Cache-Aided Non-Orthogonal Multiple Access for 5G-Enabled Vehicular Networks

Sanjeev Gurugopinath, *Member, IEEE*, Paschalis C. Sofotasios, *Senior Member, IEEE*,
Yousof Al-Hammadi, *Member, IEEE*, and Sami Muhaidat, *Senior Member, IEEE*

Abstract—The increasing demand for rich multimedia services and the emergence of the Internet-of-Things (IoT) pose challenging requirements for the next generation vehicular networks. Such challenges are largely related to high spectral efficiency and low latency requirements in the context of massive content delivery and increased connectivity. In this respect, caching and non-orthogonal multiple access (NOMA) paradigms have been recently proposed as potential solutions to effectively address some of these key challenges. In the present contribution, we introduce cache-aided NOMA as an enabling technology for vehicular networks. In this context, we first consider the full file caching case, where each vehicle caches and requests entire files using the NOMA principle. Without loss of generality, we consider a two-user vehicular network communication scenario under double Nakagami$-m$ fading conditions and propose an optimum power allocation policy. To this end, an optimization problem that maximizes the overall probability of successful decoding of files at each vehicle is formulated and solved. Furthermore, we consider the case of split file caching, where each file is divided into two parts. A joint power allocation optimization problem is formulated, where power allocation across vehicles and cached split files is investigated. The offered analytic results are corroborated by extensive results from computer simulations and interesting insights are developed. Indicatively, it is shown that the proposed caching-aided NOMA outperforms the conventional NOMA technique.

Index Terms—Caching, double Nakagami$-m$ fading channels, non-orthogonal multiple access, vehicular communications.

I. INTRODUCTION

Recent research advances in information and wireless technologies have led to growing interest in the development of intelligent transportation systems (ITS), which promise significant improvements in road safety and traffic flow, and

S. Gurugopinath is with the Department of Electronics and Communication Engineering, PES University, Bengaluru 560085, India, (email: sanjeevg@pes.edu).

P. C. Sofotasios is with the Center for Cyber-Physical Systems, Department of Electrical Engineering and Computer Science, Khalifa University of Science and Technology, PO Box 127788, Abu Dhabi, UAE, and also with the Department of Electrical Engineering, Tampere University, 33101 Tampere, Finland (email: p.sofotasios@ieee.org).

Y. Al-Hammadi is with the Department of Electrical and Computer Engineering, Khalifa University of Science and Technology, PO Box 127788, Abu Dhabi, UAE (email: yousof.alhammadi@ku.ac.ae).

S. Muhaidat is with the Center for Cyber-Physical Systems, Department of Electrical and Computer Engineering, Khalifa University of Science and Technology, PO Box 127788, Abu Dhabi, UAE and with the Institute for Communication Systems, University of Surrey, GU2 7XH, Guildford, UK, (email: muhaidat@ieee.org).

will enable new data services [1], [2]. In this context, vehicular networks consist of cooperative communication terminals that relay information with each other and also exchange information with fixed infrastructure, e.g., road side unit (RSU). Potential applications of vehicular networks are diverse and pervasive; for example, transportation can be improved through fast dissemination of road and traffic information and coordination of vehicles at critical points such as highway entries and other intersections. In addition, numerous challenging new applications of interest can be realized, e.g., high-speed internet access, cooperative downloading, network gaming among passengers of adjacent vehicles, and virtual, video-enabled meetings among co-workers in different vehicles.

As a dedicated short-range communications (DSRC) technology, wireless access for vehicular environments (WAVE), IEEE 802.11p, enables data rates in the range $6 - 27$ Mbps over short distances [3]. Furthermore, it provides high-speed vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) data transmission. In this context, the 3rd generation partnership project (3GPP) recently published V2X (vehicle-to-everything) specifications based on LTE as the underlying technology. However, in the also recently proposed internet-of-vehicles (IoV) ecosystem, the vehicles are envisioned to share and access enormous data quickly from the cloud, and are expected to be able to process them with high performance and marginal overhead. Moreover, due to the high mobility scenarios, network topology can change quickly and the handover among different base stations and vehicles becomes more frequent. This poses several challenges to the effective realization of reliable communications with low latency in both WAVE and LTE based vehicular networks. Therefore, various 5G enabling technologies, such as multiple-input multiple-output (MIMO), heterogeneous network (HN)-based communications, millimetre-wave (mmWave) communications and ultra-wideband (UWB) communications can be considered to improve the overall efficiency of a vehicular network.

A major challenge related to the above scheme is that these technologies require a significant amount of backhaul overhead. Yet, it has been recently shown that caching techniques, in which popular files are stored a priori at vehicles, are known to reduce the backhaul traffic. The basic idea of caching is to store the popular contents in different geographical locations during the off-peak time. This local storage arrangement enables a duplicate transmission of popular files, thereby increasing the spectral efficiency of the vehicular network and, subsequently, reducing latency. Also, it has been highlighted in the literature that installing memory units

at the user-end is significantly cost effective compared to increasing the backhaul overhead [4]. Based on this, numerous investigations on the integration of several caching techniques with technologies such as heterogeneous networks [5]–[7], device-to-device communications [8], [9], and with cloud- and fog-radio access networks [10] have been carried out.

It is recalled that non-orthogonal multiple access (NOMA) has been recently proposed as a promising multiple access solution to address some of the challenges in 5G networks ([11]–[14]). In particular, NOMA is envisioned to increase the system throughput and to support massive connectivity. More recently, it was shown that NOMA is capable of offering performance benefits in visible light communication systems, where enhanced performance gains in terms of metrics such as spectral and energy efficiencies were reported [15], [16]. Similarly, in the context of vehicular networks, it allows different vehicles to share the same time and frequency resources [17]. It is recalled that NOMA can be realized via two different approaches, namely, (a) power domain, and (b) code domain [18]. In the power domain NOMA (PD-NOMA), users are assigned different power levels over the same resources; on the contrary, in the code domain NOMA (CD-NOMA), multiplexing is carried out using spreading sequences for each user, similar to the code division multiple access (CDMA) technology. In this work, by NOMA, we refer to the PD-NOMA.

It is noted that NOMA can be realized by allocating different power levels to different users, according to their channel conditions. This power allocation scheme enables the efficient implementation of successive interference cancellation (SIC), which is often considered to remove multi-user interference in multi-user detection systems. In this case, SIC is carried out at users with the best channel conditions and is performed in descending order of the channel. It is also pointed out that significant performance improvements have been shown in vehicular networks due to the amalgamation of the NOMA technology with other techniques such as MIMO communications [17], [19], mmWave communications and spatial modulation [17]. In what follows, we briefly describe some of the related techniques, and then highlight the contributions of the present analysis.

A. Related Work and Motivation

1) Caching: The advantages and applications of caching for cellular systems have been extensively investigated in the recent literature [4], [20]. In this context, several metrics have been proposed and analyzed to characterize the performance of caching algorithms, such as hitting rate [21], outage/coverage probability [22], area spectral efficiency [23], and average throughput [8]. Likewise, cooperative caching in vehicular ad hoc networks (VANETs) was discussed in [24], where it was argued that cooperative caching in VANETs rely on three major components, namely discovery, management and consistency. The main task in the discovery component is to search the presence of a file requested by a vehicle in the network using, e.g., broadcasting techniques such as split caching [25] and shared caching [26]. The next task is carried

out through the management component, which is to decide on (a) whether to cache the requested file, and (b) where to cache it, for a smoother duplicate transmission. Lastly, the task driven by the consistency component is to decide on how long the cached content should be retained in the network.

2) Non-Orthogonal Multiple Access: It is recalled that NOMA is considered as one of the enabling technolgies for 5G [27], [28]. However, the use of NOMA in vehicular networks has received attention only recently, where it has been shown that it significantly outperforms the traditional orthogonal multiple access based systems [29]. Based on this, the spectral efficiency and resource allocation for NOMA based vehicle-to-everything (V2X) systems have been investigated in [30]. Then, graph theory-based encoders and reliable decoders using the belief propagation algorithm for NOMA-based V2X communication systems was addressed in [31], along with the corresponding outage capacity analysis. In [32], the performance of NOMA-enabled vehicle-to-small-cell (V2S), which constitutes a special case of V2I, has been investigated in the context of analyzing the joint optimization of cell association and power control for weighted sum rate maximization. The utility of NOMA along with spatial modulation technique was proposed for V2V MIMO communication in [17], and the achievable throughput as well as the corresponding optimal power allocation policies were derived.

3) Caching-Aided NOMA: To the best of our knowledge, caching-aided NOMA systems have not been investigated in the context of vehicular networks. However, as mentioned earlier, there are some sporadic results reported on the advantages of cache-aided NOMA in cellular systems. For example, it was shown in [33] that exploited pre-cached contents can reduce/mitigate the incurred interference, while employing SIC for NOMA. An optimum power allocation for the considered network is investigated, aiming to maximize the probability of successful decoding of files at each user. However, the simplistic Rayleigh fading conditions were assumed, which is not practically realistic in vehicular networks. Moreover, the model in [33] only considers the full file caching case; that is, the authors assume that in the caching phase, the files are cached as a whole, which is largely restrictive. On the contrary, only a split file caching framework was considered in [34]. The optimum power allocation and the performance of the proposed system are characterized by the achievable rate region. Additionally, the framework in [34] does not take into consideration the detrimental impairments due to encountered fading effects.

B. Contributions

We envision cache-aided NOMA for vehicular networks particularly useful because vehicles can exploit their own cache contents to reduce interference due to signals from other vehicles. That is, for a particular vehicle, if a file is present in its own cache and requested by another vehicle, then it can utilize it during the request phase in order to reduce the interference in the received superimposed NOMA signal. Hence, the performance of a cache-aided NOMA system is improved by exploiting its cache contents. Furthermore, the

interference reduction due to caching also helps to reduce the computational complexity of SIC. Although cache-aided NOMA for cellular networks has received attention in the recent literature [33], [34], the available framework is not directly applicable to vehicular networks. This is largely due the following challenges: i) high mobility; ii) network partitioning, leading to several isolated clusters of nodes; iii) rapidly changing network topology with intermittent connectivity; iv) the harsh propagation environments. The main focus of the present contribution is the propagation environment, where the popular Rayleigh/Nakagami-m fading assumptions are no longer valid, due to the mobility of vehicles and low elevations of their antennas. Consequently, in order to understand the full potential of cache-aided NOMA in vehicular networks, an in-depth analysis of the system performance under realistic channel models is required. To this end, contrary to the pioneering works in this area of research, we consider the cascaded (double) Nakagami-m fading channel model, which provides a realistic description of inter-vehicular channels. This can be justified by the fact that the received signal at a vehicle is due to a large number of signals reflected from statistically independent multiple scatterers [35], [36]. Accordingly, the double Nakagami-m fading model assists in studying the artifacts due to fading as a whole, in vehicular networks with high mobility.

In the present contribution, we introduce cache-aided NOMA for vehicular networks, over cascaded fading channels that follow the the double/cascaded Nakagami$-m$ distribution [35], [36]. Without loss of generality, we consider a two-user network, but extension to a multi-user network is straightforward. In this context, we consider the cases of (i) full file caching, where each vehicle can cache files as a whole during the caching phase [33]; and (ii) split file caching, where caching is performed in parts [34]. For simplicity, we also consider the case of two split files, i.e, files are split into two parts, which can be readily extended to the general case of multiple part file splitting. Under this setup, we characterize the performance of the proposed system in terms of the probability of successful decoding of intended files at both vehicles. In each case, we formulate an optimization problem to find the optimum power profile of each user such that the overall probability of successful decoding is maximized. In the case of full file caching, we analytically show that the cost function is concave. The main contributions of this paper are summarized below:

- We propose cache-aided NOMA as a promising solution to address the spectral efficiency requirements in 5G-enabled vehicular networks.
- We analytically characterize the performance of the proposed cache-aided NOMA system under realistic double Nakagami$-m$ fading conditions in terms of probability of successful decoding of files at each vehicles.
- We consider the case of full file caching, where each vehicle can cache full set of requested files during the caching phase. We formulate an optimization problem to find the optimum power profile of each vehicle, such that the probability of successful decoding of targeted files at
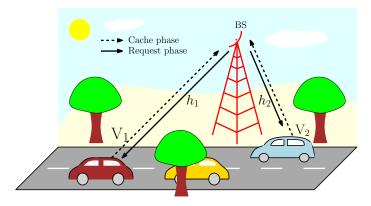

Fig. 1. System model for the proposed caching-aided NOMA vehicular system, where a BS serves vehicles V_1 and V_2 simultaneously.

each vehicle is maximized. Under the effects of double Nakagami$-m$ fading conditions, we further show that the cost function of this optimization problem is concave.
- We consider a practically relevant scenario of split-file caching, where a joint power allocation optimization problem is formulated in order to find the optimum power allocation across vehicles and within each file. Through numerical analysis, we prove that the cost function for this joint optimization problem is concave.
- We quantify the performance enhancement due to the proposed cache-aided NOMA over the conventional NOMA and cache-aided OMA. The impact of the double Nakagami$-m$ distributed fading effects on the performance is also investigated.

To the best of the authors' knowledge, the above results have not been previously reported in the open literature.

C. Organization

The remainder of this paper is organized as follows: The system setup and the cache-aided NOMA architecture are described in Section II. The optimal power allocation problem to maximize the probability of successful decoding of files in a two user vehicular network with full file caching is proposed and discussed in detail in Section III. The principle underlying the joint power allocation in the case where each requested file is split into two parts and cached in order is presented in Section IV. The corresponding numerical results and useful discussions are provided in Section V, while closing remarks are given in Section VI.

II. SYSTEM MODEL

A. Network Architecture

We consider a vehicle-to-everything (V2X) network, which incorporates V2I (vehicle-to-infrastructure), V2V (vehicle-to-vehicle), V2P (vehicle-to-pedestrian), and V2G (vehicle-to-grid) scenarios. As depicted in Fig. 1 and without loss of generality, we consider a base station (BS) serving two vehicles denoted by V_1 and V_2, respectively. Our analysis can also be extended to a network with more than two users, which can be realized by grouping them into pairs in clusters

and assigning orthogonal channels to each cluster. Also, it is assumed that h_1 and h_2 denote the complex channel coefficients between BS–V_1, and BS–V_2, respectively. It is recalled here that in vehicular communication systems[1], the channel gains $g_1 \triangleq |h_1|$ and $g_2 \triangleq |h_2|$ are accurately modeled by a product of two independent cascaded Nakagami-m distributions, i.e, $g_1 \sim \mathcal{N}^2((m_1^{(1)}, \Omega_1^{(1)}), (m_2^{(1)}, \Omega_2^{(1)}))$, and $g_2 \sim \mathcal{N}^2((m_1^{(1)}, \Omega_1^{(1)}), (m_2^{(2)}, \Omega_2^{(2)}))$, respectively, [35], [36]. To this effect, their probability density functions, denoted by $f_{g_1}(g)$ and $f_{g_2}(g)$, are given by [35]

$$f_{g_i}(g) = \frac{2 G_{0,2}^{2,0}\left(\left.\frac{m_1^{(i)} m_2^{(i)} g^2}{\Omega_1^{(i)} \Omega_2^{(i)}}\right| {\overline{} \atop m_1^{(i)}, m_2^{(i)}}\right)}{g \Gamma(m_1^{(i)}) \Gamma(m_2^{(i)})}, \qquad (1)$$

for $i = 1,\ 2$, where $\Gamma(\cdot)$ and $G_{m,n}^{p,q}(\cdot)$ denote the Euler's gamma and Meijer-G functions, respectively [37]. The double Nakagami$-m$ distribution is considered a cascaded fading model and it constitutes a generalization of the double Rayleigh distribution, which is included as a special case [38]. Moreover, the square of the double Nakagami$-m$ distributed random variables, that is, g_1^2 and g_2^2 have double Gamma distribution, and their respective densities are given by a standard transformation of random variables as:

$$f_{g_i^2}(g) = \frac{2 g^{\frac{m_1^{(i)} - m_2^{(i)}}{2} - 1}}{\Gamma(m_1^{(i)}) \Gamma(m_2^{(i)}) \left(\frac{\Omega_1^{(i)} \Omega_2^{(i)}}{m_1^{(i)} m_2^{(i)}}\right)^{\frac{m_1^{(i)} + m_2^{(i)}}{2}}}$$
$$\times \mathcal{K}_{m_1^{(i)} - m_2^{(i)}}\left(2\sqrt{\frac{m_1^{(i)} m_2^{(i)}}{\Omega_1^{(i)} \Omega_2^{(i)}} g}\right), \quad (2)$$

for $i = 1,\ 2$, where $\mathcal{K}_n(\cdot)$ is the modified Bessel function of the second kind, with order $n \in \mathbb{R}$ [37]. Based on this, the corresponding CDFs are given by

$$F_{g_i^2}(g) = \frac{G_{1,3}^{2,1}\left(\left.\frac{m_1^{(i)} m_2^{(i)} g}{\Omega_1^{(i)} \Omega_2^{(i)}}\right| {1 \atop m_1^{(i)}, m_2^{(i)}, 0}\right)}{\Gamma(m_1^{(i)}) \Gamma(m_2^{(i)})}, \qquad (3)$$

for $i = 1,\ 2$. In what follows, we introduce the model on caching.

B. Caching Model and Assumptions

We assume that V_1 and V_2 are equipped with a finite capacity cache of size κ. Let $\mathrm{F} \triangleq \{F_1, F_2, \ldots, F_T\}$ denote the finite set of files available at the BS. The popularity profile on F is then modeled by the popular Zipf distribution [4], with a skewness control parameter $\zeta > 0$.[2] In particular, the popularity of a file $F_t \in \mathrm{F}$ is given by the probability

$$q_t = \frac{1}{t^\zeta \sum_{i=1}^{T} \frac{1}{i^\zeta}}, \quad t = 1, \ldots, T, \qquad (4)$$

[1]We consider a quasi-static fading channel between the BS and each users, typically observed in peak traffic hours [17].
[2]Our analysis is valid even for any meaningful popularity profile. In general, the popularity profile is only used to dictate which files to cache at V_1 and V_2.

such that $q_1 > q_2 > \cdots > q_T > 0$, and $\sum_{t=1}^{T} q_t = 1$. In the *caching phase*, V_1 and V_2 fetch and store a set of files from F, denoted by $\mathrm{F}_{V_1} \subset \mathrm{F}$ and $\mathrm{F}_{V_2} \subset \mathrm{F}$, during the off-peak hours. With the popularity profile considered above, the optimal caching policy would be to cache from the most popular file to the least popular file. In the *requesting phase*, without loss of generality, we assume that users V_1 and V_2 request for files $F_1 \in \mathrm{F}$ and $F_2 \in \mathrm{F}$, respectively. Furthermore, it is assumed that the BS has perfect knowledge of F_{V_1} and F_{V_2}, and that V_1 and V_2 possess knowledge of the power allocation scheme at the BS. For the former assumption, the information can be obtained at the BS during the requesting phase, i.e., when V_1 and V_2 share their cache status. For the latter, the BS can broadcast the information to V_1 and V_2 before communicating their files.

C. Downlink Communication Using NOMA

Let x_1 and x_2 be the signals corresponding to F_1 and F_2. The BS uses the NOMA configuration to send a superimposed signal to both V_1 and V_2, which can decode their respective signals using the successive interference cancellation (SIC) technique. Let y_1 and y_2 be the received signals at V_1 and V_2, respectively, which are given as:

$$y_1 = h_1(\sqrt{\alpha P} x_1 + \sqrt{(1-\alpha)P} x_2) + w_1 \qquad (5)$$

and

$$y_2 = h_2(\sqrt{\alpha P} x_1 + \sqrt{(1-\alpha)P} x_2) + w_2, \qquad (6)$$

where, P is the average transmit power at the BS, $\alpha \in (0,1)$ denotes the fraction of power assigned to V_1, and w_1 and w_2 are independent, circularly symmetric complex Gaussian distributions with zero mean and variance σ_1^2 and σ_2^2, respectively. Concisely, this is denoted as $w_1 \sim \mathcal{CN}(0, \sigma_1^2)$ and $w_2 \sim \mathcal{CN}(0, \sigma_2^2)$. As mentioned earlier, $|h_1| \sim \mathcal{N}^2((m_1^{(1)}, \Omega_1^{(1)}), (m_2^{(1)}, \Omega_2^{(1)}))$ and $|h_2| \sim \mathcal{N}^2((m_1^{(2)}, \Omega_1^{(2)}), (m_2^{(2)}, \Omega_2^{(2)}))$.

It is important to observe that in the considered caching-based NOMA in vehicular networks, each user can decode its own signal by reducing or removing the interference due to the signal of the other user, by using the contents obtained during the caching phase. This increases the probability of successful decoding of the signals of each user. We assume that the files F_1 and F_2 can be decoded successfully by either V_1 or V_2, when the received SINR is greater than $\gamma_1 > 0$ and $\gamma_2 > 0$, respectively. Depending on the files cached by V_1 and V_2 during the caching phase, one of the following four scenarios can occur:

Case A: V_1 has cached F_2, and V_2 has cached F_1.
Case B: V_1 has cached F_2, but V_2 has missed caching F_1.
Case C: V_1 has missed caching F_2, but V_2 has cached F_1.
Case D: V_1 has missed caching F_2, and V_2 has missed caching F_1.

Mathematically, the above scenarios can be represented as:

Case A: $F_2 \in \mathrm{F}_{V_1}$, and $F_1 \in \mathrm{F}_{V_2}$.
Case B: $F_2 \in \mathrm{F}_{V_1}$, and $\{F_1\} \bigcap \mathrm{F}_{V_2} = \Phi$, where Φ is the null set.
Case C: $\{F_2\} \bigcap \mathrm{F}_{V_1} = \Phi$, and $F_1 \in \mathrm{F}_{V_2}$.

Case D: $\{F_2\} \bigcap F_{V_1} = \Phi$, and $\{F_1\} \bigcap F_{V_2} = \Phi$.

To this effect, we intend to find the optimal power allocation policy in each case, such that their respective probabilities of successful decoding of files F_1 and F_2 at V_1 and V_2 are maximized. In the next section, we provide the mathematical details of each case and formulate the corresponding optimization problems.

III. FULL FILE CACHING: OPTIMAL POWER ALLOCATION

In this section, we provide the mathematical details on the optimal values of α for all the above mentioned cases, such that their respective probabilities of successful decoding of F_1 and F_2 at V_1 and V_2 are maximized. Towards this end, we first derive the expressions for the probability of successful decoding of files in each case.

A. Probability of Successful Decoding

1) Case A: When both V_1 and V_2 have each others' files at their cache, irrespective of the value of α, V_1 and V_2 will be able to decode F_1 and F_2 by canceling out the signals due to F_2 and F_1, respectively. Therefore, the probabilities of successful decoding of F_1 and F_2 at V_1 and V_2 are given by

$$\mathrm{P}_{s,V_1}^{(A)} = \mathrm{P}\left\{ \frac{\alpha P g_1^2}{\sigma_1^2} > \gamma_1 \right\} \tag{7}$$

and

$$\mathrm{P}_{s,V_2}^{(A)} = \mathrm{P}\left\{ \frac{(1-\alpha) P g_2^2}{\sigma_2^2} > \gamma_2 \right\} \tag{8}$$

respectively. Recall that γ_1 and γ_2 denote the thresholds on SINR for successful decoding of files F_1 and F_2, at either V_1 or V_2.

2) Case B: As in the previous case, since $F_2 \in F_{V_1}$, V_1 can decode F_1 irrespective of α, with probability

$$\mathrm{P}_{s,V_1}^{(B)} = \mathrm{P}\left\{ \frac{\alpha P g_1^2}{\sigma_1^2} > \gamma_1 \right\}. \tag{9}$$

However, since V_2 has not cached F_1, its probability of successful decoding depends on whether it is near or far from the BS. Each of these cases are considered separately. To this end, when $\alpha \leq 0.5$, that is, when V_2 is far from the BS, it treats the signal from V_1 as interference and decodes F_2 with a success probability

$$\mathrm{P}_{s,V_2}^{(B)} = \mathrm{P}\left\{ \frac{(1-\alpha) P g_2^2}{\alpha P g_2^2 + \sigma_2^2} > \gamma_2 \right\}. \tag{10}$$

On the contrary, when $\alpha > 0.5$, that is when V_2 is closer, it first decodes F_1 and then F_2, following SIC. The success probability in this case is given by

$$\mathrm{P}_{s,V_2}^{(B)} = \mathrm{P}\left\{ \frac{\alpha P g_2^2}{(1-\alpha) P g_2^2 + \sigma_2^2} > \gamma_1 \right\}$$
$$\times \mathrm{P}\left\{ \frac{(1-\alpha) P g_2^2}{\sigma_2^2} > \gamma_2 \right\}. \tag{11}$$

3) Case C: Similar to the case B, irrespective of the value of α, V_2 decodes F_2 successfully with a probability

$$\mathrm{P}_{s,V_2}^{(C)} = \mathrm{P}\left\{ \frac{(1-\alpha) P g_2^2}{\sigma_2^2} > \gamma_2 \right\}. \tag{12}$$

On the contrary, when $\alpha > 0.5$, V_1 decodes F_1 successfully with a probability

$$\mathrm{P}_{s,V_1}^{(C)} = \mathrm{P}\left\{ \frac{\alpha P g_1^2}{(1-\alpha) P g_1^2 + \sigma_1^2} > \gamma_1 \right\}, \tag{13}$$

and with a corresponding success probability of

$$\mathrm{P}_{s,V_1}^{(C)} = \mathrm{P}\left\{ \frac{(1-\alpha) P g_1^2}{\alpha P g_1^2 + \sigma_1^2} > \gamma_2 \right\} \mathrm{P}\left\{ \frac{\alpha P g_1^2}{\sigma_1^2} > \gamma_1 \right\}, \tag{14}$$

when $\alpha \leq 0.5$.

4) Case D: In this case, when $\alpha > 0.5$, V_1 is able to decode F_1 successfully with probability

$$\mathrm{P}_{s,V_1}^{(D)} = \mathrm{P}\left\{ \frac{\alpha P g_1^2}{(1-\alpha) P g_1^2 + \sigma_1^2} > \gamma_1 \right\}, \tag{15}$$

while V_2 has a success probability of

$$\mathrm{P}_{s,V_2}^{(D)} = \mathrm{P}\left\{ \frac{\alpha P g_2^2}{(1-\alpha) P g_2^2 + \sigma_2^2} > \gamma_1 \right\}$$
$$\times \mathrm{P}\left\{ \frac{(1-\alpha) P g_2^2}{\sigma_2^2} > \gamma_2 \right\}. \tag{16}$$

Finally, when $\alpha \leq 0.5$, V_2 decodes F_2 with probability

$$\mathrm{P}_{s,V_2}^{(D)} = \mathrm{P}\left\{ \frac{(1-\alpha) P g_2^2}{\alpha P g_2^2 + \sigma_2^2} > \gamma_2 \right\}, \tag{17}$$

while the probability of success at V_1 is expressed as

$$\mathrm{P}_{s,V_1}^{(D)} = \mathrm{P}\left\{ \frac{(1-\alpha) P g_1^2}{\alpha P g_1^2 + \sigma_1^2} > \gamma_2 \right\} \mathrm{P}\left\{ \frac{\alpha P g_1^2}{\sigma_1^2} > \gamma_1 \right\}. \tag{18}$$

B. Power Allocation

The optimal power allocation problems in all the considered cases is described below, recalling that in each case, the probability of successful decoding of F_1 and F_2 is the product of individual success probabilities.

1) Case A: As seen earlier for case A, the probability of successful decoding is independent of α. Therefore, the optimization problem for case A would be

$$\mathrm{OP}_A : \max \ \mathrm{P}_{s,V_1}^{(A)} \mathrm{P}_{s,V_2}^{(A)}$$
$$\text{s.t.} \ \ 0 \leq \alpha \leq 1. \tag{19}$$

Next, we prove that the cost function in (19) is concave in $0 \leq \alpha \leq 1$, through the following proposition. The concavity of the probability of success and the details corresponding to the other cases lead to lengthy expressions, and are omitted for brevity.

Proposition 1. *The function* $\mathrm{P}_s^{(A)} \triangleq \mathrm{P}_{s,V_1}^{(A)} \mathrm{P}_{s,V_2}^{(A)}$ *is concave in* $0 \leq \alpha \leq 1$.

Proof. See Appendix. $\square$

A similar approach can be employed to establish the concavity of the cost function in the other optimization problems.

2) Case B: For case B, recall that two sub-cases occur. For the case with $\alpha > 0.5$, the requirement from $\text{P}_{s,V_2}^{(B)}$, namely,

$$\frac{\alpha P g_2^2}{(1-\alpha)P g_2^2 + \sigma_2^2} > \gamma_1 \qquad (20)$$

i.e.

$$\frac{\sigma_1^2}{\alpha P \left[\frac{1}{\gamma_1} - 1\right] - P} > 0 \qquad (21)$$

yields the condition $\alpha > \frac{\gamma_1}{1+\gamma_1}$, and $\text{P}_{s,V_2}^{(B)}$ also results in the requirement that $\alpha \leq 1$. Since $\gamma_2 > 0$, we obtain the following optimization problem:

$$\text{OP}_B^{(\alpha>0.5)} : \max \quad \text{P}_{s,V_1}^{(B)} \text{P}_{s,V_2}^{(B)}$$
$$\text{s.t.} \quad \frac{\gamma_1}{1+\gamma_1} \leq \alpha \leq 1, \qquad (22)$$

Similarly, when $\alpha \leq 0.5$, the condition

$$\frac{(1-\alpha)P g_2^2}{\alpha P g_2^2 + \sigma_2^2} > \gamma_2 \qquad (23)$$

i.e.

$$\frac{\sigma_2^2}{\frac{P}{\gamma_2^2} - \alpha P \left(\frac{1}{\gamma_2} + 1\right)} > 0 \qquad (24)$$

dictates that $\alpha \leq \frac{1}{1+\gamma_2}$. In this case, the optimization problem becomes

$$\text{OP}_B^{(\alpha\leq0.5)} : \max \quad \text{P}_{s,V_1}^{(B)} \text{P}_{s,V_2}^{(B)}$$
$$\text{s.t.} \quad 0 \leq \alpha \leq \frac{1}{1+\gamma_2}. \qquad (25)$$

with the requirement $\alpha > 0$ from $\text{P}_{s,V_1}^{(B)}$.

3) Case C: The individual optimization problems corresponding to $\alpha > 0.5$ and $\alpha \leq 0.5$ are respectively given by

$$\text{OP}_C^{(\alpha>0.5)} : \max \quad \text{P}_{s,V_1}^{(C)} \text{P}_{s,V_2}^{(C)}$$
$$\text{s.t.} \quad \frac{\gamma_1}{1+\gamma_1} \leq \alpha \leq 1, \qquad (26)$$

$$\text{OP}_C^{(\alpha\leq0.5)} : \max \quad \text{P}_{s,V_1}^{(C)} \text{P}_{s,V_2}^{(C)}$$
$$\text{s.t.} \quad 0 \leq \alpha \leq \frac{1}{1+\gamma_2}. \qquad (27)$$

4) Case D: Finally, the optimization problem for case D for $\alpha > 0.5$ and $\alpha \leq 0.5$ can be formulated as

$$\text{OP}_D^{(\alpha>0.5)} : \max \quad \text{P}_{s,V_1}^{(D)} \text{P}_{s,V_2}^{(D)}$$
$$\text{s.t.} \quad \frac{\gamma_1}{1+\gamma_1} \leq \alpha \leq 1, \qquad (28)$$

$$\text{OP}_D^{(\alpha\leq0.5)} : \max \quad \text{P}_{s,V_1}^{(D)} \text{P}_{s,V_2}^{(D)}$$
$$\text{s.t.} \quad 0 \leq \alpha \leq \frac{1}{1+\gamma_2}. \qquad (29)$$

Next, we extend the above framework to the case where the individual files F_1 and F_2 are split into two parts, and each part is transmitted simultaneously.

IV. SPLIT FILE CACHING: OPTIMAL POWER ALLOCATION

In this section, we consider the case where the files F_1 and F_2 are split further into sub-files. Without loss of generality, we consider the simplest case where each file is split into two sub-files. This analysis can then be readily extended to the case of splitting a file to greater than two sub-files. As shown in Fig. 2, the sub-files of F_1 and F_2 are denoted by $F_1^{(1)}, F_1^{(2)}$, and $F_2^{(1)}, F_2^{(2)}$, respectively. It is assumed that the BS needs β fraction of the assigned powers to transmit $F_\ell^{(1)}$, $\ell = 1, 2$ and $1 - \beta$ fraction of the powers to transmit $F_\ell^{(2)}$, $\ell = 1, 2$, respectively. Moreover, let the signals corresponding to $F_1^{(1)}, F_1^{(2)}, F_2^{(1)}$, and $F_2^{(2)}$ be denoted by $x_1^{(1)}, x_1^{(2)}, x_2^{(1)}$, and $x_2^{(2)}$, respectively. Based on this, the superimposed signal transmitted by the BS is represented as

$$x = \sqrt{\beta\alpha P}x_1^{(1)} + \sqrt{(1-\beta)\alpha P}x_1^{(2)}$$
$$+ \sqrt{\beta(1-\alpha)P}x_2^{(1)} + \sqrt{(1-\beta)(1-\alpha)P}x_2^{(2)}. \qquad (30)$$

Let the SINR constraint for decoding the files $F_1^{(1)}, F_1^{(2)}, F_2^{(1)}$, and $F_2^{(2)}$ be given by $\gamma_1^{(1)}, \gamma_1^{(2)}, \gamma_2^{(1)}$, and $\gamma_2^{(2)}$, respectively. Depending on the file portions cached by V_1 and V_2, several classes and sub-cases are conceivable. An example of one such class includes the following sub-cases:

(a) V_1 has cached $F_2^{(1)}$ and V_2 has cached $F_1^{(1)}$.
(b) V_1 has not cached any portions of F_2 and V_2 has cached $F_1^{(1)}$.
(c) V_1 has cached $F_2^{(1)}$ and V_2 has not cached any portions of F_1.
(d) Neither V_1 has cached any portions of F_2, nor V_2 has cached any portions of F_1.

In the following, we restrict our attention to the first scenario, which is illustrated in Fig. 2, among such several other cases which can be defined accordingly [34]. Our framework can also be similarly extended to the other cases. For the case at hand, the signals received at V_1 and V_2, after cancellation of pre-cached signals are given by

$$y_1 = h_1(\sqrt{\beta\alpha P}x_1^{(1)} + \sqrt{(1-\beta)\alpha P}x_1^{(2)}$$
$$+ \sqrt{(1-\beta)(1-\alpha)P}x_2^{(2)}) + w_1 \qquad (31)$$

and

$$y_2 = h_2(\sqrt{(1-\beta)\alpha P}x_1^{(2)} + \sqrt{\beta(1-\alpha)P}x_2^{(1)}$$
$$+ \sqrt{(1-\beta)(1-\alpha)P}x_2^{(2)}) + w_2. \qquad (32)$$

For the above case where V_1 has cached $F_2^{(1)}$ and V_2 has cached $F_1^{(1)}$, that is, $F_2^{(1)} \in \text{F}_{V_1}$, and $F_1^{(1)} \in \text{F}_{V_2}$, two sub-cases arise depending on whether $\alpha \leq 0.5$ or $\alpha > 0.5$. For the case when $\alpha > 0.5$, the far user V_1 will first decode $F_1^{(1)}$, with probability

$$\text{P}_{s,V_1}^{(a,1)} = \text{P}\left\{\frac{\alpha\beta P g_1^2}{(1-\beta)P g_1^2 + \sigma_1^2} > \gamma_1^{(1)}\right\}, \qquad (33)$$

and will decode $F_1^{(2)}$ later, with probability

$$\text{P}_{s,V_1}^{(a,2)} = \text{P}\left\{\frac{\alpha(1-\beta)P g_1^2}{(1-\alpha)(1-\beta)P g_1^2 + \sigma_1^2} > \gamma_1^{(2)}\right\}. \qquad (34)$$
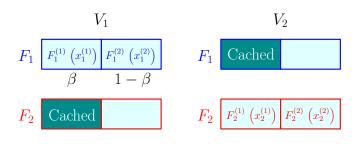
Fig. 2. File split-up model for transmission and caching. The right hand side of the figure shows the case where $F_2^{(1)} \in \mathrm{F}_{V_1}$ and $F_1^{(1)} \in \mathrm{F}_{V_2}$.

Parameters	Settings
$m_1^{(1)}, m_2^{(1)}$	1, 1
$m_1^{(2)}, m_2^{(2)}$	1, 1
$\Omega_1^{(1)}, \Omega_2^{(1)}$	2, 2
$\Omega_1^{(2)}, \Omega_2^{(2)}$	2, 2
Path loss exponent d	2
Distances of V_1 and V_2 from BS	1, 0.5
Files at BS, T	5
Cache size at V_1 and V_2	1
Noise variance σ_1^2	1
Noise variance σ_2^2	1
SINR threshold γ_1	1
SINR threshold γ_2	1
Zipf parameter ζ	0.5

On the contrary, V_2, which is the near user, employs SIC and first decodes $F_1^{(2)}$ with probability

$$P_{s,V_2}^{(a,1)} = P\left\{\frac{\alpha(1-\beta)Pg_2^2}{(1-\alpha)Pg_2^2 + \sigma_2^2} > \gamma_1^{(2)}\right\}, \qquad (35)$$

then decodes $F_2^{(1)}$ with probability

$$P_{s,V_2}^{(a,2)} = P\left\{\frac{\beta(1-\alpha)Pg_2^2}{(1-\alpha)(1-\beta)Pg_2^2 + \sigma_2^2} > \gamma_2^{(1)}\right\}, \qquad (36)$$

followed by decoding $F_2^{(2)}$ with probability

$$P_{s,V_2}^{(a,3)} = P\left\{\frac{(1-\beta)(1-\alpha)Pg_2^2}{\sigma_2^2} > \gamma_2^{(2)}\right\}. \qquad (37)$$

Similarly, for the case when $\alpha \leq 0.5$, the far user V_2 will decode $F_2^{(1)}$ first, with probability

$$P_{s,V_2}^{(a,1)} = P\left\{\frac{(1-\alpha)\beta Pg_2^2}{(1-\beta)Pg_2^2 + \sigma_2^2} > \gamma_2^{(1)}\right\}, \qquad (38)$$

followed by $F_2^{(2)}$ with probability

$$P_{s,V_2}^{(a,2)} = P\left\{\frac{(1-\alpha)(1-\beta)Pg_2^2}{\alpha(1-\beta)Pg_2^2 + \sigma_2^2} > \gamma_2^{(2)}\right\}. \qquad (39)$$

Lastly, the near user V_1 employs SIC to decode $F_2^{(2)}$ first with probability

$$P_{s,V_1}^{(a,1)} = P\left\{\frac{(1-\alpha)(1-\beta)Pg_1^2}{\alpha Pg_1^2 + \sigma_1^2} > \gamma_2^{(2)}\right\}, \qquad (40)$$

then decodes $F_1^{(1)}$ with probability

$$P_{s,V_1}^{(a,2)} = P\left\{\frac{\alpha\beta Pg_1^2}{\alpha(1-\beta)Pg_1^2 + \sigma_1^2} > \gamma_1^{(1)}\right\}, \qquad (41)$$

followed by finally decoding $F_1^{(2)}$ with probability

$$P_{s,V_1}^{(a,3)} = P\left\{\frac{\alpha(1-\beta)Pg_1^2}{\sigma_1^2} > \gamma_1^{(2)}\right\}. \qquad (42)$$

A. Power Allocation

Following the above discussion and the framework considered in Sec. III, we propose the following optimization problem to jointly design the power allocation factors α and β, such that the overall probability of successful decoding is maximized. That is, for the case when $\alpha > 0.5$, the optimization problem is formulated as

$$\mathrm{OP}_a : \max_{\alpha,\beta} \; P_{s,V_1}^{(a,1)}P_{s,V_1}^{(a,2)}P_{s,V_2}^{(a,1)}P_{s,V_2}^{(a,2)}P_{s,V_2}^{(a,3)}$$
$$\text{s.t.} \quad 0.5 \leq \alpha \leq 1,$$
$$0 \leq \beta \leq 1 \qquad (43)$$

whereas for the case of $\alpha \leq 0.5$, the optimization problem is formulated as

$$\mathrm{OP}_a : \max_{\alpha,\beta} \; P_{s,V_1}^{(a,1)}P_{s,V_1}^{(a,2)}P_{s,V_1}^{(a,3)}P_{s,V_2}^{(a,1)}P_{s,V_2}^{(a,2)}$$
$$\text{s.t.} \quad 0 \leq \alpha \leq 0.5,$$
$$0 \leq \beta \leq 1. \qquad (44)$$

It is not an easy task to analytically characterize the optimization problems given above, due to the complicated expressions in (2). Therefore, we resort to numerical techniques, where we observe that the probability of successful decoding is concave with respect to the tuple (α, β) in their respective range. The optimal pair (α^*, β^*) can be obtained through techniques such as steepest ascent, and other search algorithms. This point is further elaborated in Sec. V.

V. NUMERICAL RESULTS

In this section, we discuss the performance of the proposed cache-aided NOMA, and compare its performance to conventional NOMA and cache-aided orthogonal multiple access (OMA) systems. First, we consider the full file caching scenario studied in Sec. III. Unless stated otherwise, the parameters used for our simulations are listed in Tab. I. Since the popularity profile is modeled by the Zipf distribution, the optimal caching policy would be to cache the most popular files, depending on the cache size of V_1 and V_2.

Fig. 3 shows the performance of the probability of successful decoding, i.e., $P_{s,V_1}^{(A)}P_{s,V_2}^{(A)}$ with respect to α, for the cases A, B, C and D. It is noticed that for each case, as expected, the cost function is concave in α. In particular for case A,
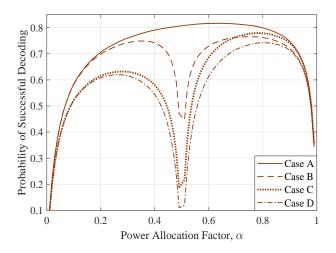
Fig. 3. The cost functions in the associated optimization problems are shown to be concave.

Fig. 5. Performance comparison of caching-aided NOMA and caching-aided OMA, in terms of variation of probability of successful decoding with ζ, for different values of cache sizes available at V_1 and V_2.
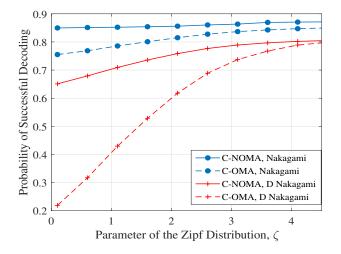

Fig. 4. Performance comparison of caching-aided NOMA (denoted by C-NOMA) and caching-aided OMA (denoted by C-OMA), in terms of variation of probability of successful decoding with different values of the Zipf distribution parameter ζ, under Nakagami and double Nakagami distributions.

Fig. 6. Performance comparison of caching-aided NOMA and caching-aided OMA, in terms of variation of probability of successful decoding with SNR, for different values of the Zipf parameter ζ.

as expected from Lemma 1, the probability of successful decoding is concave in the entire range of α. The optimal α in each case and sub-cases is found through the bisection method [39, Alg. 4.1].

Fig. 4 compares the performances of cache-aided NOMA and cache-aided OMA systems under Nakagami$-m$ and double Nakagami$-m$ fading conditions, which are considered suitable in vehicle-to-vehicle communications. To this effect, different values of ζ are considered, with an SNR of 10 dB. In both cases, NOMA exhibits a better performance compared to the OMA counterpart. It is evident though that for both NOMA and OMA, the performance under the double Nakagami$-m$ fading conditions is inferior compared to that encountered under of conventional Nakagami$-m$ fading, which is expected due to the cascaded nature of the double Nakagami$-m$ model. However, the magnitude of degradation in OMA is severe compared to NOMA, especially in the regime where $\zeta \to 0$,
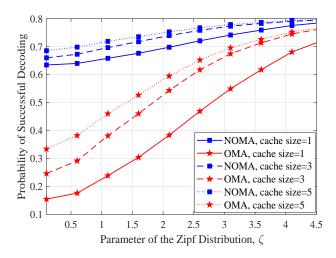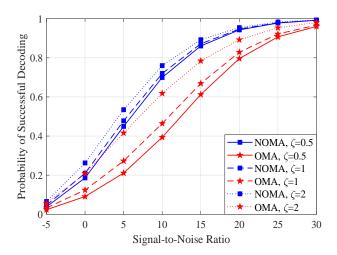
which is of practical relevance. Therefore, the assumption of Nakagami$-m$ fading model yields an unrealistic upper bound on the performances of cache-aided NOMA and cache-aided OMA, which is over-optimistic in the latter case.

Fig. 5 compares the performances of cache-aided NOMA and cache-aided OMA systems for different values of ζ and cache sizes, available at V_1 and V_2. The received SNR is fixed at 10 dB and it is observed that a large cache at V_1 and V_2 improves the performance of NOMA and OMA, since the probability of caching the files requested by another vehicle increases with the cache size. Once again, for lower values of ζ, it is observed that the NOMA system offers significant performance gain compared to the OMA counterpart, while it also exhibits a slower performance degradation with ζ.

Fig. 6 depicts the performance of proposed scheme and compare it with cache-aided OMA for different Zipf parameter ζ values. As expected, the proposed cache-aided NOMA offers
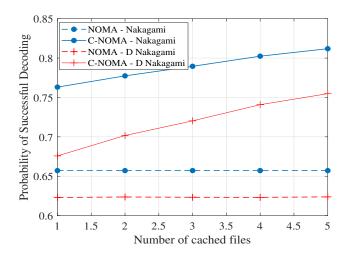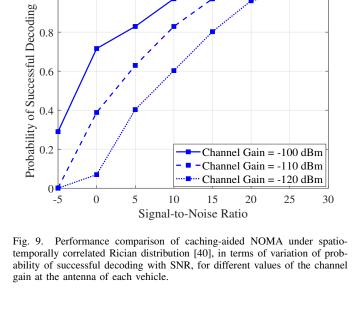
Fig. 7. Performance comparison of caching-aided NOMA, in terms of variation of probability of successful decoding for different number of cached files, under Nakagami$-m$ and double Nakagami$-m$ distributions.

Fig. 9. Performance comparison of caching-aided NOMA under spatio-temporally correlated Rician distribution [40], in terms of variation of probability of successful decoding with SNR, for different values of the channel gain at the antenna of each vehicle.
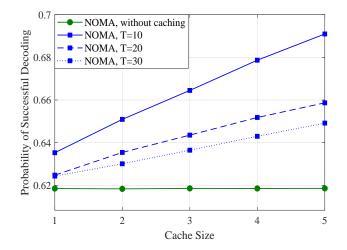

Fig. 8. Performance comparison of caching-aided NOMA and conventional NOMA, in terms of variation of probability of successful decoding with different cache sizes at each vehicle, for different number of files available at the BS.
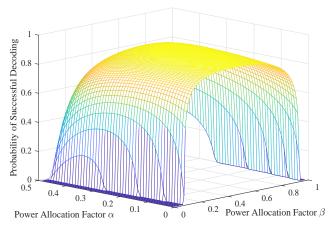

Fig. 10. Performance comparison of split file caching-aided NOMA for the cost function given in (44) with SNR, for different values of power allocation factors α and β.

a better performance for all ζ values. Note that the marginal performance improvement also matches with the behavior reported in [33]. Moreover, since a higher value of ζ makes the popularity profile more skewed towards the first file, the performances of both NOMA and OMA increase with an increase in ζ. This is because the probability of occurrence of case A increases with ζ, which dominates the average performance. The performance improvement offered due to cache-aided NOMA is significant for low ζ, in comparison to OMA.

Fig. 7 compares the performances of cache-aided NOMA and cache-aided OMA, for different number of cached files at V_1 and V_2. Once again, the achieved performance by NOMA outperforms that of OMA and, as expected, increases with an increase in the number of cached files. Also, the achieved performance due to the double Nakagami$-m$ is outperformed

by that of of the conventional Nakagami fading$-m$, as expected. Similarly, Fig. 8 shows the impact of cancellation of interference due to caching on the proposed NOMA system. It is seen that in comparison with the conventional NOMA system, the proposed cache-aided NOMA system offers a better performance as the cache size at V_1 and V_2 increases. Moreover, the performance also improves for lower values of the total files available at the BS, T. This is expected, since a lesser T and a larger cache size improves the probability of caching the requested file in the caching phase. Moreover, when the cache size at V_1 and V_2 is zero, that is, when V_1 and V_2 do not cache any files in the caching phase, the performances of conventional NOMA and the caching-aided NOMA are equal.

In Fig. 9, we investigate the performance of the caching-aided NOMA in a vehicular network with high mobility, where

the statistics of the channel is time-varying and has spatial correlation. Among several spatio-temporally correlated fading models available in the literature [17], [40], [41], we consider the spatio-temporally correlated Rician model, discussed in [40]. In particular, we assume a single antenna on the BS and on each vehicle, with multiple scatters and the covariance matrix of the observations from each clutter modeled by the approximate Gaussian local scattering model [40]. The model also includes the impairments due to the large scale fading. The Rician parameter for each vehicle is set to unity. The distances between each vehicle and the BS are chosen such that the channel gain is set to a desired value. As shown in Fig. 9, the probability of successful decoding increases with an increase in channel gain, as expected. In addition to the above result, it is observed that the probability of successful decoding depends on the value Rician parameter, and the degree of correlation over time and space. A detailed study on the impact of time-varying nature of the channel on the performance of the cache-aided NOMA is reserved as a future comprehensive analysis and contribution.
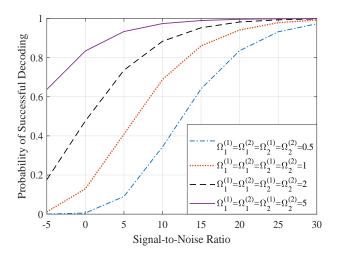

Fig. 12. Performance comparison of split file caching-aided NOMA for the cost function given in (44) with SNR, for different values of the parameters $m_1^{(1)} = m_1^{(2)} = m_2^{(1)} = m_2^{(2)}$, and $\Omega_1^{(1)} = \Omega_1^{(2)} = \Omega_2^{(1)} = \Omega_2^{(2)} = 2$.
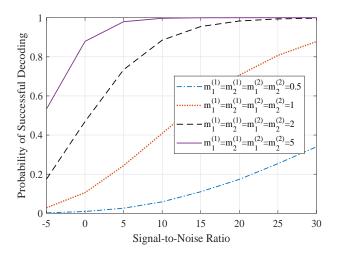

Fig. 11. Performance comparison of split file caching-aided NOMA for the cost function given in (44) with SNR, for different values of the parameters $\Omega_1^{(1)} = \Omega_1^{(2)} = \Omega_2^{(1)} = \Omega_2^{(2)}$, and $m_1^{(1)} = m_1^{(2)} = m_2^{(1)} = m_2^{(2)} = 1$.

Next, we turn our attention to the case of split file caching discussed in Sec. IV. In Fig. 10, we plot the probability of successful decoding given in the cost function of the optimization problem in (44), for different values of power allocation parameters $\alpha \in (0, 0.5)$ and $\beta \in (0, 1)$. The SINR threshold values are set to $\gamma_1^{(1)} = \gamma_2^{(1)} = \gamma_1^{(2)} = \gamma_2^{(2)} = 0.25$. It is observed that the cost function is concave in both α and β. Therefore, the corresponding optimal parameter values α^* and β^* can be found by techniques such as the bisection method [39, Alg. 4.1].

Finally, we study the effect of the fading parameters on the proposed scheme. For simplicity, we again consider the cost function in (44), where $\alpha \leq 0.5$. The performance variation of the proposed scheme is shown in Fig. 11, for different values of the Nakagami$-m$ parameter Ω. In this scenario, we assume that $\Omega_1^{(1)} = \Omega_1^{(2)} = \Omega_2^{(1)} = \Omega_2^{(2)} = \Omega$, and $m_1^{(1)} = m_1^{(2)} = m_2^{(1)} = m_2^{(2)} = 1$, which coincides with the corresponding Rayleigh case. It is also noted that since the average value of the fading distribution increases with an increase in Ω, the optimal probability of successful decoding also increases. A similar trend can be also observed in Fig. 12, where the parameter Ω is fixed to $\Omega_1^{(1)} = \Omega_1^{(2)} = \Omega_2^{(1)} = \Omega_2^{(2)} = 2$ and the parameters $m_1^{(1)} = m_1^{(2)} = m_2^{(1)} = m_2^{(2)} = m$ are varied. It is evident that, as expected, the performance of the proposed technique increases with an increase in the parameter m.

VI. Conclusions and Future Work

This work introduced cache-aided NOMA in vehicular networks and quantified the achievable performance for different scenarios of interest. First, we considered the case of full file caching, where each vehicle caches and requests the whole files. This was realized for the case of a two-user vehicular network wireless transmission over double Nakagami$-m$ fading channels. To this effect, we determined the optimum power allocation that maximizes the overall probability of successful decoding of files at each vehicle. Then, we considered the case of split file caching, where each file is partitioned into two parts. A joint power allocation optimization problem was formulated, where the power allocation across vehicles and for a single file was determined. In both cases, it was shown that the associated cost function is concave in the power allocation variables. Furthermore, a performance gain was exhibited for the cache-aided NOMA compared to the conventional NOMA counterpart.

As a part of the future work, several generalizations for the split file caching in Sec. IV are possible. For example, one may consider the case of splitting of files with sizes more than two. A more realistic model to consider is the spatio-temporally correlated fading channel [17], [40], given the mobility of both transmitter and receiver. In particular, similar to the Kronecker correlation-based model discussed in [17], and the tunnel model considered in [41], a spatio-temporally correlated cascaded Nakagami model can be developed for the setup considered in this work.

$$\frac{d^2}{d\alpha^2}\left[\alpha\, G_{2,4}^{2,2}\left(\frac{\xi_1}{\alpha}\,\middle|\,{}^{0,1}_{m_1^{(1)},m_2^{(1)},1,0}\right)\right] = \mathcal{K}_{m_2^{(1)}-m_1^{(1)}}\left(2\sqrt{\frac{\xi_1}{\alpha}}\right)\left\{2\left(\frac{\xi_1}{\alpha}\right)^{\frac{m_2^{(1)}+m_1^{(1)}}{2}} - \frac{\xi_1(m_1^{(1)}+m_2^{(1)})}{\alpha}\left(\frac{\xi_1}{\alpha}\right)^{\frac{m_1^{(1)}+m_2^{(1)}}{2}-1}\right\}$$

$$-\left(\frac{\xi_1}{\alpha}\right)^{\frac{m_1^{(1)}+m_2^{(1)}-1}{2}}\left[-\mathcal{K}_{m_2^{(1)}-m_1^{(1)}-1}\left(2\sqrt{\left(\frac{\xi_1}{\alpha}\right)}\right) - \mathcal{K}_{m_2^{(1)}-m_1^{(1)}+1}\left(2\sqrt{\left(\frac{\xi_1}{\alpha}\right)}\right)\right] \quad (46a)$$

APPENDIX: PROOF OF THEOREM 1

We show that the second derivative of $\mathrm{P}_s^{(A)}$ with respect to α is negative in $0 \le \alpha \le 1$. It is recalled that $\mathrm{P}_{s,V_1}^{(A)}$ and $\mathrm{P}_{s,V_2}^{(A)}$ are given by (2), for $g = \frac{\gamma_1\sigma_1^2}{\alpha P}$ and $g = \frac{\gamma_2\sigma_2^2}{(1-\alpha)P}$, respectively. It is also recalled that

$$\frac{d^u}{dz^u}G_{p,q}^{m,n}\left(\frac{1}{z}\,\middle|\,{}^{a_1,\dots,a_n,a_{n+1},\dots,a_p}_{b_1,\dots,b_m,b_{m+1},\dots,b_q}\right) = (-1)^u z^u$$

$$G_{p+1,q+1}^{m,n+1}\left(\frac{1}{z}\,\middle|\,{}^{1-u,a_1,\dots,a_n,a_{n+1},\dots,a_p}_{b_1,\dots,b_m,1,b_{m+1},\dots,b_q}\right), u = 1,2,\cdots. \quad (45)$$

Based on this, it is noted that the term corresponding to $\mathrm{P}_{s,V_1}^{(A)}$ can be expressed as

$$\frac{d}{d\alpha}\left[\frac{G_{1,3}^{2,1}\left(\frac{m_1^{(1)}m_2^{(1)}\gamma_1\sigma_1^2}{\Omega_1^{(1)}\Omega_2^{(1)}P\alpha}\,\middle|\,{}^{1}_{m_1^{(1)},m_2^{(1)},0}\right)}{\Gamma(m_1^{(1)})\Gamma(m_2^{(1)})}\right] = \frac{-\frac{\Omega_1^{(1)}\Omega_2^{(1)}P\alpha}{m_1^{(1)}m_2^{(1)}\gamma_1\sigma_1^2}}{\Gamma(m_1^{(1)})\Gamma(m_2^{(1)})}$$

$$\times G_{2,4}^{2,2}\left(\frac{m_1^{(1)}m_2^{(1)}\gamma_1\sigma_1^2}{\Omega_1^{(1)}\Omega_2^{(1)}P\alpha}\,\middle|\,{}^{0,1}_{m_1^{(1)},m_2^{(1)},1,0}\right). \quad (46)$$

Now, let us define

$$\xi_1 \triangleq \frac{m_1^{(1)}m_2^{(1)}\gamma_1\sigma_1^2}{\Omega_1^{(1)}\Omega_2^{(1)}P} > 0. \quad (47)$$

Also, since $m_1 > 0$ and $m_2 > 0$, it follows that

$$G_{2,4}^{2,2}\left(\frac{\xi_1}{\alpha}\,\middle|\,{}^{0,1}_{m_1^{(1)},m_2^{(1)},1,0}\right) = 2\left(\frac{\xi}{\alpha}\right)^{\frac{m_1^{(1)}+m_2^{(1)}}{2}}$$

$$\times \mathcal{K}_{m_2^{(1)}-m_1^{(1)}}\left(2\sqrt{\frac{\xi_1}{\alpha}}\right). \quad (48)$$

Evidently, by simplifying (46), substituting (48) and finding the second derivative yields a term which can be simplified as given in (46a), at the top of this page. Similar expressions can be derived for the first and second derivatives of $\mathrm{P}_{s,V_2}^{(A)}$. In addition, utilizing the Bessel function identity:

$$\mathcal{K}_{-\nu}(x) = \mathcal{K}_\nu(x), \quad \nu = 0,1,\cdots, \quad (49)$$

assists in simplifying $\mathcal{K}_{m_2^{(1)}-m_1^{(1)}}(\cdot)$ irrespective of whether $m_1^{(1)}$ is greater than or lesser than $m_2^{(1)}$. It is also noted that the following inequality holds:

$$\mathcal{K}_\mu(x) \ge \mathcal{K}_\nu(x), \quad \mu,\nu = 0,1,\cdots, \quad x \in \mathbb{R}^+. \quad (50)$$

Simplifying the first and second derivatives of both $\mathrm{P}_{s,V_1}^{(A)}$ and $\mathrm{P}_{s,V_2}^{(A)}$ using the above results, it can be shown that the second derivative of $\mathrm{P}_s^{(A)}$ is negative, and their respective first derivatives can be shown to have positive and negative values in the limiting cases as $\alpha \to 0$ and $\alpha \to 1$, respectively. Based on this and after some algebraic manipulations, $\mathrm{P}_s^{(A)}$ is shown to be concave in $0 \le \alpha \le 1$, which completes the proof.

ACKNOWLEDGMENT

REFERENCES

[1] S. Tuohy, M. Glavin, C. Hughes, E. Jones, M. Trivedi, and L. Kilmartin, "Intra-vehicle networks: A review," *IEEE Trans. Intellig. Transp. Sys.*, vol. 16, no. 2, pp. 534–545, Apr. 2015.

[2] M. Gerla, E. Lee, G. Pau, and U. Lee, "Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds," in *Proc. IEEE World Forum on Internet of Things (WF-IoT)*, Mar. 2014, pp. 241–246.

[3] A. Vinel, "3GPP LTE versus IEEE 802.11p/WAVE: Which technology is able to support cooperative vehicular safety applications?" *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 125–128, Apr. 2012.

[4] L. Li, G. Zhao, and R. S. Blum, "A survey of caching techniques in cellular networks: Research issues and challenges in content placement and delivery strategies," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1710–1732, Third Quarter 2018.

[5] M. A. Maddah-Ali and U. Niesen, "Coding for caching: fundamental limits and practical challenges," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 23–29, Aug. 2016.

[6] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 131–145, Jan. 2016.

[7] Y. Cui and D. Jiang, "Analysis and optimization of caching and multi-casting in large-scale cache-enabled heterogeneous wireless networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 250–264, Jan. 2017.

[8] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless D2D networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 849–869, Feb. 2016.

[9] M. Gregori, J. Gómez-Vilardeb, J. Matamoros, and D. Gündüz, "Wireless content caching for small cell and D2D networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1222–1234, May 2016.

[10] R. Tandon and O. Simeone, "Cloud-aided wireless networks with edge caching: Fundamental latency trade-offs in fog radio access networks," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp. 2029–2033.

[11] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[12] D. Wan, M. Wen, F. Ji, H. Yu, and F. Chen, "Non-orthogonal multiple access for cooperative communications: Challenges, opportunities, and trends," *IEEE Trans. Wireless Commun.*, vol. 25, no. 2, pp. 109–117, Apr. 2018.

[13] S. M. R. Islam, M. Zeng, O. A. Dobre, and K. Kwak, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Trans. Wireless Commun.*, vol. 25, no. 2, pp. 40–47, Apr. 2018.

[14] M. F. Kader, M. B. Uddin, S. R. Islam, and S. Y. Shin, "Capacity and outage analysis of a dual-hop decode-and-forward relay-aided NOMA scheme," *Digital Signal Processing*, vol. 88, pp. 138 – 148, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1051200418303968

[15] S. M. R. Islam, M. Zeng, and O. Dobre, "NOMA in 5G systems: Exciting possibilities for enhancing spectral efficiency," *IEEE 5G Tech Focus*, vol. 1, no. 2, Jun. 2017.

[16] H. Marshoud, S. Muhaidat, P. C. Sofotasios, S. Hussain, M. A. Imran, and B. S. Sharif, "Optical non-orthogonal multiple access for visible light communication," *IEEE Trans. Wireless Commun.*, vol. 25, no. 2, pp. 82–88, Apr. 2018.

[17] Y. Chen, L. Wang, Y. Ai, B. Jiao, and L. Hanzo, "Performance analysis of NOMA-SM in vehicle-to-vehicle massive MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2653–2666, Dec. 2017.

[18] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[19] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "On the sum rate of MIMO-NOMA and MIMO-OMA systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 534–537, Aug. 2017.

[20] C. Wang, Y. He, F. R. Yu, Q. Chen, and L. Tang, "Integration of networking, caching, and computing in wireless systems: A survey, some research issues, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 7–38, Firstquarter 2018.

[21] S. A. R. Zaidi, M. Ghogho, and D. C. McLernon, "Information centric modeling for two-tier cache enabled cellular networks," in *Proc. IEEE International Conference on Communication Workshop (ICCW)*, Jun. 2015, pp. 80–86.

[22] E. Bastug, M. Bennis, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," in *Proc. 11th International Symposium on Wireless Communications Systems (ISWCS)*, Aug. 2014, pp. 649–653.

[23] L. Zhang, H. Yang, and M. O. Hasna, "Generalized area spectral efficiency: An effective performance metric for green wireless communications," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 747–757, Feb. 2014.

[24] S. Glass, I. Mahgoub, and M. Rathod, "Leveraging MANET-based cooperative cache discovery techniques in VANETs: A survey and analysis," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2640–2661, Fourthquarter 2017.

[25] N. E. Majd, S. Misra, and R. Tourani, "Split-cache: A holistic caching framework for improved network performance in wireless ad hoc networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec. 2014, pp. 137–142.

[26] M. F. Caetano and J. L. Bordim, "A cluster based collaborative cache approach for MANETs," in *Proc. First International Conference on Networking and Computing*, Nov. 2010, pp. 104–111.

[27] L. Dai, B. Wang, Y. Yuan, S. Han, C. I, and Z. Wang, "Non-orthogonal multiple access for 5g: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[28] T. Han, J. Gong, X. Liu, S. M. R. Islam, Q. Li, Z. Bai, and K. S. Kwak, "On downlink NOMA in heterogeneous networks with non-uniform small cell deployment," *IEEE Access*, vol. 6, pp. 31 099–31 109, 2018.

[29] B. Di, L. Song, Y. Li, and G. Y. Li, "NOMA-based low-latency and high-reliable broadcast communications for 5G V2X services," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec. 2017, pp. 1–6.

[30] B. Di, L. Song, Y. Li, and Z. Han, "V2X meets NOMA: Non-orthogonal multiple access for 5G-enabled vehicular networks," *IEEE Trans. Wireless Commun.*, vol. 24, no. 6, pp. 14–21, Dec. 2017.

[31] B. W. Khoueiry and M. R. Soleymani, "An efficient noma V2X communication scheme in the internet of vehicles," in *Proc. IEEE 85th Vehicular Technology Conference (VTC Spring)*, Jun. 2017, pp. 1–7.

[32] L. P. Qian, Y. Wu, H. Zhou, and X. Shen, "Dynamic cell association for non-orthogonal multiple-access V2S networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2342–2356, Oct. 2017.

[33] K. N. Doan, W. Shin, M. Vaezi, H. V. Poor, and T. Q. S. Quek, "Optimal power allocation in cache-aided non-orthogonal multiple access systems," in *Proc. IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[34] Z. Ding, P. Fan, G. K. Karagiannidis, R. Schober, and H. V. Poor, "Noma assisted wireless caching: Strategies and performance analysis," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4854–4876, Oct. 2018.

[35] G. K. Karagiannidis, N. C. Sagias, and P. T. Mathiopoulos, "N∗Nakagami: A novel stochastic model for cascaded fading channels," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1453–1458, Aug. 2007.

[36] H. Ilhan, M. Uysal, and I. Altunbas, "Cooperative diversity for intervehicular communication: Performance analysis and optimization," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3301–3310, Sep. 2009.

[37] I. Gradshteyn and I. Ryzhik, *Tables of integrals, series and products*, 7th ed. Academic Press, 2007.

[38] J. Salo, H. M. El-Sallabi, and P. Vainikainen, "The distribution of the product of independent Rayleigh random variables," *IEEE Trans. Antennas Propag.*, vol. 54, no. 2, pp. 639–643, Feb. 2006.

[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[40] O. Ozdogan, E. Bjornson, and E. G. Larsson, "Massive MIMO with spatially correlated Rician fading channels," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2019.

[41] N. Avazov, S. M. R. Islam, D. Park, and K. S. Kwak, "Statistical characterization of a 3-D propagation model for V2V channels in rectangular tunnels," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 2392–2395, Jun. 2017.

Sanjeev Gurugopinath received the B.E. degree in electrical and electronics engineering from the Dr. Ambedkar Institute of Technology, Visvesvaraya Technological University, Bengaluru, in 2004, the M.Tech. degree in digital electronics and communication engineering from the M. S. Ramaiah Institute of Technology, Visvesvaraya Technological University, Bengaluru, in 2006, and the Ph.D. degree in signal processing for communications from the Indian Institute of Science, Bengaluru, in 2015. He is currently a Professor with the Department of Electronics and Communication Engineering, PES University, Bengaluru, and a visiting Research Scholar at Khalifa University, Abu Dhabi, UAE. He was with the PES Institute of Technology and the CMR Institute of Technology, Bengaluru. His current research interests are in the areas of cognitive radios and heterogeneous networks for 5G communication systems, powerline communications, underwater acoustics, and speech signal processing. He was a co-recipient of the Best Paper Award across all tracks at IEEE INDICON 2016.

Paschalis C. Sofotasios (S'07–M'12–SM'16) was born in Volos, Greece, in 1978. He received the MEng degree from Newcastle University, UK, in 2004, the MSc degree from the University of Surrey, UK, in 2006, and the PhD degree from the University of Leeds, UK, in 2011. His MSc studies were funded by a scholarship from UK-EPSRC and his PhD studies were sponsored by UK-EPSRC and Pace plc. He has held academic positions at the University of Leeds, UK, University of California at Los Angeles, Los Angeles, USA, Aristotle University of Thessaloniki, Greece, Tampere University of Technology, Finland, and Khalifa University, UAE, where he is currently an Assistant Professor. His research interests include broad areas of digital and optical wireless communications including topics on pure mathematics and statistics.

Dr. Sofotasios has been a member and co-chair of the Technical Program Committee of numerous IEEE conferences. He received an Exemplary Reviewer Award from the IEEE COMMUNICATIONS LETTERS in 2012 and from the IEEE TRANSACTIONS ON COMMUNICATIONS in 2015 and 2016. He is a Senior Member IEEE and was a co-recipient of the Best Paper Award at ICUFN '13. He currently serves as an Editor for the IEEE COMMUNICATIONS LETTERS.

Sami Muhaidat (S'01–M'07–SM'11) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2006. From 2007 to 2008, he was an NSERC Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, ON, Canada. From 2008 to 2012, he was an Assistant Professor with the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. He is currently an Associate Professor with Khalifa University, Abu Dhabi, United Arab Emirates, and a Visiting Professor with the Faculty of Engineering, University of Surrey, Surrey, U.K. He has authored over 150 technical papers on these topics. His research interests include wireless communications, physical-layer security, IoT with emphasis on battery-less devices, and machine learning.

Dr. Muhaidat is a member of the Mohammed Bin Rashid Academy of scientists. He was a recipient of several scholarships during his undergraduate and graduate studies and a winner of the 2006 NSERC Postdoctoral Fellowship Competition. He served as an Associate Editor and Senior Editor of the IEEE COMMUNICATIONS LETTERS and he currently serves as an Area Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.

Yousof Al Hammadi received the bachelors degree in computer engineering from Khalifa University of Science and Technology (previously known as Etisalat College of Engineering), UAE, in 2000, the MSc degree in Telecommunications Engineering from the University of Melbourne, Australia, in 2003, and the PhD degree in Computer Science and Information Technology from the University of Nottingham, United Kingdom, in 2009. He is currently an Acting Dean of Graduate Studies and Assistant Professor in the Electrical & Computer Engineering department, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates. His main research interests include the area of information security which include intrusion detection, botnet/bots detection, viruses/worms detection, artificial immune systems, machine learning and artificial intelligence, RFID security and mobile security.