Layered Reception for Heterogeneous Traffics From Mobile Cloud Applications

Jinho Choi
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST), Korea
Email: jchoi0114@gist.ac.kr

***Abstract*—In this paper, we consider a new architecture for mobile systems to effectively accommodate mobile cloud computing traffics in conjunction with conventional delay-limited traffics (e.g., voice traffics). It is assumed that small base stations (BSs) do not decode cloud data streams from a user, but forward the received signals from the users to a cloud server (CS), while small BSs are to decode conventional delay-limited data streams. The resulting approach is called layered reception for uplink transmissions as decoding takes place in different layers. For a dense deployment of small BSs, we can exploit a macro-diversity gain. For the power allocation, we derive a closed-form expression for the outage probability of delay-limited traffics over a block-fading channel. The impact of the macro-diversity gain on the performances has also be shown by the outage probability analysis.**

***Index Terms*—uplink transmissions, layered reception, mobile cloud**

I. INTRODUCTION AND AN ARCHITECTURE FOR MOBILE CLOUD COMPUTING

Mobile phones (i.e., smart phones) are devices not only for voice and text communications, but also for advanced computing [1]. Due to the advance of mobile phones, mobile cloud computing will play a crucial role in 5th generation (5G) systems for better resource utilization in terms of computing and energy efficiency [2], [3]. Thus, it is necessary to consider different architectures for 5G radio access networks to include cloud servers (or centers) [4].

In mobile cloud computing and other mobile cloud services, there might be data traffics that are different from conventional data traffics such as voice traffics. A base stations (BS) might be used as a gateway for those traffics for mobile cloud applications. In this case, the BS would decode the received signals from a user regardless of types of traffics. Once the signals are decoded, they might be divided to re-send to different networks. For example, voice packets are to be sent to mobile switching centers (MSCs) and exchanged through their network, while cloud data streams are to be forwarded to Internet where cloud servers (CSs) are connected.

In this paper, we consider a different architecture for radio access network from the conventional one, which might be suitable for mobile cloud computing in 5G. We assume that a mobile terminal or user has heterogeneous traffics. In particular, there are two different types of traffics: *primary* and *cloud data streams*. The primary data stream is the conventional one that is to deliver voice packets with low latency, but possibly less reliability. On the other hand, the cloud data stream is to upload codes for cloud computing and/or data files such as documents, images, and video clips that are to be stored in a cloud storage system [5], [6]. Furthermore, we assume that there are multiple small BSs that can receive the signal from a user, which only decode primary data streams and forward them to MSCs. Note that in 5G, it is expected to have a dense deployment of small BSs [4], [7], [8]. Thus, the signal from a user would be received by multiple nearby small BSs. Small-BSs do not decode cloud data streams, but forward their received signals to cloud-servers (CSs) that decode cloud streams. Throughout the paper, we assume that the connection between small BSs and cloud-BS is reliable and has a sufficiently wide-bandwidth. For example, this connection can be established by optical cables. In Fig. 1, an illustration of the system is shown. The signals from the user are received by multiple small BSs and those small BSs are connected to a CS.

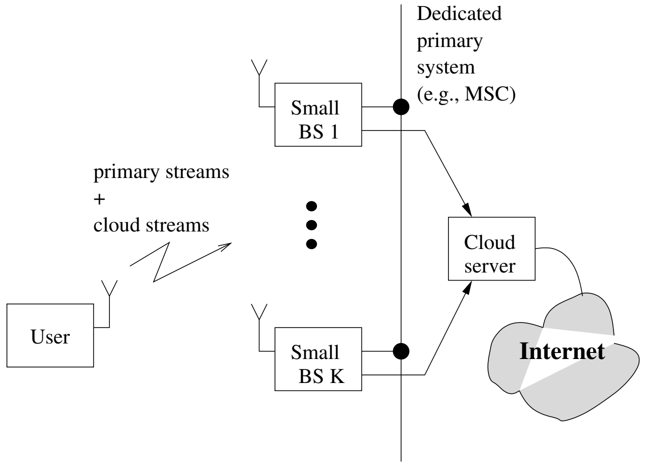

Fig. 1. An illustration of the radio access network with mobile cloud computing.

The main focus of the paper is on layered reception for the novel architecture of radio access network in Fig. 1 and the determination of power and rate for primary and cloud data streams at a user.

Notation: Matrices and vectors are denoted by upper- and lower-case boldface letters, respectively. The superscripts T and H denote the transpose and complex conjugate, respectively. The 2-norm of $\mathbf{a}$ is denoted by $||\mathbf{a}||$. $\mathbb{E}[\cdot]$ denotes the statistical expectation. $\mathcal{CN}(\mathbf{a}, \mathbf{R})$ represents the distribution of circularly symmetric complex Gaussian (CSCG) random vectors with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{R}$.

II. LAYERED RECEPTION FOR SUPERPOSITION CODED SIGNALS

In this section, we present the main idea of layered reception and consider superposition coding to allow a user to transmit primary and cloud data streams simultaneously. According to layered reception, primary data streams are to be decoded at small BSs and cloud data streams are decoded at a CS.

A. Layered Reception

Suppose that X denotes the signal from the user. In addition, we assume that X consists of X_1 and X_2 that represent the primary and cloud data symbols, respectively, i.e.,

$$X = (X_1, X_2). \tag{1}$$

As shown in Fig. 1, we assume that there are K small BSs that can receive the signal from the user. Let the received signal at the kth small BS be

$$Y_k = h_k X + N_k, \tag{2}$$

where h_k represents the channel coefficient from the user to the kth small BS and $N_k \sim \mathcal{CN}(0, 1)$ denotes the background noise at the kth small BS. For convenience, the variance of the background noise has been normalized. Furthermore, we assume that

$$\mathbb{E}[X] = 0 \text{ and } \mathbb{E}[|X|^2] = P,$$

where P is the total signal power.

It might be expected that the K small BSs cooperate to decode X_1. However, we do not assume any cooperative decoding for X_1 to avoid computational overhead between small BSs. Rather, we consider the macro-diversity where at least one small BS is to decode X_1 (a better performance is expected by increasing K or a denser deployment of small BSs). Thus, K can be seen as the macro-diversity gain. Furthermore, all the K small BSs are not to decode X_2. Each small BS forwards its received signal to the CS and the CS has $(Y_1, \ldots, Y_K)$. Overall, we can see that there are no information exchanges between small BSs, while the received signals are forwarded to the CS.

In order to save the computational complexity at the CS, we assume that the small BS that succeeds to decode X_1 also forwards X_1 to the CS. From this, the CS does not need to decode X_1. For convenience, the resulting system will be referred to as the layered reception system for cloud applications (LRS-CA) in this paper as decoding takes place in different layers (decoding of primary data symbols takes place in the layer of small BSs, while decoding of cloud data symbols takes place in the layer of CS).

B. Superposition Coding and Rates

Since X_1 is to be decoded by at least one small BS, the achievable rate for X_1, denoted by R_1, can be given by

$$R_1 = \max_k \mathsf{I}(Y_k; X_1), \tag{3}$$

where $\mathsf{I}(X; Y)$ represents the mutual information between X and Y. All the K small BSs forward their received signals to the CS and X_1 is also available at the CS, the achievable rate for X_2, denoted by R_2, becomes

$$R_2 = \mathsf{I}(\mathbf{y}; X_2 \,|\, X_1), \tag{4}$$

where $\mathbf{y} = [Y_1 \ \ldots \ Y_K]$.

For these achievable rates, we consider superposition coding in this paper. Each signal is independently encoded and modulated. Thus, X is given by

$$X = X_1 + X_2.$$

Let $P_m = \mathbb{E}[|X_m|^2]$, $m = 1, 2$. Clearly, we have $P = P_1 + P_2$. Furthermore, we assume that Gaussian codebooks are used for X_m in order to achieve the channel capacity, i.e., $X_m \sim \mathcal{CN}(0, P_m)$, $m = 1, 2$. Then, it can be shown that

$$\mathsf{I}(Y_k; X_1) = \log_2 \left(1 + \frac{\alpha_k P_1}{\alpha_k P_2 + 1}\right), \tag{5}$$

where $\alpha_k = |h_k|^2$.

As mentioned earlier, X_1 is available at the CS. Since it can be subtracted from Y_k, we have

$$\begin{aligned} \mathsf{I}(\mathbf{y}; X_2 \,|\, X_1) &= \mathsf{I}(\tilde{\mathbf{y}}; X_2 \,|\, X_1) \\ &= \log_2 (1 + U P_2), \end{aligned} \tag{6}$$

where $U = \sum_{k=1}^K \alpha_k$, $\tilde{\mathbf{y}} = [\tilde{Y}_1 \ \ldots \ \tilde{Y}_K]$, and $\tilde{Y}_k = Y_k - h_k X_1$.

III. DETERMINATION OF POWER AND RATE

In this section, we consider the power allocation to X_1 and X_2 under a block-fading channel environment. In addition, we derive the outage probability of X_1 and the achievable rate of X_2 that allow us to see the impact of K (or the macro-diversity gain) on the performances.

Suppose that the length of codewords for the primary data stream is not long enough and it has to be decoded without re-transmissions for a latency constraint. In this case, the maximum latency is the same as the duration of a codeword. With a short length of codewords, we can assume block-fading channels, where the channel coefficients remain unchanged over the transmission of a codeword, and define the outage probability as follows:

$$\begin{aligned} P_{\mathrm{out}}(B_1) &= \Pr\left(\max_k \mathsf{I}(Y_k; X_1) \le B_1\right) \\ &= \Pr\left(\beta \le \frac{1 - 2^{-B_1}}{P_1 - (1 - 2^{B_1})P}\right), \end{aligned} \tag{7}$$

where B_1 becomes the ϵ-outage capacity [9] and $\beta = \max\{\alpha_1, \ldots, \alpha_K\}$.

According to (7), P_1 has to be bounded as

$$P(1 - 2^{-B_1}) < P_1 \le P \tag{8}$$

in order to have a reasonably low outage probability. Furthermore, $P_{\rm out}(B_1)$ decreases with P_1. Let

$$\nu(B_1, P_1) = \frac{1 - 2^{-B_1}}{P_1 - (1 - 2^{-B_1})P}.$$

Then, we have

$$\begin{aligned} P_{\rm out}(B_1) &= \Pr\left(\beta \le \nu(B_1, P_1)\right) \\ &= \left(F_\alpha(\nu(B_1, P_1))\right)^K, \end{aligned} \tag{9}$$

where $F_\alpha(x) = \Pr(\alpha_k \le x)$ is the cdf of α_k.

For further analysis, we consider the following assumption.

A1) The α_k's are iid and their distribution is

$$f_\alpha(\alpha_k) = \frac{1}{\bar\alpha}\exp\left(-\frac{\alpha_k}{\bar\alpha}\right), \ \alpha_k \ge 0,$$

where $\bar\alpha = \mathbb{E}[\alpha_k] = \mathbb{E}[|h_k|^2]$.

That is, the channels from the user to K small BSs are assumed to be independent Rayleigh fading channels.

Under **A1)**, since $F_\alpha(x) = 1 - \exp\left(-\frac{x}{\bar\alpha}\right)$, we have

$$\begin{aligned} P_{\rm out}(B_1) &= \left(1 - e^{-\frac{\nu(B_1,P_1)}{\bar\alpha}}\right)^K \\ &\le \left(\frac{\nu(B_1, P_1)}{\bar\alpha}\right)^K, \end{aligned} \tag{10}$$

where the upper-bound is tight when $\frac{\nu(B_1,P_1)}{\bar\alpha} \ll 1$. It can be seen that the outage probability can quickly decrease with K if $\frac{\nu(B_1,P_1)}{\bar\alpha} < 1$.

With the closed-form expression for the outage probability in (10), it is straightforward to decide the value of P_1 if the target outage probability, denoted by ϵ, is given. That is,

$$P_{\rm out}(B_1) = \epsilon.$$

After some manipulations, for given B_1 and ϵ, we can determine the power of X_1, P_1, to meet the target outage probability ϵ as follows:

$$P_1(B_1, \epsilon) = (1 - 2^{-B_1})\left(P + \frac{1}{\psi(\epsilon)}\right), \tag{11}$$

where

$$\psi(\epsilon) = -\bar\alpha \ln(1 - \epsilon^{1/K}).$$

For given B_1 and ϵ, if $P_1(\epsilon) > P$, then there is no feasible solution. Thus, a large P is required. In particular, since $P_1(\epsilon) \le P$, we need to have

$$P \ge P_{\min} = \frac{2^{B_1} - 1}{\psi(\epsilon)} \tag{12}$$

for the feasibility condition under **A1)**, where $P_{\min}$ denotes the minimum required power to support at least primary data stream with an outage probability of ϵ.

For $\epsilon \ll 1$ and not too large K (i.e., $\epsilon^{1/K} \ll 1$), we have the following approximation:

$$\psi(\epsilon) \approx \bar\alpha \epsilon^{1/K}.$$

Then, we can show that the minimum required power decreases quickly with K.

Suppose that cloud data streams do not have any stringent latency constraints. Thus, we may need to consider the average achievable rate. Note that the outage event in decoding X_1 affects on the the average achievable rate of X_2. If an outage event happens, X_1 is not available and it becomes the interference at the CS. Consequently, the average achievable rate of X_2 becomes

$$\begin{aligned} \Omega_2(P_1, P_2) &= \mathbb{E}[\mathsf{I}(\mathbf{y}; X_2)\mathbb{1}(\beta \le \nu) \\ &\quad + \mathsf{I}(\mathbf{y}; X_2 \,|\, X_1)(1 - \mathbb{1}(\beta \le \nu))] \\ &= \mathbb{E}[\mathsf{I}(\mathbf{y}; X_2 \,|\, X_1)] - \mathbb{E}[\Psi(U)\mathbb{1}(\beta \le \nu)], \end{aligned} \tag{13}$$

where $\nu = \nu(B_1, P_1)$, $\mathbb{1}(\cdot)$ is the indicator function, and

$$\Psi(U) = \mathsf{I}(\mathbf{y}; X_2 \,|\, X_1) - \mathsf{I}(\mathbf{y}; X_2).$$

Since U and β are correlated, we need to have the joint pdf of U and β to find $\mathbb{E}[\Psi(U)\mathbb{1}(\beta \le \nu)]$. However, for an approximation, we can have

$$\begin{aligned} \mathbb{E}[\Psi(U)\mathbb{1}(\beta \le \nu)] &\approx \mathbb{E}[\Psi(U)]\mathbb{E}[\mathbb{1}(\beta \le \nu)] \\ &= \mathbb{E}[\Psi(U)]\epsilon. \end{aligned}$$

Thus, we have

$$\Omega_2(P_1, P_2) \approx \mathbb{E}[\mathsf{I}(\mathbf{y}; X_2 \,|\, X_1)](1 - \epsilon) + \mathbb{E}[\mathsf{I}(\mathbf{y}; X_2)]\epsilon. \tag{14}$$

Since

$$\mathbb{E}[\mathsf{I}(\mathbf{y}; X_2 \,|\, X_1)] > \mathbb{E}[\mathsf{I}(\mathbf{y}; X_2)],$$

from (14), we can see that the successful decoding of X_1 by any small BS plays a crucial role in increasing the rate of cloud data streams in the proposed architecture of radio access network.

Note that the length of codewords has to be sufficiently long to achieve $\Omega_2(P_1, P_2)$. That is, the length of codewords for X_2 might be much longer than that for X_1 so that a codeword for X_2 can experience a number of channel realizations over block-fading channels.

Using the chain rule [10], we have

$$\begin{aligned} \mathbb{E}[\mathsf{I}(\mathbf{y}; X_2 | X_1)] &= \mathbb{E}[\mathsf{I}(\mathbf{y}; X_2, X_1)] - \mathbb{E}[\mathsf{I}(\mathbf{y}; X_1)] \\ &= \mathbb{E}\left[\log_2\left(1 + UP_2\right)\right], \end{aligned} \tag{15}$$

and

$$\mathbb{E}[\mathsf{I}(\mathbf{y}; X_2)] = \mathbb{E}\left[\log_2\left(1 + \frac{UP_2}{UP_1 + 1}\right)\right]. \tag{16}$$

Here, the signal-to-interference-plus-noise ratio (SINR) at the CS is $\frac{UP_2}{UP_1+1}$ in the absence of knowledge of X_1. If any one of small BSs can decode X_1, X_1 is forwarded to the CS and the CS can remove X_1 from $\mathbf{y}$. In this case, the signal-to-noise ratio (SNR), not SINR, becomes UP_2.

Under **A1)**, from [11], the approximation of $\Omega_2(P_1, P_2)$ in (14) becomes

$$\Omega_2(P_1, P_2) \approx \epsilon(W_K(P) - W_K(P_1)) + (1 - \epsilon)W_K(P_2), \tag{17}$$

where

$$W_K(x) = \frac{e^{\frac{2}{\bar{\alpha}x}}}{\ln 2} \sum_{q=0}^{K-1} E_{q+1}\left(\frac{2}{\bar{\alpha}x}\right). \tag{18}$$

Here, $E_q(x) = \int_1^\infty e^{-xy} y^{-q} dy$.

Consequently, if B_1 and ϵ are given for X_1, we can decide the power for X_1 as in (11). Provided that the total power, P, is sufficiently high to satisfy (12), the rest power, i.e., $P_2 = P - P_1(B_1, \epsilon)$, is allocated to X_2. In this case, the code rate for X_2 can be decided according to (17), i.e.,

$$B_2 \le \Omega_2(P_1, P_2),$$

where B_2 is the rate of X_2.

IV. SIMULATION RESULTS

In this section, we present simulation results under **A1)** (i.e., independent Rayleigh fading channels are assumed for simulations) with $\bar{\alpha} = 1$.

Fig. 2 shows the outage probability of X_1 and the achievable rate of X_2 for various values of B_1 when $P = 6$ dB, $K = 4$, and $\epsilon = 10^{-2}$. We can see that if B_1 is too high, it is not possible to satisfy the target outage probability, $\epsilon = 10^{-2}$. In addition, as B_1 decreases, the achievable rate of X_2 can increase as more power can be assigned to X_2 (i.e., P_2 can increase). Note that the approximation of the achievable rate in (14) is shown as the solid line in the right-hand-side (RHS) figure in Fig. 2.

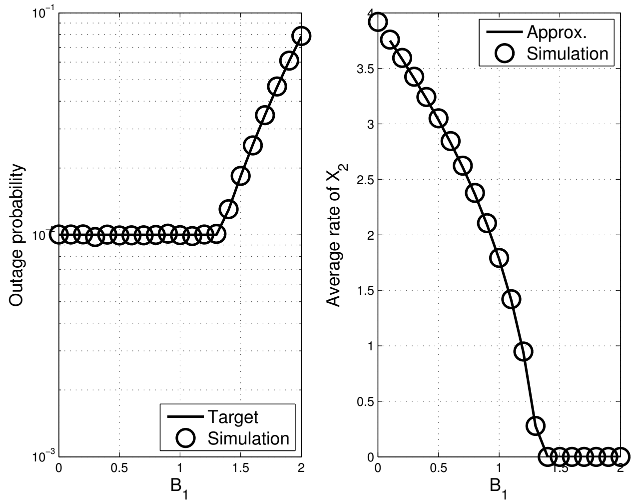

Fig. 2. The outage probability of X_1 and the achievable rate of X_2 for various values of B_1 when $P = 6$ dB, $K = 4$, and $\epsilon = 10^{-2}$.

Fig. 3 shows the outage probability of X_1 and the achievable rate of X_2 for different values of P when $B_1 = 1$, $K = 4$, and $\epsilon = 10^{-2}$. A large P is beneficial to X_2 as more power can be assigned to X_2 once the target outage probability of X_1 is met.

In order to see the impact of K on the performances, we consider different values of K when $P = 6$ dB, $B_1 = 1$, and $\epsilon = 10^{-2}$ and show the outage probability of X_1 and the achievable rate of X_2 in Fig. 4. We can see that a denser deployment of small cells or the increase of K results in better performances for X_1 and X_2 in terms of the outage probability and the achievable rate, respectively. The gain is due to the macro-diversity from more small BSs that can receive the signals from a user. This gain can also improve the performance for mobile cloud computing as the achievable rate of X_2 increases with K.

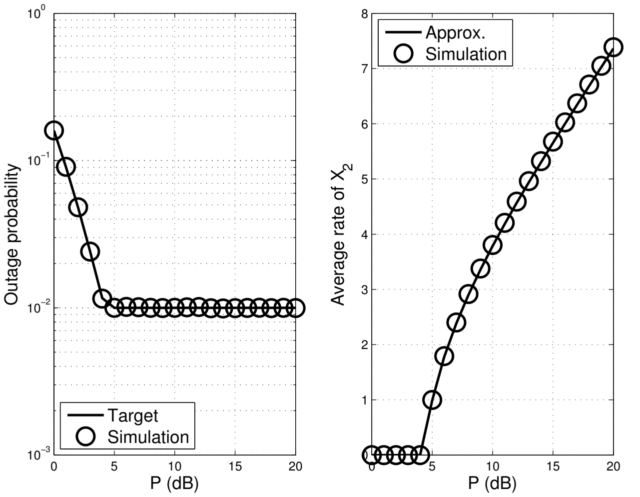

Fig. 3. The outage probability of X_1 and the achievable rate of X_2 for different values of P when $B_1 = 1$, $K = 4$, and $\epsilon = 10^{-2}$.

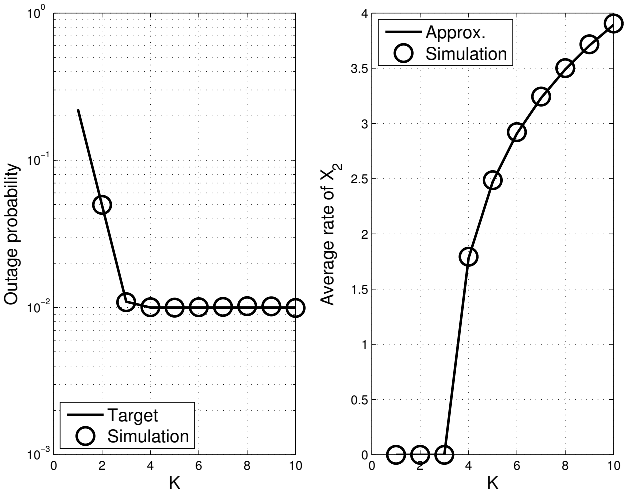

Fig. 4. The outage probability of X_1 and the achievable rate of X_2 for various values of K when $P = 6$ dB, $B_1 = 1$, and $\epsilon = 10^{-2}$.

The target outage probability of X_1, ϵ, can vary depending on the required quality of service (QoS) for X_1. To see this impact of ϵ on the performances, we consider different values of ϵ when $P = 10$ dB, $B_1 = 1$, and $K = 4$ and show the outage probability of X_1 and the achievable rate of X_2 in Fig. 5. We can see that with a sufficiently high total power, P, the constraint of the outage probability of X_1 can be satisfied. It is interesting to see that the achievable rate of X_2 increases with ϵ and then decreases. In other words, it is not necessarily true that the decrease of ϵ results in the increase of the achievable rate of X_2. If ϵ is too high, there might be more outage events that result in unknown interference in $\mathbf{y}$ at the CS. This results in the decrease of the achievable rate of X_2.

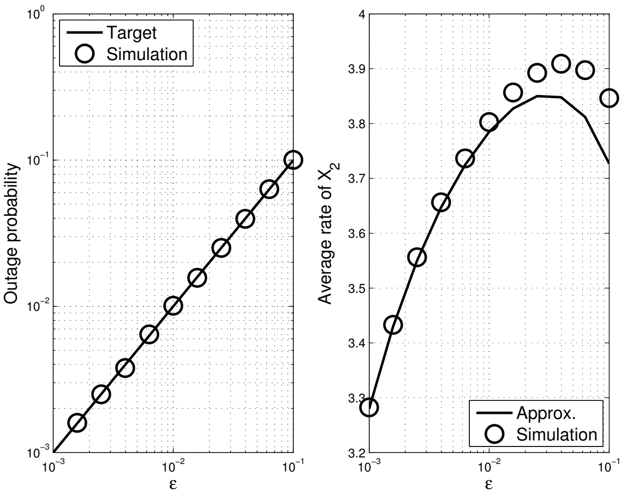

Fig. 5. The outage probability of X_1 and the achievable rate of X_2 for various values of ϵ when $P = 10$ dB, $B_1 = 1$, and $K = 4$.

It is noteworthy that the approximation of $\Omega(P_1, P_2)$ in (14) in Figs. 2, 3, and 4 is reasonably good. On the other hand, in Fig. 5, we can see that the approximation of $\Omega(P_1, P_2)$ in (14) becomes a lower-bound when ϵ is high.

V. CONCLUSIONS

To efficiently accommodate mobile cloud computing applications, we studied a new architecture for mobile systems with a dense deployment of small BSs and considered layered reception where decoding of conventional delay-limited traffics such as voice packets is performed at small BSs and decoding of mobile cloud traffics takes place at a CS. We derived a closed-form expression for the outage probability of primary data symbol (for conventional traffics), which had been used for the power allocation to satisfy a constraint of outage probability. Based on the proposed architecture and analysis, we can see that a dense deployment of small BSs can help improve the performance of mobile cloud computing.

REFERENCES

[1] S. Abolfazli, Z. Sanaei, E. Ahmed, A. Gani, and R. Buyya, "Cloud-based augmentation for mobile devices: Motivation, taxonomies, and open challenges," *Communications Surveys Tutorials, IEEE*, vol. 16, pp. 337–368, First 2014.

[2] K. Kumar and Y.-H. Lu, "Cloud computing for mobile users: Can offloading computation save energy?," *Computer*, vol. 43, pp. 51–56, April 2010.

[3] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5g heterogeneous networks," *Signal Processing Magazine, IEEE*, vol. 31, pp. 45–55, Nov 2014.

[4] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *Network, IEEE*, vol. 29, pp. 6–14, March 2015.

[5] P. Bahl, R. Y. Han, L. E. Li, and M. Satyanarayanan, "Advancing the state of mobile cloud computing," in *Proceedings of the Third ACM Workshop on Mobile Cloud Computing and Services*, MCS '12, pp. 21–28, ACM, 2012.

[6] M. Barbera, S. Kosta, A. Mei, and J. Stefa, "To offload or not to offload? the bandwidth and energy costs of mobile cloud computing," in *INFOCOM, 2013 Proceedings IEEE*, pp. 1285–1293, April 2013.

[7] J. Xu, J. Wang, Y. Zhu, Y. Yang, X. Zheng, S. Wang, L. Liu, K. Horneman, and Y. Teng, "Cooperative distributed optimization for the hyper-dense small cell deployment," *Communications Magazine, IEEE*, vol. 52, pp. 61–67, May 2014.

[8] E. Hossain and M. Hasan, "5G cellular: key enabling technologies and research challenges," *Instrumentation Measurement Magazine, IEEE*, vol. 18, pp. 11–21, June 2015.

[9] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. NJ: John Wiley, second ed., 2006.

[11] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2636–2647, Oct. 2003.