User Detection Performance Analysis for Grant-Free Uplink Transmission in Large-Scale Antenna Systems

Jonghyun Kim, Kyung Lin Ryu, and Kwang Soon Kim

School of Electrical and Electronic Engineering, Yonsei University

50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Korea.

Email: jonghyun.kim@yonsei.ac.kr, kyunglin.ryu@yonsei.ac.kr, ks.kim@yonsei.ac.kr

Abstract—In this paper, user detection performance of a grant-free uplink transmission in a large scale antenna system is analyzed, in which a general grant-free multiple access is considered as the system model and Zadoff-Chu sequence is used for the uplink pilot. The false alarm probabilities of various user detection schemes under the target detection probabilities are evaluated.

Index Terms—Grant-free uplink transmission, large-scale antenna system, Neyman-Pearson decision criterion, ultra-reliable low-latency communication.

I. Introduction

To enable the diverse future wireless services, the 5th generation (5G) mobile communication has been extensively discussed, in which the 5G services are categorized as enhanced mobile broadband (eMBB), massive machine type communications (mMTC), and ultra-reliable low-latency communications (URLLC) according to the target requirements for the services. Among them, the URLLC can support so called *mission-critical control* services, including automated driving, tele-surgery (e-health), video-driven interaction, augmented reality applications, and factory automations [1], which typically requires high reliability level of 99.999% and 1 ms end-to-end latency. Such services require various kinds of packets with different traffic characteristics, including packet size, arrival rate and model, etc. In this paper, we focus on *sporadic event-driven uplink transmission* in which traffics occur randomly but require high reliability, low latency, and medium to high throughput, which would be one of the most challenging tasks.

In the previous 4G communication (LTE-Advanced), the uplink random access is *grant-based* so that a 4-way hand-shaking must be completed before a data transmission from user equipments (UEs) to a base station (BS). The 4-way handshaking includes 1) random access preambles from UEs, 2) random access responses from a BS, 3) scheduling requests or buffer status reports from UEs, and then 4) uplink scheduling grants from a BS [2]. Through theses 4 steps, a BS recognizes the requests for uplink transmissions from UEs, identifies which UEs will transmit, and allocates resources to them to establish uplink connections. But this procedures cause severe delays at least 4 times of the transmission time interval (TTI) and decrease the spectral efficiency, especially as the latency requirement becomes tight. Also, the hybrid automatic
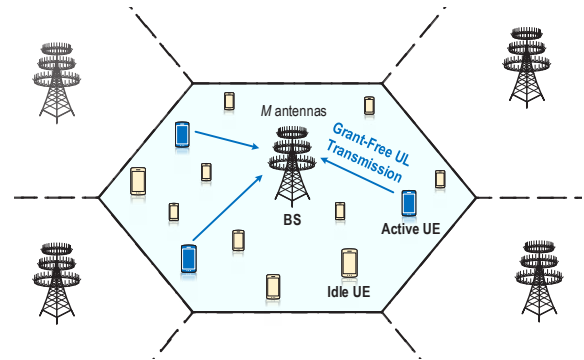

Fig. 1. A grant-free multiple access system model

repeat request (H-ARQ) employed for achieving reliability cannot work appropriately due to the low latency requirement. To overcome these limitations, a *grant-free* multiple access (GFMA) has been proposed and discussed [3]. For such a GFMA, one important task is to design a highly reliable user detection scheme [4] because a BS does not have any prior information about which UEs will transmit and also its performance should be reliable enough to meet the high reliability requirements of a URLLC.

Recently, a large-scale antenna system (LSAS), which employs massive antennas at a BS, has been in a great interest due to its inherit merit of providing massive spatial dimension [5]. To enhance the user detection performance in a GFMA, a large spatial multiplexing gain as well as a large diversity gain can be utilized. Moreover, as the number of antennas increases, due to the *channel hardening effect* [6], we can predict the achievable data rate without considering the instantaneous channel state information and therefore it is possible to schedule users by using only their long-term channel states [7], [8].

II. System Model

A. Grant-Free Multiple Access

We consider an uplink large-scale antenna system as illustrated in Fig. 1, where each of BS is equipped with M antennas and each of U UEs is equipped with a single antenna. Also, we configure the time-frequency sub-channels with a sub-frame with a duration of T_s and the sub-band bandwidth W_s. Within

each sub-channel, the channel state is assumed to be static so that the $M \times 1$ SIMO spatial small scale block fading, denoted by $\mathbf{h}_j$, does not change, while the large scale fading β_j only depends on the location of UE and does not change also within the scheduling period. We assume a perfect non line of sight (NLOS) propagation so that the small scale fading is modeled by a Rayleigh distribution. The UEs are divided into scheduling groups and then the BS allocates the sub-channels to the scheduling groups. The UEs allocated in the same sub-channel access to the BS by using space division multiple access (SDMA) and every UE with a new packet arrival gets activated and transmits by using the allocated sub-channel with a size of $N = W_s T_s$ symbols. For analytical tractability, it is assumed that the packet arrival follows a Poisson distribution with the average arrival rate of P_A. The BS and UE are assumed to operate in a perfect synchronization.

B. Channel and Signal Model

Let $\mathcal{O}$ be the set of users scheduled at the sub-channel of the sub-frame t and the sub-band f. Denote $\mathbf{x}_j$ the transmission block of user j in $\mathcal{O}$, then the $M \times N$ received signal matrix $\mathbf{Y}$ at the BS of interest can be expressed as

$$\mathbf{Y} = \sum_{j \in \mathcal{O}} \sqrt{\beta_j} \mathbf{h}_j \mathbf{x}_j^{\mathrm{H}} \mathbb{1}_j + \mathbf{V}, \tag{1}$$

where $\mathbb{1}_j$ is the indicator function for the activation of user j and $\mathbf{V}$ is the $M \times N$ matrix for the thermal noise and other-cell interference signals. The transmission block $\mathbf{x}_j$ has L symbols for its pilot sequence and $N - L$ symbols for data so that we can write $\mathbf{x}_j$ as $\mathbf{x}_j = \left[\sqrt{p_j^{\mathrm{tr}}} \, \boldsymbol{\psi}_j^{\mathrm{H}}, \ \sqrt{p_j^{\mathrm{dt}}} \, \mathbf{d}_j^{\mathrm{H}} \right]^{\mathrm{H}}$, where $\boldsymbol{\psi}_j$ denotes the pilot sequence of length L and $\mathbf{d}_j$ denotes the $N - L$ data symbol vector. Also, p_j^{tr} and p_j^{dt} denote the powers of the pilot sequence and the data symbols, respectively. Here we assume that p_j^{tr} is determined by the full channel inversion power control for a given common target received power $\bar{p}^{\mathrm{tr}}$, i.e., $p_j^{\mathrm{tr}} = \bar{p}^{\mathrm{tr}} \beta_j^{-1}$. Let $\mathbf{Y} = \left[\mathbf{Y}^{\mathrm{tr}}, \mathbf{Y}^{\mathrm{dt}} \right]$ and $\mathbf{V} = \left[\mathbf{V}^{\mathrm{tr}}, \mathbf{V}^{\mathrm{dt}} \right]$, then $\mathbf{Y}^{\mathrm{tr}}$ can be written

$$\mathbf{Y}^{\mathrm{tr}} = \sum_{j \in \mathcal{O}} \sqrt{\bar{p}^{\mathrm{tr}}} \mathbf{h}_j \boldsymbol{\psi}_j^{\mathrm{H}} \mathbb{1}_j + \mathbf{V}^{\mathrm{tr}}. \tag{2}$$

For the pilot sequence, we use Zadoff-Chu (ZC) sequences currently used in LTE-A [9]. With a length of a prime number L, there are sequences with $L - 1$ different roots and L circularly shifted versions for each root so that there are $L^2 - L$ sequences available. Note that the cross correlation value is zero between the ZC sequences of circularly shifted versions in the same root and is a constant of $\frac{1}{L}$ between sequences from different roots. As the cross correlation causes interferences and degrades the user detection performance, the sequence assignment must maximize the number of sequences of the same root. Therefore, when $K = |\mathcal{O}|$ and the sequence length is L, the number of used roots is minimized to $\left\lceil \frac{K}{L} \right\rceil$. For example, the pilot symbol for user j is assigned as $\psi_j(l) = \left\langle \exp\left(-\mathrm{j}\pi r_j l(l+1)/L\right) \right\rangle_{s_j}$, where the root is given as $r_j = 1 + \mod\left(j - 1, \left\lceil \frac{K}{L} \right\rceil\right)$, the circularly shift is given

as $s_j = 1 + \left\lfloor (j-1)/\left\lceil \frac{K}{L} \right\rceil \right\rfloor$, and $\langle \cdot \rangle_s$ denotes the circular shift by s symbols. Let define $\mathcal{O}_{r_j}$ be the set of users in $\mathcal{O}$ whose pilot sequences have root r_j and let $K_{r_j} = |\mathcal{O}_{r_j}|$.

III. THE NEYMAN-PEARSON DECISION CRITERION

The number of activated users in $\mathcal{O}$ at a time follows a binomial distribution $\mathcal{B}(K, P_A)$ because we assume a Poisson random packet arrival. At the BS side, for every sub-frame duration, it tries to detect which users transmitted signals by using the sequences and a correlation detector. Let define $Z_j = \left\| \mathbf{Y}^{\mathrm{tr}} \boldsymbol{\psi}_j \right\|^2$, then Z_j is the sufficient statistic for detecting user j from the received signal $\mathbf{Y}$ by using sequence $\boldsymbol{\psi}_j$. Also, define $\widetilde{\mathbf{v}}_j^{\mathrm{tr}} = \mathbf{V}^{\mathrm{tr}} \boldsymbol{\psi}_j$. Then the sufficient statistic Z_j can be written as

$$Z_j = L \bar{p}^{\mathrm{tr}} \left\| \mathbf{h}_j \mathbb{1}_j \right\|^2 + \sum_{k \in \mathcal{O} \setminus \mathcal{O}_{r_j}} \bar{p}^{\mathrm{tr}} \left\| \mathbf{h}_k \mathbb{1}_k \right\|^2 + \left\| \widetilde{\mathbf{v}}_j^{\mathrm{tr}} \right\|^2, \tag{3}$$

where $\boldsymbol{\psi}_j^{\mathrm{H}} \boldsymbol{\psi}_j = L$, $\left\| \boldsymbol{\psi}_j^{\mathrm{H}} \boldsymbol{\psi}_k \right\| = 1$ if $r_j \neq r_k$, and $\left\| \boldsymbol{\psi}_j^{\mathrm{H}} \boldsymbol{\psi}_k \right\| = 0$ if $r_j = r_k$, $j \neq k$. For the second term in (3), let define $J = \sum_{k \in \mathcal{O} \setminus \mathcal{O}_{r_j}} \mathbb{1}_k$, then $J \sim \mathcal{B}(K - K_{r_j}, P_A)$. Because $\mathbf{h}_j, \mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ and $\widetilde{\mathbf{v}}_j^{\mathrm{tr}} \sim \mathcal{CN}(\mathbf{0}, L^2 \mathbf{I}_M)$, Z_j follows a chi-square distribution with the degree of freedom (DoF) of $2M$ if J is determined. More specifically, $Z_j \sim \sigma_{J,\text{active}}^2 \chi_{2M}^2$ if $\mathbb{1}_j = 1$ and $Z_j \sim \sigma_{J,\text{idle}}^2 \chi_{2M}^2$ if $\mathbb{1}_j = 0$, where $\sigma_{J,\text{active}}^2 = (L\bar{p}^{\mathrm{tr}} + J\bar{p}^{\mathrm{tr}} + L^2)/2$ and $\sigma_{J,\text{idle}}^2 = (J\bar{p}^{\mathrm{tr}} + L^2)/2$. In fact, due to the large value of M, Z_j can be treated as a Gaussian random variable by using the central limit theorem.

By the NP decision criterion under a target detection probability P_D, the detection threshold Ω is the minimum value satisfying the detection probability condition as

$$
\begin{aligned}
1 - P_D &\geq P(Z_j \leq \Omega | \mathbb{1}_j = 1) \\
&= \sum_{q=0}^{K-K_{r_j}} P\left(Z_j \leq \Omega | \mathbb{1}_j = 1, J = q\right) P\left(J = q\right) \\
&= \sum_{q=0}^{K-K_{r_j}} Q\left(\frac{\Omega/\sigma_{q,\text{active}}^2 - 2M}{\sqrt{4M}}\right) P\left(J = q\right),
\end{aligned}
\tag{4}
$$

where $Q(\cdot)$ denotes the tail probability of the standard normal distribution. Note that the decision criterion with a threshold of Ω in (4) can cause false alarm and the false alarm probability P_{FA} can be calculated as

$$
\begin{aligned}
P_{FA} &= P(Z_j > \Omega | \mathbb{1}_j = 0) \\
&= \sum_{q=0}^{K-K_{r_j}} P\left(Z_j > \Omega | \mathbb{1}_j = 0, J = q\right) P\left(J = q\right) \\
&= 1 - \sum_{q=0}^{K-K_{r_j}} Q\left(\frac{\Omega/\sigma_{q,\text{idle}}^2 - 2M}{\sqrt{4M}}\right) P\left(J = q\right).
\end{aligned}
\tag{5}
$$

Note that there always exists a threshold that satisfies the detection probability condition. However, it is generally impossible to satisfy both the detection probability condition and the false alarm probability condition simultaneously. The resulted false alarm probability by the threshold from the
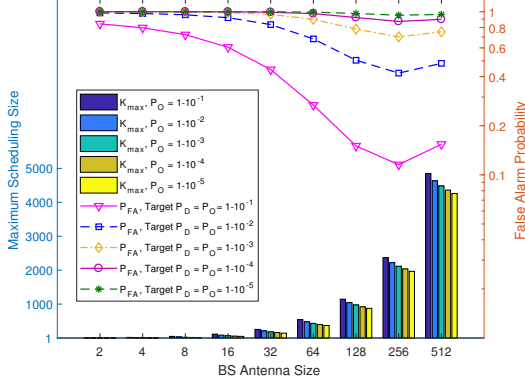
Fig. 2. The false alarm probability and the maximum scheduling size versus the BS antenna size when $L = 97$ and $P_A = 0.1$

Fig. 3. The false alarm probability and the maximum scheduling size versus the arrival rate when $M = 128$ and $L = 47$

detection probability condition shows the distance between $P(Z_j = z | \mathbb{1}_j = 1)$ and $P(Z_j = z | \mathbb{1}_j = 0)$ and thus it can evaluate a trade-off between the detection probability and the false alarm probability.

IV. NUMERICAL EVALUATIONS

The false alarm probabilities of various detection schemes under various target detection probabilities are evaluated numerically in this section. For one sub-channel, the sub-frame duration is 1 ms and the sub-band bandwidth is 125 KHz. Using orthogonal frequency division multiplexing (OFDM) waveform with 25% cyclic prefix (CP) overhead, $N = 100$ symbols are used in a transmission block. The target received power $\overline{p}^{tr}$ is set to be 15 dB over the thermal noise and other-cell interference. Here, we consider the maximum scheduling size constrained by the packet arrival rate P_A, the BS antenna size M, and the outage probability P_O. Because, when ZF receiver is used at the receiver, the maximum supportable number of users is at most $M - 1$ to have non-zero rates [6]. Thus, the maximum scheduling size is $K_{max} = \max K$, subject to $P(J \leq M - 1) \leq P_O$, where $J \sim \mathcal{B}(K, P_A)$. In addition, we used sufficiently large pilot length L so that $L^2 - L \geq K_{max}$ for the simulations.

In Fig. 2, the false alarm probability and the maximum scheduling size are plotted. Here, we set the pilot length $L = 97$ because K_{max} increases up to 4846 as M increases to 512. In this figure, the false alarm probability dose not always decrease with the number of BS antennas. The reason is that as the number of antennas increases, the number of interferers also increases very rapidly. From this, the number of BS antennas at a given target detection probability can be seen to have some optimal value for a given false alarm probability. In Fig. 3, the false alarm probability and the maximum scheduling size are shown when the arrival rate P_A changes from 0.1 to 0.9. Here, since the maximum scheduling size decreases significantly while the arrival rate increases, the number of active users at a time does not change a lot, and thus the false alarm probability decreases very slightly.
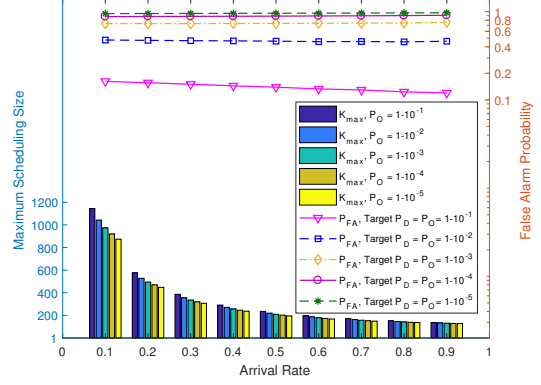
V. CONCLUSION

In this paper, user detection schemes are analyzed in a general GFMA system model. Assuming a Poisson packet arrival and using ZC sequences, the detection threshold for the target detection probability and the resulted false alarm probability are derived. By employing a number of antennas at a BS, the resulted false alarm probability can be lowered generally to an acceptable level. These results show that highly reliable detection schemes in GFMA for the future URLLC services are feasible in realistic LSAS environments.

REFERENCES

[1] "5G Innovations for new Business Opportunities," White Paper, 5G-PPP, Mar. 2017. [Online]. Available: https://5g-ppp.eu/roadmaps/
[2] *Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification*, 3GPP TS 36.321, Rev. 10.5.0, Mar. 2012.
[3] CATT, "Consideration on Grant-Free Transmission," 3GPP TSG RAN WG1 Meeting #86bis, R1-1608757, Oct. 2016.
[4] Huawei, "Advanced Multi-User Detectors for Grant-Free Transmissions," 3GPP TSG RAN WG1 Meeting #86bis, R1-1608855, Oct. 2016.
[5] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
[6] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-Antenna Channel Hardening and Its Implications for Rate Feedback and Scheduling," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1893–1909, Sept. 2004.
[7] J. H. Kim, *et al.*, "Grant-free Multiple Access for Ultra-Reliable Low-Latency Communications in a Large-Scale Antenna System," in *Proc. 2016 Int. Conf. on Information and Communication Technology Convergence (ICTC)*, Oct. 2016, pp. 466–470.
[8] K. J. Choi and K. S. Kim, "Optimal Semi-Persistent Uplink Scheduling Policy for Large-Scale Antenna Systems," *IEEE Access*, vol. 5, pp. 22 902–22 915, Oct. 2017.
[9] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation*, 3GPP TS 36.211, Rev. 10.0.0, Jan. 2011.