

Coded Caching and Spatial Multiplexing Gains in MIMO Interference Networks

Antonious M. Girgis*, Ozgur Ercetin[†], Mohammed Nafie*[‡], and Tamer ElBatt^{§‡}

*Wireless Intelligent Networks Center (WINC), Nile University, Cairo, Egypt

[†] Faculty of Engineering and Natural Sciences, Sabanci University, Turkey.

[‡] Electronics and Communications Engineering Dept., Faculty of Engineering, Cairo University, Egypt.

[§] Computer Science and Engineering Dept., The American University in Cairo, Egypt.

Abstract—This paper studies the Multi-Input-Multi-Output (MIMO) interference networks with arbitrary number of transmitters and receivers, where both the transmitters and receivers are equipped with caches. Our objective is to propose content placement and delivery schemes that minimize the worst case normalized delivery time (NDT). First, we design a delivery scheme for the cache-aided Single-Input-Multiple-Output (SIMO) interference networks. Then, we obtain the achievable NDT of the cache-aided MIMO interference networks by using the decomposition property. The numerical results show the superiority of our proposed scheme over the state-of-the-art schemes in the literature. Furthermore, we show that increasing the receiver-cache sizes achieves a higher gain than increasing the number of receive-antennas. In other words, the coded caching gain has a more significant contribution in reducing the transmission latency than the spatial multiplexing gain.

I. INTRODUCTION

Multiple antennas at both transmitters and receivers can be used to send simultaneously different data streams over the Multi-Input-Multi-Output (MIMO) wireless channel increasing the capacity of the wireless networks (or equivalently reducing the transmission latency). The gain from the multiple antennas at the network nodes is referred to as the spatial multiplexing gain measured by the number of independent data streams multiplexed in space. Recently, caching systems have received considerable attention due to the significant performance gain obtained from the availability of caches at the network nodes. In this work, we study the MIMO interference network with caches at both transmitters and receivers. Our main objective is to determine how the transmission latency varies as a function of the system parameters. Furthermore, we address the question which technique is more effective in reducing the transmission latency of the MIMO interference networks: the spatial multiplexing or the coded caching.

Cache-aided interference networks have been first studied in [1], where the authors considered the 3×3 Single-Input-Single-Output (SISO) interference network with transmitter-side caches only. It has been shown that increasing the transmitter-cache sizes leads to virtual cooperation among transmitters which in turn increases the degrees of freedom (DoF) of the interference networks. In [2], the authors have defined the normalized delivery time (NDT) as a performance

metric for the 2×2 fog radio access network (F-RAN) with caches at transmitters only. In [3], the authors have derived a lower bound on the NDT of the SISO cache-aided interference networks. The authors in [4] have studied an F-RAN with transmitters equipped with caches and multiple antennas. The work in [5] has studied an F-RAN with caches at both transmitters and receivers under the assumption that the content placement is applied in a decentralized manner. In [6], the authors have characterized both the peak NDT and the expected NDT under uniform popularity distribution for F-RAN with caches at transmitters and receivers. In addition, the authors in [6] have proposed a coding scheme for the F-RAN with caches only at the transmitters and shown that this scheme is order optimal for any popularity distribution. The work in [7]–[9] studied SISO interference networks with caches at both transmitters and receivers. The work in [10] studied MIMO interference networks with caches at both transmitters and receivers; however, the authors considered only the case of three transmitters and three receivers.

In this paper, we consider a $K_T \times K_R$ cache-aided interference network with a library of multiple files, where both transmitters and receivers are equipped with an isolated cache memory and multiple antennas. We study the fundamental limits on the normalized delivery time (NDT) as a function of the transmitter-cache size, receiver-cache size, the number of transmit-antennas, and the number of receive-antennas. Unlike [10], our work is general for arbitrary number of transmitters and receivers. Our main contributions in this work are as follows: We propose an achievable scheme to obtain the NDT of the Single-Input-Multiple-Output (SIMO) cache-aided interference network. We obtain the NDT of the MIMO cache-aided interference networks by using the decomposition property by splitting each multi-antenna transmitter into multiple single antenna transmitters. We show that our proposed scheme achieves a lower NDT than the scheme proposed in [10]. Our scheme demonstrates that increasing the receiver-cache size achieves a gain higher than the gain obtained by increasing the receive-antennas. Hence, caching is a more effective tool than the spatial multiplexing in reducing the transmission latency.

II. SYSTEM MODEL

A Multi-Input-Multi-Output (MIMO) interference network comprising K_T transmitters connected to K_R receivers over a

This work was supported in part by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 690893 and a grant from the Egyptian National Telecommunications Regulatory Authority (NTRA).

time-varying Gaussian channel is studied. There is a content library of N files, $\mathcal{W} \triangleq \{W_1, \dots, W_N\}$, each of size F bits, where each file $W_n \in \mathcal{W}$ is chosen independently and uniformly from the set $[2^F]$ at random. Each transmitter TX_i , $i \in [K_T]$, has A_T antennas and a local cache memory \mathcal{V}_i of size $M_T F$ bits, where $\mu_T = M_T/N$ refers to the normalized transmitter-cache size. Moreover, each receiver RX_j , $j \in [K_R]$, has A_R antennas and a local cache memory \mathcal{Z}_j of size $M_R F$ bits, where $\mu_R = M_R/N$ refers to the normalized receiver-cache size in files.

The system consists of two separate phases: a placement phase and a delivery phase. In the placement phase, each network node fills its cache memory as an arbitrary function of the content library \mathcal{W} under its cache size constraint. We emphasize that the caching decisions are taken without any prior knowledge of the future receiver demands and channel coefficients between the transmitters (TXs) and the receivers (RXs). In the delivery phase, receiver RX_j requests a file W_{d_j} out of the N files of the library. We consider $\mathbf{d} = [d_1, \dots, d_{K_R}] \in [N]^{K_R}$ as the vector of receiver demands. The transmitters are aware of all receiver demands \mathbf{d} . Thus, transmitter TX_i , $i \in [K_T]$, responds by sending a message $\mathbf{X}_i \triangleq (\mathbf{X}_i(t))_{t=1}^T$ of block length T over the interference channel, where $\mathbf{X}_i(t) \in \mathbb{C}^{A_T}$ is the transmitted vector of transmitter TX_i at time $t \in [T]$. We impose a transmit power constraint over the channel input $\|\mathbf{X}_i(t)\|^2 \leq P$. In this phase, each transmitter has only access to its own cache content, so, codeword \mathbf{X}_i is determined by an encoding function of the receiver demands \mathbf{d} , the cache contents \mathcal{V}_i , and the channel coefficients between TXs and RXs. Afterwards, each receiver RX_j implements a decoding function to estimate the requested file \hat{W}_{d_j} from its cache contents \mathcal{Z}_j , and the received signal $\mathbf{Y}_j \triangleq (\mathbf{Y}_j(t))_{t=1}^T$ which is given by

$$\mathbf{Y}_j(t) = \sum_{i=1}^{K_T} \mathbf{H}_{ji}(t) \mathbf{X}_i(t) + \mathbf{N}_j(t), \quad (1)$$

where $\mathbf{Y}_j(t) \in \mathbb{C}^{A_R}$ is the received signal by receiver RX_j at time $t \in [T]$, and $\mathbf{N}_j(t) \in \mathbb{C}^{A_R}$ is the complex Gaussian noise vector at receiver RX_j at time $t \in [T]$. Let $\mathbf{H}_{ji}(t) \in \mathbb{C}^{A_R \times A_T}$ represent the channel matrix between transmitter TX_i and receiver RX_j at time t . We assume that all channel coefficients are drawn independently and identically distributed (i.i.d.) from a continuous distribution. For a given caching, encoding, and decoding functions, the probability of error is obtained by

$$\text{Pe} = \max_{\mathbf{d} \in [N]^{K_R}} \max_{j \in [K_R]} \mathbb{P}(\hat{W}_{d_j} \neq W_{d_j}), \quad (2)$$

which is the worst-case probability of error over all possible demands $\mathbf{d} \in [N]^{K_R}$ and over all receivers. The coding scheme is said to be feasible if each receiver can decode its requested file with vanishing probability of error $\text{Pe} \rightarrow 0$ as $F \rightarrow \infty$. In the following, we define the normalized delivery time (NDT) first discussed in [2] as a performance metric for any coding scheme.

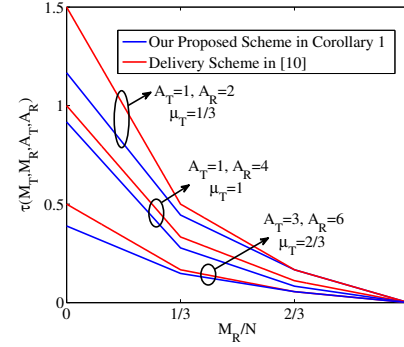


Fig. 1: The achievable NDT of the 3×3 cache-aided MIMO interference network.

Definition 1. The normalized delivery time (NDT) for a given feasible coding scheme with transmitter cache size $M_T F$ bits, and receiver cache size $M_R F$ bits is defined as

$$\tau(M_T, M_R, A_T, A_R) = \lim_{P \rightarrow \infty} \lim_{F \rightarrow \infty} \sup_{\mathbf{d}} \frac{T(\mathbf{d})}{F/\log(P)}. \quad (3)$$

Moreover, we define the minimum NDT for a given tuple (M_T, M_R, A_T, A_R) as

$$\tau^*(M_T, M_R, A_T, A_R) \triangleq \inf \{\tau : \tau \text{ is feasible}\}. \quad (4)$$

Note that the $\tau(M_T, M_R, A_T, A_R)$ represents the delivery time to serve the worst case user demands $\mathbf{d} \in [N]^{K_R}$ normalized with respect to an interference-free baseline system, where $F/\log(P)$ refers to the time for delivering F bits with transmission rate $\log(P)$ in the high signal-to-noise-ratio.

Remark 1. (The relation between the NDT and the DoF [9]): Let R_j denote the number of bits normalized by file size F that is required to be delivered for receiver RX_j . Thus, the degrees of freedom (DoF) for receiver RX_j is defined as $d_j = \lim_{P \rightarrow \infty} \lim_{F \rightarrow \infty} \frac{R_j F}{\log(P) T} = \frac{R_j}{T}$, where $R_j F/T$ represents the transmission rate of receiver RX_j . Furthermore, the sum-DoF of the network is defined by

$$\text{sDoF} = \lim_{P \rightarrow \infty} \lim_{F \rightarrow \infty} \frac{R F}{\log(P) T} = \frac{R}{T},$$

where $R = \sum_{j=1}^{K_R} R_j$ denotes the total number of bits delivered for all receivers normalized by the file size F . The sum-DoF is a performance metric that defines the pre-log capacity or the multiplexing gain of the network. In other words, the capacity can be expressed by $\text{sDoF} \log(P) + o(\log(P))$ at the high SNR regime, where the $o(\log(P))$ term vanishes as $P \rightarrow \infty$. Therefore, the NDT is inversely proportional to the sum-DoF for a fixed receiver cache size, and characterizing one of these metrics leads to the other.

III. MAIN RESULTS

In this section, we introduce the main results and insights of this paper.

Theorem 1. (Achievable NDT of SIMO interference networks): For the $K_T \times K_R$ cache-aided SIMO interference

network with A_R antennas at each receiver, transmitter-cache size $M_T \in [N/K_T : N]$, receiver-cache size $M_R \in [0 : N]$, and parameter $t_R = K_R M_R / N$, the achievable NDT is given by

$$\tau(M_T, M_R, A_R) = \frac{K_R(1 - \mu_R)}{\text{sDoF}(t_R, A_R)}, \quad (5)$$

for $t_R \in \{0, \dots, K_R\}$, where $\text{sDoF}(t_R, A_R)$ is the sum-DoF given by

$$\text{sDoF}(t_R, A_R) = \begin{cases} K_T(t_R + 1) & \text{if } K_T \leq A_R \\ \frac{(t_R + 1)A_R K_T K_R}{(t_R + 1)K_T + A_R(K_R - t_R - 1)} & \text{if } K_T > A_R \end{cases}. \quad (6)$$

Moreover, the lower convex envelope of these integer points is also achievable.

The proof of Theorem 1 is presented in Section V. To prove this theorem, we define a new communication problem called *SIMO multicast X-channel* which is a generalization of the SISO multicast X-channel defined in [8] and the SIMO unicast X-channel defined in [11]. In the SIMO multicast X-channel, each transmitter has a dedicated message for each group of the multi-antenna receivers. We can obtain the achievable NDT of the cache-aided MIMO interference network from this theorem as stated in the following corollary.

Corollary 1. The achievable NDT of the $K_T \times K_R$ MIMO interference channel with A_T antennas at each transmitter, A_R antennas at each receiver, transmitter-cache size $M_T F$ bits, receiver-cache size $M_R F$ bits, and parameter $t_R = K_R M_R / N$ is obtained by

$$\tau(M_T, M_R, A_R, A_T) = \frac{K_R(1 - \mu_R)}{\text{sDoF}(t_R, A_T, A_R)}, \quad (7)$$

for $t_R \in \{0, \dots, K_R\}$, where $\text{sDoF}(t_R, A_T, A_R)$ is the sum-DoF given by

$$\text{sDoF} = \begin{cases} A_T K_T(t_R + 1) & \text{if } A_T K_T \leq A_R \\ \frac{(t_R + 1)A_R A_T K_T K_R}{(t_R + 1)A_T K_T + A_R(K_R - t_R - 1)} & \text{if } A_T K_T > A_R \end{cases}. \quad (8)$$

Moreover, the lower convex envelope of these integer points is also achievable.

The proof of this corollary is obtained directly from Theorem 1 by using the decomposition property, where each transmitter with A_T antennas is decomposed into A_T distributed transmitters with a single antenna. Note that the achievable NDTs in Theorem 1 and Corollary 1 are not functions of the transmitter-cache size, since our delivery scheme neglects the gains obtained from increasing the transmitter-cache size. However, increasing the transmitter-cache size can provide a higher performance gain as we proposed in the extended version in [12]. Now, we compare the achievable NDT of our proposed scheme with the delivery scheme proposed in [10] for the 3×3 cache-aided MIMO interference network. In Figure 1, we plot the achievable NDT of our proposed scheme and the delivery scheme in [10]. We can see that our proposed scheme outperforms the delivery scheme in [10] for

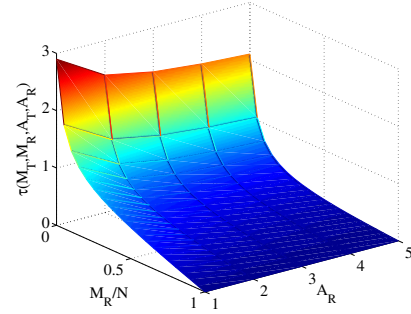


Fig. 2: The NDT of the MIMO interference network as a function of A_R and M_R for $K_T = 5$ transmitters, $A_T = 2$ antennas, and $K_R = 20$ receivers.

different system parameters. The main reason for this is that the scheme of [10] is restricted to linear delivery schemes with finite symbol extensions, while we do not impose any restrictions on the delivery scheme. The achievable NDT is a function of mainly two gains. The first one is the coded caching gain that is obtained from the coding opportunities due to the availability of caches at transmitters and/or receivers. The second gain is the spatial multiplexing gain that is obtained from the coding opportunities due to the availability of multiple antennas at both transmitters and receivers. One of the intriguing questions is to determine which one of these two gains dominates performance in terms of reducing the achievable NDT. Figure 2 depicts the achievable NDT of the MIMO cache-aided interference network as a function of M_R and A_R . From Figure 2, it is clear that increasing the receiver cache with NF/K_R bits achieves a higher gain than increasing the number of the receive-antennas by one for $K_T \leq K_R$. Thus, the coded caching achieves a higher gain compared to the gain attributed to the spatial multiplexing. Due to the space limitations, we left the other comparisons and discussions to the extended version in [12].

IV. CONTENT PLACEMENT

In this section, we present the content placement strategy for cache-aided interference networks, where we follow the same strategy as in [8]. Let $t_R = K_R M_R / N$. In the following, we focus on the cache placement and the delivery scheme for integer points $t_R \in \{0, \dots, K_R\}$. The achievable schemes for general t_R can be obtained by memory-time sharing between these integer points, since the NDT is a convex function of M_R . In the placement phase, we split each file $W_n \in \mathcal{W}$ into $K_T \binom{K_R}{t_R}$ disjoint subfiles, each of size $F/K_T \binom{K_R}{t_R}$ bits. As a result, file W_n is represented by

$$W_n = \{W_{n, S_R, i} : S_R \subseteq [K_R], |S_R| = t_R, i \in [K_T]\}. \quad (9)$$

For every file $W_n \in \mathcal{W}$, the subfile $W_{n, S_R, i}$ is stored at the cache of transmitter TX_i , and at the cache of receiver RX_j if $j \in S_R$. Thus, the cache content of transmitter TX_i and receiver RX_j is expressed by $\mathcal{V}_i \triangleq \{W_{n, S_R, i} : n \in [N]\}$ and $\mathcal{Z}_j \triangleq \{W_{n, S_R, i} : j \in S_R, n \in [N]\}$. Therefore, each transmitter stores $N \binom{K_R}{t_R}$ subfiles, and each receiver stores the

$NK_T \binom{K_R-1}{t_R-1}$ subfiles. Accordingly, the number of bits stored at each transmitter is equal to $N \binom{K_R}{t_R} \frac{F}{K_T \binom{K_R}{t_R}} \leq M_T F$ bits. Similarly, the number of bits stored at each receiver is equal to $NK_T \binom{K_R-1}{t_R-1} \frac{F}{K_T \binom{K_R}{t_R}} = M_R F$ bits. Thus, this placement strategy satisfies the cache size constraint for each transmitter and each receiver. Moreover, we emphasize that the content placement is performed without any prior knowledge of the receiver demands or channel gains in the delivery phase, which is a practically relevant assumption, since there is a large time separation between the placement and delivery phases.

Remark 2. Note that for a given demand \mathbf{d} , receiver RX_j has subfiles $\{W_{d_j, S_R, i}\}_{i \in S_R}$. Hence, the transmitters have to deliver the remaining subfiles $\{W_{d_j, S_R, i}\}_{i \notin S_R}$ with a total of $K_T \binom{K_R-1}{t_R}$ subfiles to receiver RX_j , where each subfile has a size of $F / \binom{K_T}{t_T} \binom{K_R}{t_R}$ bits. Hence, the total number of bits desired to be delivered to receiver RX_j is given by $R_j F = \frac{K_R - t_R}{K_R} F$ bits. Accordingly, the total number of bits required to be delivered for all receivers is equal $RF = (K_R - t_R) F = K_R (1 - M_R/N) F$ bits.

V. ACHIEVABLE SCHEME OF SIMO INTERFERENCE NETWORKS

In this section, we prove the achievable scheme of the NDT for SIMO interference networks introduced in Theorem 1. Let receiver RX_j request file W_{d_j} , and hence, the transmitters need to deliver subfiles $\{W_{d_j, S_R, i}\}_{i \notin S_R, i \in [K_T]}$ that are not stored at the cache of receiver RX_j .

A. When $K_T \leq A_R$

In this case, the delivery is applied into $\binom{K_R}{t_R+1}$ time slots. At each time slot, the transmitters serve a set $\mathcal{K} \subseteq [K_R]$ of $|\mathcal{K}| = t_R + 1$ receivers, in which the subfiles $\{W_{d_j, \mathcal{K} \setminus \{j\}, i} : i \in \mathcal{K}, i \in [K_T]\}$ would be delivered correctly with interference-free to receivers \mathcal{K} . We can easily verify that each receiver $\text{RX}_j, j \in [K_R]$, will receive all the required subfiles $\{W_{d_j, S_R, i} : i \notin S_R, i \in [K_T]\}$ by the end of the transmission. Consider an arbitrary time slot to serve a set \mathcal{K} of receivers. Each transmitter TX_i broadcast the subfiles that are available at its cache memory, i.e., $X_i = \sum_{l \in \mathcal{K}} W_{d_i, \mathcal{K} \setminus \{l\}, i}$ for $i \in [K_T]$. Thus, the received signal at receiver $\text{RX}_j, j \in \mathcal{K}$, can be expressed by

$$\begin{aligned} \mathbf{Y}_j &= \sum_{i=1}^{K_T} \mathbf{H}_{ji} X_i + \mathbf{N}_j = \sum_{i=1}^{K_T} \sum_{l \in \mathcal{K}} \mathbf{H}_{ji} W_{d_i, \mathcal{K} \setminus \{l\}, i} + \mathbf{N}_j \\ &= \sum_{i=1}^{K_T} \mathbf{H}_{ji} W_{d_j, \mathcal{K} \setminus \{j\}, i} + \sum_{i=1}^{K_T} \sum_{\substack{l \in \mathcal{K} \\ l \neq j}} \mathbf{H}_{ji} W_{d_i, \mathcal{K} \setminus \{l\}, i} + \mathbf{N}_j, \end{aligned} \quad (10)$$

where $\mathbf{Y}_j \in \mathbb{C}^{A_R \times 1}$ denotes the received vector at receiver RX_j , and $\mathbf{H}_{ji} \in \mathbb{C}^{A_R \times 1}$ is the channel vector between transmitter TX_i and receiver RX_j . Observe that the received signal in (10) consists of two terms. The first term represents the desired signals, while the second term denotes interference signals. Since receiver RX_j has subfiles $\{W_{d_i, \mathcal{K} \setminus \{l\}, i} : l \in$

$\mathcal{K}, i \in [K_T]\}$ at its cache, receiver can subtract the interference signals from the receiver signal \mathbf{Y}_j , and hence, the received signal is given by

$$\tilde{\mathbf{Y}}_j = \sum_{i=1}^{K_T} \mathbf{H}_{ji} W_{d_j, \mathcal{K} \setminus \{j\}, i} + \mathbf{N}_j. \quad (11)$$

It remains to prove that receiver RX_j can decode the desired subfiles correctly. Note that the K_T desired subfiles are received with vectors $[\mathbf{H}_{j1}, \dots, \mathbf{H}_{jK_T}]$ which has a full rank of K_T when $A_R \geq K_T$ almost surely. In other words, the desired subfiles arrive at receiver RX_j over linearly independent directions. Therefore, receiver RX_j can decode the desired subfiles correctly from the received signal. As a result, each receiver can decode $K_T \binom{K_R-1}{t_R}$ desired messages over $\binom{K_R}{t_R+1}$ time slots. Thus, the DoF per-receiver and the sum-DoF are given by

$$\begin{aligned} d_j &= \frac{K_T \binom{K_R-1}{t_R}}{\binom{K_R}{t_R+1}} = \frac{K_T (t_R + 1)}{K_R} \\ \text{sDoF} &= \frac{K_R K_T \binom{K_R-1}{t_R}}{\binom{K_R}{t_R+1}} = K_T (t_R + 1) \end{aligned} \quad (12)$$

From Remark 1, the NDT $\tau(M_T, M_R, A_R) = (1 - \frac{M_R}{N}) / d_j = K_R (1 - \frac{M_R}{N}) / \text{sDoF}$ is achievable.

B. When $K_T > A_R$

In this case, there are $K_R K_T \binom{K_R-1}{t_R}$ subfiles required to be delivered to receivers, where receiver RX_j requires subfiles $\{W_{d_j, S_R, i} : i \in [K_T], j \notin S_R\}$. First, we implement the multicast coding scheme as in [13] at each transmitter. For a given a subset $\mathcal{K} \subseteq [K_R]$ of $|\mathcal{K}| = t_R + 1$ receivers, each transmitter TX_i generates a multicast message $W_{\mathcal{K}, i}$ as follows

$$W_{\mathcal{K}, i} = \bigoplus_{j \in \mathcal{K}} W_{d_j, \mathcal{K} \setminus \{j\}, i}, \quad (13)$$

where the single message $W_{\mathcal{K}, i}$ of size $F / K_T \binom{K_R}{t_R}$ bits combines $t_R + 1$ different subfiles, in which each subfile is desired by a single receiver $\text{RX}_j, j \in \mathcal{K}$. Since receiver $\text{RX}_j, j \in \mathcal{K}$, has subfiles $\{W_{d_{j'}, \mathcal{K} \setminus \{j'\}, i}\}_{j' \in \mathcal{K}}$, it can recover its desired subfile $W_{d_j, \mathcal{K} \setminus \{j\}, i}$ from the multicast message $W_{\mathcal{K}, i}$. Therefore, there are $K_T \binom{K_R}{t_R+1}$ coded messages

$$\{W_{\mathcal{K}, i} : i \in [K_T], \mathcal{K} \subseteq [K_R], |\mathcal{K}| = t_R + 1\}$$

required to be delivered to receivers, where each transmitter TX_i has a dedicated message $W_{\mathcal{K}, i}$ to every subset \mathcal{K} of $t_R + 1$ receiver. In the following, we define a new communication problem called *SIMO multicast X-channel*.

Definition 2. In the SIMO multicast X-channel, there are K_T single-antenna transmitters, K_R receivers each with A_R antennas, and a total of $K_T \binom{K_R}{\sigma}$ messages, where σ denotes the multicast order. Each transmitter has a dedicated message for every multicast group of σ receivers. The following lemma provides the achievable DoF of the SIMO multicast X-channel.

Lemma 1. For a SIMO multicast X-channel with K_T transmitters, K_R receivers, each equipped with A_R antennas,

and multicast order $\sigma = t_R + 1$, the DoF per message is given by

$$d(t_R, A_R) = \frac{A_R}{K_T \binom{K_R-1}{t_R} + A_R \binom{K_R-1}{t_R+1}}. \quad (14)$$

Proof. The proof is presented in Subsection V-C. ■

Remark 3. The SIMO multicast X-channel is a generalization of the SIMO X-channel studied in [11] and a generalization of the SISO multicast X-channel studied in [8]. When $t_R = 0$, Lemma 1 gives the DoF per message of SIMO X-channel which is the same result as in [11, Theorem 2]. When $A_R = 1$, Lemma 1 gives the DoF per message for the SISO multicast X-channel which is the same result as in [8, Theorem 2].

Note that there are a total of $K_T \binom{K_R}{t_R+1}$ messages, and hence, the sum-DoF of the SIMO multicast X-channel is given by

$$\begin{aligned} \text{sDoF}_{M-X}(t_R, A_R) &= \frac{A_R K_T \binom{K_R}{t_R+1}}{K_T \binom{K_R-1}{t_R} + A_R \binom{K_R-1}{t_R+1}} \\ &= \frac{A_R K_T K_R}{(t_R + 1) K_T + A_R (K_R - t_R - 1)}. \end{aligned} \quad (15)$$

Moreover, each receiver RX_j is interested in $K_T \binom{K_R-1}{t_R}$ messages out of the total $K_T \binom{K_R}{t_R+1}$ messages. Thus, the DoF per receiver of the SIMO multicast X-channel is obtained

$$\begin{aligned} d_j(t_R, A_R) &= \frac{A_R K_T \binom{K_R-1}{t_R}}{K_T \binom{K_R-1}{t_R} + A_R \binom{K_R-1}{t_R+1}} \\ &= \frac{(t_R + 1) A_R K_T}{(t_R + 1) K_T + A_R (K_R - t_R - 1)}. \end{aligned} \quad (16)$$

Observe that the sum-DoF for the SIMO multicast X-channel is equal to the DoF per receiver multiplied by $K_R / (t_R + 1)$, since each message is desired by $t_R + 1$ receivers. However, for the cache-aided SIMO interference network, each message $W_{\mathcal{K},i}$ combines $(t_R + 1)$ different messages, a dedicated message to every receiver RX_j , $j \in \mathcal{K}$. As a result, the sum-DoF of the cache-aided SIMO interference network is equal $\text{sDoF} = K_R d_j = (t_R + 1) \text{sDoF}_{M-X}$. Thus, we can obtain the achievable NDT of the cache-aided SIMO interference network from remark 1 as $\tau = \frac{(1 - \frac{M_R}{N})}{d_j} = \frac{K_R (1 - \frac{M_R}{N})}{\text{sDoF}}$. This completes the proof of Theorem 1.

C. Proof of Lemma 1

In this Subsection, we introduce the achievable scheme of the SIMO multicast X-channel. We propose an interference alignment scheme which generalizes the idea of the many-to-many alignment introduced in [11] for the SIMO X-channel. The transmission occurs over $T_n = \frac{K_T}{A_R} \binom{K_R-1}{t_R} L_n + \binom{K_R-1}{t_R+1} L_{n+1}$ symbol extensions of the original channel¹, where $L_n = n^\Gamma$ for arbitrary $n \in \mathbb{N}^+$ and $\Gamma = A_R K_T (K_R - t_R - 1)$. The input-output relation of the original channel over T_n -symbol extensions can be expressed by

¹In order to insure that T_n is integer, the element n is chosen as a multiple of A_R such that L_n / A_R is integer.

$$\begin{aligned} \mathbf{Y}_j &= \sum_{i=1}^{K_T} \mathbf{H}_{ji} \mathbf{X}_i + \mathbf{N}_j \\ &= \sum_{i=1}^{K_T} \begin{bmatrix} \mathbf{H}_{ji}^1 \\ \vdots \\ \mathbf{H}_{ji}^{A_R} \end{bmatrix} \mathbf{X}_i + \mathbf{N}_j, \end{aligned} \quad (17)$$

where \mathbf{Y}_j and \mathbf{N}_j are $A_R T_n \times 1$ column vectors of the received signal and the Gaussian noise at the receiver RX_j over T_n -symbol extension, respectively. \mathbf{X}_i is a $T_n \times 1$ column vector representing the transmitted vector of transmitter TX_i . \mathbf{H}_{ji} is the $A_R T_n \times T_n$ channel matrix from transmitter TX_i to RX_j over T_n symbol extension, where \mathbf{H}_{ji}^r is the $T_n \times T_n$ diagonal channel matrix from transmitter TX_i to the r -th receive-antenna of receiver RX_j .

$$\mathbf{H}_{ji}^r = \begin{bmatrix} h_{ji}^r(1) & 0 & \cdots & 0 \\ 0 & h_{ji}^r(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{ji}^r(T_n) \end{bmatrix}. \quad (18)$$

Each message $W_{\mathcal{K},i}$ is encoded into L_n independent streams represented by a $L_n \times 1$ column vector $\mathbf{X}_{\mathcal{K},i}$. We use the same beamforming matrix $\mathbf{V}_{\mathcal{K}}$ to encode the vectors $\{\mathbf{X}_{\mathcal{K},i}\}$ desired by the set \mathcal{K} receivers from all transmitters, where $\mathbf{V}_{\mathcal{K}}$ is a $T_n \times L_n$ matrix. Hence, we can describe the transmitted vector of transmitter TX_i by

$$\mathbf{X}_i = \sum_{\substack{\mathcal{K} \subseteq [K_R] \\ |\mathcal{K}|=t_R+1}} \mathbf{V}_{\mathcal{K}} \mathbf{X}_{\mathcal{K},i}.$$

Furthermore, the received signal at receiver RX_j is given by

$$\begin{aligned} \mathbf{Y}_j &= \sum_{i=1}^{K_T} \mathbf{H}_{ji} \mathbf{X}_i + \mathbf{N}_j = \sum_{i=1}^{K_T} \mathbf{H}_{ji} \sum_{\substack{\mathcal{K} \subseteq [K_R] \\ |\mathcal{K}|=t_R+1}} \mathbf{V}_{\mathcal{K}} \mathbf{X}_{\mathcal{K},i} + \mathbf{N}_j \\ &= \sum_{i=1}^{K_T} \sum_{\substack{\mathcal{K} \subseteq [K_R] \\ \mathcal{K} \ni j}} \mathbf{H}_{ji} \mathbf{V}_{\mathcal{K}} \mathbf{X}_{\mathcal{K},i} + \sum_{i=1}^{K_T} \sum_{\substack{\mathcal{K} \subseteq [K_R] \\ \mathcal{K} \not\ni j}} \mathbf{H}_{ji} \mathbf{V}_{\mathcal{K}} \mathbf{X}_{\mathcal{K},i} + \mathbf{N}_j, \end{aligned} \quad (19)$$

where the first term represents the desired signals, while the second term represents the interference signals. Our objective is to design the beamforming matrices $\{\mathbf{V}_{\mathcal{K}}\}$ to reduce the dimensional space spanned by the interference signals at each receiver. Note that each receiver is equipped with A_R antennas. Thus, the signals from any A_R transmitters cannot be aligned over each other at any receiver, since the channel matrix between any A_R transmitters and any receiver is invertible almost surely. As a result, the precoder $\mathbf{V}_{\mathcal{K}}$ will occupy a space of dimension $A_R L_n$ at the unintended receivers $[K_R] \setminus \mathcal{K}$. Thus, we design precoders $\mathbf{V}_{\mathcal{K}}$ such that the signals intended to receivers \mathcal{K} from all transmitters are aligned into a vector space of dimension $A_R L_{n+1}$ at the unintended receivers $[K_R] \setminus \mathcal{K}$. More specific, we choose $\mathbf{V}_{\mathcal{K}}$ such that

$$\begin{bmatrix} \mathbf{H}_{ki}^1 \\ \vdots \\ \mathbf{H}_{ki}^{A_R} \end{bmatrix} \mathbf{V}_{\mathcal{K}} \prec \begin{bmatrix} \tilde{\mathbf{V}}_{\mathcal{K}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \tilde{\mathbf{V}}_{\mathcal{K}} \end{bmatrix} \quad \forall i \in [K_T], \forall k \in [K_R] \setminus \mathcal{K}, \quad (20)$$

where $\mathbf{P} \prec \mathbf{Q}$ means that the column space of matrix \mathbf{P} is a subspace of the column space of the matrix \mathbf{Q} . In (20), there are $\Gamma = A_R K_T (K_R - t_R - 1)$ conditions between \mathbf{V}_K and $\tilde{\mathbf{V}}_K$ required to be satisfied. To achieve these conditions, we generate a random vector $\mathbf{a}_K = [a_K(1), \dots, a_K(T_n)]^T$, where its elements are drawn independently from a continuous distribution bounded between a non-zero minimum value and a finite maximum value. Then, we choose

$$\mathbf{V}_K = \left\{ \prod_{i=1}^{K_T} \prod_{r=1}^{A_R} \prod_{k \in [K_R] \setminus \mathcal{K}} (\mathbf{H}_{ki}^r)^{\alpha_{ki}^r} \mathbf{a}_K : 0 \leq \alpha_{ki} \leq n-1 \right\}$$

$$\tilde{\mathbf{V}}_K = \left\{ \prod_{i=1}^{K_T} \prod_{r=1}^{A_R} \prod_{k \in [K_R] \setminus \mathcal{K}} (\mathbf{H}_{ki}^r)^{\alpha_{ki}^r} \mathbf{a}_K : 0 \leq \alpha_{ki} \leq n \right\} \quad (21)$$

Thus, we can verify that the conditions in (20) are satisfied. The received signal at receiver RX_j after alignment design is given by

$$\mathbf{Y}_j = \sum_{i=1}^{K_T} \sum_{\substack{\mathcal{K} \subseteq [K_R] \\ \mathcal{K} \ni j}} \mathbf{H}_{ji} \mathbf{V}_K \mathbf{X}_{K,i}$$

$$+ \sum_{\substack{\mathcal{K} \subseteq [K_R] \\ \mathcal{K} \not\ni j}} \begin{bmatrix} \tilde{\mathbf{V}}_K & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \tilde{\mathbf{V}}_K \end{bmatrix} \tilde{\mathbf{X}}_K + \mathbf{N}_j \quad (22)$$

where $\tilde{\mathbf{X}}_K$ is the sum of interference data streams received in the same direction. We index the sets $\mathcal{K} \subseteq [K_R]$ for $j \in \mathcal{K}$ with \mathcal{K}_s for $s \in [\varrho]$, where $\varrho = \binom{K_R-1}{t_R}$. Furthermore, we index the sets $\mathcal{K} \subseteq [K_R]$ for $j \notin \mathcal{K}$ with \mathcal{K}_s for $s \in [\varrho+1 : \gamma]$, where $\gamma = \binom{K_R}{t_R+1}$. The desired streams of receiver RX_j arrive with direction

$$\mathbf{D}_j = \begin{bmatrix} \mathbf{D}_j^1 \\ \vdots \\ \mathbf{D}_j^{A_R} \end{bmatrix}$$

$$\mathbf{D}_j^r = [\mathbf{H}_{j1}^r \mathbf{V}_{K_1} \cdots \mathbf{H}_{jK_T}^r \mathbf{V}_{K_1} \cdots \mathbf{H}_{j1}^r \mathbf{V}_{K_\varrho} \cdots \mathbf{H}_{jK_T}^r \mathbf{V}_{K_\varrho}], \quad (23)$$

where $j \in \mathcal{K}_s$ for $s \in [\varrho]$. Meanwhile the interference signals have arrived after alignment with directions.

$$\mathbf{I}_j = [\mathbf{I}_{K_{\varrho+1}} \cdots \mathbf{I}_{K_\gamma}]$$

$$\mathbf{I}_{K_s} = \begin{bmatrix} \tilde{\mathbf{V}}_{K_s} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \tilde{\mathbf{V}}_{K_s} \end{bmatrix} \quad (24)$$

where $j \notin \mathcal{K}_s$ for $s \in [\varrho+1 : \gamma]$. To ensure that the receiver RX_j can decode the desired streams, we should maintain that the directions of all desired streams are linearly independent on each other and independent on all directions of the interference streams. This can be ensured if the following matrix

$$\mathbf{F}_j = [\mathbf{D}_j \quad \mathbf{I}_j]$$

has full rank of $A_R T_n$ almost surely for almost all channel realizations. The proof is skipped due to the space limitations, but it can be found in [12]. As a result, each receiver RX_j can decode its desired data streams over T_n -symbol extensions.

Since each message $W_{K,i}$ is encoded into L_n data streams, the achievable DoF per message is given by

$$d(t_R, A_R) = \lim_{n \rightarrow \infty} \frac{n^\Gamma}{\frac{K_T}{A_R} \binom{K_R-1}{t_R} n^\Gamma + \binom{K_R-1}{t_R+1} (n+1)^\Gamma}$$

$$= \frac{A_R}{K_T \binom{K_R-1}{t_R} + A_R \binom{K_R-1}{t_R+1}}$$

This completes the proof of Lemma 1.

VI. CONCLUSION

We have studied the cache-aided interference networks, where both transmitters and receivers are equipped with multiple antennas. First, we proposed an achievable scheme for minimizing the normalized delivery time (NDT) of the cache-aided SIMO interference network. Then, we obtain the NDT of the MIMO cache interference network by using the decomposition property. Our results show that the gain obtained from coded caching is higher than that obtained from spatial multiplexing. Hence, increasing the receiver-cache sizes achieves a higher gain than increasing the receive antennas. Moreover, we have shown that our proposed scheme outperforms the state-of-the-art schemes in the literature.

REFERENCES

- [1] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in *International Symposium on Information Theory (ISIT)*, pp. 809–813, IEEE, 2015.
- [2] R. Tandon and O. Simeone, "Cloud-aided wireless networks with edge caching: Fundamental latency trade-offs in fog radio access networks," in *International Symposium on Information Theory (ISIT)*, pp. 2029–2033, IEEE, 2016.
- [3] A. M. Girgis, O. Ercetin, M. Nafie, and T. ElBatt, "A converse bound for cache-aided interference networks," in the *52nd Asilomar Conference on Signals, Systems, and Computers*, CA, IEEE, Oct. 2018.
- [4] J. Zhang and O. Simeone, "Fundamental limits of cloud and cache-aided interference management with multi-antenna base stations," *arXiv preprint arXiv:1712.04266*, 2017.
- [5] A. M. Girgis, O. Ercetin, M. Nafie, and T. ElBatt, "Decentralized coded caching in wireless networks: Trade-off between storage and latency," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 2443–2447, June 2017.
- [6] A. M. Girgis, O. Ercetin, M. Nafie, and T. ElBatt, "Fundamental limits of memory-latency tradeoff in fog radio access networks under arbitrary demands," in *submitted*, 2018.
- [7] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr, "Fundamental limits of cache-aided interference management," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 3092–3107, 2017.
- [8] J. Hachem, U. Niesen, and S. N. Diggavi, "Degrees of freedom of cache-aided wireless interference networks," *IEEE Transactions on Information Theory*, vol. 64, pp. 5359–5380, July 2018.
- [9] F. Xu, M. Tao, and K. Liu, "Fundamental tradeoff between storage and latency in cache-aided wireless interference networks," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7464–7491, 2017.
- [10] Y. Cao, M. Tao, F. Xu, and K. Liu, "Fundamental storage-latency tradeoff in cache-aided mimo interference networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5061–5076, 2017.
- [11] H. Sun, T. Gou, and S. A. Jafar, "Degrees of freedom of mimo x networks: Spatial scale invariance and one-sided decomposability," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8377–8385, 2013.
- [12] A. M. Girgis, O. Ercetin, M. Nafie, and T. ElBatt, "Fundamental limits of cache-aided mimo wireless networks," in *preparation*, 2018.
- [13] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2856–2867, 2014.