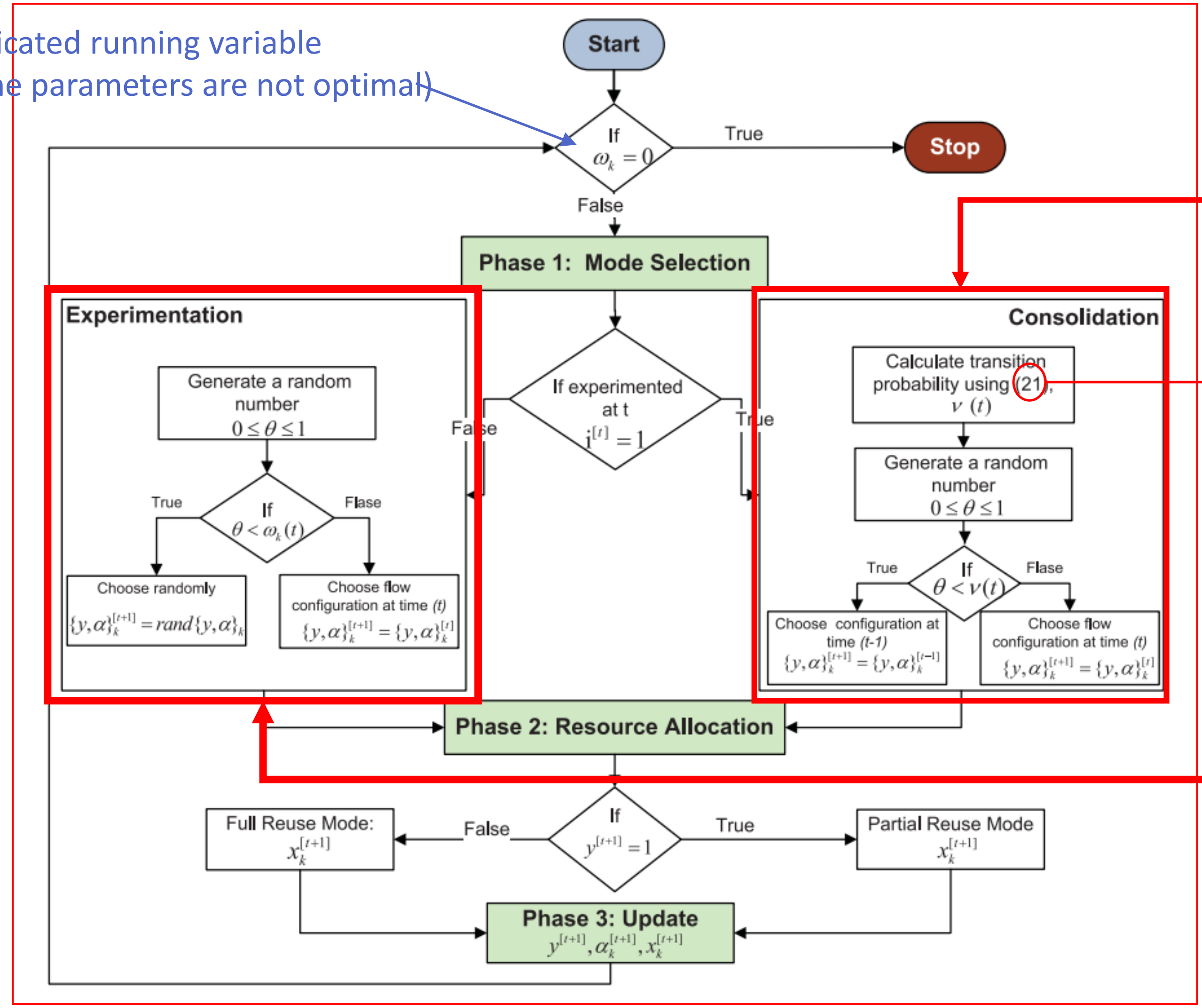
$$y \in \{0,1\}.$$

Mode selection variable: Partial-reuse or Full-reuse

S. M. A. Kazmi, **CS Hong** *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

Mode Selection: Matching Game Approach

Dedicated running variable
(if the parameters are not optimal)



During consolidation, the current utility obtained at time slot t is compared with the previously achieved utility at time slot t-1 by all players

$$q_{(f_k \to f'_k)} = (1 + \exp[\beta_k(U_{k,f_k} - U_{k,f'_k})])^{-1}, \qquad (21)$$

In Experimentation: each player executes one of the two actions, i.e., a new random configuration is chosen (exploration) or it stays with the current configuration (exploitation) with probability

S. M. A. Kazmi, **CS Hong** *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

Mode Selection: Matching Game Approach

Algorithm 1. Mode Selection and Resource Allocation

Phase 1: Initialization:
input: $\boldsymbol{\alpha}$, $\mathcal{P}_k$, $\mathcal{P}_r$, $\forall r, k$.
initialize: $t = 1$, $\mu^{(1)} \triangleq \{\mu(k)^{(1)}, \mu(r)^{(1)}\}_{k \in \mathcal{K}, r \in \mathcal{R}} = \emptyset$, $\mathcal{L}_r^{(1)} = \emptyset$
  $\mathcal{P}_k^{(1)} = \mathcal{P}_k$, $\mathcal{P}_r^{(1)} = \mathcal{P}_r$, $I^r_{\max}$, $\forall r, k$.
Phase 2: Matching:
repeat
  $t \leftarrow t + 1$.
  for $k \in \mathcal{K}$, propose r according to $\mathcal{P}_k^{(t)}$ **do**
    while $k \notin \mu(r)^{(t)}$ and $\mathcal{P}_k^{(t)} \neq \emptyset$ **do**
      if $I^r_{\max} \geq I^r_k$ **then**
        if $k \succ_r \mu(r)^{(t)}$ **then**
          $\mu(r)^{(t)} \leftarrow \mu(r)^{(t)} \setminus k'$.
          $\mu(r)^{(t)} \leftarrow k$.
          $\mathcal{P}'^{(t)}_r = \{k' \in \mu(r)^{(t)} | k \succ_r k'\}$.
        else
          $\mathcal{P}''^{(t)}_r = \{k \in \mathcal{K} | \mu(r)^{(t)} \succ_r k\}$.
      else
        $\mathcal{P}'''^{(t)}_r = \{k \in \mathcal{K} | I^r_{\max} \leq I^r_k\}$.
        $\mathcal{L}_r^{(t)} = \{\mathcal{P}'^{(t)}_r\} \cup \{\mathcal{P}''^{(t)}_r\} \cup \{\mathcal{I}'^{(t)}_r\}$.
        for $l \in \mathcal{L}_r^{(t)}$ **do**
          $\mathcal{P}_l^{(t)} \leftarrow \mathcal{P}_l^{(t)} \setminus \{r\}$.
          $\mathcal{P}_r^{(t)} \leftarrow \mathcal{P}_r^{(t)} \setminus \{l\}$.
until $\mu^{(t)} = \mu^{(t-1)}$.
Phase 3: Resource Allocation:
output: $\mu^{(t)}$.

Preference list of D2D Users/Resource Blocks (BS)

Set of matching pairs

Set of blocking pairs: containing all pair of D2D users and resource blocks that not satisfied the interference condition

Matching candidate set of resource block r

Each unassigned D2D pair k proposes to its most preferred RB r according to $\mathcal{P}_k$

The BS checks the maximum interference constraint if assigned RB r to D2D pair k

If maximum interference constraint is not violated, the D2D pair k is accepted

If maximum interference constraint is violated, the D2D pair k is rejected from RB r.

Matching is stable – convergence to a stationary point
The matching at time t is the at time t - 1

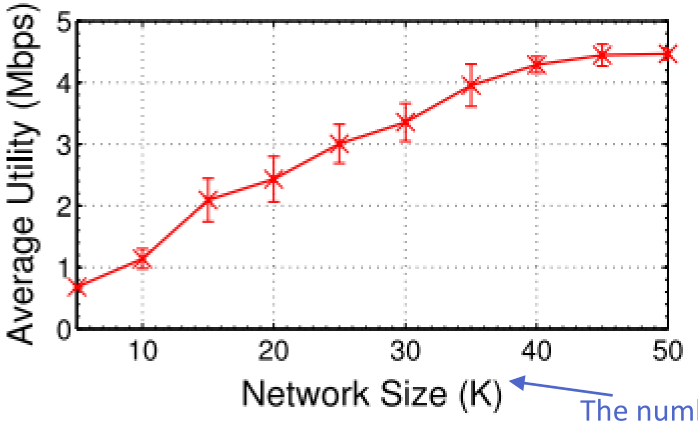
Simulation Setup for Mode Selection

Simulation Parameters	Values
Radius of MBS	500 m
Carrier frequency (f)	2 GHz
Frame Structure	Type 1 (FDD)
Transmission Time Interval (TTI)	1 ms
Total transmit power of BS	46 dBm
Total transmit power of D2Ds	23 dBm
System bandwidth	3 MHz
Bandwidth of each RB (W)	180 kHz
Number of subcarriers per RB	12
Neighboring subcarrier spacing	15 kHz
Modulation and coding scheme (MCS) [46]	QPSK: $1/12, 1/9, 1/6, 1/3, 1/2, 3/5$ 16QAM: $1/3, 1/2, 3/5$
Path loss (cellular link)	$128.1 + 37.6\ \log(\mathrm{d})$, d[km]
Path loss (D2D links) [47]	$32.45 + 20\ \log(\mathrm{f}) + 20\ \log(\mathrm{d})$, f[MHz]
Shadow fading standard deviation [47]	3 dB
Proximity of D2Ds ($R2$)	random $\{20 \sim 30\}$ m
Thermal noise for 1 Hz at 20 °C	-174 dBm

S. M. A. Kazmi, **CS Hong** *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

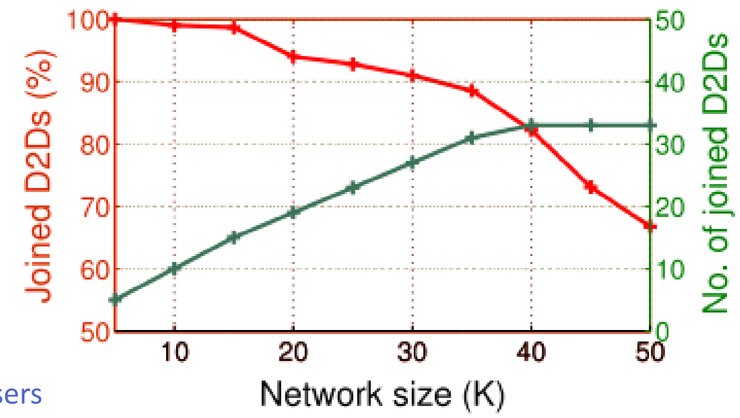
Mode Selection: Matching Game Approach

1



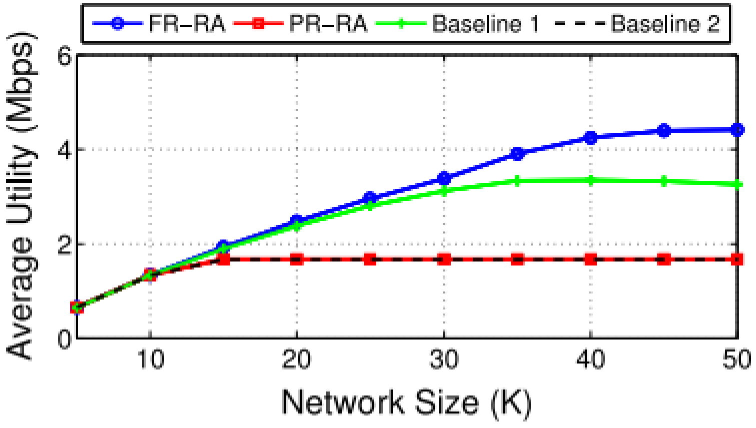
The number of D2D Users

(a) Average Utility

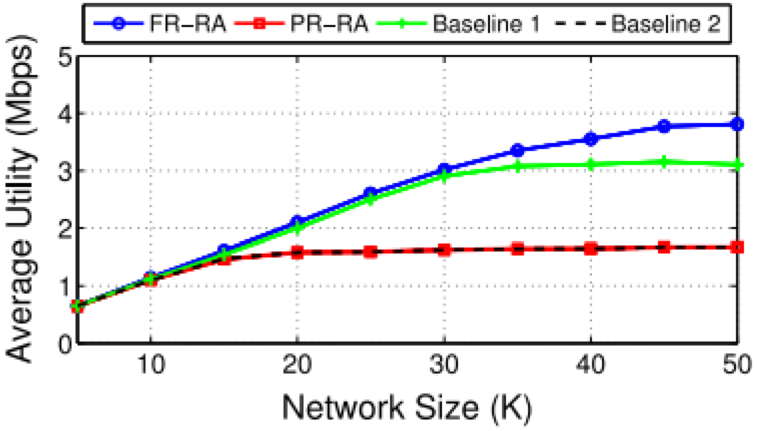


(b) Average successfully joined D2Ds

2



(a) $I^r_{\max} = -80$ dBm



(b) $I^r_{\max} = -100$ dBm

S. M. A. Kazmi, **CS Hong** *et al.*, "Mode Selection and Resource Allocation in Device-to-Device Communications: A Matching Game Approach," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3126-3141, 1 Nov. 2017, doi: 10.1109/TMC.2017.2689768.

KYUNG HEE UNIVERSITY

FR-RA: Full-reuse Resource Allocation

PR-RA: Partial-reuse Resource Allocation

NETWORKING INTELLIGENCE LAB

- D2D with Non-Orthogonal Multiple Access(NOMA)
  - Successive Convex Approximation Approach

System Model of D2D with NOMA

- Consider the downlink transmission of a macro base station (MBS)
- The MBS serves a set of M cellular users (CUs) denoted by $\mathcal{M}$
- And, a set $\mathcal{N}$ of N D2D user pairs that sharing the same system bandwidth with the CUs

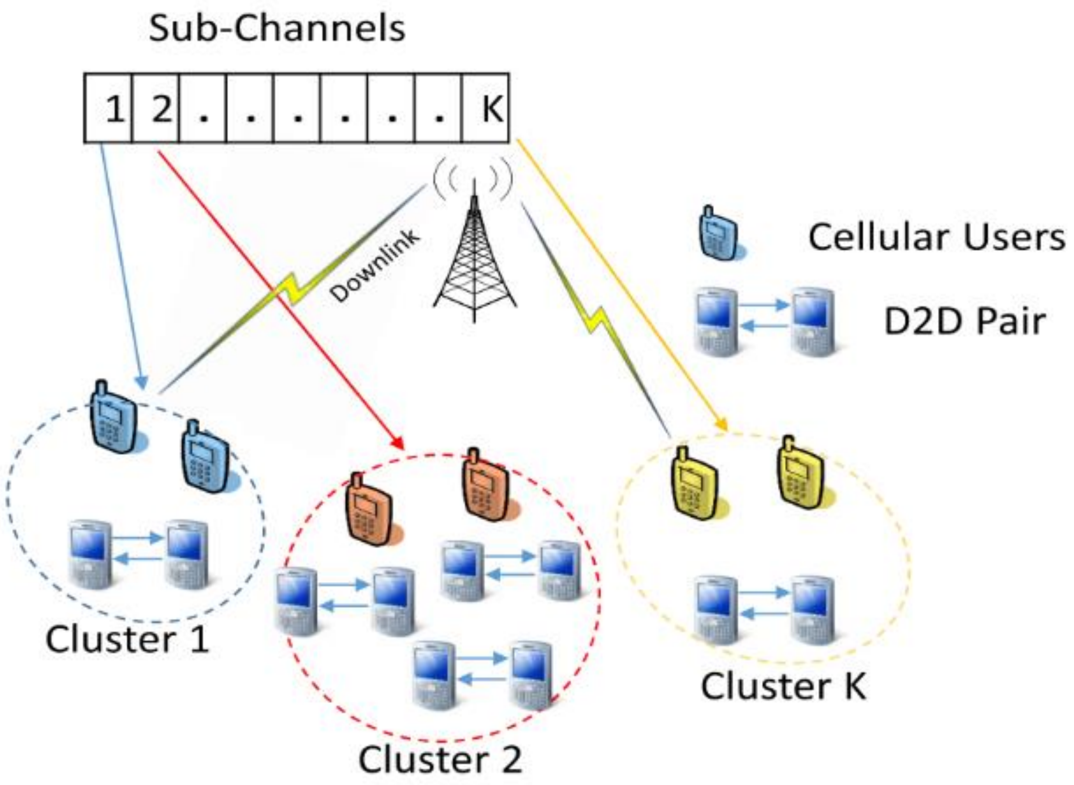


Figure: System model.

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

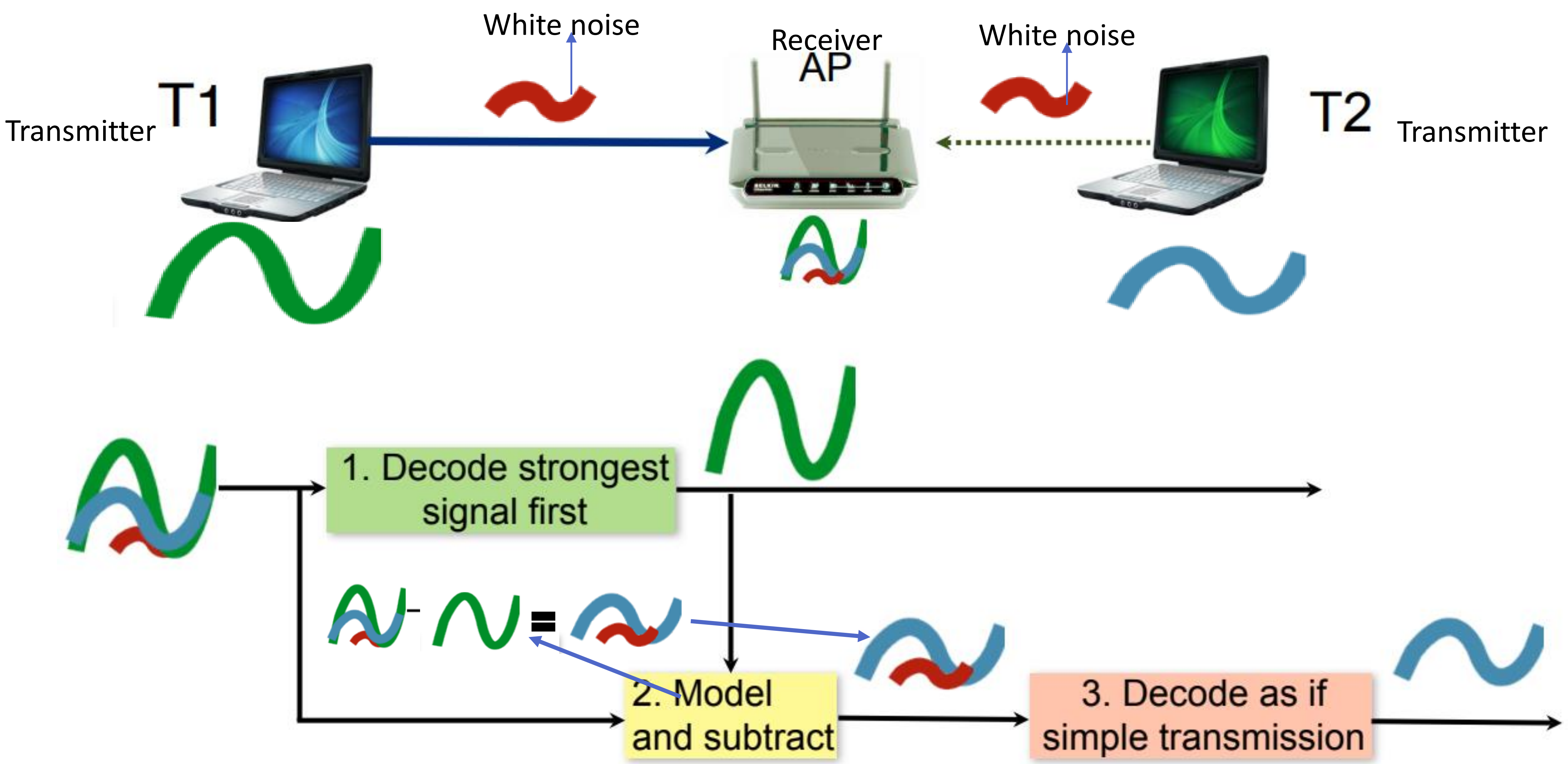
System Model of D2D with NOMA

- Assuming that each cluster operates in an orthogonal manner on a different RB
- The total number of users per cluster can range between 2 and $|\mathcal{N}| + |\mathcal{M}|$
- Let $\mathcal{S}$ be the set of clusters(i.e., each RB block is allocated to each cluster)
- Set $\mathcal{M}_k, \mathcal{N}_k$ represent the scheduled CUs and D2D users on the k-th cluster, respectively
- Note that all CUs and D2D users scheduled on a cluster utilize the same RB
- Thus, each NOMA-based CU needs to perform SIC[(*)] after receiving the superposed signals in order to demodulate its target message
- SIC is achieved by the receiver decoding the stronger signal first
- And, subtracting it from the combined signal and then decoding the difference as the weaker signal

(*)SIC: Successive Interference Cancellation

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

Successive Interference Cancellation



Patel, Pulin, and Jack Holtzman. "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system." IEEE journal on selected areas in communications 12.5 (1994): 796-807.

- Moreover, CUs treat the signals of D2D users as interference and are not decoded
- Each D2D user has to control its transmit power in order to maintain the interference to CUs in the same cluster under a predefined threshold
- Let β_m^k, α_n^k be the scheduling variables for the CU m and D2D user n, respectively
- $$\beta_m^k = \begin{cases} 1, & \text{if CU } m \text{ is scheduled into cluster } k, \\ 0, & \text{otherwise,} \end{cases} \tag{19}$$

  and,

  $$\alpha_n^k = \begin{cases} 1, & \text{if D2D } n \text{ is scheduled into cluster } k, \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

SINR Model for D2D Users

- The achievable rate for a D2D n scheduled on the k-th cluster is given as

$$R^k_{n,D2D} = \log_2\left(1 + \frac{|h_n|^2 P^d_n}{I^k_{n,CU} + I^k_{n,D2D} + z_n}\right) \quad (21)$$

- $I^k_{n,CU}$ is the interference that D2D user n experiences from the transmissions of all the CUs in k-cluster

$$I^k_{n,CU} = \sum_{m \in \mathcal{M}} \beta^k_m |h_{BS,n}|^2 P_m \quad (22)$$

- $I^k_{n,D2D}$ is the co-tier interference produced by the other scheduled D2D users on D2D user n in k-cluster

$$I^k_{n,D2D} = \sum_{n' \in \mathcal{N}/\{n\}} \alpha^k_n |h_{n',n}|^2 P^d_{n'} \quad (23)$$

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

SINR Model for Cellular Users

- The achievable rate for a CU m scheduled on the k-th cluster is given as

$$R^k_{m,CU} = \log_2\left(1 + \frac{|h_m|^2 P_m}{I^k_{m,CU} + I^k_{m,D2D} + z_m}\right) \quad (24)$$

- $I^k_{m,CU}$ is the interference that CU user m experiences from the transmissions of the other CUs in k-cluster

$$I^k_{m,CU} = \sum_{m' \in \mathcal{M} | \frac{|h_{m'}|^2}{z_{m'}} > \frac{|h_m|^2}{z_m}} \beta^k_{m'} |h_{BS,m}|^2 P_{m'} \quad (25)$$

- $I^k_{m,D2D}$ is the interference produced by the other scheduled D2D users on CU user m in k-cluster

$$I^k_{m,D2D} = \sum_{n \in \mathcal{N}} \alpha^k_n |h_{n,m}|^2 P^d_n \quad (26)$$

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

Problem Formulation

JUCPA :

$$\max_{\boldsymbol{\beta},\boldsymbol{\alpha},P} \sum_{k\in\mathcal{S}} \left(\sum_{m\in\mathcal{M}} \beta_m^k R_{m,CU}^k + \sum_{n\in\mathcal{N}} \alpha_n^k R_{n,D2D}^k \right)$$

Objective: maximize sum-rate CU&D2D

$$\text{s.t.: } C_1 : \sum_{k\in\mathcal{S}} \sum_{m\in\mathcal{M}} \beta_m^k P_m \le P_T,$$

Transmit power threshold/max. Tx. power

$$C_2 : \sum_{n\in\mathcal{N}} \sum_{m\in\mathcal{M}} \alpha_n^k |h_{n,m}|^2 P_n^d \le \theta^k, \quad \forall k \in \mathcal{S},$$

Interference threshold/QoS constraint

$$C_3 : \sum_{k\in\mathcal{S}} \beta_m^k = 1, \quad \forall m \in \mathcal{M},$$

a CU **belong to at most 1 cluster**

$$C_4 : \sum_{k\in\mathcal{S}} \alpha_n^k = 1, \quad \forall n \in \mathcal{N},$$

a D2D pair **belong to at most 1 cluster**

$$C_5 : \beta_m^k \in \{0,1\}, \alpha_n^k \in \{0,1\}, \quad \forall m \in \mathcal{M}, n \in \mathcal{N},$$

Binary decision variables of CU/D2D pair

$$C_6 : P_m \ge 0, P_n^d \ge 0, \quad \forall m \in \mathcal{M}, n \in \mathcal{N},$$

Transmit power variable CU&D2D

NP-Hard Problem

*JUCPA: Joint User Clustering And Power Assignment

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

Algorithm 1 Joint User Clustering and power assignment (JUCPA)

1: Solving the problem User Clustering (UC): Problem. **UC**

2: Solving the problem of power assignment for CU: Problem. $\mathbf{PA1_{CU}}$

3: Solving the problem of power assignment for D2D: Problem. $\mathbf{PA1_{D2D}}$

Data rate of CUs

Data rate of D2D

UC :

$$\max_{\boldsymbol{\beta},\boldsymbol{\alpha}} \sum_{k\in\mathcal{S}} \left(\sum_{m\in\mathcal{M}} \boxed{\beta_m^k R_{m,CU}^k} + \sum_{n\in\mathcal{N}} \boxed{\alpha_n^k R_{n,D2D}^k} \right)$$

$$\text{s.t.: } C_1 : \sum_{k\in\mathcal{S}} \sum_{m\in\mathcal{M}} \beta_m^k P_m \le P_T,$$

$$C_2 : \sum_{n\in\mathcal{N}} \sum_{m\in\mathcal{M}} \alpha_n^k |h_{n,m}|^2 P_n^d \le \theta^k, \quad \forall k \in \mathcal{S},$$

$$C_3 : \sum_{k\in\mathcal{S}} \beta_m^k = 1, \quad \forall m \in \mathcal{M},$$

$$C_4 : \sum_{k\in\mathcal{S}} \alpha_n^k = 1, \quad \forall n \in \mathcal{N},$$

$$C_5 : \underbrace{\beta_m^k \in \{0,1\}, \alpha_n^k \in \{0,1\}}, \quad \forall m \in \mathcal{M}, n \in \mathcal{N}.$$

Transmit power threshold/maximum Tx power

Interference threshold/QoS constraint

a CU **belongs to at most 1 cluster**

a D2D pair **belongs to at most 1 cluster**

Binary decision variables of CU/D2D pair

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

Successive Convex Approximation

$\mathbf{PA1_{CU}}$:

$$\max_{P} \sum_{k\in\mathcal{S}} \sum_{m\in\mathcal{M}_k} \log_2\left(1 + \frac{P_m|h_m|^2}{I_m^k + \boxed{\sum_{n\in\mathcal{N}_k} |h_{n,m}|^2 P_n^d} + z_m}\right)$$

$$\text{s.t.: } C_1 : \sum_{k\in\mathcal{S}} \sum_{m\in\mathcal{M}_k} P_m \le P_T,$$

$$C_2 : P_m \ge 0, \quad \forall m \in \mathcal{M}.$$

Treat D2D interference as noise

Total transmit power constraint

Tx. Power

Obtaining solution by Successive Convex Approximation (SCA) method

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

Successive Convex Approximation for PA1CU

Algorithm 2 Pseudo code of Successive Convex Approximation Method for Solving $\mathbf{PA1}_{\mathbf{CU}}$

Find a feasible power at $t = 0$ is P^0 in $\mathbf{PA1}_{\mathbf{CU}}$
Choose a step size $\gamma \in (0, 1]$, iteration $t = 0$,
repeat
    $t \leftarrow t + 1$

$$\boxed{\tilde{P}^t = \arg\max \left\{ \sum_{k \in \mathcal{S}} \sum_{n \in \mathcal{N}_k} \log_2 \left(1 + \frac{P_n^{t-1} |h_n|^2}{I_m^{k+} \sum_{n' \in \mathcal{N}_k / \{n\}} |h_{n',n}|^2 P_{n'}^{t-1} + \tilde{z}_n} \right) \right\}}$$

Approximate the original functions with some upper-bounds of them which have the same first order behavior

    $\boxed{P^t \leftarrow \gamma \tilde{P}^t + (1 - \gamma) P^{t-1}}$

until Some convergence criterion is met

Update the solution by taking trade-off between the pervious solution and the current approximation solution

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

KYUNG HEE UNIVERSITY

NETWORKING INTELLIGENCE LAB

Successive Convex Approximation

$$\mathbf{PA1_{D2D}}:$$

$$\max_{P} \sum_{k\in\mathcal{S}} \sum_{n\in\mathcal{N}_k} \log_2\left(1+\frac{P_n^d|h_n|^2}{\sum_{n'\in\mathcal{N}_k\setminus\{n\}} |h_{n',n}|^2 P_{n'}^d + \boxed{\tilde{z}_n}}\right)$$

$$\text{s.t.: } C_2: \sum_{n\in\mathcal{N}_k} \sum_{m\in\mathcal{M}_k} |h_{n,m}|^2 P_n^d \le \theta^k, \quad \forall k\in\mathcal{S},$$

$$C_5: P_n^d \ge 0, \quad \forall n\in\mathcal{N},$$

Treat interference of CU as noise

QoS constraint to protect CUs

D2D transmit power variable

Obtaining solution by Successive Convex Approximation (SCA) method

Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

Algorithm 3 Pseudo code of Successive Convex Approximation Method for Solving $\mathbf{PA1_{D2D}}$

1: Find a feasible power at $t = 0$ is P^0 in $\mathbf{PA1_{D2D}}$
2: Choose a step size $\gamma \in (0, 1]$, iteration $t = 0$,
3: **repeat**
4: $t \leftarrow t + 1$
5:

$$\boxed{\tilde{P}^t = \arg\max \left\{ \sum_{k \in \mathcal{S}} \sum_{n \in \mathcal{N}_k} \log_2 \left(1 + \frac{P_m^{t-1} |h_m|^2}{\sum_{n \in \mathcal{N}_k} |h_{n,m}|^2 P_m^{t-1} + z_m} \right) \right\}}$$

6: $\boxed{P^t \leftarrow \gamma \tilde{P}^t + (1 - \gamma) P^{t-1}}$
7: **until** Some convergence criterion is met

Approximate the original functions with some upper-bounds of them which have the same first order behavior

Update the solution by taking trade-off between the pervious solution and the current approximation solution

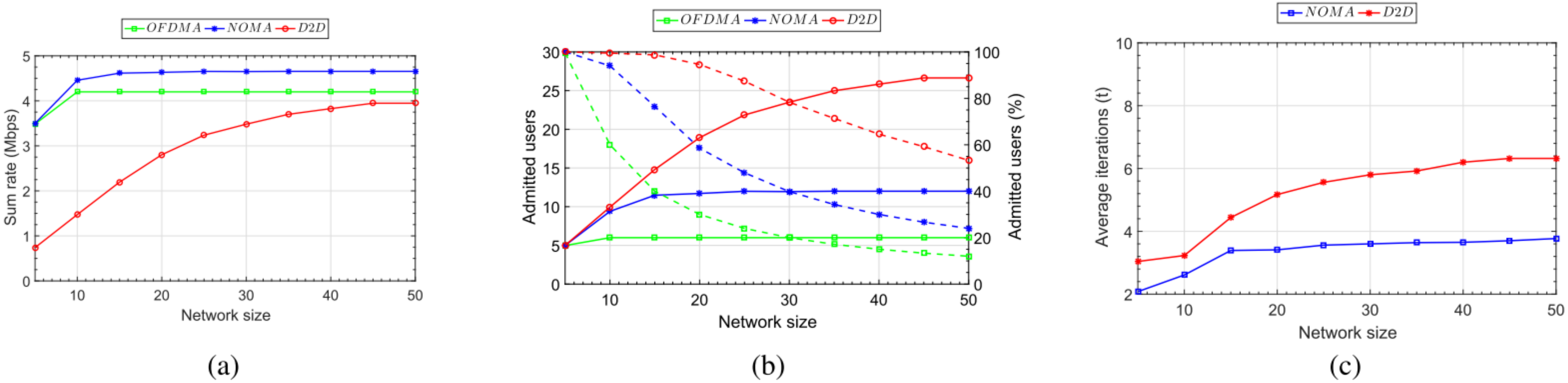
Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

Successive Convex Approximation

Simulation Parameters	Values
Radius of MBS	500 m
Carrier frequency (f)	2 GHz
Frame structure	Type 1 (FDD)
Transmission Time Interval (TTI)	1 ms
Total transmit power of BS	46 dBm
Total transmit power of D2D users	23 dBm
System bandwidth	1.4 MHz
Bandwidth of each RB (W)	180 kHz
Number of subcarriers per RB	12
Neighboring subcarrier spacing	15 kHz
Modulation and coding scheme (MCS) [43]	QPSK: $1/12, 1/9, 1/6, 1/3, 1/2, 3/5$ 16QAM: $1/3, 1/2, 3/5$
Path loss (cellular link)	$128.1 + 37.6\log(\mathrm{d})$, d[km]
Path loss (D2D links) [44]	$32.45 + 20\log(\mathrm{f}) + 20\log(\mathrm{d})$, f[MHz]
Shadow fading standard deviation [44]	3 dB
Proximity of D2D user ($R2$)	random $\{20 \sim 30\}$ m
Thermal noise for 1 Hz at 20 °C	-174 dBm

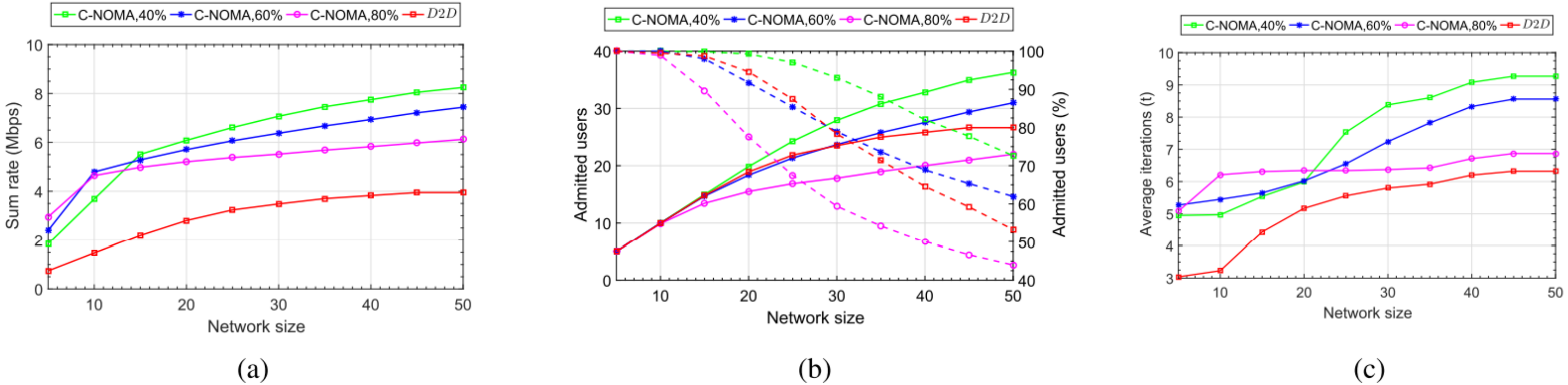
Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

1



Performance comparison of the OFDMA, NOMA and D2D schemes under various network sizes. (a) Average sum-rate. (b) Average number of admitted users. (c) Average number of iterations.

2



Kazmi, SM Ahsan, CS Hong et al. "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks." *IEEE Access* 6 (2018): 39860-39875.

D2D RRM: Deep Reinforcement Learning

D2D RRM: Multi Agent DRL

State

- The state observed by the D2D link D_i (agent i) for characterizing the environment consists of several parts
- The instant channel information of the D2D corresponding link $G_i^{d,t}$
- The channel information of the cellular link $G_i^{c,t}$
- The previous interference to the link I_i^{t-1}
- The RB selected by the D2D link in the previous time slot K_i^{t-1}
- The state if D2D user i can be represented as $s_i^t = \left[G_i^{d,t}, G_i^{c,t}, I_i^{t-1}, K_i^{t-1}\right]$

*RRM: Radio Resource Management
*DRL: Deep Reinforcement Learning

Li, Zheng, Caili Guo, and Yidi Xuan. "A multi-agent deep reinforcement learning based spectrum allocation framework for D2D communications." *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019.

D2D RRM: Multi Agent DRL

Action

- At each time t, the agent i takes an action $a_i^t \in \mathcal{A}$, which represents the agent selects a RB, according to the current state, $s_i^t \in \mathcal{S}$, based on the decision policy π_i
- The dimension of the action space is K if there are K RBs
- Note that the action selection of each agent should satisfy the constraint $\xi^c_{m,k} > \xi_{\min}$, where $\xi_{\min}$ is the SINR threshold of the CUE

Reward function

- The reward function relates to two parts: the D2D link rate, and the SINR constraints of CUE

$$r_i^t = \begin{cases} \log\left(1 + \boxed{\xi_i^{d,t}}\right), & \xi^c_{m,k} > \xi_{\min}, \\ r_{neg}, & \text{otherwise.} \end{cases}$$

Transmit power(P) / Channel gain(g)

$$\xi^d_{n,k} = \frac{\boxed{P^d g_n^{t,r}}}{\boxed{P^b g_n^{b,r}} + \boxed{\sum_{i \in \mathbf{D}_k, i \neq n} P^d g_{i,n}^{t,r}} + \boxed{\sigma^2}}$$

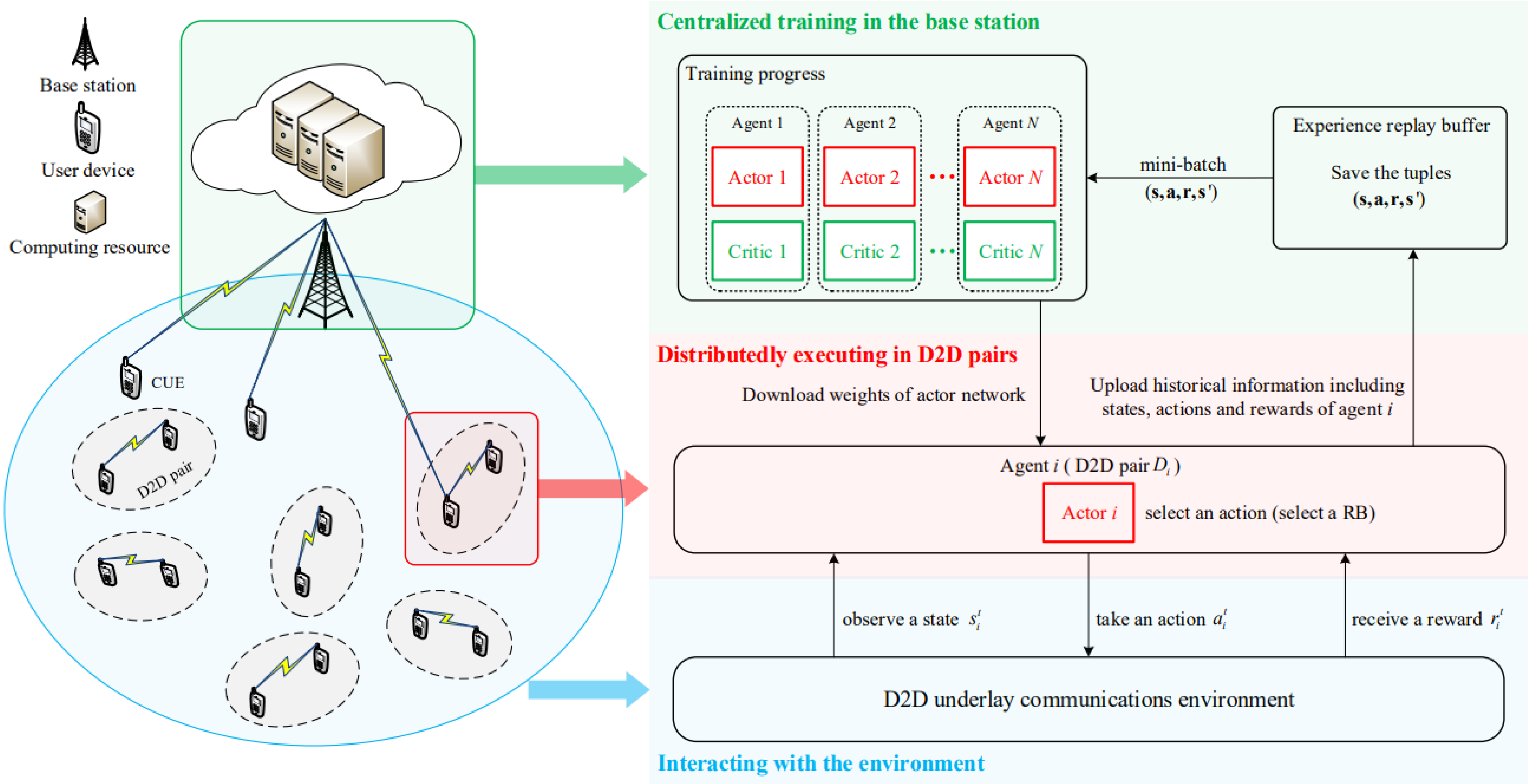
White noise

Interference from cellular user

Co-tier interference from D2D users

Li, Zheng, Caili Guo, and Yidi Xuan. "A multi-agent deep reinforcement learning based spectrum allocation framework for D2D communications." *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019.

D2D RRM: Multi Agent DRL

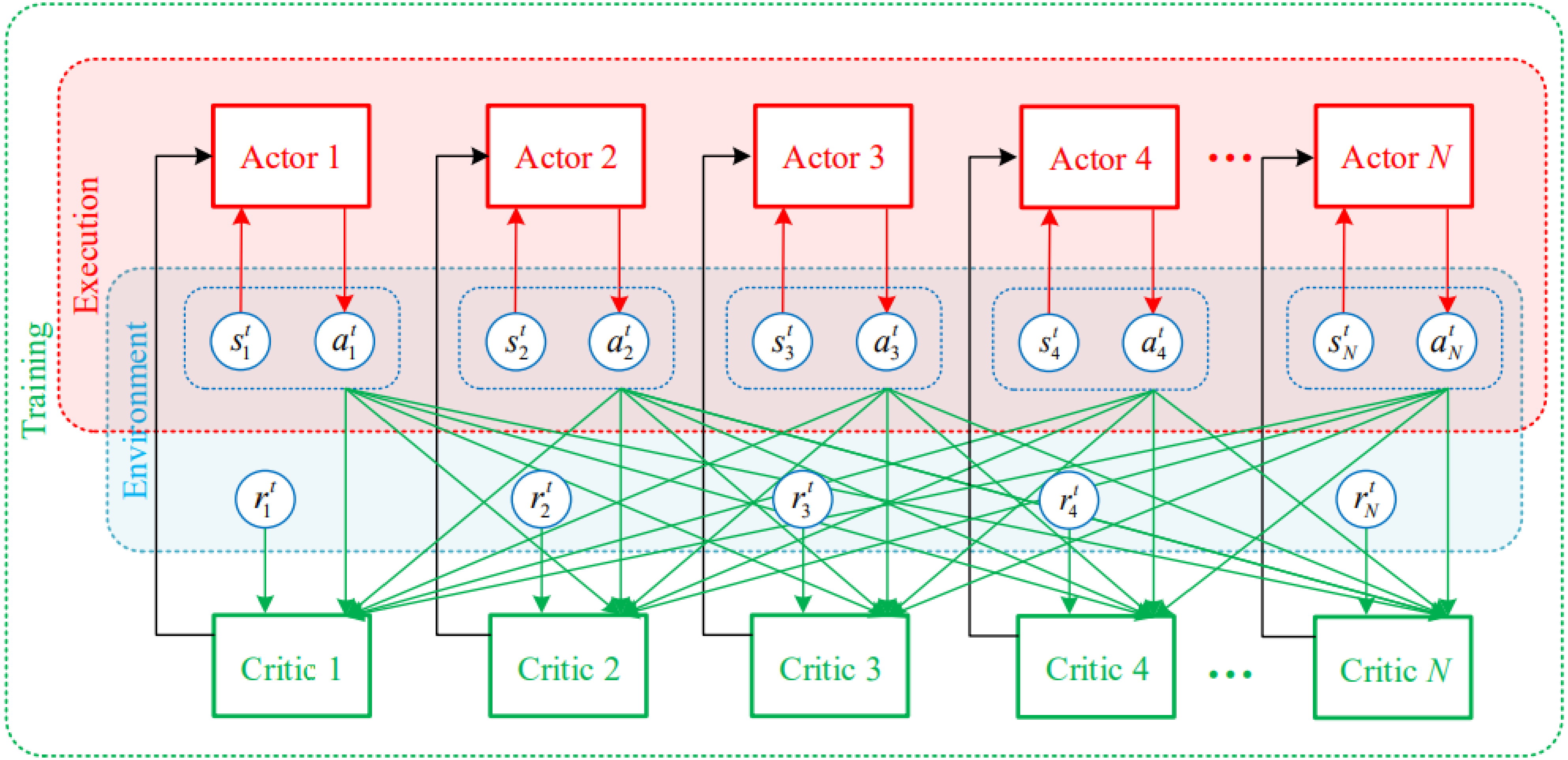


Li, Zheng, Caili Guo, and Yidi Xuan. "A multi-agent deep reinforcement learning based spectrum allocation framework for D2D communications." *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019.

KYUNG HEE UNIVERSITY

NETWORKING INTELLIGENCE LAB

D2D RRM: Multi Agent DRL



Li, Zheng, Caili Guo, and Yidi Xuan. "A multi-agent deep reinforcement learning based spectrum allocation framework for D2D communications." *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019.

Multi Agent DRL Algorithm

Algorithm 1 Training Algorithm

Input: Actor network structure, critic network structure.
Output: The weights of target actor network $\theta^{\mu'}$.
Training:

1: Random initialize actor network $\boldsymbol{\mu}$ and critic network $\mathbf{Q}$ with the weights θ^{μ} and $\theta^{\mathbf{Q}}$.
2: Initialize the target actor network $\boldsymbol{\mu}'$ and target critic network $\mathbf{Q}'$ with weights $\theta^{\mu'} \leftarrow \theta^{\mu}$, $\theta^{\mathbf{Q}'} \leftarrow \theta^{\mathbf{Q}}$.
3: Initialize the experience replay buffer $\mathcal{D}$
4: All agents (D2D pairs) receive initial observation states $\mathbf{s}^0 = \{s_1^0, ..., s_N^0\}$
5: **for** each time slot $t = 0, 1, ...T$ **do**
6: All agents select actions $\mathbf{a}^t = \{a_i^t = \mu_i(s_i^t), i \in \mathcal{N}\}$ according to the current policy.
7: All agents execute actions $\mathbf{a}^t$, observe rewards $\mathbf{r}^t = \{r_1^t, ..., r_N^t\}$ and observe new states $\mathbf{s}^{t+1}$.
8: Save the tuples $(\mathbf{s}^t, \mathbf{a}^t, \mathbf{r}^t, \mathbf{s}^{t+1})$ in $\mathcal{D}$
9: Sample a random mini-batch of tuples $(\mathbf{s}, \mathbf{a}, \mathbf{r}, \mathbf{s}')$ from the $\mathcal{D}$.
10: Set $y_i = r_i + \gamma Q_i'(\mathbf{s}', \boldsymbol{\mu}'(s'|\theta^{\mu'})|\theta_i^{Q'})$.
11: Update critic by minimizing the loss in equation (14).
12: Update the actor policy using the sampled policy gradient according to equation (17).
13: Update the target networks:
$\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}$,
$\theta^{\mathbf{Q}'} \leftarrow \tau\theta^{\mathbf{Q}} + (1-\tau)\theta^{\mathbf{Q}'}$.
14: **end for**

Return: $\theta^{\mu'}$

Initialization step

- The agent takes an action that depends on the current policy
- If there is no policy available yet, it would take a random action

- Update the reward according to gradients

$$Loss(\theta_i^Q) = \mathbb{E}_{\mathbf{s},\mathbf{a},r_i,\mathbf{s}'}[(Q_i(\mathbf{s},\mathbf{a}|\theta_i^Q) - y_i)^2], \qquad (14)$$

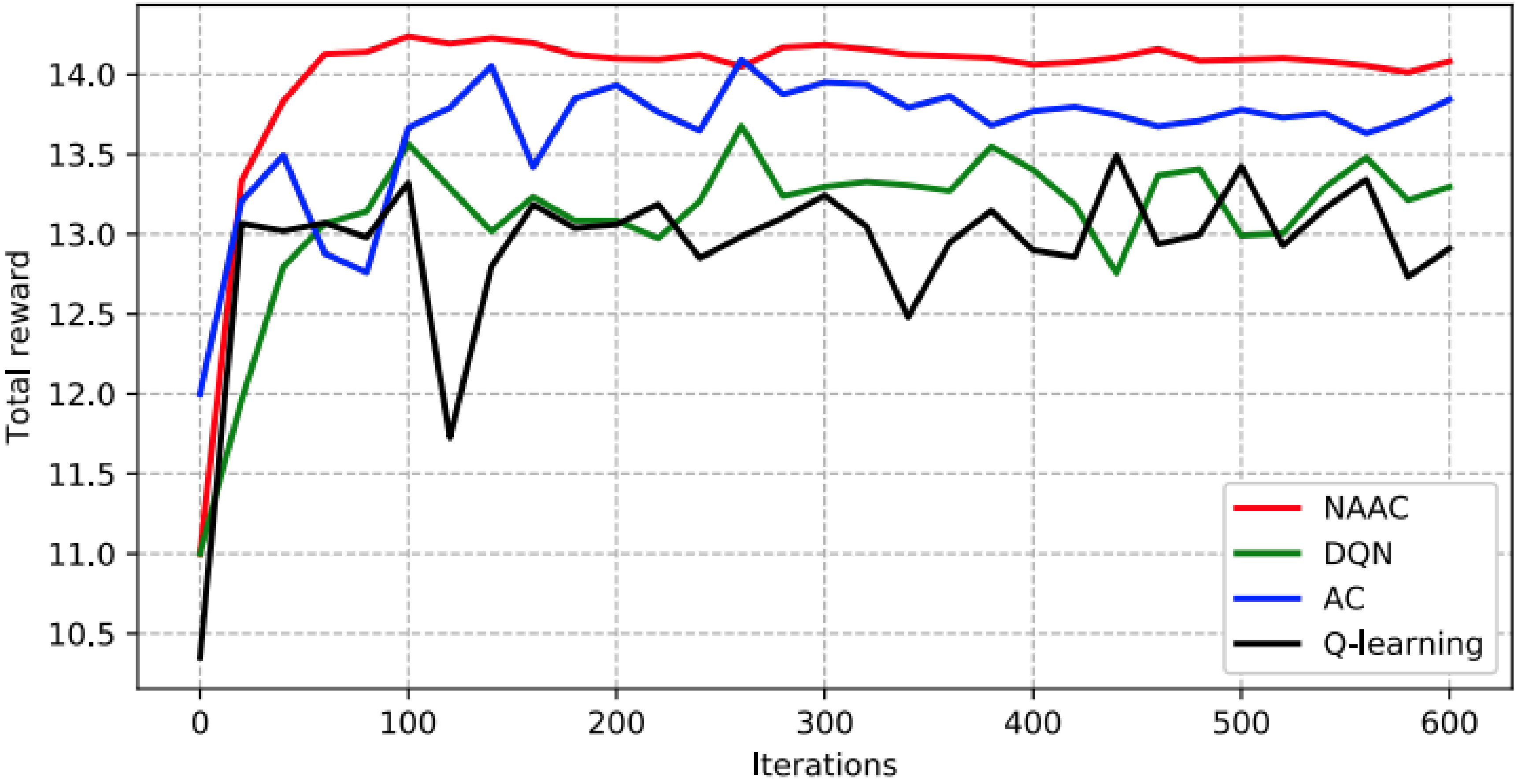
$$\begin{aligned}\nabla_{\theta_i^{\mu}} J_i &\approx \mathbb{E}_{\mathbf{s},\mathbf{a}\sim\mathcal{D}}[\nabla_{\theta_i^{\mu}} Q_i(\mathbf{s},\mathbf{a}|\theta_i^Q)|_{a_i=\mu(s_i|\theta_i^{\mu})}] \\ &= \mathbb{E}_{\mathbf{s},\mathbf{a}\sim\mathcal{D}}[\nabla_{a_i} Q_i(\mathbf{s},\mathbf{a}|\theta_i^Q)|_{a_i=\mu_i(s_i)}] \nabla_{\theta_i^{\mu}} \mu_i(s_i|\theta_i^{\mu}).\end{aligned} \qquad (17)$$

Li, Zheng, Caili Guo, and Yidi Xuan. "A multi-agent deep reinforcement learning based spectrum allocation framework for D2D communications." *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019.

D2D RRM: Multi Agent DRL

Parameter	**Value**
RB bandwidth	180 KHz
Number of CUEs	10
Number of RBs	10
BS transmission power (P^b)	46 dBm
D2D transmission power (P^d)	13 dBm
Cellular link pathloss	$128.1 + 37.6 \log_{10}(d[km])$
D2D link path loss exponent	4
UE thermal noise density	-174 dBm/Hz
CUE target SINR threshold (ξ^c_{min})	0 dB
Negative reward (r_{neg})	-1

Li, Zheng, Caili Guo, and Yidi Xuan. "A multi-agent deep reinforcement learning based spectrum allocation framework for D2D communications." *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019.



Comparison of total reward performance during training process

*NAAC: Neighbor-Agent Actor Critic
*DQN: Deep Q Network
*AC: Actor Critic

Li, Zheng, Caili Guo, and Yidi Xuan. "A multi-agent deep reinforcement learning based spectrum allocation framework for D2D communications." *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019.

Challenges And Ongoing Research

Challenges And Ongoing Research

- D2D for Augmented Reality, Virtual Reality, Mixed Reality
  - High bandwidth, Low-latency
  - Power control for dense network setting
  - Improve user experiences
- DRL is required global information such as channel state information(CSI) of the other D2D as well as cellular users
  - Reduce message overhead
  - Improve accuracy with local information such as CSI of D2D itself

Thanks for your attention!