

Article

Joint Congestion Control and Resource Allocation in Cache-Enabled Sensor Networks

Yuan Ren , Guangyue Lu ^{*} and Changyin Sun

Shaanxi Key Laboratory of Information Communication Network and Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, China

^{*} Correspondence: renyuan@xupt.edu.cn (Y.R.); tonylugy@163.com (G.L.)

Received: 14 June 2019; Accepted: 2 July 2019; Published: 5 July 2019



Abstract: In this paper, we investigate the optimal beamforming design to achieve joint congestion control and energy-efficient resource allocation in cache-enabled sensor networks. The network of interest works in the time-slotted mode. The dynamic buffering queue for each node is introduced to reflect the degree of network congestion and service delay. Then, a time-averaged sum rate maximization problem is proposed under the constraints of queue stability, instantaneous power consumption, average power consumption, and the minimum quality of service requirements. By introducing the method of Lyapunov optimization, the importance of buffering queue backlogs and sum rate maximization can be traded off, then the original queue-aware and time-averaged optimization problem is transformed into a weighted sum rate maximization problem at each time slot. It can be further converted into a second-order cone-programming problem by successive convex approximation, which is convex and can be efficiently solved by off-the-shelf solvers. Numerical results validate that wireless caching can greatly relieve the network congestion by reducing the buffering backlogs, and show that the proposed scheme can trade off the average queue length and time-averaged sum rate by selecting different control parameters.

Keywords: beamforming; congestion control; resource allocation; cache-enabled sensor networks; successive convex approximation; Internet of Things

1. Introduction

In recent years, the data traffic generated by mobile users has experienced explosive growth [1]. Along with the development of mobile networks, the Internet of Things (IoT) and other wireless techniques are expected to bring us a wide variety of mobile applications and even more mobile traffic will be generated [2,3]. Notably, it is predicted that most of the traffic will result from the multimedia video services, which require higher network throughput and stricter network latency. To meet these unprecedented traffic demands and challenges, the standardization process of the fifth-generation (5G) network is accelerated in aspects of the network capacity and latency [4,5]. However, the limited capacity of backhaul links becomes the bottleneck of the wireless networks for large-scale video transmissions. Moreover, much data traffic is produced minute by minute, and there are many repeated contents among it, which will be transmitted more than once during the traffic-peak time periods. Confronting such a severe situation, proactive caching is regarded as one of the most promising techniques in 5G communication system to effectively alleviate the severe backhaul burden and improve the service delay, which has drawn tremendous attention from the industries and academia [6–9].

Proactive caching refers to the technique which prefetches popular video files to the local cache of base station/cluster head (CH) via backhaul links during traffic-free periods [10]. Currently, there are many caching schemes, which can be roughly divided into two categories, namely coded

caching [11,12] and uncoded caching [13,14], and both of them can significantly relieve network burden and reduce service delay. Caching has shown great potentials in heterogeneous networks [15], device-to-device (D2D) communications [16], small cell networks [17,18] and cloud radio access networks (C-RANs) [19,20] to make the wireless networks perform in an energy-efficient manner.

With caching in sensor networks, the repeated end-to-end content deliveries can be avoided. The sensor nodes with limited battery will consume less power for data transmission, especially for data re-transmission for loss recovery [21]. Thus, the life span of the entire sensor network can be prolonged. The caching schemes have been investigated in the literature. Specifically, Sun et al. [22] proposed a novel energy-aware and latency-guaranteed dynamic resource caching strategy. The cache equipment was able to cache suitable data resources and the energy savings from servers were maximized in the meanwhile. In [23], they also proposed the re-cache and re-allocation scheme for popular data resources among load-unbalanced cache equipment to improve the delay performance. Though proactive caching can relieve the traffic burden to a certain degree, it is of great necessity and importance to observe the fluctuation of data traffic to relieve network congestion in the time periods of traffic surge. However, the current research ignored the time-varying characteristic of the dynamic traffic changes in the cache-enabled networks. Moreover, the issues of network congestion and service delay were not considered in a long-term observation, which were usually studied in an analytical manner based on the theory of stochastic geometry, e.g., [24,25].

In this paper, we study the optimal beamforming design to achieve joint congestion control and resource allocation in cache-enabled sensor networks. We propose a centralized scheme to solve this problem. Compared to this centralized method, the distributed approach can also be adopted to solve the joint optimization problems, such as those proposed in [26,27] to solve the joint problem of uplink power control and transmission rate allocation. However, this distributed approach is not appropriate in the considered downlink beamforming design. This is because the joint information of nodes is needed [13,14,28], and the battery and computational capacity of the sensor nodes, rather than the CH, is limited [29]. The degree of network congestion and service delay can be reflected by the length of dynamic buffering queue for each node, and it is obvious that smaller queue length will lead to smaller service delay and less network congestion. Then, under the constraints of queue stability, instantaneous/average power consumption, and minimum quality of service (QoS) requirements, the queue-aware and time-averaged sum rate optimization problem is proposed to improve the QoS provided to end nodes in the long term. The main contributions of this paper are summarized in the following aspects.

- By selecting different control parameters, control decisions are made to assign different importance levels to network congestion and sum rate maximization. In the case with severe traffic burden, more importance is allocated to network congestion control; while there is little congestion, the network operators will pay more attention to improve the sum rate performance.
- We observe the time-varying characteristic of the traffic fluctuations in a long time period instead of the instantaneous observation. The traffic queues of mobile nodes are established to reflect the congestion conditions of the entire network, which will be beneficial for network operators to monitor the dynamic network conditions and make the proper control decisions.
- The method of Lyapunov optimization is introduced to transform the proposed time-averaged maximization problem into a weighted sum rate maximization problem at each time slot. Then this problem is further converted into a second-order cone-programming (SOCP) problem via successive convex approximation (SCA), which owns lower computational complexity and can be efficiently solved.

The rest of this paper is organized as follows. In Section 2, we present the system model and formulate the time-averaged sum rate maximization problem. Then, the optimization algorithm based on Lyapunov optimization and SCA is proposed in Section 3. Simulation results are presented in Section 4, and this paper concludes in Section 5.

2. System Model and Problem Formulation

In this section, we present the system model, including the physical channel model, video cache model, and dynamic queue model. Then a queue-aware and time-averaged sum rate maximization problem is formulated.

2.1. Physical Channel Model

We consider a cluster with a CH to provide the on-demand video services to K single-antenna nodes, as shown in Figure 1. The cache-enabled sensor network works in the time-slotted mode, where the time slot is indexed by $t \in \{1, 2, 3, \dots\}$. The CH is equipped with R_T transmit antennas, and its transmit power is P_T . In each time slot, a block Rayleigh fading channel is considered. Let $\mathbf{h}_k(t) \in \mathbb{C}^{R_T \times 1}$ denote the channel from the CH to the k -th node. Define $\mathbf{w}_k(t) \in \mathbb{C}^{R_T \times 1}$ the beamforming vector from the CH to the k -th node at time slot t . The received signal for the k -th node can be written as

$$y_k(t) = \mathbf{h}_k^H(t) \mathbf{w}_k(t) x_k(t) + \sum_{n=1, n \neq k}^K \mathbf{h}_k^H(t) \mathbf{w}_n(t) x_n(t) + n_k(t), \quad (1)$$

where $x_k(t)$ is the required data symbol of the required video file with unit power, and $n_k(t)$ is the additive white Gaussian noise following $\mathcal{CN}(0, \sigma_k^2)$. Based on the received signal, the signal-to-interference-plus-noise ratio (SINR) of the k -th node can be shown as

$$\text{SINR}_k(t) = \frac{|\mathbf{h}_k^H(t) \mathbf{w}_k(t)|^2}{\sum_{n=1, n \neq k}^K |\mathbf{h}_k^H(t) \mathbf{w}_n(t)|^2 + \sigma_k^2}. \quad (2)$$

Therefore, the rate of the k -th node can be expressed as

$$R_k(t) = \log_2(1 + \text{SINR}_k(t)). \quad (3)$$

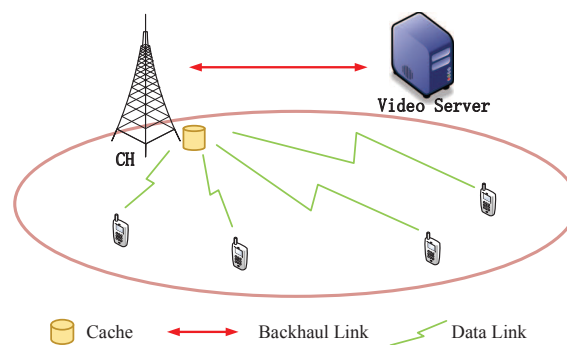


Figure 1. The CH with local cache capacity can provide K nodes with on-demand video services. If the required videos are cached in the local storage, the CH will directly serve the nodes with its local cache. Otherwise, the CH will deliver the required videos from the video server via backhaul links.

2.2. Video Cache Model

Note that the CH is equipped with local storage capacity, and can cache the most popular video files in advance. Assume that there are M files in the video server. Owing to the limited cache size, the CH can only cache a fraction, i.e., η , of these video files, which means the $\lfloor \eta M \rfloor$ most popular videos are locally cached. At the beginning of each time slot, each node is supposed to request only one video file, and the CH will collect these demands periodically. All videos are sorted in the descending order of popularity where more popular videos are ranked with smaller indices. The probability of

video requests of mobile nodes follows the Zipf's law [30]. To this end, the probability of the video file requested by the k -th node, i.e., m_k , can be obtained as

$$p_{m_k} = \frac{i_{m_k}^{-\alpha}}{\sum_{j=1}^M j^{-\alpha}}, \quad i_{m_k} \in \{1, 2, \dots, M\}, \quad (4)$$

where i_{m_k} is the index of video popularity order requested by the k -th node and α is the skewness parameter. In general, a larger α indicates most node requests can be satisfied by a few video files. Moreover, the time scale of video update is much larger than that of video delivery, e.g., 2–3 h for news with short videos and one week for movies [31], so this issue is not considered in this paper.

2.3. Dynamic Queue Model

In the considered cache-enabled sensor network, we intend to supervise and control the data traffic generated from the on-demand video services, such that the network congestion can be avoided, and the service delay is reduced. It is a common sense that longer buffering queue length will lead to larger service delay and more serious network congestion. Therefore, it is of vital significance to control the length of buffering data traffics of each node to achieve a stable and robust cache-enabled sensor network. If the network is in heavy traffic load, sufficient attention should be paid to the network congestion and proper decisions should be made to reduce the traffic backlogs. In this context, traffic buffering queues are introduced to reflect the fluctuation of data traffic, where the traffic arrival rate is denoted as $A_k(t)$ and the departure rate is $R_k(t)$ as defined in (3). Please note that the CH is equipped with local storage capacity, thus the traffic arrival rate $A_k(t)$ will be zero if the requested video file m_k can be obtained locally. Then the binary constant c_{m_k} is introduced to indicate whether the requested video m_k is cached in the CH or not. Therefore, the arrival rate $A_k(t)$ can be obtained as

$$A_k(t) = \begin{cases} \log_2(1 + \gamma_k), & c_{m_k} = 0, \\ 0, & c_{m_k} = 1, \end{cases} \quad (5)$$

where γ_k is the target QoS requirement to fetch the requested video file from the video server through backhaul links for the k -th node. Then, at the $(t + 1)$ -th time slot, the traffic buffering queue for the k -th node can be denoted as

$$Q_k(t + 1) = \max \{Q_k(t) - R_k(t), 0\} + A_k(t). \quad (6)$$

The queue length in the subsequent time slot only depends on the arrival rate, departure rate, and queue length in the previous time slot. To describe the stability of these queues, we can resort to the following definition [32].

Definition 1. The discrete time process $Q(t)$ is mean rate stable if

$$\lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E} \{|Q(t)|\} = 0. \quad (7)$$

Though the actual buffering queues are non-negative, we regulate $Q(t)$ in the form of absolute value so that this definition can be extended to the virtual queues whose values are likely to be negative.

2.4. Problem Formulation

We aim to observe the queue stability of the cache-enabled network in the long term, and consider the time-varying characteristics of the multimedia data traffic in the meanwhile. Therefore, a queue-aware and time-averaged sum rate maximization problem is proposed, which can be shown as

$$\max_{\{\mathbf{w}_k(t)\}} \lim_{T \rightarrow \infty} \sum_{t=1}^T \sum_{k=1}^K \frac{1}{T} \mathbb{E} \{R_k(t)\} \quad (8a)$$

$$\text{s.t. } Q_k(t) \text{ is mean rate stable, } \forall k, t, \quad (8b)$$

$$\sum_{k=1}^K \|\mathbf{w}_k(t)\|^2 \leq P_T, \forall t, \quad (8c)$$

$$\lim_{T \rightarrow \infty} \sum_{t=1}^T \sum_{k=1}^K \frac{1}{T} \|\mathbf{w}_k(t)\|^2 \leq P_{avg}, \quad (8d)$$

$$\text{SINR}_k(t) \geq \gamma_k, \forall k, t, \quad (8e)$$

where P_{avg} is the average power consumption. The stability of the buffering queues can be satisfied in constraint (8b). Constraint (8c) restricts the budget of instantaneous power consumption, while (8d) is the constraint of average power consumption observed in an infinite time period. Meanwhile, the minimum QoS requirements are guaranteed in the last constraint.

3. Congestion Control and Resource Allocation Optimization

In this section, the method of Lyapunov optimization is introduced to transform the original optimization problem into a weighted sum rate maximization problem at each time slot, where the control decisions can be made to trade off the importance of network congestion and sum rate maximization. Then, this problem is converted into an SOCP problem, which requires less computation effort and can be easily solved.

3.1. Lyapunov Optimization

Firstly, we construct a virtual queue related to the constraint (8d), which is shown in the following proposition.

Proposition 1. $S(t)$ is a virtual queue with $\mathbb{E} \{|S(0)|\} < \infty$. At time slot $(t+1)$, the virtual queue can be denoted as

$$S(t+1) = \max \{S(t) - P_{avg}, 0\} + P(t), \quad (9)$$

where $P(t) = \sum_{k=1}^K \|\mathbf{w}_k(t)\|^2$. If the virtual queue $S(t)$ is mean rate stable, the constraint (8d) holds.

Proof. The details for the proof are provided in Appendix A. \square

From Proposition 1, the constraint (8d) regarding to average power consumption is transformed into the issue of virtual queue stability. To this end, there are two kinds of queues, i.e., buffering queue $Q_k(t)$ and virtual queue $S(t)$. Denote $\Xi(t) = (Q_1(t), Q_2(t), \dots, Q_K(t), S(t))$ the mixed queue vector of actual traffic queues and virtual queue. According to [32], the scalar Lyapunov function can be defined as

$$L(\Xi(t)) = \frac{1}{2} \left\{ \sum_{k=1}^K Q_k^2(t) + S^2(t) \right\}. \quad (10)$$

This function can roughly reflect the degree of network congestion. Specifically, if the value of Lyapunov function $L(\Xi(t))$ is low, it means that all the queue lengths are small, and the system is stable and robust. However, the value of this function will increase if there is at least one queue whose traffic backlog is large. Next, the concept of Lyapunov drift is adopted to force the Lyapunov

function to a lower value, so as to obtain smaller queue length and buffering backlogs [32], which can be expressed as

$$\Delta(\Xi(t)) = \mathbb{E} \{L(\Xi(t+1)) - L(\Xi(t))\}. \quad (11)$$

To jointly control the network congestion and maximize the time-averaged sum rate, we can resort to the drift-plus-penalty expression, as given by

$$\Delta(\Xi(t)) - V\mathbb{E} \{R_k(t)|\Xi(t)\}, \quad (12)$$

where V is a non-negative control parameter to trade off the traffic backlogs and sum rate maximization. A larger V will put more emphasis on sum rate maximization, while a smaller V emphasizes more on network stability. So, we can make control decisions to achieve different requirements under different control parameters. It is obvious that a smaller value of this expression can achieve a smaller queue length and a larger sum rate. Then, the following proposition is recommended to present the upper bound of the drift-plus-penalty expression.

Proposition 2. *At time slot t , under the observed queue state, the drift-plus-penalty expression of joint congestion control and sum rate maximization satisfies*

$$\begin{aligned} \Delta(\Xi(t)) - V\mathbb{E} \{R_k(t)|\Xi(t)\} &\leq A + \sum_{k=1}^K Q_k(t)\mathbb{E} \{A_k(t) - R_k(t)|\Xi(t)\} \\ &\quad + S(t)\mathbb{E} \{P(t) - P_{avg}|\Xi(t)\} - V\mathbb{E} \{R_k(t)|\Xi(t)\}, \end{aligned} \quad (13)$$

where A is a finite constant satisfying

$$A \geq \frac{1}{2}\mathbb{E} \left\{ \sum_{k=1}^K R_k^2(t) + A_k^2(t)|\Xi(t) \right\} + \frac{1}{2}\mathbb{E} \left\{ P_{avg}^2 + P^2(t)|\Xi(t) \right\}. \quad (14)$$

Proof. The details for the proof of this proposition are presented in Appendix B. \square

To minimize the drift-plus-penalty expression, we can focus on minimizing the right-hand side of (13), which is the upper bound of the expression (12). To this end, the original optimization problem (8) can be transformed into

$$\min_{\{\mathbf{w}_k(t)\}} S(t)P(t) - \sum_{k=1}^K (Q_k(t) + V) R_k(t) \quad (15a)$$

$$\text{s.t. (8c), (8e).} \quad (15b)$$

3.2. Weighted Sum Rate Maximization

After the transformation based on the method of Lyapunov optimization, the joint congestion control and time-averaged sum rate maximization problem has been converted into a weighted sum rate maximization problem. Then, this problem can be further transformed into an SOCP problem via SCA. For simplicity, we denote

$$a(t) = S(t), \forall t, \quad (16)$$

$$b_k(t) = Q_k(t) + V, \forall k, t. \quad (17)$$

Firstly, the slack variable $z_k(t)$ is introduced, satisfying $z_k(t) \leq \text{SINR}_k(t)$. Then, the approximation method used in [33,34] is adopted to obtain the lower bound of $\log_2(1 + z_k(t))$, which can be presented as

$$\log_2(1 + z_k(t)) \geq \theta_k(t)\log_2 z_k(t) + \mu_k(t), \quad (18)$$

where

$$\theta_k(t) = \frac{\tilde{z}_k(t)}{1 + \tilde{z}_k(t)}, \quad (19)$$

$$\mu_k(t) = \log_2(1 + \tilde{z}_k(t)) - \frac{\tilde{z}_k(t)}{1 + \tilde{z}_k(t)} \log_2 \tilde{z}_k(t). \quad (20)$$

It is obvious that the equality holds in (18) when $\tilde{z}_k(t) = z_k(t)$. The selection of the point $\tilde{z}_k(t)$ will affect the optimal value of the optimization problem, so $\tilde{z}_k(t)$ can be updated by $z_k^{(n)}(t)$ iteratively where the superscript (n) means the optimal solution obtained in the n -th iteration. Then the slack variable $d_k(t)$ is introduced, which satisfies $\sqrt{\sum_{n=1, n \neq k}^K |\mathbf{h}_k^H(t) \mathbf{w}_n(t)|^2} + \sigma_k^2 \leq d_k(t)$. After these manipulations, problem (15) can be converted as follows.

$$\min_{\{\mathbf{w}_k(t), z_k(t), d_k(t)\}} a(t)P(t) - \sum_{k=1}^K b_k(t) (\theta_k(t) \log_2 z_k(t) + \mu_k(t)) \quad (21a)$$

$$\text{s.t.} \quad \mathbf{h}_k^H(t) \mathbf{w}_k(t) \geq \sqrt{z_k(t)} d_k(t), \quad \forall k, t, \quad (21b)$$

$$\text{Im}(\mathbf{h}_k^H(t) \mathbf{w}_k(t)) = 0, \quad \forall k, t, \quad (21c)$$

$$\sqrt{\sum_{n=1, n \neq k}^K |\mathbf{h}_k^H(t) \mathbf{w}_n(t)|^2} + \sigma_k^2 \leq d_k(t), \quad \forall k, t, \quad (21d)$$

$$(8c), (8e). \quad (21e)$$

As shown in (21c), the imaginary part of $\mathbf{h}_k^H(t) \mathbf{w}_k(t)$ is set to zero and this will have no effect on the optimal value of the optimization problem since phase rotations of beamforming vector $\mathbf{w}_k(t)$ can bring out the same objective function while satisfying all constraints [35].

(1). Transformation of (21b). Please note that $\sqrt{z_k(t)} d_k(t)$ is not in the convex form, so it needs to be converted into the convex form. Referring to [35,36], the convex upper bound of $\sqrt{z_k(t)} d_k(t)$ can be shown as

$$\sqrt{z_k(t)} d_k(t) \leq \frac{x_k}{2} d_k^2(t) + \frac{1}{2x_k} z_k(t), \quad (22)$$

when $x_k = \frac{\sqrt{z_k(t)}}{d_k(t)}$ the equality holds. Then (21b) can be approximated as

$$\mathbf{h}_k^H(t) \mathbf{w}_k(t) \geq \frac{x_k^{(n)}}{2} d_k^2(t) + \frac{1}{2x_k^{(n)}} z_k(t), \quad (23)$$

and its second-order cone (SOC) form is

$$||[\sqrt{2x_k^{(n)}} d_k(t); (\mathbf{h}_k^H(t) \mathbf{w}_k(t) - \frac{1}{2x_k^{(n)}} z_k(t) - 1)]|| \leq \mathbf{h}_k^H(t) \mathbf{w}_k(t) - \frac{1}{2x_k^{(n)}} z_k(t) + 1, \quad (24)$$

where

$$x_k^{(n)} = \frac{\sqrt{z_k^{(n)}(t)}}{d_k^{(n)}(t)}. \quad (25)$$

(2). Transformation of (21d) and (8e). The SOC form of (21d) can be easily obtained as

$$||[\mathbf{h}_k^H(t) \mathbf{w}_1(t); \dots; \mathbf{h}_k^H(t) \mathbf{w}_{k-1}(t); \mathbf{h}_k^H(t) \mathbf{w}_{k+1}(t); \dots; \mathbf{h}_k^H(t) \mathbf{w}_K(t)]|| \leq d_k(t). \quad (26)$$

As for the constraint (8e), it can be approximated as

$$\mathbf{h}_k^H(t) \mathbf{w}_k(t) \geq \sqrt{\gamma_k} d_k(t). \quad (27)$$

To this end, the optimization problem (15) has been transformed into an SOCP problem, as presented as follows

$$\min_{\{\mathbf{w}_k(t), z_k(t), d_k(t)\}} a(t)P(t) - \sum_{k=1}^K b_k(t) (\theta_k(t) \log_2 z_k(t) + \mu_k(t)) \quad (28a)$$

$$\text{s.t.} \quad (8c), (21c), (24), (26), (27). \quad (28b)$$

The pseudo-code to solve the optimization problem (28) is presented in Algorithm 1. Note that the initial point of this algorithm is generated randomly until all constraints are satisfied. This has been validated a feasible method in practical simulations.

Algorithm 1: The proposed iterative algorithm for solving (28)

- 1: Initialization: Set $n = 0$, and find a point $(z_k^{(0)}(t), d_k^{(0)}(t), \mathbf{w}_k^{(0)}(t))$ randomly that is feasible to problem (15).
 - 2: Repeat:
 - 3: Solve the SOCP problem (28) with $(z_k^{(n)}(t), d_k^{(n)}(t), \mathbf{w}_k^{(n)}(t))$ and obtain the optimal point $(z_k^*(t), d_k^*(t), \mathbf{w}_k^*(t))$.
 - 4: Update $(z_k^{(n)}(t), d_k^{(n)}(t), \mathbf{w}_k^{(n)}(t)) = (z_k^*(t), d_k^*(t), \mathbf{w}_k^*(t))$.
 - 5: Update $\theta_k(t)$ using (19), $\mu_k(t)$ using (20), $x_k^{(n)}$ using (25), then $n = n + 1$.
 - 6: Until convergence or reach the requested number of iterations, update $a(t)$ using (16) and $b_k(t)$ using (17), then $t = t + 1$.
-

3.3. Convergence and Complexity Analysis

At time slot t , supposing $(z_k^{(n)}(t), d_k^{(n)}(t), \mathbf{w}_k^{(n)}(t))$ is the optimal point in the n -th iteration, which is adopted as the starting feasible point in the $(n + 1)$ -th iteration. It is apparent that the optimal value of the objective function in the $(n + 1)$ -th iteration must be larger than or at least equal to that obtained in the n -th iteration for a maximization problem. Furthermore, the optimization problem (28) is bounded by the budget of transmit power consumption. Therefore, the convergence point is bound to be obtained after a sequence of non-decreasing points. After enough iterations, the difference of the optimal values between two adjacent iterations can be ignored. According to the conclusion presented in [36], the optimal solution also satisfies the Karush–Kuhn–Tucker (KKT) conditions of the optimization problem (21). After the transformation based on SCA, Algorithm 1 aims to solve the SOCP problem (28), which requires lower computation effort. To be more specific, by referring to [37], the maximum number of iterations is restricted by $\mathcal{O}(\sqrt{5K + 2}) \ln(1/\zeta)$ when interior method is employed, where ζ is the accuracy threshold. The required computational cost for each iteration is bounded by $\mathcal{O}(K^3(R_T^3 + 7R_T^2 + 16R_T + 12) + (3K + 1)(R_T + 2))$.

4. Simulation Results

In this section, the numerical results are presented. In our simulations, we consider a small-scale sensor network, where one CH serves six nodes. This is because serving too many nodes for a CH will lead to unbearable computational complexity [20], and thus the QoS may not be guaranteed. When there are multiple nodes to be served, they can be divided into many clusters. Then, the beamforming design for each cluster can be realized by the same methods presented in this paper. MATLAB is

used as a simulation tool, and SeDuMi is employed as an optimization solver [38]. Suppose the channels are independent and identically distributed with elements following the complex Gaussian distribution with zero mean and unit variance, i.e., $\mathcal{CN}(0, 1)$ at each time slot. Assume that there are $M = 1000$ video files in the video server and the skewness parameter α is set to 1 unless stated otherwise. The values of simulation parameters are summarized in Table 1.

Table 1. Values of Simulation Parameters.

Parameters	Values
K	6
P_T	0.2 W
P_{avg}	0.18 W
R_T	16
M	1000
η	0.1
α	1
α_k	2 dB
σ^2	1

Convergence Property of Algorithm 1: It is worth noting that the convergence property is validated in a random time slot. As shown in Figure 2, the proposed algorithm to solve the weighted sum rate maximization problem can converge to the local optimal point within a few number of iterations. So, Algorithm 1 shows great potentials in practical implementations and can be extended to the large-scale cache-enabled sensor networks to achieve joint congestion control and time-averaged sum rate maximization in an energy-efficient manner.

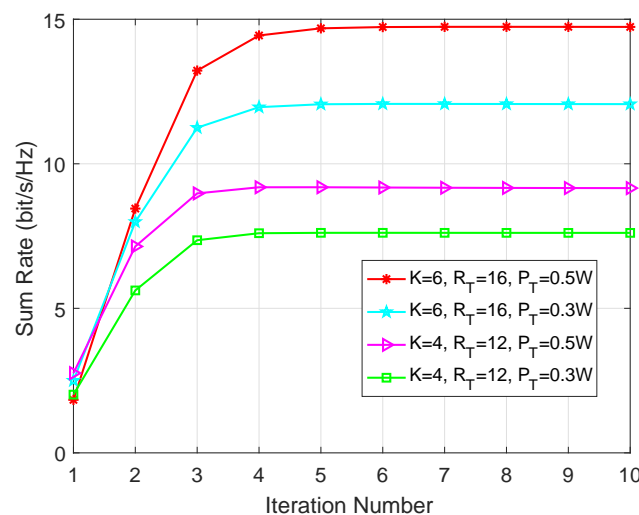


Figure 2. The convergence property of Algorithm 1.

Analysis of Average Queue Length: In Figure 3, we present how the average queue length changes in the observed time slots under the given control parameter V , where the average queue length is defined as $\bar{Q}(t) = \frac{1}{t} \sum_{\tau=1}^t \sum_{k=1}^K Q_k(\tau)$. Here the benchmark scheme for comparison is the CH without cache capacity. Firstly, it can be concluded that the proposed caching scheme can greatly reduce the average queue length under the given control parameter and arrival rate. Secondly, a larger control parameter V will lead to a longer average queue length, which can be intuitively understood that more emphasis is placed on the sum rate maximization. Finally, all curves tend to the stable levels when the observed time period is long enough, which strongly validates the stable characteristic of the traffic buffering queues.

Analysis of the Time-Averaged Sum Rate with Different Control Parameters: As shown in Figure 4, the average sum rate, defined as $\bar{R} = \sum_{t=1}^T \sum_{k=1}^K \frac{1}{T} \mathbb{E}\{R_k(t)\}$, is increasing with the control parameter V , which can be explained that a larger V will emphasize more on the sum rate maximization and lead to the larger value of \bar{R} . Furthermore, under the given V , \bar{R} will increase with the skewness parameter α . It is because that a larger α means that fewer videos will satisfy the majority of node requirements and more of them can be served by local cache, thus a smaller amount of arriving data traffics will be allowed to admit to the cache-enabled sensor networks, which can effectively relieve the network congestion and improve the sum rate of end nodes.

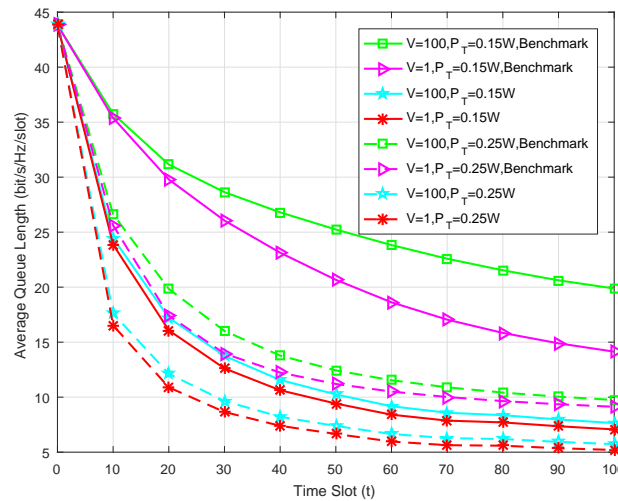


Figure 3. Average queue length at the observed time slot for $K = 6$ and $R_T = 16$.

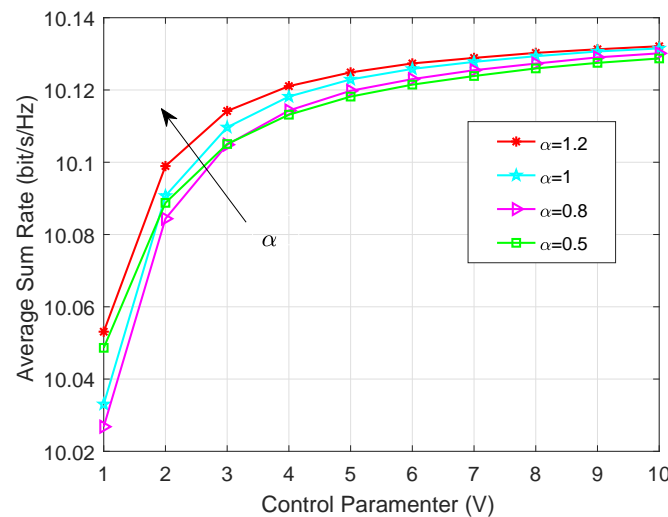


Figure 4. Time-averaged sum rate under different skewness parameter for $K = 6$, $R_T = 16$ and $P_T = 0.2$ W.

5. Conclusions

In this paper, we jointly considered the congestion control and resource allocation optimization in cache-enabled sensor networks. Firstly, we showed the system model and then proposed a time-averaged sum rate maximization problem. With the help of Lyapunov optimization, the original maximization problem was transformed into a weighted sum rate maximization problem at each time slot. Then, it was converted into an SOCP problem based on SCA. Simulation results demonstrated the

fast convergence property of the proposed algorithm and the characteristic of dynamic queue stability under the proposed scheme.

Author Contributions: Conceptualization, Y.R., G.L. and C.S.; methodology, Y.R.; investigation, Y.R. and G.L.; writing—original draft preparation, Y.R.; writing—review and editing, Y.R. and G.L.; supervision, G.L. and C.S.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grants 61801382 and 61701399, in part by the Research Program of Education Bureau of Shaanxi Province (18JK0718), in part by the Innovation Team Project of Shaanxi Province (2017KCT-30-02), in part by the Key Industrial Chain Project of Shaanxi Province (2019ZDLGY07-06), and in part by the Natural Science Foundation Key Project of Shaanxi Province (2019JZ-06).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IoT	Internet of Things
5G	Fifth Generation
CH	Cluster Head
D2D	Device-to-Device
C-RAN	Cloud Radio Access Network
QoS	Quality of Service
SOCP	Second-Order Cone Programming
SCA	Successive Convex Approximation
SINR	Signal-to-Interference-plus-Noise Ratio
SOC	Second-Order Cone
KKT	Karush–Kuhn–Tucker

Appendix A

If the virtual queue $S(t)$ is mean rate stable, it indicates that $\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \{|S(T)|\} = 0$. Calculate the sum of (9) from $t = 0$ to $T - 1$, and we can get

$$\begin{aligned} \sum_{t=0}^{T-1} S(t+1) &= \sum_{t=0}^{T-1} [\max \{S(t) - P_{avg}, 0\} + P(t)] \\ &\geq \sum_{t=0}^{T-1} [S(t) - P_{avg} + P(t)], \end{aligned}$$

which is equivalent to

$$S(T) - S(0) \geq \sum_{t=0}^{T-1} P(t) - TP_{avg}. \quad (A1)$$

Take expectation operation of (A1) and divide it by $T \rightarrow \infty$, then the left-hand side of (A1) can be shown as

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \{|S(T)|\} - \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \{|S(0)|\} = 0.$$

Therefore, $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T P(t) - P_{avg} \leq 0$ and the constraint (8d) is satisfied. This completes the proof of Proposition 1.

Appendix B

Based on the definition of $Q_k(t+1)$, we can get

$$\begin{aligned} Q_k^2(t+1) &= (\max \{Q_k(t) - R_k(t), 0\} + A_k(t))^2 \\ &\leq Q_k^2(t) + R_k^2(t) + A_k^2(t) - 2Q_k(t)(R_k(t) - A_k(t)). \end{aligned} \quad (A2)$$

Taking sum of both sides in inequality (A2) for $k = 1, 2, \dots, K$, we can obtain

$$\sum_{k=1}^K \left(Q_k^2(t+1) - Q_k^2(t) \right) \leq \sum_{k=1}^K [R_k^2(t) + A_k^2(t) - 2Q_k(t)(R(t) - A_k(t))].$$

Similarly, we can conclude that the virtual queue $S(t)$ satisfies

$$S^2(t+1) - S^2(t) \leq P_{avg}^2 + P^2(t) - 2S(t)(P_{avg} - P(t)).$$

Therefore, the Lyapunov drift is bounded by

$$\begin{aligned} \Delta(\Xi(t)) &= \mathbb{E} \{ L(\Xi(t+1)) - L(\Xi(t)) \} \\ &= \frac{1}{2} \left(\sum_{k=1}^K \left(Q_k^2(t+1) - Q_k^2(t) \right) + \left(S^2(t+1) - S^2(t) \right) \right) \\ &\leq \frac{1}{2} \sum_{k=1}^K \left(R_k^2(t) + A_k^2(t) - 2Q_k(t)(R_k(t) - A_k(t)) \right) \\ &\quad + \frac{1}{2} \left(P_{avg}^2 + P^2(t) - 2S(t)(P_{avg} - P(t)) \right). \end{aligned} \quad (A3)$$

Subtract the term $V\mathbb{E}\{R_k(t)|\Theta(t)\}$ in both sides of (A3), then it can be shown as the inequality (13) where A meets (14). Therefore, the proof of Proposition 2 is completed.

References

1. Cisco. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper*; Cisco: San Jose, CA, USA, 2017.
2. Li, L.; Zhao, G.; Blum, R.S. A Survey of Caching Techniques in Cellular Networks: Research Issues and Challenges in Content Placement and Delivery Strategies. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 1710–1732. [CrossRef]
3. Duan, P.; Jia, Y.; Liang, L.; Rodriguez, J.; Huq, K.M.S.; Li, G. Space-Reserved Cooperative Caching in 5G Heterogeneous Networks for Industrial IoT. *IEEE Trans. Ind. Inform.* **2018**, *14*, 2715–2724. [CrossRef]
4. Zhang, K.; Leng, S.; He, Y.; Maharjan, S.; Zhang, Y. Cooperative Content Caching in 5G Networks with Mobile Edge Computing. *IEEE Wirel. Commun.* **2018**, *25*, 80–87. [CrossRef]
5. Parvez, I.; Rahmati, A.; Guvenc, I.; Sarwat, A.I.; Dai, H. A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 3098–3130. [CrossRef]
6. Bastug, E.; Bennis, M.; Debbah, M. Living on the Edge: The Role of Proactive Caching in 5G Wireless Networks. *IEEE Commun. Mag.* **2014**, *52*, 82–89. [CrossRef]
7. Paschos, G.; Bastug, E.; Land, I.; Caire, G.; Debbah, M. Wireless Caching: Technical Misconceptions and Business Barriers. *IEEE Commun. Mag.* **2016**, *54*, 16–22. [CrossRef]
8. Wang, X.; Chen, M.; Taleb, T.; Ksentini, A.; Leung, V.C. Cache in the Air: Exploiting Content Caching and Delivery Techniques for 5G Systems. *IEEE Commun. Mag.* **2014**, *52*, 131–139. [CrossRef]
9. Bilal, M.; Kang, S.G. A Cache Management Scheme for Efficient Content Eviction and Replication in Cache Networks. *IEEE Access* **2017**, *5*, 1692–1701. [CrossRef]
10. Giatsoglou, N.; Ntontin, K.; Kartsakli, E.; Antonopoulos, A.; Verikoukis, C. D2D-Aware Device Caching in mmWave-Cellular Networks. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 2025–2037. [CrossRef]
11. Xu, F.; Liu, K.; Tao, M. Cooperative Tx/Rx Caching in Interference Channels: A Storage-Latency Tradeoff Study. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 2034–2038.
12. Chen, Z.; Lee, J.; Quek, T.Q.; Kountouris, M. Cooperative Caching and Transmission Design in Cluster-Centric Small Cell Networks. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 3401–3415. [CrossRef]

13. Tao, M.; Chen, E.; Zhou, H.; Yu, W. Content-Centric Sparse Multicast Beamforming for Cache-Enabled Cloud RAN. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 6118–6131. [[CrossRef](#)]
14. Zhou, H.; Tao, M.; Chen, E.; Yu, W. Content-Centric Multicast Beamforming in Cache-Enabled Cloud Radio Access Networks. In Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, 6–10 December 2015; pp. 1–6.
15. Tran, T.D.; Hoang, T.D.; Le, L.B. Caching for Heterogeneous Small-Cell Networks With Bandwidth Allocation and Caching-Aware BS Association. *IEEE Wireless Commun. Lett.* **2018**, *8*, 49–52. [[CrossRef](#)]
16. Chen, B.; Yang, C.; Molisch, A.F. Cache-Enabled Device-to-Device Communications: Offloading Gain and Energy Cost. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 4519–4536. [[CrossRef](#)]
17. Tamoor-ul Hassan, S.; Bennis, M.; Nardelli, P.H.; Latva-Aho, M. Caching in Wireless Small Cell Networks: A Storage-Bandwidth Tradeoff. *IEEE Commun. Lett.* **2016**, *20*, 1175–1178. [[CrossRef](#)]
18. Krishnan, S.; Afshang, M.; Dhillon, H.S. Effect of Retransmissions on Optimal Caching in Cache-Enabled Small Cell Networks. *IEEE Trans. Veh. Technol.* **2017**, *66*, 11383–11387. [[CrossRef](#)]
19. Zhao, Z.; Peng, M.; Ding, Z.; Wang, W.; Poor, H.V. Cluster Content Caching: An Energy-Efficient Approach to Improve Quality of Service in Cloud Radio Access Networks. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 1207–1221. [[CrossRef](#)]
20. Peng, M.; Yu, Y.; Xiang, H.; Poor, H.V. Energy-Efficient Resource Allocation Optimization for Multimedia Heterogeneous Cloud Radio Access Networks. *IEEE Trans. Multimed.* **2016**, *18*, 879–892. [[CrossRef](#)]
21. Zirak, M.; Yaghmaee, M.H.; Tabbakh, S.R.K. A Distributed Cache Points Selection Scheme for Reliable Transport Protocols with Intermediate Caching in Wireless Sensor Networks. In Proceedings of the 16th International Conference on Advanced Communication Technology, Pyeongchang, Korea, 16–19 February 2014; pp. 1–4.
22. Sun, X.; Ansari, N. Dynamic Resource Caching in the IoT Application Layer for Smart Cities. *IEEE Internet Things J.* **2018**, *5*, 606–613. [[CrossRef](#)]
23. Sun, X.; Ansari, N. Traffic Load Balancing among Brokers at the IoT Application Layer. *IEEE Trans. Netw. Serv. Manag.* **2018**, *15*, 489–502. [[CrossRef](#)]
24. Yang, C.; Yao, Y.; Chen, Z.; Xia, B. Analysis on Cache-Enabled Wireless Heterogeneous Networks. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 131–145. [[CrossRef](#)]
25. Wu, L.; Zhang, W. Caching-Based Scalable Video Transmission over Cellular Networks. *IEEE Commun. Lett.* **2016**, *20*, 1156–1159. [[CrossRef](#)]
26. Tsiropoulou, E.E.; Vamvakas, P.; Katsinis, G.K.; Papavassiliou, S. Combined Power and Rate Allocation in Self-Optimized Multi-Service Two-Tier Femtocell Networks. *Comput. Commun.* **2015**, *72*, 38–48. [[CrossRef](#)]
27. Musku, M.R.; Chronopoulos, A.T.; Popescu, D.C.; Stefanescu, A. A Game-Theoretic Approach to Joint Rate and Power Control for Uplink CDMA Communications. *IEEE Trans. Commun.* **2010**, *58*, 923–932. [[CrossRef](#)]
28. Xiang, Z.; Tao, M. Robust Beamforming for Wireless Information and Power Transmission. *IEEE Wirel. Commun. Lett.* **2012**, *1*, 372–375. [[CrossRef](#)]
29. Jia, D.; Zhu, H.; Zou, S.; Hu, P. Dynamic Cluster Head Selection Method for Wireless Sensor Network. *IEEE Sens. J.* **2016**, *16*, 2746–2754. [[CrossRef](#)]
30. Breslau, L.; Cao, P.; Fan, L.; Phillips, G.; Shenker, S. Web Caching and Zipf-Like Distributions: Evidence and Implications. In Proceedings of the IEEE Conference on Computer Communications (INFOCOM), New York, NY, USA, 21–25 March 1999; pp. 126–134.
31. Golrezaei, N.; Molisch, A.F.; Dimakis, A.G.; Caire, G. Femtocaching and Device-to-Device Collaboration: A New Architecture for Wireless Video Distribution. *IEEE Commun. Mag.* **2013**, *51*, 142–149. [[CrossRef](#)]
32. Neely, M.J. Stochastic Network Optimization with Application to Communication and Queueing Systems. *Synth. Lect. Commun. Netw.* **2010**, *3*, 1–211. [[CrossRef](#)]
33. He, S.; Huang, Y.; Jin, S.; Yang, L. Energy Efficient Coordinated Beamforming Design in Multi-Cell Multicast Networks. *IEEE Commun. Lett.* **2015**, *19*, 985–988. [[CrossRef](#)]
34. Papandriopoulos, J.; Evans, J.S. SCALE: A Low-Complexity Distributed Protocol for Spectrum Balancing in Multiuser DSL Networks. *IEEE Trans. Inf. Theory* **2009**, *55*, 3711–3724. [[CrossRef](#)]
35. Tran, L.N.; Hanif, M.F.; Tolli, A.; Juntti, M. Fast Converging Algorithm for Weighted Sum Rate Maximization in Multicell MISO Downlink. *IEEE Signal Process. Lett.* **2012**, *19*, 872–875. [[CrossRef](#)]
36. Beck, A.; Ben-Tal, A.; Tretushvili, L. A Sequential Parametric Convex Approximation Method with Applications to Nonconvex Truss Topology Design Problems. *J. Glob. Optim.* **2010**, *47*, 29–51. [[CrossRef](#)]

37. Wang, K.Y.; So, A.M.C.; Chang, T.H.; Ma, W.K.; Chi, C.Y. Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks: Tractable Approximations by Conic Optimization. *IEEE Trans. Signal Process.* **2014**, *62*, 5690–5705. [[CrossRef](#)]
38. Grant, M.; Boyd, S. CVX: Matlab Software for Disciplined Convex Programming, Version 2.1. 2016. Available online: <http://cvxr.com/cvx/> (accessed on 4 July 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).