Cooperative Wireless Edges with Composite Resource Allocation in Hierarchical Networks

Nan Li, Xiping Hu, Edith Ngai and Erol Gelenbe

***Abstract*—With the expansion of the IoT, it is important to optimize available bandwidth to reliably support edge to device communications. Thus we propose a wireless network where each edge server communicates with its end devices using its wireless band as a primary channel, assisted by a secondary edge server that can relay communications via its own wireless band as a secondary channel. The network can optimize capacity by balancing load between primary and secondary wireless bands, and we analyze the geometry of achievable rate regions, depending on the state of bands modeled as Rayleigh fading channels. The allocation of a connection to the primary or secondary band is formulated as an optimization problem which is then solved, and illustrated with numerical examples.**

***Index Terms*—Edge computing, cooperation, non-orthogonal multiple access, resource allocation.**

I. Introduction

With the growth of the IoT for homes [1], e-Health [2] and other applications [3], secure, energy efficient and quality of service (QoS) aware networks [4] need to be deployed to support a myriad of different end devices. Real-time and reliable connectivity is needed in such systems, for instance in the medical domain [5] for the health monitoring of patients both in and outside hospitals and care facilities [6]. these real-time reliable needs can be met with the help of edge computing [7], [8], QoS aware edge servers [9], and fast and dependable wireless connectivity.

Existing work on edge computing for the IoT focuses on the distribution of workload at the edges, especially for content caching [9]–[11], and little attention has been paid to extensions in which wireless communications may enhance the system at the edge. Indeed, due to the broadcast nature of wireless communications, nodes that overhear the transmitted message from a source to a destination may help re-establish or enhance the communication by acting as relays that collaboratively forward messages [12], [13]. Orthogonal multiple access (OMA) allows resource sharing in the time and frequency domain, while non-orthogonal multiple access (NOMA) superimposes the signals of multiple users at the transmitter with distinct power allocation factors, and successive interference cancellation (SIC) at the receiver. In this regard, all users can co-exist in the same time, frequency or code domain, leading to extensions of NOMA to power domain resource sharing [14]–[16].

In this work, we consider a 3-tier hierarchy architecture that can help support IoT and other applications, composed of mobile edge devices or IoT edge devices, edge servers, and a control center. Edge devices can access the resources of the corresponding edge server through a wireless local area network. Each edge server is typically attached to an access point (e.g. wireless router or base station), the wireless access point is close enough for wireless communication with the edge devices or an a IoT gateway. We also assume that the edge servers are interconnected through backhaul to the control center, and possibly to a remote cloud.

We focus on data delivery to and from the control center to the edge servers and edge devices, and investigate the wireless transmission efficiency between edge servers and edge devices. Thus edge servers may cooperate to offer more reliable service to the edge devices, using NOMA for data transmission between edge servers and edge devices. We propose the use of primary and secondary channels of each connection, so as to provide more flexibility, dependability and QoS as compared to simple intra-cell models [17], [18]. In the network that we consider, if the network control center attempts to deliver a message to an edge device through a given server and the channel is weak, another edge server with a better channel can act as a relay while also carrying out its own transmissions as a primary channel.

Specifically, we suggest a novel hybrid multiple access strategy combining OMA and NOMA that improves system performance and resource efficiency, and derive the achievable rate regions depending on channel conditions. Furthermore, we solve the optimal joint resource allocation problem at both the control center and the edge servers. The results are illustrated through numerical examples based on a geometrical representation that highlights the impact of node geometry.

II. System Model

Consider a hierarchical network model consisting of a control center (CC), two edge servers (ES1 and ES2) and one e-Health device (DE) correlated to ES2, as shown in Fig. 1. The ESs attempt to receive and process messages

N. Li and X. Hu are with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China (email: {nan.li3, xp.hu}@siat.ac.cn).

E. Ngai is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China (email: chngai@eee.hku.hk).

E. Gelenbe is with Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Gliwice, Poland (email: gelenbe.erol@orange.fr). This author's research was supported by the SerIoT Research and Innovation Action, funded by the European Commission (EC) under the H2020-IOT-2017 Program, Grant Agreement 780139. The EC's support does not constitute an endorsement of this paper, which reflects the views only of the authors.

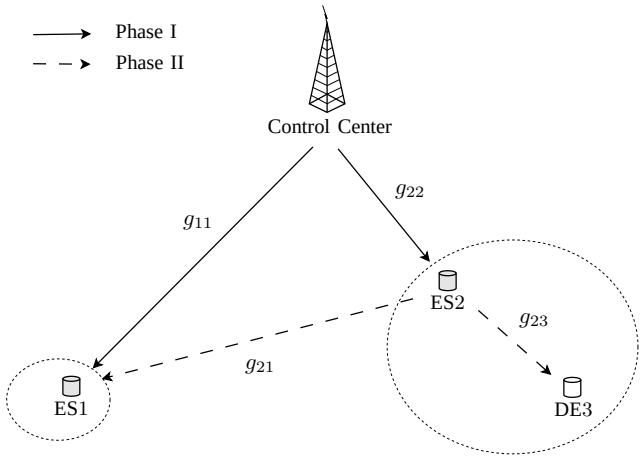


Fig. 1: System model of cooperative edge nodes with non-orthogonal multiple transmissions.

delivered by the CC, denoted as $\mathcal{I}_1$ to ES1 and $\mathcal{I}_2$ to ES2. In case of the geometry location of ES1 being far away from the CC, or the weak channel condition caused by varying network environment, ES2 takes on the role of a potential relay for the transmission to ES1. Meanwhile, it may utilize the opportunity of cooperation to transmit a message $\mathcal{I}_3$ to DE3. With NOMA signaling at the CC, transmission of $\mathcal{I}_1$ and $\mathcal{I}_2$ are accomplished simultaneously. Similar to the relaying of $\mathcal{I}_1$ and the transmission of $\mathcal{I}_3$ at ES2. The transmit power at the CC and the ESs are limited by P_c and P_e, respectively.

Denote by d_{11}, d_{22}, d_{21}, and d_{23} the distances from the CC to ES1, CC to ES2, ES2 to ES1, and ES2 to DE3, respectively. All channels experience independent but not necessarily identical (i.n.i.d) Rayleigh fading, with the corresponding channel coefficients g_{ij} where $ij \in \{11, 22, 21, 23\}$. The path-loss exponent $\alpha > 0$ and g_{ij} is a complex Gaussian variable with $g_{ij} \sim \mathcal{CN}(0, d_{ij}^{-\alpha})$. We assume that perfect channel information is always available at the receiver side. All channels are quasi-static such that channel coefficients remain constant within one protocol cycle. For the ease of exposition, the additive white Gaussian noise z_0 at each user is modeled as independent and identically (i.i.d) distributed $\mathcal{CN}(0, \sigma^2)$.

The transmission process consists of two phases where a time-division multiple access scheme is exploited for the CC transmitting in Phase I and ES transmitting in Phase II. Assume that the CC transmits $\mathcal{I}_1$ intended for UE1 and $\mathcal{I}_2$ for ES2 with a split of transmit power P_c, in a fraction of the time in one protocol cycle and share the remaining time to ES2. ES2 needs to extract its own message before relaying to ES1. In Phase II, ES2 broadcasts the superimposed signals with a fraction of its total power P_e for relaying $\mathcal{I}_1$ and the remaining of P_e for transmitting $\mathcal{I}_3$. For the whole transmission process, the protocol we designed is a joint OMA and NOMA.

In this paper, we study decode-and-forward and analog network coding-based forward applied at ES2 for the relaying scheme. Two schemes are proposed as Decode-and-Forward based NOMA (DF-NOMA) and Analog-Network-Coding based NOMA (ANC-NOMA). In the DF-NOMA scheme, ES2 needs to decode the signal it received before relaying. To alleviate this problem, ANC-NOMA allows ES2 to simply forward the signal it receives without decoding required. Aside from the power allocation at ES2 in Phase II, jointly considering the power split at CC in Phase I introduces a complex resource allocation strategy.

III. Achievable Rate Regions

In this section, we denote by β the factor of time sharing for OMA and θ, ω the factor of transmit power allocation at CC and ES2 for NOMA, respectively. $x_{\{1,2,3\}}$ is the signal of $\mathcal{I}_{\{1,2,3\}}$ with zero mean and unit-power $E\left[x^{\star}_{\{1,2,3\}} x_{\{1,2,3\}}\right] = 1$. The CC transmits x_1 to ES1 and x_2 to ES2 in Phase I. With power split parameter θ, the signal transmitted by CC is a mixture of x_1 and x_2,

$$x_{12} = \sqrt{\theta P_c} x_1 + \sqrt{(1-\theta)P_c} x_2.$$

Accordingly, the signals $y_{1,\mathrm{I}}$ received at ES1 and y_2 at ES2 are

$$y_{1,\mathrm{I}} = g_{11} x_{12} + z_0 \quad \text{and} \quad y_2 = g_{22} x_{12} + z_0.$$

Based on the channel quality, ES2 has two decoding strategies. When $|g_{22}|^2 \geq |g_{11}|^2$, ES2 is able to decode x_1 first, and then extract the interference out of the received signal before decoding x_2; while ES1 treats x_2 as noise. For decoding x_1, the SNR at the ES2 is

$$\Gamma_{1,2} = \frac{\theta \rho_c |g_{22}|^2}{(1-\theta)\rho_c |g_{22}|^2 + 1},$$

where $\rho_c = P_c/\sigma^2$ represents the transmit SNR of CC. Conditioning on successfully decoding x_1, ES2 subtracts x_1 and decodes x_2. Then, the SNR at ES1 to detect x_1 and at ES2 to detect x_2 is given by

$$\Gamma_{11} = \frac{\theta \rho_c |g_{11}|^2}{(1-\theta)\rho_c |g_{11}|^2 + 1}, \tag{1}$$

$$\Gamma_{22} = (1-\theta)\rho_c |g_{22}|^2. \tag{2}$$

On the other hand, when $|g_{22}|^2 < |g_{11}|^2$, ES2 treats x_1 as noise but ES1 will decode x_2 before x_1. In this case, the SNR at ES1 to decode x_2 is

$$\Gamma_{2,1} = \frac{(1-\theta)\rho_c |g_{11}|^2}{\theta \rho_c |g_{11}|^2 + 1}.$$

Thus, the SNR at ES1 to detect x_1 and at ES2 to detect x_2 is given by

$$\Gamma_{11} = \theta \rho_c |g_{11}|^2, \tag{3}$$

$$\Gamma_{22} = \frac{(1-\theta)\rho_c |g_{22}|^2}{\theta \rho_c |g_{11}|^2 + 1}. \tag{4}$$

An interesting observation here is, when the receiving channel of ES2 performs better than ES1, ES2 is able to decode both messages from the CC. As a result, ES2 can exploit the decode-and-forward scheme to relay the message to ES1 in the next phase. When the receiving channel of ES1 outperforms

ES2, although ES2 could not decode x_1, it can extract x_2 out of the received signal and then relay the remaining part to ES1. Obviously the analog network coding-based scheme can be exploited for relaying. Through the analysis above, we find that the transmission process can be classified as two schemes, the DF-NOMA scheme and the ANC-NOMA scheme. Therefore, we re-denote the SNRs above as Γ_{11}^{D} in (1), Γ_{22}^{D} in (2), Γ_{11}^{A} in (3) and Γ_{22}^{A} in (4).

In Phase II, ES2 broadcasts the superimposed signals with a fraction ω $(0 \leq \omega \leq 1)$ of its total power P_e for relaying $\mathcal{I}_1$ to ES1 and $1-\omega$ of P_e for transmitting $\mathcal{I}_3$ to DE3. In DF-NOMA scheme, the decoded signal x_1 is combined with the transmit signal x_3. In ANC-NOMA scheme, the signal extracted from y_2 is combined with x_3. Hence, the transmit signal x_{13} encoded by x_1 and x_3 is different in each case, given by

$$x_{13}^{\mathrm{D}} = \sqrt{\omega P_e}x_1 + \sqrt{(1-\omega)P_e}x_3, \tag{5}$$

$$x_{13}^{\mathrm{A}} = \sqrt{\omega P_e}\frac{\sqrt{\theta P_c}g_{22}x_1 + z_0}{\gamma} + \sqrt{(1-\omega)P_e}x_3, \tag{6}$$

where γ is the normalization factor used to ensure that the transmit power for the signal of $\mathcal{I}_1$ always equals ωP_e. Thus, we have $\gamma = \sqrt{\theta P_c|g_{22}|^2 + \sigma^2}$. Correspondingly, the received signal $y_{1,\mathrm{II}}$ at ES1 and y_3 at DE3 can be denoted as

$$y_{1,\mathrm{II}} = g_{21}x_{13} + z_0 \quad \text{and} \quad y_3 = g_{23}x_{13} + z_0.$$

When $|g_{21}|^2 \geq |g_{23}|^2$, ES1 can decode x_3 first, and then strip this interference off the received signal before decoding x_1; while DE3 will treat x_1 as noise and decode x_3 directly. Vice versa when $|g_{21}|^2 < |g_{23}|^2$.

A. Decode-and-Forward NOMA

For DF-NOMA, since ES2 has decoded x_1 in Phase I, the transmit signal x_{13}^{D} is directly superimposed with x_1 and x_3 in Phase II, and the received signal at ES1 is degraded by g_{21},

$$y_{1,\mathrm{II}}^{\mathrm{D}} = \sqrt{\omega P_e}g_{21}x_1 + \sqrt{(1-\omega)P_e}g_{21}x_3 + z_0;$$

and the received signal at DE3 is the same but degraded by g_{23}. The corresponding SNR at both receivers for two decoding strategies based on the quality of channels are derived as

$$\begin{cases} \Gamma_{21}^{\mathrm{D}} = \omega\rho_e|g_{21}|^2 \\ \Gamma_{23}^{\mathrm{D}} = \frac{(1-\omega)\rho_e|g_{23}|^2}{\omega\rho_e|g_{23}|^2+1} \end{cases} \quad \text{if } |g_{21}|^2 \geq |g_{23}|^2; \tag{7}$$

$$\begin{cases} \Gamma_{21}^{\mathrm{D}} = \frac{\omega\rho_e|g_{21}|^2}{(1-\omega)\rho_e|g_{21}|^2+1} \\ \Gamma_{23}^{\mathrm{D}} = (1-\omega)\rho_e|g_{23}|^2 \end{cases} \quad \text{if } |g_{21}|^2 < |g_{23}|^2. \tag{8}$$

$\rho_e = P_e/\sigma^2$ represents the transmit SNR of ES2.

Proposition 1. *The achievable rates $(R_1^{\mathrm{D}}, R_2^{\mathrm{D}}, R_3^{\mathrm{D}})$ given in the following region:*

$$R_1^{\mathrm{D}} = \min\left\{R_{1,2}^{\mathrm{D}}, \bar{R}_1^{\mathrm{D}}\right\} \tag{9a}$$

$$R_2^{\mathrm{D}} = \beta\log\left(1+\Gamma_{22}^{\mathrm{D}}\right) \tag{9b}$$

$$R_3^{\mathrm{D}} = (1-\beta)\log\left(1+\Gamma_{23}^{\mathrm{D}}\right) \tag{9c}$$

for every possible time sharing factor β and power allocation factors θ and ω.

Proof. The transmission rate of the CC is limited by the link to ES2 as

$$R_{1,2}^{\mathrm{D}} = \beta\log\left(1+\Gamma_{1,2}\right). \tag{10}$$

$\bar{R}_1^{\mathrm{D}}$ represents the maximum rate at which ES1 can reliably decode $\mathcal{I}_1$ given repeated transmissions from the CC and ES2. Requiring both the ES2 and ES1 to decode the signal without error in two phases results in the minimum of the two maximum rates in (9a). By exploiting maximum-ratio combining (MRC) at the receivers, $\bar{R}_1^{\mathrm{D}}$ is given by

$$\begin{aligned}\bar{R}_1^{\mathrm{D}} =& \beta\log\left(1+\Gamma_{11}^{\mathrm{D}}+\Gamma_{21}^{\mathrm{D}}\right) \\ &+ (1-2\beta)\log\left(1+\Gamma_{21}^{\mathrm{D}}\right), \quad \beta \in [0, 0.5]; \end{aligned} \tag{11a}$$

$$\begin{aligned}\bar{R}_1^{\mathrm{D}} =& (1-\beta)\log\left(1+\Gamma_{11}^{\mathrm{D}}+\Gamma_{21}^{\mathrm{D}}\right) \\ &+ (2\beta-1)\log\left(1+\Gamma_{11}^{\mathrm{D}}\right), \quad \beta \in [0.5, 1]. \end{aligned} \tag{11b}$$

Additionally, the achievable rate for ES2 reliably decoding x_2 is determined by the transmission in Phase I only as in (9b), while for DE3 it is determined by the transmission in Phase II only as in (9c). □

B. Analog-Network-Coding NOMA

For ANC-NOMA, ES2 only decodes x_2 and subtracts it from y_2 in Phase I. With x_3 superimposed, the received signal $y_{1,\mathrm{II}}^{\mathrm{A}}$ at ES1 in Phase II is impacted by the propagated noise from ES2 as

$$\begin{aligned} y_{1,\mathrm{II}}^{\mathrm{A}} =& \frac{\sqrt{\theta\omega P_c P_e}g_{22}g_{21}}{\gamma}x_1 \\ &+ \sqrt{(1-\omega)P_e}g_{21}x_3 + \frac{\sqrt{\omega P_e}g_{21}}{\gamma}z_0 + z_0, \end{aligned}$$

and similar to the received signal y_3^{A} at DE3. The corresponding SNR at ES1 and DE3 for each decoding strategy is

$$\begin{cases} \Gamma_{21}^{\mathrm{A}} = \frac{\theta\rho_c|g_{22}|^2\omega\rho_e|g_{21}|^2}{\omega\rho_e|g_{21}|^2+\theta\rho_c|g_{22}|^2+1} \\ \Gamma_{23}^{\mathrm{A}} = \frac{(1-\omega)\rho_e|g_{23}|^2}{\omega\rho_e|g_{23}|^2+1} \end{cases} \quad \text{if } |g_{21}|^2 \geq |g_{23}|^2 \tag{12}$$

$$\begin{cases} \Gamma_{21}^{\mathrm{A}} = \frac{\theta\rho_c|g_{22}|^2\omega\rho_e|g_{21}|^2/(\theta\rho_c|g_{22}|^2+1)}{(1-\omega)\rho_e|g_{21}|^2+\omega\rho_e|g_{21}|^2/(\theta\rho_c|g_{22}|^2+1)+1} \\ \Gamma_{23}^{\mathrm{A}} = \frac{(1-\omega)\rho_e|g_{23}|^2}{\omega\rho_e|g_{23}|^2/(\theta\rho_c|g_{22}|^2+1)+1} \end{cases} \quad \text{if } |g_{21}|^2 < |g_{23}|^2 \tag{13}$$

Proposition 2. *The achievable rates $(R_1^{\mathrm{A}}, R_2^{\mathrm{A}}, R_3^{\mathrm{A}})$ are given in the following region:*

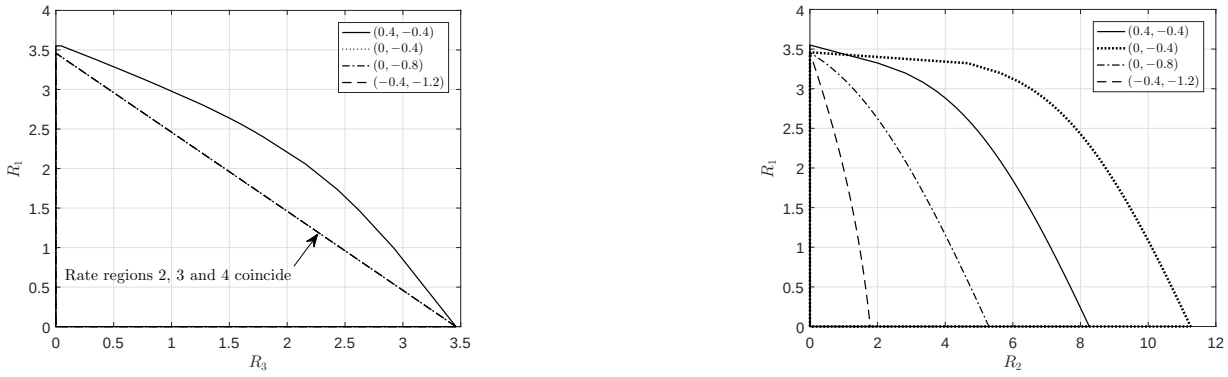
$$R_1^{\mathrm{A}} \quad (\textit{see } (15)) \tag{14a}$$

$$R_2^{\mathrm{A}} = \beta\log\left(1+\Gamma_{22}^{\mathrm{A}}\right) \tag{14b}$$

$$R_3^{\mathrm{A}} = (1-\beta)\log\left(1+\Gamma_{23}^{\mathrm{A}}\right) \tag{14c}$$

for every possible time sharing factor β and power allocation factors θ and ω.

Proof. Different from DF-NOMA, ES2 is not able to decode the message for ES1 for the poor channel condition here. In this case, the transmission rate of the CC is not limited by the link to ES2. The ES2 extracts its own signal x_2 out of the received signal and then superimposes the remaining part



(a) Rate regions on (R_1, R_3)

(b) Rate regions on (R_1, R_2)

Fig. 2: Rate regions on different rate pairs with varying positions of ES2. For each position, the corresponding $d_{22} = \{0.57, 0.4, 0.8, 1.26\}$ and $d_{21} = \{0.72, 1.08, 1.28, 1.34\}$.

to x_3 aimed for DE3. As shown in (6), the relayed signal is normalized by γ with the additive noise. The achievable rate for ES1 is therefore impacted by the propagated noise as

$$R_1^{\mathrm{A}} = \beta \log\left(1 + \Gamma_{11}^{\mathrm{A}} + \Gamma_{21}^{\mathrm{A}}\right) + (1 - 2\beta)\log\left(1 + \Gamma_{21}^{\mathrm{A}}\right), \quad \beta \in [0, 0.5]; \quad \text{(15a)}$$

$$R_1^{\mathrm{A}} = (1 - \beta)\log\left(1 + \Gamma_{11}^{\mathrm{A}} + \Gamma_{21}^{\mathrm{A}}\right) + (2\beta - 1)\log\left(1 + \Gamma_{11}^{\mathrm{A}}\right), \quad \beta \in [0.5, 1]. \quad \text{(15b)}$$

For R_2^{A} and R_3^{A}, the derivation is similar to that of the DF-NOMA scheme. □

C. Comparison of Two NOMA Schemes

Based on the achievable rate regions derived above, we compare the performance of two NOMA schemes in an experimental environment. We assume ES2 and DE3 as mobile while the CC and ES1 as stationary at coordinates $(0, 1)$ and $(1, 1)$. All distances between users are normalized w.r.t the CC-ES1 distance. Without loss of generality, the position of ES2 and DE3 is considered relatively stationary and the distance between is fixed as 1. Experimental assumptions are as follows: $\alpha = 3$ as the path-loss exponent; $\rho_c = \rho_e = 10$dB with power constraint $P_c = P_e = 10$ and unit-power additive white Gaussian noise; $T = 1$ as normalized time period of each protocol cycle. Each possible set of the resource allocation factors (β, θ, ω) yields a pentagon-shaped rate region. We vary the parameters over a sufficiently fine grid and take the convex hull over all corresponding regions.

In Fig. 2, we plot the rate regions for varying positions of ES2 on two rate pairs. Fig. 2a depicts the achievable rate regions on (R_1, R_3) while Fig. 2b on (R_1, R_2). According to each user position, we calculate the distance of d_{22} and compare it to d_{11}, since channel quality is w.r.t distance in our channel model. As previously defined, when $d_{22} \leq d_{11}$, ES2 is able to decode x_1 so that DF-NOMA scheme is applied. In our experiment, rate regions 1, 2 and 3 represent the performance when exploiting DF-NOMA as herein $d_{22} < d_{11}$; while region 4 represents the performance when exploiting ANC-NOMA with $d_{22} > d_{11}$. There are three rate regions coincided in Fig. 2a even though different schemes are exploited. It is also the boundary of rate regions achievable with optimal orthogonal schemes. When d_{22} increases, ES2 will choose ANC-NOMA to avoid decoding x_1 so that it is not limited by the quality of channel between the CC and ES2. Meanwhile, with d_{21} increasing, the relay channel does not sustain a better performance than the direct transmission to ES1, the performance of NOMA schemes is bounded by the optimal OMA schemes but not influenced by the channel with d_{22}. Thus in Fig. 2b, we observe that R_2 is simply determined by the quality of channel between CC and ES2. As shown in the plot, rate region 2 performs the best on R_2 because d_{22} is the shortest in this case.

Through the analysis above, we get insights into the impact of channel qualities on the achievable rate of each user. Moreover, considering varying channel qualities, the resource allocation strategy should be adjusted.

IV. Optimal Resource Allocation on Achievable Rate Maximization

We form an optimization problem on maximizing the achievable rate of DE3 to find an appropriate resource allocation strategy in certain application scenarios.

$$(\mathbf{OP}): \max_{\beta,\theta,\omega} R_3(\beta, \theta, \omega) \quad (16)$$

$$\text{s.t. } R_1, R_2, R_3 \geq R_0, \quad (17)$$

$$0 \leq \beta, \theta, \omega \leq 1, \quad (18)$$

where for a target rate R_0, ES1, ES2 and DE3 can successfully decode the received signal respectively, as in (17).

We notice that the objective function R_3 is neither a convex nor a concave function of β, θ and ω, and not amendable to a convex formulation. Considering that the resource allocation factors are independent with each other, we decompose this problem into sub-optimization problems on each factor. Moreover, the objective function and constraint functions are mostly monotonic on each factor. All these properties enable us to derive a tractable result through decomposition of variables. One of the sub-optimization problems is to optimize time

sharing of the whole transmission process under a given power allocation, as **OP1** on $R_3(\beta)$; **OP2** is to optimize the power splitting ratio at CC on $R_3(\theta)$ within a given time-sharing period; and **OP3** is to optimize the transmit power allocation at ES2 on $R_3(\omega)$. To approach a joint optimal solution on $(\beta^*, \theta^*, \omega^*)$, we try to find local solutions of **OP1**, **OP2** and **OP3** first and a mapping method can be applied. For any pair of two optimized factors in these three sub-optimization problems, denote the mapping from A to B by function f as $A \overset{f}{\mapsto} B$, the joint optimal time-sharing and power allocation strategy $(\beta^*, \theta^*, \omega^*)$ satisfies $\beta^* \overset{\textbf{(OP1)}}{\longmapsto} (\theta^*, \omega^*)$, $\theta^* \overset{\textbf{(OP2)}}{\longmapsto} (\beta^*, \omega^*)$ and $\omega^* \overset{\textbf{(OP3)}}{\longmapsto} (\theta^*, \beta^*)$.

A two-dimensional linear search algorithm is developed on the basis of solving each sub-optimization problem with mapping. Firstly, β^* is derived from **OP1** and the corresponding local optimum $R_3(\beta^*)$ and $(\theta(\beta^*), \omega(\beta^*))$ can be obtained. Secondly, θ^* is derived from **OP2**, so as $R_3(\theta^*)$ and $(\beta(\theta^*), \omega(\theta^*))$. Thirdly, ω^* from **OP3** and $R_3(\omega^*)$ and $(\beta(\omega^*), \theta(\omega^*))$ are obtained accordingly. Finally, the maximum $R_3^* = \max\{R_3(\beta^*), R_3(\theta^*), R_3(\omega^*)\}$ is determined with the corresponding optimal solution $(\beta^*, \theta^*, \omega^*)$.

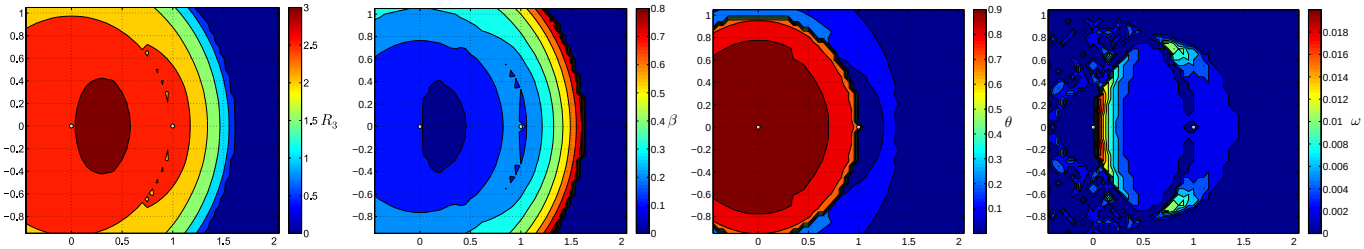
Analysis on Optimization for DF-NOMA and ANC-NOMA Systems: Since the relay scheme is determined by the channel quality corresponding to CC, the system exploits DF-NOMA scheme and ANC-NOMA scheme in different use cases. When $|g_{22}|^2 \geq |g_{11}|^2$, ES2 is able to decode x_1 and forward it, which is then defined as a DF-NOMA system. As $R_1^{\mathrm{D}} = \min\left\{R_{1,2}^{\mathrm{D}}, \bar{R}_1^{\mathrm{D}}\right\}$, constraint (17) can be rewritten as two inequalities with $R_{1,2}^{\mathrm{D}}$ and $\bar{R}_1^{\mathrm{D}}$ separately. Firstly, we solve **OP1**. When (θ, ω) is fixed, we find that R_3^{D} is monotonically decreasing with β while R_2^{D} increasing; for R_1^{D}, $R_{1,2}^{\mathrm{D}}$ is increasing with β too, yet $\bar{R}_1^{\mathrm{D}}$ varies in different conditions. The feasible interval of the optimal solution is determined by analyzing the monotonicity of the functions. Same to the other two sub-optimization problems **OP2** and **OP3**. In ANC-NOMA systems with $|g_{22}|^2 < |g_{11}|^2$, we observe that the objective function R_3^{A} is w.r.t (β, θ) when $|g_{21}|^2 \geq |g_{23}|^2$ but on (β, θ, ω) when $|g_{21}|^2 < |g_{23}|^2$, while R_1^{A} and R_2^{A} are the same as in DF-NOMA systems. To obtain a general solver for the optimization problem, we hold the scenario when $|g_{21}|^2 < |g_{23}|^2$ in ANC-NOMA systems as a paragon and provide a brief analysis. When $|g_{21}|^2 < |g_{23}|^2$, R_3^{A} is monotonically decreasing with β. Meanwhile, R_2^{A} is monotonically increasing with β. On the other hand, R_1^{A} for $\beta \in [0, 0.5]$ is monotonically increasing when $(c1): u1 - 2u2 > 0$ and decreasing when $(c2): u1 - 2u2 < 0$ with $u1 = \log\left(1 + \Gamma_{11}^{\mathrm{A}} + \Gamma_{21}^{\mathrm{A}}\right)$ and $u2 = \log\left(1 + \Gamma_{21}^{\mathrm{A}}\right)$. For $\beta \in [0.5, 1]$, R_1^{A} is monotonically increasing when $(c3): 2u3 - u1 > 0$ and decreasing when $(c4): 2u3 - u1 < 0$ with $u3 = \log\left(1 + \Gamma_{11}^{\mathrm{A}}\right)$. The feasible interval of β under each condition $(c1)$, $(c2)$, $(c3)$ and $(c4)$ can be derived respectively. Within which, β^* locates at the minimum value. Similar to the power allocation factor θ at CC and ω at ES2. It is worth noting that the methodology to solve this problem is through analyzing the boundary for each monotonic function. With all the results derived above, a joint optimal $(\beta^*, \theta^*, \omega^*)$ can be found by the two-dimensional linear search algorithm.

V. Numerical Results

In this section, we evaluate the performance on joint optimal resource allocation in the transmission rate maximization problem above. In particular, we are interested in how the system behaves for varying locations of the cooperative ES2 and depict the results on its coordinates. We discover the performance altering with the position of ES2 and the corresponding receiver DE3, which determines the quality of relay-related channels w.r.t the distance between corresponding users. On the basis of the solution to the optimization problem derived in the previous sections, the numerical results are demonstrated for each use case correlated with relations between user positions.

To keep consistency of the numerical experiments, the environmental parameters are set the same as in Section III-C. We build a geometrical model as a spatial area ranging from $(-0.5, -1)$ to $(2, 1)$ in coordinate system. ES2 is assumed moving within the spatial region with the distance to its receiver DE3 set as 1. The target rate for receivers successfully decoding the received signal is set as $R_0 = 0.5$ bits/s/Hz, where feasible transmission could be done at a reasonable sensitivity. At each position of ES2, we find the optimal transmission efficiency, such that a contour figure will be depicted showing the variations of the achievable rate and the resource fractions in geometry. Fig. 3 compares the achievable rate R_3 and the optimal time sharing and power allocations when ES2 transmits to ES1 and DE3 simultaneously. Each plot depicts the same spatial region, where the positions of CC and ES1 are marked by white circles while the position of ES2 varies. Fig. 3a shows the achievable rate R_3, Fig. 3b corresponds to the time fraction β allocated to the CC and Fig. 3c and 3d to the power fractions θ and ω allocated for transmission to ES1 at CC and relaying to ES1 at ES2. Note that the relay scheme is determined naturally on the location of cooperative users. Therefore, the plots indicate the output of an intelligent scheme selection.

In Fig. 3a, the transmission could obtain a higher rate once ES2 locates within the area closer to the CC than ES1. Especially in a small area between the CC and ES1, the peak performance is achieved as both the receiving channel and the relay channel are in good conditions, w.r.t small (d_{22}, d_{21}). Moreover, when ES2 locates closer to ES1 but further from CC, since it needs to receive and decode its own message $\mathcal{I}_2$, such that the CC has to allocate more resource for the transmission of $\mathcal{I}_2$ and thereby affects the transmission efficiency for other users. The tendency to time sharing for the transmission phase of the CC is altering reversely to the transmission rate of ES2 to DE3 in Fig. 3b. Obviously, when the cooperative users are closer to the CC, less resource is required for the transmission of $\mathcal{I}_1$ and $\mathcal{I}_2$, as ES2 is able to decode both messages and compensate by relaying in the next phase. For power allocation in Fig. 3c, CC allocates the transmit power in accordance with the channel quality. If the



(a) R_3^* (b) β^* (c) θ^* (d) ω^*

Fig. 3: Optimal solution of (a) transmission rate and resource sharing fractions in (b) time domain and (c),(d) power domain.

ES2 has a better receiving channel, the CC will leave more power for transmission to ES1. It implies a strategy that in the cooperation phase, ES2 is able to allocate more power for transmission to DE3 as less is demanded for relaying to satisfy the requirement of ES1. Therefore, in general the power allocation for relaying at ES2 is small as most of the power can be allocated for its own transmission, as shown in Fig. 3d.

VI. Conclusions

In this paper, we have investigated composite resource allocation through hybrid multiple access in hierarchical networks supporting cooperation between edge servers. To enhance the system performance, we proposed a joint OMA and NOMA protocol in a system model where transmissions share the resource in time domain orthogonally and in power domain non-orthogonally. Compared to decode-and-forward based relaying, we exploit analog network coding to avoid decoding the received message in accordance with varying network environment. We derived the achievable regions for both relaying schemes and the results consolidate the advantage of applying adaptive schemes. We then formulated an optimization problem which jointly optimizes the strategy on time sharing and transmit power allocation at the control center and the cooperative edge server. We thoroughly demonstrated the significance of applying NOMA jointly with OMA in possible application scenarios and proposed a resource allocation strategy, showing the efficiency and practicality of intelligent cooperation in hierarchical networks. While the approach we propose may result in a larger energy consumption both for processing and wireless transmission [19], the power consumption analysis will be carried out in future work.

References

[1] J. Augusto-Gonzalez et al., "From internet of threats to Internet of Things: A cyber security architecture for smart homes," in *24th IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2019, pp. 1–6.

[2] M. Staffa et al., "Konfido: An open NCP-based secure e-health data exchange system," in *Recent Cybersecurity Research in Europe: Proceedings of the 2018 ISCIS Security Workshop, Imperial College London. Lecture Notes CCIS No. 821*, ser. Communications in Computer and Information Science. Springer Verlag, 2018, pp. 11–27.

[3] A. Frötscher et al., "Improve cybersecurity of C-ITS road side infrastructure installations: the serIoT-secure and safe IoT approach," in *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)*. IEEE, November 2019, pp. 1–5.

[4] E. Gelenbe, J. Domanska, P. Fröhlich, M. P. Nowak, and S. Nowak, "Self-aware networks that optimize security, QoS, and energy," *Proceedings of the IEEE*, no. 7, pp. 1150–1167, July 2020.

[5] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwal, "The Internet of Things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, Jun. 2015.

[6] S. Movassaghi, M. Abolhasan, J. Lipman, D. Smith, and A. Jamalipour, "Wireless body area networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 3, pp. 1658–1686, 2014.

[7] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Latief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[8] J. Pan and J. McElhannon, "Future edge cloud and edge computing for Internet of Things applications," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 439–449, Feb. 2018.

[9] P. Fröhlich and E. Gelenbe, "Optimal fog services placement in SDN IoT network using random neural networks and cognitive network map," in *the 19th International Conference on Artificial Intelligence and Soft Computing*. Springer, Cham, 2020.

[10] G. Qiao, S. Leng, S. Maharjan, Y. Zhang, and N. Ansari, "Deep reinforcement learning for cooperative content caching in vehicular edge computing and networks," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 247–257, Jan. 2020.

[11] C. Gong, F. Lin, X. Gong, and Y. Lu, "Intelligent cooperative edge computing in the Internet of Things," *IEEE Internet of Things Journal*, 2020.

[12] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[13] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[14] J.-B. Kim and I.-H. Lee, "Non-orthogonal multiple access in coordinated direct and relay transmission," *IEEE Communications Letters*, vol. 19, no. 11, pp. 2037–2040, Nov. 2015.

[15] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Communiations Letters*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[16] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Communiations Letters*, vol. 18, no. 2, pp. 313–316, Feb. 2014.

[17] M. Shahab and S. Shin, "Time shared half/full-duplex cooperative NOMA with clustered cell edge users," *IEEE Communications Letters*, vol. 22, no. 9, pp. 1794–1797, Sep. 2018.

[18] Y. Liu, "Exploiting NOMA for cooperative edge computing," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 99–103, Oct. 2019.

[19] E. Gelenbe, "A diffusion model for packet travel time in a random multihop medium," *ACM Transactions on Sensor Networks*, vol. 3, no. 2, p. 10, June 2007.