Resource Allocation for Enabled-Network-Slicing in Cooperative NOMA-Based Systems with Underlay D2D Communications

Asmaa Amer, Sahar Hoteit, Jalel Ben Othman

Resource Allocation for Enabled-Network-Slicing in Cooperative NOMA-based Systems with Underlay D2D Communications

Asmaa Amer*, Sahar Hoteit*, Jalel Ben Othman*†
*Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France
†Université Sorbonne Paris Nord, Paris, France
Emails: {asmaa.amer, sahar.hoteit, jalel.benothman}@centralesupelec.fr

***Abstract*—Non-orthogonal multiple access technique (NOMA) has appeared at the forefront as a viable solution capable of improving spectral and energy efficiency in fifth-generation (5G) and beyond-5G networks. This study aims at evaluating the benefits of adopting network slicing in cooperative NOMA-based systems with underlay Device-to-Device (D2D) communications. We formulate an optimization problem that maximizes the overall system's throughput while guaranteeing slices' technical requirements. We decouple the problem into two sub-problems: first, assigning the cellular users to resource blocks allocated to each NOMA group and then assigning D2D pairs to NOMA groups. A two-stage resource allocation solution by swapping-based matching theory is implemented. Numerical results show that the proposed scenario outperforms other ones in terms of the overall system's throughput and the number of admitted D2D pairs.**

***Index Terms*—Cooperative NOMA, Network Slicing, Matching Theory**

I. Introduction

The rapid expansion of communication systems is revolutionizing today's world. As a result, new applications emerge with various heterogeneous service types and technical requirements in terms of data rate, latency, and energy consumption. The one-size-fits-all architecture of the previous generation networks can not fully support this heterogeneity. Fifth-generation (5G) and beyond-5G (B5G) networks are envisioned to support this heterogeneity using network slicing to provide customized services satisfaction. These services are classified into three types: enhanced mobile broadband (eMBB), which refers to services requiring high data rates, ultra-reliable and low latency communications (URLLC) which refers to mission-critical applications with low latency, and massive machine-type communications (mMTC), which stands for massively connected and energy-constrained services [1]. The primary idea behind network slicing is to divide the network into several flexible logical slices. Each is designed to provide one of the aforementioned services while sharing a common physical infrastructure. However, introducing network slicing faces numerous challenges in terms of efficient and fair resource allocation and slice isolation. We note that slice isolation refers to having separate slices so that any change in one does not result in the required services of another slice being unsatisfied [1]. Slice isolation can be achieved by either assuming a fixed subset of resources for each slice to which its users have restricted access, or by defining requirements for each slice to be guaranteed without restricting access to resources [2].

Lately, the concept of non-orthogonal multiple access (NOMA) has appeared at the forefront as a viable solution capable of improving spectral and energy efficiency in B5G networks [3]. Unlike orthogonal multiple access (OMA) schemes used in the previous generations of mobile cellular networks, NOMA serves multiple users using the same resource blocks (RBs). This can be done by superposition coding (SC) at the transmitter level and successive interference cancellation (SIC) at the receiver level. In other words, in the downlink, the base station (BS) transmits a combined signal, which is a superposition of multiple users' signals with different assigned power coefficients. Then, each user executes the SIC process until its signal is decoded. Furthermore, in some scenarios, users can act as relays to help one another, extend coverage, and improve reception reliability; these scenarios are known as cooperative NOMA (CNOMA) [4].

Since B5G systems aim to more support network heterogeneity and diversity, Device-to-Device (D2D) communications are much used in these systems. They enable direct communication between close devices without passing the base station (BS) [5]. Therefore, they serve the demands of proximity-based services and improve throughput and latency. To avoid the uncontrolled nature of the unlicensed spectrum and for better spectral efficiency at the cost of interference [5], the majority consider the in-band underlay D2D, i.e., share licensed RBs of cellular users (CUs) [6, 7].

In this paper, the benefits of adopting network slicing in cooperative NOMA-based systems with underlay D2D communications are evaluated. To the best of our knowledge, this is the first that has studied all of the aforementioned concepts. In particular, we first enable the grouping of CUs having similar requirements in the same slice. Then, within each slice, we propose a two-step matching-theory-based algorithm to determine the CUs that share the same RBs (referred to as a NOMA group) and the underlay D2D pairs of each NOMA group. The performance of the proposed scenario is analyzed and compared to the case where network slicing is not adopted in terms of the overall system throughput and the number of admitted D2D pairs in the system.

The rest of the paper is structured as follows. Section II presents the state-of-the-art. Section III defines the system model. The problem formulation and the proposed solution are presented in Sections IV and V, respectively. Section VI presents the numerical results. Finally, Section VII concludes our work and gives some perspectives.

II. Related Works

Authors in [6, 7] study resource allocation in NOMA-based systems with D2D communications. However, they do not consider network slicing nor the existence of heterogeneous technical requirements, but this study does. Authors in [8–10] investigate resource allocation for network slicing-enabled NOMA-based system with mobile edge computing (MEC). In [8, 9], users are grouped into slices based on their latency requirements. They use NOMA scheme to offload their computing tasks to the MEC server at the base station, due to its higher computational capability. The goal in [8] is to minimize energy consumption, while [9] deals with a trade-off between reducing energy consumption and minimizing latency using a decision scheme between local computing and offloading. Authors in [10] study network slicing with MEC in vehicular systems. They use deep reinforcement learning for vehicle's slice and coverage range selection, and for power allocation. Authors in [11] combine network slicing and NOMA to send medical data in a mobile hospital system. They assume two eMBB and URLLC slices; they implement power optimization to maximize the system throughput. However, these works have neither studied cooperative NOMA nor the underlay of D2D communications. Thus, they do not exploit the redundant information existence at the strong users due to SIC to improve the weaker users' performance, nor consider the system's heterogeneity with the existence of proximity-based D2D communications, which is considered in this study.

According to the review of the literature, the research about the network slicing synergy in cooperative NOMA-based systems with underlay D2D communications, specifically for resource allocation, is still in its early stages. Our major work contributions in this paper are resumed as follows:

- A CNOMA-based cellular network with underlay D2D communications is proposed. Two slices coexist, eMBB and URLLC; their technical requirements are data rate and latency, respectively.
- An efficient radio resource allocation algorithm based on swapping-based matching theory is defined. Our algorithm determines which cellular users are allocated to the same RBs (i.e., to form a NOMA group), and which D2D pairs share the RBs of each NOMA group.
- We aim to maximize the overall system throughput and to increase the number of admitted D2D pairs while guarantying slices' technical requirements.

Our proposed system model is detailed in Section III.

III. System Model

Consider the downlink transmission scenario of a cellular network with one infrastructure provider represented by BS and a set of radio resource blocks RBs $\mathcal{R} = \{r|1 \leq r \leq |\mathcal{R}|\}$, where each RB has a bandwidth B. The BS serves a set of cellular users $\mathcal{U}= \{i|i = 1, ..., |\mathcal{U}|\}$. Denote by $\mathcal{D}= \{d|d = 1, ..., |\mathcal{D}|\}$ the set of underlay D2D pairs. The network includes a set of slices $\mathcal{S} = \{s|1 \leq s \leq |\mathcal{S}|\}$ where each slice s requires a different service. Several cellular users and D2D pairs are associated with each slice. They are denoted by $\mathcal{U}_{\mathcal{S}}= \{\mathcal{U}_s|1 \leq s \leq |\mathcal{S}|\}$ and $\mathcal{D}_{\mathcal{S}}= \{\mathcal{D}_s|1 \leq s \leq |\mathcal{S}|\}$, where $\mathcal{U}_s$ and $\mathcal{D}_s$ are the sets of cellular users and D2D pairs associated to slice $s \in \mathcal{S}$, respectively. Within each slice s, the system's scenario is detailed as follows:

- Cellular users $\mathcal{U}_s$ are divided into different NOMA groups. In each group, they share the same RBs. As in [4], we assume that each group is formed by only two cellular users (to reduce SIC error propagation), a weak user with a small channel gain, and a strong user with a higher channel gain. Denote by $\mathcal{N}_s = \{k|k = 1, ..., K\}$, the set of NOMA groups for slice $s \in \mathcal{S}$ and by $\mathcal{U}_s^{st}$ and $\mathcal{U}_s^{we}$, the sets of strong and weak cellular users, respectively.
- We consider that, at most, q underlay D2D pairs share the RBs with the cellular users in each NOMA group.
- Within each NOMA-group k for slice $s \in \mathcal{S}$, we consider a cooperative NOMA scenario where the strong cellular user acts as an in-band full duplex relay to assist the weak user. Due to the imperfect self-interference (SI) cancellation techniques [12], the strong user will still be affected by a specific SI level quantified by $\rho \in [0, 1]$.[1]

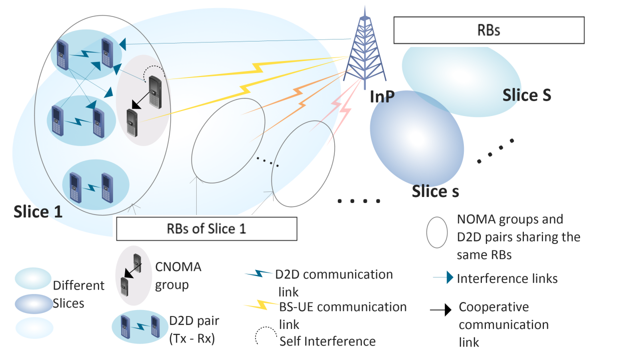

Fig. 1: Multi-slices CNOMA-based cellular network with underlay D2D communications

The proposed system scenario is depicted in Fig. 1. For the sake of clarity, not all interference links are shown. Only those from and to one D2D pair are shown.

In the following equations, denote by h_*: the channel gain coefficient between BS and the cellular user, or that between transmitter and receiver of the D2D pair; and by $h_{*,*'}$: the channel coefficient between any transmitter $*$ and receiver $*'$.

Within a NOMA group k of each slice $s \in \mathcal{S}$, the received signal by the strong cellular user, $u \in \mathcal{U}_s^{st}$ is given by:

$$y_u = \underbrace{h_u(\sqrt{P_u}x_u + \sum\nolimits_{v\in\mathcal{U}_s^{we}} \eta_v^k \sqrt{P_v}x_v)}_{\text{Superposed signal from BS}} + \underbrace{n}_{\text{noise}} + \underbrace{\sqrt{\rho}h_{u,u}\sqrt{P_r}x_r}_{\text{Self Interference}} + \underbrace{\sum\nolimits_{d\in\mathcal{D}_s} \eta_d^k \sqrt{P_d}h_{d,u}x_d}_{\text{Interference from D2D pairs}} . \quad (1)$$

Equation (1) includes the superposed signal from BS, noise, SI, and interference from D2D pairs sharing the same RBs, respectively. x_u, P_u, x_v, and P_v denote the intended messages and allocated power for the strong user $u \in \mathcal{U}_s^{st}$ and weak user $v \in \mathcal{U}_s^{we}$ of each NOMA group k, respectively. x_r and P_r are the decoded and relayed message from u to v and its relaying power, respectively. The binary variable η_i^k indicates if the cellular user $i \in \mathcal{U}_s$ belongs to the NOMA group $k \in \mathcal{N}_s$, and η_d^k indicates if D2D pair d shares RBs of k. x_d and P_d denote the transmit signal and power of the D2D pair, respectively.

[1] ρ indicates the ratio of the residual SI level after SI cancellation; a $\rho = 0$ indicates perfect SI cancellation, and $\rho = 1$ indicates no SI cancellation at all.

User $u \in \mathcal{U}_s^{st}$ applies SIC and first decodes x_v. The received signal to interference plus noise ratio (SINR) at u to decode x_v is given by:

$$\gamma_{u,v} = \frac{|h_u|^2 P_v}{|h_u|^2 P_u + \sum_{d\in\mathcal{D}_s} \eta_d^k |h_{d,u}|^2 P_d + \rho |h_{u,u}|^2 P_r + \sigma^2 + \sigma_c^2}. \quad (2)$$

$\sigma^2 = BN_0$ and σ_c^2 refer to the additive white Gaussian noise (AWGN) and circuit hardware noise power, respectively. N_0 is the AWGN power spectral density. Then, it decodes x_u. The received SINR at u to decode x_u is given by:

$$\gamma_{u,u} = \frac{|h_u|^2 P_u}{\sum_{d\in\mathcal{D}_s} \eta_d^k |h_{d,u}|^2 P_d + \rho |h_{u,u}|^2 P_r + \sigma^2 + \sigma_c^2}. \quad (3)$$

For the weak user $v \in \mathcal{U}_s^{we}$ in NOMA group k, it receives its data from two sides: directly from BS and from the strong user u acting as a relay in the same NOMA group k. It is affected by D2D interference, so its received signal y_v , received SINR $\gamma_{v,v}$ to decode the data transmitted by BS, and received SINR $\gamma_{v,u}$ to decode data forwarded by u are respectively given by:

$$\begin{aligned} y_v =& \underbrace{h_v(\sqrt{P_v}x_v + \sum_{u\in\mathcal{U}_s^{st}} \eta_u^k \sqrt{P_u} x_u)}_{\text{Superposed signal from BS}} + n \\ &+ \underbrace{h_{u,v}\sqrt{P_r}x_r}_{\text{Relayed signal from strong user}} + \underbrace{\sum_{d\in\mathcal{D}_s} \eta_d^k \sqrt{P_d} h_{d,v} x_d}_{\text{Interference from D2D pairs}}, \end{aligned} \quad (4)$$

$$\gamma_{v,v} = \frac{|h_v|^2 P_v}{|h_v|^2 \sum_{u\in\mathcal{U}_s^{st}} \eta_u^k P_u + \sum_{d\in\mathcal{D}_s} \eta_d^k |h_{d,v}|^2 P_d + \sigma^2 + \sigma_c^2}, \quad (5)$$

$$\gamma_{v,u} = \frac{|h_{u,v}|^2 P_r}{\sum_{d\in\mathcal{D}_s} \eta_d^k |h_{d,v}|^2 P_d + \sigma^2 + \sigma_c^2}. \quad (6)$$

User v combines the signals received from BS and the strong user u using maximal ratio combining technique[13], so its received SINR is $\gamma_{MRC} = \gamma_{v,u} + \gamma_{v,v}$.

Within the same NOMA group k, the data rates of a strong user u and a weak user v can be given, respectively, by:

$$R_u = B \sum_{r\in\mathcal{R}} \lambda_r^k \log_2\left(1 + \gamma_{u,u}\right), \quad (7)$$

$$R_v = B \sum_{r\in\mathcal{R}} \lambda_r^k min(\log_2(1+\gamma_{MRC}), \log_2(1+\gamma_{u,v})), \quad (8)$$

where λ_r^k is the binary variable that indicates if RB $r \in \mathcal{R}$ is allocated to NOMA group k or not. Furthermore, for a D2D receiver d sharing RBs of k, its received signal y_d, received SINR γ_d, and data rate R_d are respectively given in (9), (10), and (11):

$$\begin{aligned} y_d =& \underbrace{\sqrt{P_d} h_d x_d}_{\text{D2D transmitter signal}} + \underbrace{h_{BS,d}(\sqrt{P_u}x_u + \sqrt{P_v}x_v)}_{\text{Interference from BS}} + n \\ &+ \underbrace{h_{u,d}\sqrt{P_r}x_r}_{\text{Interference from the strong user}} + \underbrace{\sum_{d'\in\mathcal{D}_s\setminus\{d\}} \eta_{d'}^k \sqrt{P_{d'}} h_{d',d} x_{d'}}_{\text{Interference from other D2D pairs}}, \end{aligned} \quad (9)$$

$$\gamma_d = \frac{|h_d|^2 P_d}{|h_{BS,d}|^2 P + \sum_{d'\in\mathcal{D}_s\setminus\{d\}} \eta_{d'}^k |h_{d',d}|^2 P_{d'} + |h_{u,d}|^2 P_r + \sigma^2 + \sigma_c^2} \quad (10)$$

$$R_d = B \sum_{r\in\mathcal{R}} \lambda_r^k log_2(1+\gamma_d), \quad (11)$$

where $P{=}P_u{+}P_v$. Note that the D2D receiver d receives the transmit signal from its D2D transmitter, and is affected by interference from the BS, the strong user u and the other D2D pairs sharing the RBs allocated to NOMA group k.

In the next section, the optimization problem is formulated.

IV. Problem Formulation

In our system, we aim to maximize the overall system throughput and satisfy slices' technical requirements. In this section, the slices' technical requirements and the total system throughput are defined. Then, the optimization problem is formulated.

For the eMBB slice, the throughput of cellular users and D2D pairs must exceed a minimum threshold. So, the technical requirement of this slice is expressed by :

$$R_j \geq \begin{cases} R_{CUs}^{min}, & \forall j \in \mathcal{U} \\ R_d^{min}, & \forall j \in \mathcal{D}, \end{cases} \quad (12)$$

For the URLLC slice, the latency of cellular users and D2D pairs must not exceed a given threshold. Denote by X_j the data size. So, the technical requirement of this slice is:

$$T_j = \frac{X_j}{R_j} \leq \begin{cases} T_{CUs}^{max}, & \forall j \in \mathcal{U} \\ T_d^{max}, & \forall j \in \mathcal{D}. \end{cases} \quad (13)$$

The overall system throughput R_{sum} is given by:

$$R_{sum} = \sum_{s\in\mathcal{S}} \sum_{k\in\mathcal{N}_s} \left(\sum_{u\in\mathcal{U}_s^{st}} \eta_u^k R_u + \sum_{v\in\mathcal{U}_s^{we}} \eta_v^k R_v + \sum_{d\in\mathcal{D}_s} \eta_d^k R_d \right) \quad (14)$$

Intuitively, the two objectives of maximizing R_{sum} and satisfying each slice's technical requirement depend on:

- Intra-NOMA group interference (inferred by $\eta_i^k, k \in \mathcal{N}_s$, $i \in \mathcal{U}$),
- Interference between the D2D pairs and the cellular users of a NOMA group (inferred by η_d^k, $d \in \mathcal{D}$).

Our optimization problem can be formulated as follows:

$$\begin{aligned} \textbf{P1:} \max_{\eta} \quad & R_{sum}(\eta) \\ \text{s.t.} \quad & \mathbf{C}_1: \eta_d^k, \eta_i^k \in \{0,1\}, \quad \forall i \in \mathcal{U}; d \in \mathcal{D}; k \in \mathcal{N}_s; s \in \mathcal{S}, \\ & \mathbf{C}_2: \sum_{k\in\mathcal{N}_s} \eta_d^k \leq 1, \quad \forall d \in \mathcal{D}; s \in \mathcal{S}, \\ & \mathbf{C}_3: \sum_{k\in\mathcal{N}_s} \eta_i^k = 1, \quad \forall i \in \mathcal{U}; s \in \mathcal{S}, \\ & \mathbf{C}_4: \sum_{d\in\mathcal{D}} \eta_d^k \leq q, \quad \forall k \in \mathcal{N}_s; s \in \mathcal{S}, \\ & \mathbf{C}_5: \sum_{i\in\mathcal{U}} \eta_i^k = 2, \quad \forall k \in \mathcal{N}_s; s \in \mathcal{S}, \\ & \mathbf{C}_6: \sum_{u\in\mathcal{U}_s^{st}} \eta_u^k = 1, \sum_{v\in\mathcal{U}_s^{we}} \eta_v^k = 1, \forall k \in \mathcal{N}_s; s \in \mathcal{S}, \\ & \mathbf{C}_7: \sum_{s\in\mathcal{S}} \sum_{k\in\mathcal{N}_s} \sum_{r\in\mathcal{R}} \lambda_r^k \leq \mid \mathcal{R} \mid, \\ & \mathbf{C}_8: \gamma_{u,v} \geq \gamma_{v,v}, \\ & \mathbf{C}_9: (12), (13). \end{aligned} \quad (15)$$

Constraint $\mathbf{C}_1$ ensures binary values to indicate whether the CU i and D2D pair d are allocated the RBs of NOMA group k or not. Constraints $\mathbf{C}_2$ and $\mathbf{C}_3$ ensure that each D2D pair d and CU i are associated to only one NOMA group k. Constraint $\mathbf{C}_4$ guarantees that at most q D2D pairs are associated to each NOMA group k. Constraints $\mathbf{C}_5$ and $\mathbf{C}_6$ guarantee that there are only 2 CUs in each NOMA group k: one strong CU $u \in \mathcal{U}_s^{st}$ and one weak CU $v \in \mathcal{U}_s^{we}$. Constraint $\mathbf{C}_7$ ensures that the number of allocated RBs to the NOMA groups in all slices does not exceed the total number of RBs. Constraint $\mathbf{C}_8$

ensures the SIC success within each NOMA group k. Finally, $\mathbf{C}_9$ ensures the satisfaction of technical requirements of each slice $s \in \mathcal{S}$. Problem **P1** is a non-convex problem introducing computational complexity to reach an optimal solution. So, we decouple it into two sub-problems: i) NOMA-grouping of CUs, ii) D2D-NOMA groups matching. In each sub-problem, we consider the slices' technical requirements and the total system throughput. We solve them using matching theory [14] as explained in Section V.

V. Proposed Solution

In this section, we present the proposed solution to the two sub-problems presented above in two stages: 1) Assigning CUs to RBs of NOMA groups ($\eta_i^k, i \in \mathcal{U}, k \in \mathcal{N}_s$); 2) Assigning D2D pairs to RBs of NOMA groups ($\eta_d^k, d \in \mathcal{D}, k \in \mathcal{N}_s$), as detailed in Algorithms 1 and 2, respectively. They are solved using swapping-based matching theory.

In the first sub-problem, $\eta_d^k, k \in \mathcal{N}_s$, is randomly fixed, while CUs and NOMA groups are the two player sets to be matched. Each CU is matched to one NOMA group, and each NOMA group is matched to two CUs. This forms a many-to-one matching ψ_g with externalities (interference) [14].

Definition 1. *A many-to-one matching ψ_g is defined over the set $\mathcal{U}_s \cup \mathcal{N}_s$ and is characterized by:*

- $|\psi_g(i)| = 1, \forall i \in \mathcal{U}_s$ *and* $\psi_g(i) \in \mathcal{N}_s$;
- $|\psi_g(k)| = 2, \forall k \in \mathcal{N}_s$ *and* $\psi_g(k) \subset \mathcal{U}_s$;
- $\psi_g(i) = k$ *if* $i \in \psi_g(k)$;
- $\psi_g(k) = \{i \in \mathcal{U}_s : \psi_g(i) = k\}$,

where $|.|$ denotes the cardinality. Each player in a set has different preferences for the players of the opposite set. This preference is determined by its satisfaction when matched with players from the opposite set. Utility functions can be used to quantify these preferences.

For any CU i, its preference on a NOMA group k is quantified by the utility function $U_i(k)$. And for a NOMA group k, its preference on a set of CUs (i' and i'') is quantified in $U_k(i', i'')$. They are formulated as below:

$$U_i(k) = b_i R_i \tag{16}$$

$$U_k(i', i'') = b_{i'} b_{i''} \left(R_{i'} + R_{i''} + \sum\nolimits_{d \in \mathcal{D}_{s,k}^*} R_d \right), \tag{17}$$

where R_i, $R_{i'}$, and $R_{i''}$ in (16) and (17) are related to (7) and (8) depending on whether it is strong or weak CU. R_d relates to (11), and $\mathcal{D}_{s,k}^*$ denotes D2D pairs that are initially matched to share RBs of NOMA group k (randomly fixed η_d^k). b_i is a binary variable that indicates if the technical requirement of the CU $i \in \mathcal{U}_s$ is satisfied or not; it is related to (12) and (13).

For the swapping iteration, two CUs, i and i', matched to different NOMA groups k and k', respectively, swap their matched groups. This changes the matching ψ_g (where $k = \psi_g(i), k' = \psi_g(i')$) to $\psi_g^{i',i}$ (where $k = \psi_g(i'), k' = \psi_g(i)$). If they can swap, they are called a swap-blocking pair. The swap-blocking pair is defined over these Pareto conditions:

Definition 2. *(i, i') is a swap-blocking pair if and only if*

- $\forall z \in \{i, i', k, k'\}$, $U_z(\psi_g^{i',i}) \geq U_z(\psi_g)$
- $\exists z \in \{i, i', k, k'\}$ *such that* $U_z(\psi_g^{i',i}) > U_z(\psi_g)$,

where $U_z(\psi_g)$ and $U_z(\psi_g^{i',i})$ are the utilities of element z in the case of matching ψ_g and $\psi_g^{i',i}$, respectively.

Algorithm 1 shows how NOMA-grouping in each slice s is implemented using swapping-based matching theory. It is initialized with a fixed η_d^k. Then, it runs swapping between CUs. Swapping is approved if it increases the total throughput and satisfies Pareto conditions, the slice technical requirements, and other **P1** constraints. Algorithm 1 runs until no swap-blocking pairs exist, achieving Pareto efficiency and thus reaching the final stable matching ψ_g^{final}.

Algorithm 1: NOMA-grouping algorithm.

```
Input  : U_s, D_s, D*_{s,k}, λ_k, ψ_g = φ, ∀k ∈ N_s, ∀s ∈ S
foreach s ∈ S do
    For initial matching ψ_g^0, randomly pair CUs to
      form NOMA groups; ψ_g ← ψ_g^0
end
foreach s ∈ S do
    repeat
        foreach i ∈ U_s do
            foreach i' ∈ U_s do
                if (i, i') satisfies Pareto conditions and
                  constraints of P1 then
                    i and i' swap their matches,
                    ψ_g ← ψ_g^{i',i}
                else
                    i and i' do not swap; ψ_g ← ψ_g
                end
            end
        end
    until Pareto efficiency is achieved.
end
Output : η_i^k, ψ_g^{final}
```

The second sub-problem aims to find the D2D pairs that share the RBs of each NOMA group. So, D2D pairs and NOMA groups are the two players' sets. Each D2D pair d can be matched to only one NOMA group k, and each NOMA group k can be matched to a subset of D2D pairs $D' \subset \mathcal{D}_s$ so that $|D'| \leq q$. A similar swapping-based matching algorithm is implemented. What differs is that some D2D pairs might not be admitted into the system (i.e., they are not associated to any NOMA group). The utilities of d on k and that of k on D' are formulated, respectively, as:

$$U_d(k) = b_d R_d \tag{18}$$

$$U_k(D') = b_u b_v \left(R_u + R_v + \sum\nolimits_{d' \in \mathcal{D}'} R_{d'} \right) \prod\nolimits_{d' \in \mathcal{D}'} b_d, \tag{19}$$

where b_u and b_v refer to the binary variables that indicate the satisfaction of the technical requirements of the strong user u and weak user v that are already associated to NOMA group k as an output of Algorithm 1; and b_d refers to the binary variable indicating the technical requirement satisfaction of D2D pair d. Similar to Algorithm 1, in each slice $s \in \mathcal{S}$, we aim a stable matching ψ_d with no swap-blocking pairs (d, d') where d and $d' \in \mathcal{D}_s$. Algorithm 2 takes as an input the output of Algorithm 1, it iterates until no swap-blocking pairs exist, and outputs ψ_d^{final}.

Algorithm 2: NOMA groups-D2D pairs matching algorithm.

Input : $\mathcal{N}_s$, $\mathcal{D}_s$, λ_k, ψ_d, ψ_g^{final}, $\forall k \in \mathcal{N}_s$, $\forall s \in \mathcal{S}$
foreach $s \in \mathcal{S}$ **do**
  repeat
    foreach $d \in \mathcal{D}_s$ **do**
      foreach $d' \in \mathcal{D}_s$ **do**
        if *(d, d′) satisfies Pareto conditions and constraints of* ***P1*** **then**
          d and d' swap their matches,
          $\psi_d \leftarrow \psi_d^{d',d}$
        else
          d and d' do not swap; $\psi_d \leftarrow \psi_d$
        end
      end
    end
  until *Pareto efficiency is achieved.*
end
Output : η_d^k, ψ_d^{final}

The numerical results of implementing these two algorithms are presented in the next section.

VI. Numerical Results

In this section, we analyze the system's performance using MATLAB. We assume a uniform cellular distribution with one BS at the center and a cell radius of 300 m. D2D pairs are randomly located over the cell coverage. Strong CUs are randomly located such that their distances to BS is at maximum 50% of the cell radius, and for weak CUs, at least 85% of it. An example of network topology with 12 CUs and 14 D2D pairs is shown in Fig. 2. As in [13], the following path loss model is adopted: $PL(distance) = distance^{-\tau}$, where $\tau = 2$ refers to the path loss exponent. Table I shows the system parameters used unless otherwise specified. The results are averaged and plotted over 1000 random cellular distributions, and the co nfidence intervals of 95% are computed.

TABLE I: Simulation Parameters

NOMA groups, D2D pairs / slice	11, 23
Maximum power of base station	46 dBm
Transmit power of D2D pair	15dBm
Transmit power of strong CU	10dBm
Bandwidth, Number of RBs	20 MHz, 100
Channel Model	Rayleigh Fading & Path loss
AWGN power spectral density	-174 dBm/Hz
Circuit Hardware Noise power	-80dBm
Residual SI factor ρ	0.02
$R^{min}_{CU_s}, R^{min}_d$	1Mbps, 0.5Mbps
$T^{max}_{CU_s}, T^{max}_d, X_j$	5 ms, 10 ms, 7000 bits

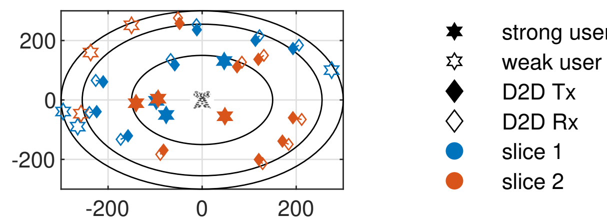

Fig. 2: Cellular Distribution

We start by evaluating the performance of the proposed algorithm in terms of convergence rate. Fig. 3 shows the CDF of the number of swapping iterations for different numbers of CUs and D2D pairs in both slices. We note that the number of swapping iterations is the sum of the total number of swappings in algorithms 1 and 2. Intuitively, the higher the number of CUs and D2D pairs, the higher the probability of having swap-blocking pairs. This is clearly shown in Fig. 3, with a maximum of ≈37 iterations to reach the convergence.

In the following, the influence of the transmit power of the underlay D2D communications on the overall system throughput and the number of admitted D2D pairs is studied. The latter quantifies the number of D2D pairs admitted to underlay the cellular communication (i.e., share the RBs allocated to the NOMA groups). As these D2D pairs introduce additional interference, their admission depends on the satisfaction of the technical requirements of the CUs in each slice.

Fig. 4 studies the effect of varying the SI cancellation level on the performance of the proposed algorithm. It shows that less residual SI (better SI cancellation) achieves higher total throughput and number of admitted D2D pairs. As P_d increases, the interference on CUs increases, so the number of admitted D2D pairs decreases. While the total throughput initially increases due to the increase of the D2D data rates, it reaches a peak and then starts to decrease due to increased interference level. This also applies to figures 5 and 6.

Fig. 5 compares the performance of different matching schemes between D2D pairs and NOMA groups. In particular, the many-to-one matching (i.e., considered in our paper) is compared to the one-to-one matching (i.e., only one D2D pair can share the RBs allocated to one NOMA group), and random many-to-one (i.e., many D2D pairs can randomly share the RBs of one NOMA group without using matching theory). Numerical results show that the many-to-one matching scheme outperforms the other two in terms of the overall throughput and the number of admitted D2D pairs. The random many-to-one scheme presents the worst performance.

Fig. 6 studies the effect of network slicing on our system's performance. For comparison, we consider a scenario where the network slicing is disabled (i.e., users of heterogeneous service types are not separated into different slices. They may share the same RBs (different service type CUs can be in the same NOMA group k, and different service type D2D pairs can share the RBs allocated to k), and their diverse technical requirements are not considered, but a common quality of service requirement is considered). Numerical results show that the enabled-network slicing scenario outperforms the other in terms of the overall throughput and the number of admitted

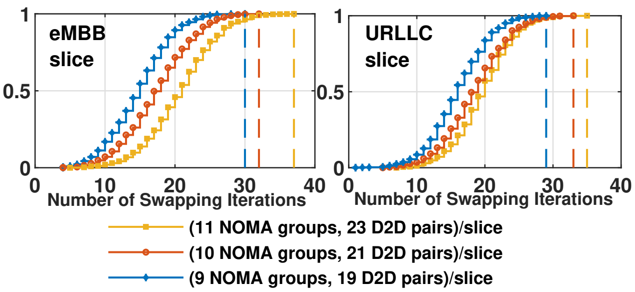

Fig. 3: CDF of Swapping Iterations

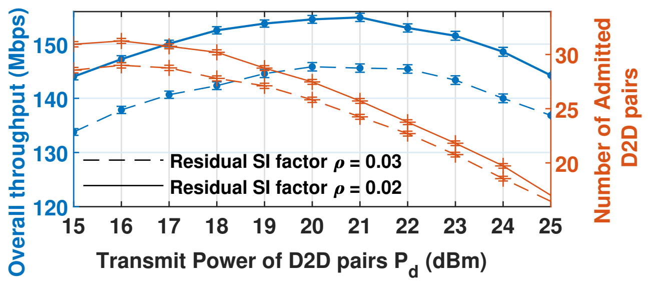

Fig. 4: Comparison of our proposed algorithm's performance versus D2D transmit power for different residual SI levels

D2D pairs. This is due to the different interference tolerance of different service type users. The gap between the two scenarios increases with P_d because the increase in interference level highlights the different interference tolerance and the need to consider different technical requirements separately.

VII. Conclusion

In this paper, the benefits of adopting network slicing in full-duplex CNOMA-based networks with underlay D2D communications are investigated. An optimization problem that maximizes the overall system's throughput while respecting slices' technical requirements is formulated. A two-phases heuristic based on matching theory is proposed. In particular, the convergence of the proposed solution is proved, and its performance is analyzed as a function of different residual SI levels. Furthermore, we have compared it with another scenario from the state of the art where network slicing is not adopted. It is shown that the proposed solution outperforms its counterpart in terms of the overall throughput and the number of admitted D2D pairs. As a future work, we plan to extend the proposed solution by varying NOMA-groups sizes.

Acknowledgement

The authors would like to thank Dr. Abdel-Mehsen Ahmad from the Lebanese International University, Lebanon for his helpful and insightful comments about the system model.

References

[1] Kazmi S.M.A et al. *Network slicing for 5G and beyond networks*. Vol. 1. Springer, 2019.

[2] Kim D. H. et al. "Distributed Radio Slice Allocation in Wireless Network Virtualization: Matching Theory Meets Auctions". *IEEE Access* 8, 73494–73507, 2020.

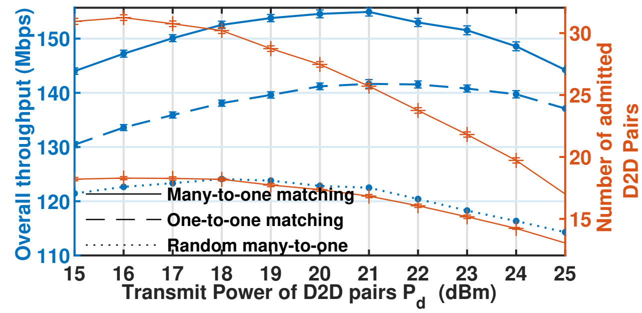

Fig. 5: Comparison of different matching theory schemes

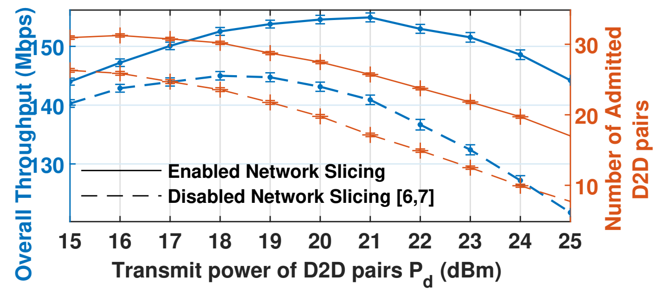

Fig. 6: Comparison of enabled vs. disabled network slicing

[3] Liaqat M et al. "Power-Domain Non Orthogonal Multiple Access (PD-NOMA) in Cooperative Networks: An Overview". *Wireless Netw* 26.1, 181–203, 2020.

[4] Ding Z et al. "Cooperative Non-Orthogonal Multiple Access in 5G Systems". *IEEE Commun. Lett.* 19.8, 1462–1465, 2015.

[5] Asadi A et al. "A Survey on Device-to-Device Communication in Cellular Networks". *IEEE Commun. Surv. Tutor.* 16.4, 1801–1819, 2014.

[6] Amer A et al. "Resource Allocation for Downlink Full-Duplex Cooperative NOMA-Based Cellular System with Imperfect SI Cancellation and Underlaying D2D Communications". *Sensors* 21.8, 2768, 2021.

[7] Diao X et al. "Joint Computing Resource, Power, and Channel Allocations for D2D-Assisted and NOMA-Based Mobile Edge Computing". *IEEE Access* 7, 9243–9257, 2019.

[8] Hossain M.A. et al. "Network Slicing for NOMA-Enabled Edge Computing". *IEEE Trans. Cloud Comput.*, 1–1, 2021.

[9] Hossain M.A. et al. "Energy Aware Latency Minimization for Network Slicing Enabled Edge Computing". *IEEE Trans. Green Commun. Netw.* 5.4, 2150–2159, 2021.

[10] Zoubeir M et al. "Network Slicing with MEC and Deep Reinforcement Learning for the Internet of Vehicles". *IEEE Netw.* 35.3, 132–138, 2021.

[11] Tebe P.I. et al. "Using 5G Network Slicing and Non-Orthogonal Multiple Access to Transmit Medical Data in a Mobile Hospital System". *IEEE Access* 8, 189163–189178, 2020.

[12] Wu S et al. "In-Band Full Duplex Wireless Communications and Networking for IoT Devices:Progress, Challenges and Opportunities". *Future Gener. Comput. Syst* 92, 705–714, 2019.

[13] Liu Y et al. "Cooperative Non-orthogonal Multiple Access With Simultaneous Wireless Information and Power Transfer". *IEEE J. Sel. Areas Commun.* 34.4, 938–953, 2016.

[14] Gu Y et al. "Matching Theory for Future Wireless Networks: Fundamentals and Applications". *IEEE Commun. Mag.* 53.5, 52–59, 2015.