Proactive Multicasting with Predictable Demands

John Tadrous, Atilla Eryilmaz, and Hesham El Gamal
Department of Electrical and Computer Engineering
The Ohio State University, Columbus, OH, USA
E-mail: {tadrousj, eryilmaz, helgamal}@ece.osu.edu

Abstract—In a recent work, we have introduced the notion of proactive resource allocation in wireless networks whereby the predictability of user demands are leveraged to significantly enhance the spectral efficiency of the network in outage limited regimes. In this paper, we expand the horizon to the important scenario of multicast traffic. Our analysis reveals two additional types of gains that can be leveraged in this proactive multicast scenario. The first can be attributed to the basic nature of multicast traffic in which each request would represent a data source rather than a user, as it would in the unicast case. The second is the *demand alignment* phenomenon whereby the predictive network would wait to gather as much requests as possible and serve them altogether using the same resources. We analytically derive the impact of these advantages on the system diversity gain, which quantifies the exponential decay rate of the outage probability, and further illustrate the resulting gains via numerical results.

I. Introduction

In [1], [2], we have introduced a novel scheme for resource allocation that exploits the predictability of the wireless network traffic to significantly reduce the capacity required to maintain a target level of Quality-of-Service (QoS) defined as the outage probability or improve the QoS performance given a certain capacity limit. We have addressed scenarios when the network is unicast and characterized the asymptotic decay rate of the outage probability with the network capacity which we have defined as the diversity gain. We have shown that the new proactive resource allocation technique improves the diversity gain of the network by a factor of $T + 1$, where T is a prediction interval by which the wireless network can anticipate users' requests in advance.

In this paper, we extend our focus to consider wireless networks with multicast capabilities where the wireless network can serve all requests in the system targeting the same data source by the same resources required to serve a single request. This indeed offers a potential for additional gains that can be leveraged in such proactive networks. In particular, we show that there are two types of enhancements to the diversity gain of multicast networks compared to the unicast case. The first type is that, in multicast networks, the input traffic differs completely from the input traffic to the unicast. In fact, the unicast traffic can be modeled independently of the number of data sources in the system, this is because each unicast request is served individually, whereas the multicast traffic

is highly dependent on the number of data sources of the network since all requests targeting a certain data source can be served together by the same amount of resources that are sufficient to serve a single request. For instance, consider a network that has C resources and each request requires only a single unit of these resources, hence, if this network is unicast with random input traffic that can, at some instants, exceed the maximum capacity C, then at some point in time the network may encounter more requests than what it can serve concurrently, and hence run into an outage. However, if this network is multicast with a single data source, then all of the input traffic is targeting such data source, and hence, serving one request of the arriving requests at any point in time implies that all of the arriving requests are served implicitly by a single unit of resources, and there will be zero outage probability.

The second kind of the potential multicast diversity gain is based on the alignment property offered by prediction. If the multicast network can predict the arriving requests by an interval of time T before their actual arrival, then it does not need to serve these requests once they are predicted. Instead, it can wait and gather as much requests as possible in that interval T and then serve them altogether by a single unit of resources. This, consequently, would save network resources that can be harnessed to serve other traffic flows that are non-predictive, or unicast, sharing the same network. In this paper, we address these two types of additional gains offered by multicast proactive networks separately, each in a different scenario. We also compare the obtained results with the unicast case whose results derived in [2]. Then, we show numerical simulations to demonstrate the analytical results.

The rest of this paper is organized as follows. In Section II, we define the system model. In Section III, we analyze the diversity gain of the multicast networks to demonstrate the improved gain offered by the multicast traffic model. In Section IV, we introduce the notion of multicast traffic alignment when the network is predictive and derive a sufficient condition that implies the superiority of alignment policy. In Section V, we present numerical results, then the paper is concluded in Section VI.

II. System Model

We consider a time-slotted wireless network where requests arrive only at the beginning of each slot. The random number of arriving requests at time slot $n > 0$ is denoted by $Q(n)$, which includes both the unicast and multicast traffic types. The network is assumed to have a fixed capacity of C units

per slot and that, unless otherwise stated, each served request in the network consumes *one* unit of the network capacity.

We distinguish two types of wireless resource allocation: **reactive** and **proactive**. In reactive (or traditional) resource allocation, the wireless network responds to user requests right after they are initiated by the user, whereas in proactive resource allocation, the network can track, learn and then *predict* the user requests ahead of time, and hence possesses more flexibility in scheduling these requests before their actual time of arrival. We refer to the networks that perform reactive and proactive resource allocation, respectively, as **non-predictive** and **predictive** networks.

We model the predictive capabilities of the network by introducing a prediction window duration of T slots, where T is a non-negative integer. Thus, letting $q(n) \in \{1, \cdots, Q(n)\}$ denote the identifier of requests arriving in time slot n and T denotes the number of slots allowed to serve those requests so that request $q(n)$ arriving in slot n has a deadline slot of $D_{q(n)} = n + T$. Hence, scheduling the requests of the predictive network can be considered as a scheduling problem with deadlines, e.g., [3], [4].

We note that the requests for *unicast* and *multicast* traffic differ in the following sense. Each unicast traffic request is for *new* data and hence must be served separately from all other requests. On the other hand, multicast traffic requests are for data from a common pool of L data sources (e.g. files, packets, movies, podcasts, etc.) and hence can be simultaneously served to multiple interested users.

We distinguish the unicast and multicast requests arriving at slot n respectively by $Q^u(n)$ and $Q^m(n)$ so that $Q(n) = Q^u(n) + Q^m(n)$ for each n. Since the multicast requests are for particular data sources out of L, we can further decompose the total number of multicast requests $Q^m(n)$ arriving in slot n into the number of requests for particular data sources as

$$Q^m(n) = \sum_{l=1}^{L} Q^{m,[l]}(n), \quad \text{for all } n > 0,$$

where $Q^{m,[l]}(n)$ denotes the number of multicast requests for data source l arriving in slot n.

In the rest of the paper, we assume that the unicast and individual multicast request arrivals are independent from each other and across time slots. Moreover, following the natural Poisson approximation of call arrivals to service centers, we assume that $Q^u(n)$ is Poisson distributed with mean λ^u, and $Q^{m,[l]}(n)$ is Poisson distributed with mean $\lambda^{m,[l]} \triangleq p^{[l]} \lambda^m$, where $\mathbf{p} \triangleq (p^{[l]})_{l=1}^{L}$ is a valid probability distribution capturing the potentially asymmetric multicast demands over the pool of L data sources. Hence, the total number of multicast data requests $Q^m(n)$ in slot n is also be Poisson distributed with mean λ^m.

In this work, we consider outage probability as the metric of QoS where the *outage event* is defined as.

The time scale of a slot is determined by prediction window and is typically large enough to average out the time variations in the wireless channels.

$\mathbf{p}$ is a valid distribution if $0 \le p^{[l]} \le 1$ and $\sum_{l=1}^{L} p^{[l]} = 1$.

Definition 1: Let $N_0^u(n)$ be the number of awaiting unicast requests at the beginning of time slot n that expire in slot n, i.e., with deadline n. Similarly, let $N_0^{m,[l]}(n)$ denote the indicator that there is *at least* one awaiting multicast request for data source $l \in \{1, \cdots, L\}$ that expires in slot n. Then, letting $N_0^m(n) \triangleq \sum_{l=1}^{L} N_0^{m,[l]}(n)$, the *outage event* is

$$\mathcal{O} \triangleq \{n > 0 : N_0^u(n) + N_0^m(n) > C\}, \quad (1)$$

collecting the slots when at least one urgent unicast or multicast request must be dropped.

We further let $\mathcal{O}_u$ and $\mathcal{O}_m$ respectively denote the subset events of $\mathcal{O}$ when unicast and multicast outages occur. Thus, $\mathcal{O}_u$ and $\mathcal{O}_m$ respectively excludes slots when only unicast and only multicast outages occur. ◇

The binary parameter $X^{m,[l]}(n)$ for each multicast data source $l \in \{1, \cdots, L\}$ is also defined as

$$X^{m,[l]}(n) \triangleq \begin{cases} 1, & \text{if } Q^{m,[l]}(n) > 0, \\ 0, & \text{if } Q^{m,[l]}(n) = 0, \end{cases} \quad l = 1, \cdots, L, \quad (2)$$

which gives the indicator of at least one multicast request for data source l arrives at slot n. And, under the aforementioned Poisson assumptions on the multicast requests, $X^{m,[l]}(n)$ is a simple Bernoulli random variable with parameter

$$A^{m,[l]} = 1 - e^{-p^{m,[l]} \lambda^m}, \quad l \in \{1, \cdots, L\}. \quad (3)$$

Finally, we denote the total number of distinct multicast data requests arriving in slot n as $S^m(n)$, defined as

$$S^m(n) \triangleq \sum_{l=1}^{L} X^{m,[l]}(n). \quad (4)$$

In the remainder of the paper, we focus on the *diversity gain* d of the system, defined generically as

$$d \triangleq \lim_{C \to \infty} -\frac{\log P(\mathcal{O})}{C}, \quad (5)$$

which extracts the exponential decay rate of the outage probability event, and hence is a natural measure of system performance. In the next sections, we shall study this metric under different scenarios when the total traffic load λ scales proportionally with the capacity C of the system as $\lambda = \gamma C$, for some positive real γ.

III. MULTICAST TRAFFIC WITH SYMMETRIC DEMANDS

In this section, we restrict our attention to the system performance in the presence of only multicast traffic where the number of data sources in the network scales linearly with capacity as $L = \theta C$, for some $\theta > 1$. In all cases, we study the performance of the asymptotically optimal well-known Earliest Deadline First (EDF) policy (e.g. see [5]) under the symmetric demand rates, i.e., when $p^{m,[l]} = 1/L$ for all l, to facilitate the asymptotic analysis. This analysis reveals the improvement resulting from the multicast nature of the underlying traffic. We further omit the superscript m introduced in Section II to signify multicast traffic, as it is the only traffic present in the system. For easy reference, the
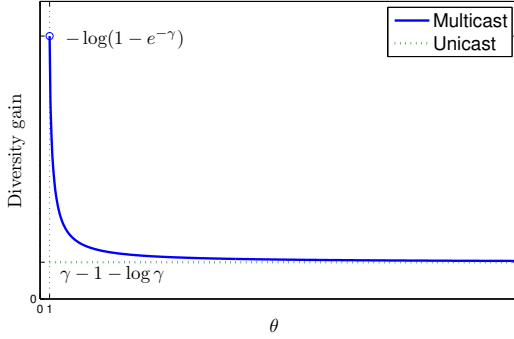
Fig. 1: Diversity gain of the non-predictive multicast network monotonically decreases with θ. However, it is lower bounded by the diversity gain of non-predictive unicast networks.

results on non-predictive and predictive unicast diversity gains are respectively repeated here [1], [2]:

$$d_N(\gamma) = \gamma - 1 - \log\gamma, \quad 0 < \gamma < 1. \tag{6}$$

$$d_P(\gamma) \geq (1+T)(\gamma - 1 - \log\gamma), \quad 0 < \gamma < 1. \tag{7}$$

A. Non-Predictive Multicasting

For the non-predictive network, it is straightforward to see that the outage probability is given by $P_N(\mathcal{O}_m) = P(S(n) > C)$. The following result characterizes the exponential decay rate of this probability.

Theorem 1: The diversity gain of non-predictive multicasting, denoted by $d_N(\gamma, \theta)$, is given by

$$d_N(\gamma, \theta) = (\theta - 1)\log(\theta - 1) - \theta\log\theta + \gamma\left(\frac{\theta - 1}{\theta}\right)$$
$$- \log\left(1 - e^{-\frac{\gamma}{\theta}}\right), \quad 0 < \gamma < 1, \quad \theta > 1. \tag{8}$$

Proof: (Sketch) We use Cramer's theorem [6] to determine $d_N(\gamma, \theta)$ as

$$\lim_{C \to \infty} \frac{-1}{\theta C} \log P\left(\frac{S(n)}{\theta C} > \frac{1}{\theta}\right) = \inf_{r > 0} \Lambda_{X^{[l]}}(r) - \frac{r}{\theta}, \tag{9}$$

where $\Lambda_{X^{[l]}}(r) = \log\left(Ae^r + (1 - A)\right)$, $l = 1, \cdots, L$. Then $r^* = \log\left(\frac{1-A}{A(\theta-1)}\right)$ and the conditions $0 < \gamma < 1$, $\theta > 1$ ensure that $r^* > 0$ holds. We refer the interested reader to [8] for the detailed proof. ∎

Fig. 1 depicts the diversity gains of non-predictive multicast and unicast networks (cf. [2]). It is obvious from the figure that $d_N(\gamma, \theta)$ is monotonically decreasing in θ. As θ increases, the number of data sources in the network grows faster with C, and hence, from (8),

$$\lim_{\theta \to \infty} d_N(\gamma, \theta) = \gamma - \log\gamma - 1 = d_N(\gamma). \tag{10}$$

That is, multicast diversity gain $d_N(\gamma, \theta)$ is strictly greater than its unicast counterpart $d_N(\gamma)$, and converges to it in the limit as $\theta \to \infty$. In fact, a much stronger result is obtained in [8] where the outage event of such multicast network converges in distribution to its unicast counterpart as $\theta \to \infty$.

B. Predictive Multicasting

Now we assume that the multicast network is predictive with fixed look-ahead interval $T > 0$. Hence, each request will have a deadline T slots after its arrival. We also assume that the network schedules requests according to EDF scheduling policy. The predictive network in this case has outage probability denoted by $P_P(\mathcal{O}_m)$ and diversity gain $d_P(\gamma, \theta)$. Here we want to emphasize an important fact that such a network has a **zero** outage probability as long as $L \leq C(T+1)$. This holds because the network can send data from C sources per slot and the arriving requests at any slot can remain in the system for T slots. Hence, being predictive multiplies the network capacity by a factor of at least $T+1$. As a result, in the sequel we will consider the case where $L > C(T+1)$.

Theorem 2: The optimal prediction diversity gain of a multicast predictive network satisfies

$$d_P(\gamma, \theta) \geq (T+1)d_N(\gamma, \theta), \quad 0 < \gamma < 1, \quad \theta > T+1. \tag{11}$$

Proof: (Sketch) Since the proof of Lemma 1 in [2] does not depend on the distribution of the arrival process $Q(n)$, $n > 0$. Then, replacing $Q(n)$ of $\mathcal{U}_d$ in [2] with $S(n)$, we will have $\mathcal{U}'_d$ such that $P_P(\mathcal{O}_m) \leq P(\mathcal{U}'_d)$. The proof of the theorem then follows straightforward as the proof of Theorem 2 in [2], with the important formulas found in the proof of Theorem 1 of this paper. ∎

The predictive network offers an improved diversity gain by a factor of at least $T+1$ in addition to multiplying the number of data sources for which the outage probability is zero by $T+1$. In fact, from (7) and (8), the unicast predictive network has a larger diversity gain than the multicast non-predictive network if

$$T > \frac{d_N(\gamma, \theta)}{d_N(\gamma)} - 1. \tag{12}$$

Moreover, if

$$T > \frac{-\log\left(1 - e^{-\gamma}\right)}{\gamma - 1 - \log\gamma}, \tag{13}$$

then $d_P(\gamma) > d_N(\gamma, \theta) \quad \forall \theta > 1$.

IV. DEMAND ALIGNMENT

Multicast alignment refers to the fact that the predictive multicast network does not serve all multicast requests targeting a certain data source upon their arrival, but since those requests have a deadline after $T+1$ slots, the network will wait for the next T slots to gather more requests to the same data source and then serve them all by a single unit of resources. This concept was not manifested in the previous section because the data sources were all multicast and having the same input traffic. So, to demonstrate the potential multicast alignment gain, we consider here another scenario where the wireless network has the following multicast and unicast input traffic flows.

The multicast flow is $Q^m(n), n > 0$ which is assumed to be Poisson distributed with rate $\lambda^m = \gamma^m C, \gamma^m > 0$ and is

---

This is an implication of the alignment property that a predictive multicast network has.

supposed to target a fixed multicast data source, i.e., $L = 1$. To serve the requests to this data source at any time slot, a μC amount of resources is consumed where $0 < \mu < 1$. The other traffic flow is $Q^u(n), n > 0$, which is also Poisson distributed with rate $\lambda^u = \gamma^u C$, where $0 < \gamma^u < 1 - \mu$, but this flow is assumed to target unicast sources, and hence, each request of this flow consumes one unit of the network capacity.

We also assume that the unicast flow is predictable by a prediction window of T slots where $T > 0$, and that the network schedules its requests in a way such that there are no outages occurring on the multicast traffic, i.e., $P(\mathcal{O}_m) = \phi$. This is motivated by the fact that the capacity required to serve the multicast requests is deterministic, smaller than C, and is independent of the number of such requests, which is expected to be much larger than the number of unicast requests as γ^m unbounded above, theoretically. Therefore, the cost of a multicast outage would be severe, if occurred. Conversely, for the unicast traffic, the required capacity to serve its requests is random and depends on the number of arriving requests per slot and hence it is natural to have unicast outages. So, an outage event for such network turns out to be

$$\mathcal{O} = \mathcal{O}_u = \{n > 0 : N_0^u(n) > C(n)\}. \tag{14}$$

Here, we emphasize the dependency of the unicast capacity on the time slots in (14) since $C(n) = C$ if the network does not serve the multicast traffic at slot n and $C(n) = (1 - \mu)C$ if the network serves the multicast traffic at slot n. The activity of the multicast data source can be captured through the process $X^m(n), n > 0$, where $X^m(n) = 1$ if $Q^m(n) > 0$ and $X^m(n) = 0$ if $Q^m(n) = 0$.

A. Non-predictable Multicast Traffic

In this case, the network follows the following service policy at the beginning of time slot $n > 0$: the network serves the multicast requests by amount of resources equals to $\mu C X^m(n)$, and serves the unicast according to FCFS by a capacity $C(n) = C - \mu C X^m(n)$.

Theorem 3: The diversity gain of the given network with non-predictable multicast traffic is denoted by $d_N(\gamma^u, \mu)$ and satisfies

$$d_N(\gamma^u, \mu) \leq (T + 1)\left(\gamma^u - 2(1 - \mu)\right. \\ \left. - 2(1 - \mu)\log\left(\frac{\gamma^u}{2(1 - \mu)}\right)\right). \tag{15}$$

Proof: (Sketch) We use the asymptotic inequality

$$P(\mathcal{O}_u) \geq P\left(\sum_{i=0}^{T} Q^u(n - i - T) > 2(1 - \mu)(T + 1)C\right)$$

to construct the upper bound on $d_N(\gamma^u, \mu)$ using Cramer's theorem. The proof of why this inequality holds is omitted for space limitations and the detailed proof can be found in [8]. The condition $0 < \gamma^u < 1 - \mu$ ensures that $d_N(\gamma^u, \mu)$ is positive. ∎

B. Predictable Multicast Traffic

In this case, the multicast traffic is also predictable with the same window of T slots exactly as the unicast traffic. Based on this, the network can align the arriving multicast requests by employing the following service policy at the beginning of time slot n: if there is a multicast request with deadline of n, the network serves all multicast requests with μC resources, and then serves the unicast requests with a capacity of $C(n) = (1 - \mu)C$. Otherwise, the network serves the unicast requests first from the available resources C, and then if the remaining resources are greater than μC, it serves the existing multicast requests, if any.

Under the above service policy, we conclude the following theorem on the diversity gain.

Theorem 4: The diversity gain of the given network with predictable multicast traffic is denoted by $d_P(\gamma^u, \mu)$ and satisfies

$$d_P(\gamma^u, \mu) \geq (T + 1)\left[\gamma^u - 1 + \frac{\mu}{T + 1}\right. \\ \left. + \left(1 - \frac{\mu}{T + 1}\right)\log\left(\frac{1}{\gamma^u}\left(1 - \frac{\mu}{T + 1}\right)\right)\right]. \tag{16}$$

Proof: (Sketch) We use the asymptotic inequality

$$P_N(\mathcal{O}_u) \leq P\left(\sum_{i=0}^{T} Q^u(n - i - T) > (T + 1)C - \mu C\right)$$

and Cramer's theorem to determine the lower bound on $d_P(\gamma^u, \mu)$. The complete proof is omitted for space limitations and can be found in [8]. ∎

Note that, the right hand side of (16) requires $\gamma^u < 1 - \frac{\mu}{T+1}$ to be positive, which is indeed greater than $1 - \mu$, and hence it enhances the maximum unicast traffic rate that can be sustained by the network while attaining positive diversity gain.

Now, having obtained upper and lower bounds on $d_N(\gamma^u, \mu)$ and $d_P(\gamma^u, \mu)$ respectively, we compare the two bounds to show the potential alignment gain offered by the predictability of the multicast traffic. Let the right hand side of (15) be denoted by U_N and the right hand side of (16) be denoted by L_P, then $L_P - U_N > 0$ if

$$\mu < \frac{(T + 1)\log\left(\frac{T - \mu + 1}{\gamma^u T + \gamma^u}\right) + T + \log\left(\frac{2 - 2\mu}{\gamma^u}\right)(-2T - 2) + 1}{\log\left(\frac{T - \mu + 1}{\gamma^u T + \gamma^u}\right) + 2T + \log\left(\frac{2 - 2\mu}{\gamma^u}\right)(-2T - 2) + 1}. \tag{17}$$

The above condition is sufficient to ensure that $L_P > U_N$, which implies the superiority of the multicast alignment over the non-predictable multicast traffic. The expression of the condition, however, is quite complicated so that a closed form range of μ satisfying (17) is not obtainable. Nevertheless, we will show by a numerical example in Section V that there exists μ and T satisfying (17) for all $0 < \gamma^u < \min\{2(1 - \mu), 1 - \frac{\mu}{T+1}\}$.
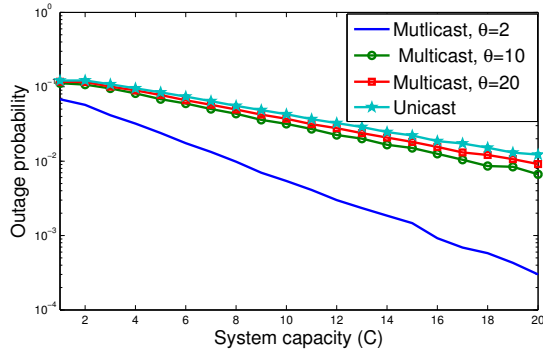
Fig. 2: Outage probability vs. C for non-predictive networks. The outage probability of multicast approaches the unicast as θ increases. In this simulation, $\gamma = 0.6$.
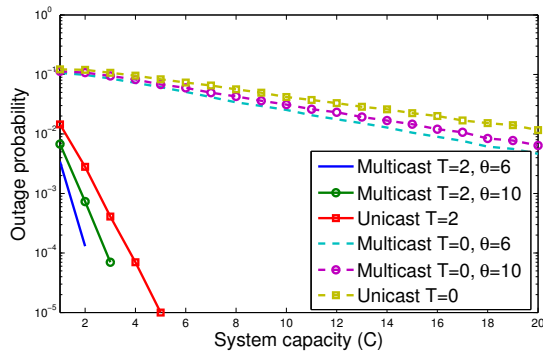

Fig. 3: Outage probability vs. C. Predictive network with multicast has better performance than non-predictive, and unicast networks. In this simulation, $\gamma = 0.6$.

V. NUMERICAL RESULTS

Fig. 2 depicts the outage probability of the multicast and unicast non-predictive networks versus C under the symmetric traffic scenario. For the multicast network, different values of θ are considered and the gain compared to the unicast scenario is exhibited. Eventually, as $\theta \to \infty$, the outage probability curves of both networks coincide.

The outage probability of the predictive multicast and unicast networks of the symmetric input traffic is compared numerically to that of non-predictive network and is plotted in Fig. 3. The figure shows the significant enhancement to the outage probability of the multicast network when prediction is employed. Moreover, we can see that the outage probability of the unicast predictive network is better than that of the multicast non-predictive network.

Finally, in Fig. 4 we show by an example that the alignment of the predictable multicast traffic can strictly improve the diversity gain over the case of non-predictable multicast traffic; i.e., the lower bound L_P is strictly greater than the upper bound U_N for the whole range of possible γ^u.
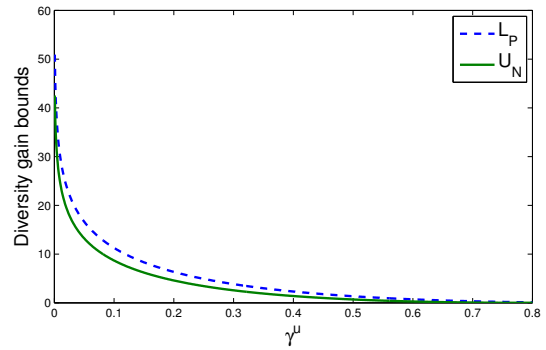

Fig. 4: L_P and L_N versus γ^u given $\mu = 0.6$ and $T = 8$. The alignment of multicast requests results in a diversity gain that is strictly larger than the diversity gain when the network can not predict the multicast traffic.

VI. CONCLUSION

We have extended our novel technique of proactive resource allocation to wireless multicast networks. We have investigated the potential performance improvement and shown that the multicast network can achieve infinite diversity gain as long as the number of data sources in the network is not greater than $T + 1$ times the network capacity, where T is the prediction window of the network. Moreover, if the number of data sources exceeds that limit, the multicast nature of the traffic results in an improved diversity gain as compared to unicast networks. Our analysis also reveals the phenomenon of multicast traffic alignment where the predictive network can collect and align the multicast traffic in a sequence of $T + 1$ slots to serve them altogether using the same spectral resources. Finally, our theoretical claims were validated by numerical results derived in certain representative scenarios.

REFERENCES

[1] H. El-Gamal, J. Tadrous and A. Eryilmaz, "Proactive resource allocation: turning predictable behavior into spectral gain", *Forty-Eighth Annual Allerton Conference on Communication, Control, and Computing* , to appear.

[2] H. El-Gamal, J. Tadrous and A. Eryilmaz, "Proactive resource allocation and scheduling", to be submitted to *IEEE Transactions on Information Theory*.

[3] J. Lee, and N. Jindal, "Asymptotically optimal policies for hard-deadline scheduling over fading channels", Submitted to *IEEE Transactions on Information Theory*, June 2009.

[4] S. Kittipiyakul, P. Elia, and T. Javidi, "High-SNR analysis of outage-limited communications with bursty and delay-limited information", *IEEE Transactions on Information Theory*, vol.55, no.2, pp.746-763, Feb. 2009.

[5] S. S. Panwar, D. Towsley and J. K. Wolf, "Optimal scheduling policies for a class of queues with customer deadlines to the beginning of service", *ACM SIGMETRICS Performance Evaluation Review*, volume 18, issue 3, Nov. 1990.

[6] A. Ganesh, N. O'Connell and D. Wischik, "Big queues", Lecture Notes in Mathematics, volume 1838, 2004.

[7] R. G. Gallager, "Discrete stochastic processes", Kluwer, Boston, 1996.

[8] J. Tadrous, A. Eryilmaz, and H. El-Gamal, "On proactive resource allocation for predictive unicast and multicast traffic," *Technical Report*, Feb. 2011.
http://www2.ece.ohio-state.edu/~eryilmaz/ISIT11_TechReport.pdf