Online Estimation and Adaptation for Random Access with Successive Interference Cancellation

Sang-Woon Jeon, Member, IEEE and Hu Jin, Senior Member, IEEE

Abstract

This paper proposes an adaptive transmission algorithm for slotted random access systems supporting the successive interference cancellation (SIC) at the access point (AP). When multiple users transmit packets simultaneously in a slot, owing to the SIC technique, the AP is able to decode them through SIC resolve procedures (SRPs), which may occupy multiple consequent slots. While such an SRP could potentially improve the system throughput, how to fully exploit this capability in practical systems is still questionable. In particular, the number of active users contending for the channel varies over time which complicates the algorithm design. By fully exploiting the potential of SIC, the proposed algorithm is designed to maximize the system throughput and minimize the access delay. For this purpose, an online estimation is introduced to estimate the number of active users in real-time and controls their transmissions accordingly. It is shown that the throughput of the proposed algorithm can reach up to 0.693 packets/slot under practical assumptions, which is the first result achieving the throughput limit proved by Yu–Giannakis. It is further shown that the system throughput of 0.559 packets/slot (80.6% of the throughput limit) is still achievable when the SIC capability is restricted by two.

Index Terms

Online estimation, random access, successive interference cancellation, adaptive transmission control.

I. INTRODUCTION

For various Internet-of-Things (IoT) applications, a massive number of devices are expected to access wireless networks via random access or machine-to-machine (M2M) communication protocols [1], [2]. In order to support such rapidly growing demands, an efficient random access protocol providing enhanced throughput should be essentially developed, which is operated in a distributed and decentralized manner to cope with excessive uplink accesses for future wireless networks [3], [4].

Traditionally, it has been well known that the tree (or splitting) algorithm [5] and its variants [6]–[8] provide improved throughputs. In particular, the first-come first-serve (FCFS) splitting algorithm [6] has been known as the best one so far, which yields the maximum throughput of 0.487 (packets/slot). Splitting algorithms in [5]–[8] have been adopted for various wireless systems [9], [10], [12]–[15], as the physical layer technologies have evolved. They have been particularly considered for a cellular system with smart antennas [9] and with successive interference cancellation (SIC) [10], [12]. In particular, for the systems with SIC, the splitting algorithms have been further investigated with several power control schemes incorporated in [16]. Note that in general, when more than one packets can be decoded in a slot with advanced signal processing techniques, it is called multipacket reception (MPR) capable channel and thus it encompasses the systems discussed in [9], [10], [12], [16].

In principle, SIC-enabled techniques can provide substantial throughput improvement of random access systems by utilizing collided packets in the decoding procedure after subtracting interference from decoded packets in a sequential manner. Because of such potential improvement, SIC techniques have been widely adopted and studied in the literature [10], [17]–[20]. A tree-based splitting algorithm has been developed in [10] in order to resolve collision efficiently by the help of a SIC-enabled receiver and its modified algorithms have been proposed robust to error-prone networks [11], [12], achieving the maximum throughput of 0.693 (packets/slot). By allowing more refined feedback for collision resolve beyond tree-based splitting algorithms or allowing a partially centralized control or cooperation between users [17]–[20], it was shown that the maximum throughput can reach up to one, depending on the types of feedback information and network environment [19], [20]. Decentralized power control

S.-W. Jeon is with the Department of Military Information Engineering, Andong Hanyang University, Ansan 15588, South Korea (e-mail: sangwoonjeon@hanyang.ac.kr).

Hu Jin is with the Division of Electrical Engineering, Hanyang University, Ansan 15588, South Korea (e-mail: hjin@hanyang.ac.kr).

and allocation methods have been proposed in [21]–[23] for random access with capabilities of MRP and SIC in order to maximize the sum rate of random access systems by optimally allocating the transmit power of packets and their transmission rates.

Recently, more flexible and generalized random access or multiple access frameworks beyond the traditional slot-based user control paradigm have been considered to support massive M2M or IoT traffic demands in an efficient manner [27]. Coded random access incorporating with packet erasure codes or rateless-type codes with SIC decoding has been designed in [24]–[26], which provide flexibility and robustness for network dynamics such as fading, bursty noise, and user interference, yielding improved throughput compared to the slotted random access systems. More recently, non-orthogonal random access (NOMA) has been studied as a promising solution to improve the spectral efficiency for future cellular systems, e.g., fifth generation (5G) systems [28]–[30]. For various different types of NOMA systems, SIC is the most well-known detection techniques and actively studied to boost the spectral efficiency in NOMA systems [28]–[30].

In spite of recently developed random access or multiple access frameworks and advanced physical layer coding schemes, in this paper, we focus on slotted random access systems with SIC decoding and tree-based splitting algorithm to revolve collision via simple feedback information. It is worthwhile to emphasize that the maximum throughput of 0.693 (packets/slot) in [10] is still quite challenging to achieve in practical environment. There might be three major difficulties to be addressed: 1) SIC capabilities can be limited in practice due to imperfect physical layer processing or excessive complexity; 2) the backlog size is hard to obtain in real-time, which is needed for optimal transmission control; 3) the event of SIC failure can occur because of channel estimation error. Therefore, the primary aim of this paper to address those constraints and provide throughput close to 0.693 (packets/slot) in more practical random access environment. Our main contributions are summarized as follows:

- A universal online learning method for estimating the current backlog size inspired by the concept of Bayes' rule and an adaptive transmission control algorithm operated with the estimated backlog size are proposed.
- A novel analytical methodology to derive the maximum throughput limit and the related optimal transmission
 probabilities are presented, considering practical constraints of bounded SIC capabilities and SIC failure events.
- A practical random access system with SIC incorporated with a tree-based splitting algorithm is proposed, achieving the maximum throughput limit by optimally controlling and adapting its transmission under bounded SIC capabilities and SIC failure events.

Even though we present the proposed online learning and adaptation method in a standard slotted random access system, the developed online learning framework proposed in this paper is universally applicable for other types of random access systems.

II. SYSTEM MODEL

We consider a wireless communication system in which an access point (AP) is located at the center of a cell and time is slotted of equal length for one packet transmission. Users arrive to the system randomly and each user is assumed to have one packet to transmit to the AP. Therefore, a packet and a user are indistinguishable and they are used interchangeably in this paper. Such a system scenario was frequently discussed in the literature and is also suitable for the machine type communications to support internet of things (IoT) where most of the machine devices transmit one small packet to the AP. It is worthy to note that we *do not restrict the traffic arrival to be Poisson* which is frequently assumed in the literature on random access systems. Due to the random arrival of the users, the number of active users in the system, i.e., the backlog size, also varies over time. Such a time-varying backlog size is the main difficulty of designing efficient protocols for random access systems.

The backlogged users *randomly* transmit their packets with a certain probability broadcasted by the AP and consequently the number of packets arrived at the AP also varies over time. When a single packet is received at the AP, it can decode the packet successfully. On the other hand, if the AP receives multiple packets simultaneously at a slot, it is possible to perform SIC for decoding. Considering the difficulty of implementing SIC practically at the physical layer, in this paper we abstract the *SIC capability* by M. That is, if there are more than M users' simultaneous transmissions, the AP is not able to perform SIC due to some technical limitations of physical layer decoding and this slot is denoted as a collision slot. Alternatively, if the number of transmitting users k is between [2: M], the AP initiates an *SIC resolve procedure (SRP)*, denoted by SRP(k), from the next slot. Within SRP(k), only the k transmitted users are allowed to retransmit their packets under the *SIC-enabled collision resolve algorithm*.

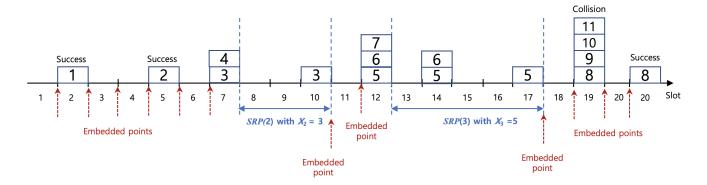


Fig. 1. Example of transmission behavior over time when M = 3.

which will be introduced in Section V. Note that when initiating SRP(k), the AP is aware of the k value which is the minimum information of performing SIC. With this information, SRP(k) can be terminated when all the k packets are successfully received. We denote X_k ($k \in [2 : M]$) by the duration for SRP(k) and $E[X_k]$ by its mean. For differentiation, we define the slots that are not in any SPR by the *normal* slots. We assume that the AP provides the feedback information of success, collision and, initiating and terminating the SRP at the end of each slot instantaneously.

Fig. 1 shows an example of packet transmissions over time when the SIC capability is M = 3. At Slots 2, 5, and 20, only one packet is transmitted which would be successfully received by the AP. At Slot 7, two packets are transmitted and the AP initiates an SRP(2) which lasts for 3 slots, i.e., $X_2 = 3$. In terms of decoding at the AP, as it receives only *Packet 3* at Slot 10 and it is able to decode *Packet 4* by performing SIC for the signal received previously at Slot 7. How those packets are retransmitted in the SRP(2) will be detailed in Section V. Similarly, at Slot 13, an SRP(3) is initiated and, in this case, the duration T_3 is 5. Finally, at Slot 19, as there are more than M(=3) users' transmissions, a collision occurs and the AP cannot initiate an SRP. We define the *embedded points* as the end of each idle slot, success slot, collision slot and the last slot of each SRP, while we do not count the idle and success slots occurred within any SRP for the embedded points. As a representative example, in Fig. 1, all the embedded points are tagged. The embedded points are always located in front of the *normal* slots.

For such a system, our primary purpose is to optimally control the transmissions of the backlogged users so that the system could provide maximum throughput as well as minimum access delay for each user. Note that as the communication of the system bases on the random access, the AP has no information about users, such as who have packets to transmit and what the current backlog size is. To achieve such a goal, we design a transmission control algorithm by clearing the following missions one by one.

- 1) How to design the SIC-enabled collision resolve algorithm so that the duration of SRP(k) is minimized for each $k (\in [2 : M])$. Note that the users involved in SRP(k) should transmit their packets in a distributed manner, i.e., the AP cannot schedule specific user's transmission as it has no information about the users who have packets.
- 2) How to control the transmissions of the backlogged users in *normal* slots (other than the slots in an SRP) so that the system operates optimally while exploiting the potentials of the SIC-enabled collision resolve algorithm?

We should comment here that the random access systems with SIC was also analyzed in [10] and it showed that the maximum throughput could reach 0.693. However, [10] considered an extreme scenario of *unbounded* SIC capability of $M = \infty$, i.e., no matter how many packets that are transmitted simultaneously, the AP is always able to perform SIC. This assumption also simplifies the optimal transmission control of the backlogged users: whenever an SRP is completed, all the backlogged users if exist in the system should transmit with probability one (which is called *Gated Access* in [10]). Unfortunately, such an operation no longer the optimal for the random access systems with *bounded* SIC capability M. For example, if the backlog size is larger than M at a slot, the transmission probability of one would always result in collisions in the consequent slots. Therefore, alternative transmission control algorithm should be designed when it comes to the bounded SIC capability. We shall design the SIC-enabled collision resolve algorithm in Section V to solve the above mentioned first mission. In Section III, we start our discussion by assuming that the backlog size is known at each normal slot and investigate the optimal system performance. This discussion further motivates us to design a transmission control algorithm that can estimate the the backlog size online while also accommodating the advantages of the SIC-enabled collision resolve algorithm in Sections IV.

III. SYSTEM OPTIMIZATION ASSUMING KNOWN BACKLOG SIZE

We first investigate an *ideal* scenario where the backlog size n is assumed to be known at each normal slot. In practice, such information cannot obtained in practice as the traffic arrivals are random. Investigation on such an ideal scenario enables us to observe a performance upper bound and also gives us useful insight for designing a transmission control algorithm.

Suppose that there are n backlogged users in a normal slot and they transmit their packet with probability p. Then, according to Renewal Theory, the expected service rate can be derived as

$$R_{n,M}(p) = \frac{B_n^p(1) + \sum_{k=2}^M k B_n^p(k)}{1 - \sum_{k=2}^M B_n^p(k) + \sum_{k=2}^M B_n^p(k)(1 + E[X_k])} = \frac{\sum_{k=1}^M k B_n^p(k)}{1 + \sum_{k=2}^M B_n^p(k) E[X_k]},$$
(1)

where $B_n^p(k) = \binom{n}{k}p^k(1-p)^{n-k}$ denotes the binomial distribution with parameters p and n. The numerator indicates the average number of packets successfully transmitted between two consecutive embedded points and the denominator indicates the corresponding duration.

Once the value of n is known, based on (1), we can find the optimal transmission probability p that maximizes the service rate. If such an optimal probability could be controlled slot by slot, we can maximize the system throughput and minimize each user's access delay. While we can optimize the system operation if the backlog size is known prior, it is hard to obtain such information in practice. Motivated by this, in the next Section IV, we propose a transmission control algorithm that has the capability of estimating the backlog size online. It is worthy to highlight that online estimation allows the algorithm to adapt to network dynamics.

IV. PROPOSED TRANSMISSION CONTROL ALGORITHM

In order to estimate the backlog size in each slot, we adopt the concept of Bayes' rule. As well known, based on *a priori* distribution, Bayes' rule introduces *a posteriori* distribution given that an event is observed. Therefore, when estimating the backlog size, instead of giving an constant number for the estimated backlog size, Bayes' rule estimates it through introducing a distribution. If we define the observed event as the number of users transmitted at each slot, we can update the estimated distribution of the backlog size slot by slot recursively. That is, the *a posteriori* distribution of the current slot becomes the a priori distribution of the next slot. Unfortunately, to implement such a process, we need to calculate the distribution at each slot which accompanies high computational complexity. In order to reduce the computational burden that occurs at each slot, we assume the backlog size follows the Poisson distribution. Such an assumption was also made in [32] and was shown to be effective when applying to random access systems.

Let the distribution of the backlog size n to be Poisson with mean ν , i.e,

$$\Phi_{\nu}(n) = \frac{\nu^n}{n!} e^{-\nu}.$$
(2)

Then, for the estimation of the backlog size, we need to update ν recursively slot by slot.

TABLE IOptimal parameter values.

M	x_M^*	\hat{S}_M^*	C_M
1	1	0.3678	1.3922
2	1.378	0.5586	2.0458
3	1.739	0.6352	2.7020
4	2.060	0.6665	3.3833
5	2.3762	0.6802	4.0681
10	3.8734	0.6926	7.5626

A. Optimal Transmission Probability

For a given backlog size n, the service rate is given in (1). If n is a random variable with distribution $\Phi_{\nu}(n)$, the average service rate is expressed as

$$\hat{S}_{\nu,M}(p) = \frac{\sum_{n=0}^{\infty} \sum_{k=1}^{M} k B_n^p(k) \Phi_{\nu}(n)}{\sum_{n=0}^{\infty} \left[1 + \sum_{k=2}^{M} B_n^p(k) E[X_k] \right] \Phi_{\nu}(n)}$$

$$= \frac{\sum_{k=1}^{M} k \Phi_{p\nu}(k)}{1 + \sum_{k=2}^{M} E[X_k] \Phi_{p\nu}(k)}$$

$$= \frac{\sum_{k=1}^{M} k \Phi_x(k)}{1 + \sum_{k=2}^{M} E[X_k] \Phi_x(k)},$$
(3)

where we have applied the fact that $\sum_{n=0}^{\infty} B_n^p(k) \Phi_{\nu}(n) = \Phi_{p\nu}(k)$ and $x = p\nu$ denotes the average number of transmissions per slot. Once M and $E[X_k]$ are given, we can numerically search for the optimal x_M^* that maximizes the above service rate $\hat{S}_{\nu,M}(p)$. Table I summarizes the optimal x_M^* and the corresponding maximum service rate \hat{S}_M^* , where the values of $E[X_k]$ are obtained from Section V. Consequently, the optimal transmission probability for a given mean backlog size ν is derived as

$$p^* = \frac{x_M^*}{\nu}.\tag{4}$$

From Table I, we observe that if the SIC capability of 2 or 3 can be obtained, the service rate of 0.5586 or 0.6352 is achievable, which corresponds to 80.6% and 91.6% of the maximum throughput of 0.693 obtained with unbounded SIC capability [10]. Therefore, from the engineering viewpoint, one can expect that implementing the SIC capability of 3 at the physical layer is reasonably good. It is notable that higher SIC capability requires more complicated advanced signal processing techniques, which in general accompany higher implementation complexity and cost.

B. Estimation on the Backlog Size

Suppose that the transmission probability is p. Then, the probability that m among n users are transmitting is obtained as

$$\Pr(m|n) = B_n^p(m), \quad \text{for } m \in [0:n].$$
(5)

Moreover, given the *a priori* distribution of $\Phi_n(\nu)$, the unconditional probability of *m* transmitting users is be derived as

$$Pr(m) = \sum_{n=0}^{\infty} B_n^p(m) \Phi_{\nu}(n)$$

= $\sum_{n=0}^{\infty} \Phi_{p\nu}(m) \Phi_{(1-p)\nu}(n-m)$
= $\Phi_{p\nu}(m).$ (6)

If there are m users' simultaneous transmissions in a slot, by applying Bayes' rule, we obtain

$$\Pr(n|m) = \frac{\Pr(m|n)\Phi_{\nu}(n)}{\Pr(m)} = \frac{B_{n}^{p}(m)\Phi_{\nu}(n)}{\Phi_{p\nu}(m)} = \Phi_{(1-p)\nu}(n-m),$$
(7)

which is the *a posteriori* distribution of the backlog size n. Interestingly, it is still Poisson but with an updated mean by

$$E[n|m] = (1-p)\nu + m.$$
(8)

If the optimal transmission probability $p^* = x_M^* / \nu$ shown in (4) is applied, we have

$$E[n|m] = \nu - x_M^* + m,$$
 (9)

where x_M^* is given in Table I. Now we are ready to discuss different channel observations such as idle, success, SRP, and collision.

1) Idle event: An idle event occurs when m = 0. Then, the mean backlog size of the *a posteriori* distribution is obtained as

$$E[n|\text{Idle}] = \nu - x_M^*. \tag{10}$$

2) Success event: A success event occurs when m = 1. Consequently, the mean backlog size of the *a posteriori* distribution is

$$E[n|\operatorname{Success}] = \nu - x_M^* + 1. \tag{11}$$

Since one packet is successfully received by the at the AP after a success event, the estimation on the mean backlog size should be further subtracted by 1.

3) SRP event: When $m \in [2 : M]$, the AP initiates an SRP(m). For such an event, the mean backlog size of the *a posteriori* distribution is obtained as

$$E[n|SRP(m)] = \nu - x_M^* + m.$$
⁽¹²⁾

Since those m packet would be eventually received by the AP at the end of the SRP(m), the estimation on the mean backlog size should be further subtracted by m at the end of the SRP.

4) Collision event: When m > M, a collision occurs and the AP is not able to know the number of users involved in the collision. The probability that a collision occurs when there are n backlogged users in the system is obtained as

$$\Pr(C|n) = \sum_{m=M+1}^{n} B_n^p(m).$$
(13)

Consequently, the unconditional probability of a collision event is obtained as

$$Pr(C) = \sum_{n=M}^{\infty} \left(\sum_{m=M+1}^{n} B_n^p(m) \right) \Phi_{\nu}(n)$$

$$= \sum_{m=M+1}^{\infty} \Phi_{p\nu}(m)$$

$$= 1 - \sum_{m=0}^{M} \Phi_{p\nu}(m).$$
 (14)

Then, by applying Bayes' rule, we obtain the conditional probability of n given a collision event by

$$\Pr(n|C) = \frac{\Pr(C|n) \Pr(n)}{\Pr(C)} = \frac{\Phi_n(\nu) \sum_{m=M+1}^n B_n^p(m)}{1 - \sum_{m=0}^M \Phi_{p\nu}(m)} = \frac{\sum_{m=M+1}^n \Phi_{(1-p)\nu}(n-m) \Phi_{p\nu}(m)}{1 - \sum_{m=0}^M \Phi_{p\nu}(m)}.$$
(15)

Algorithm 1 Estimation and Adaptation Algorithm at the AP with the SIC capability M

1: Initialize $\nu = 10$, $p^* = x_M^*/M$, $\lambda_E = 0.5$ and repeat the following at each embedded point. 2: if Idle then 3: $\lambda_E \leftarrow \theta \lambda_E$ $\nu \leftarrow \nu - x_M^* + \lambda_E$ 4: 5: else if Success then $\lambda_E \leftarrow \theta \lambda_E + (1 - \theta) \cdot 1$ 6: $\nu \leftarrow \nu - x_M^* + \lambda_E$ 7: 8: else if SRP(m) with $m \in [2:M]$ then Initiates an SRP(m) with the involved m users. 9: Record the SRP duration X_m . 10: $\begin{array}{l} \lambda_E \leftarrow \frac{\theta \lambda_E + (1 - \theta)m}{\theta \cdot 1 + (1 - \theta)(1 + X_m)} \\ \nu \leftarrow \nu - x_M^* + \lambda_E T_m \end{array}$ 11: 12: 13: else if Collision then $\lambda_E \leftarrow \theta \lambda_E$ 14: $\nu \leftarrow \nu + C_M + \lambda_E$ 15: 16: **end if** 17: $p^* = \min\left(1, \frac{x_M^*}{\nu}\right)$ 18: Broadcast \hat{p}^* for the next time slot.

It is not difficult to check that the above distribution is no longer the Poisson while we still approximate the distribution by Poisson with its mean as

$$E[n|C] = \sum_{n=0}^{\infty} n \Pr(n|C)$$

= $(1-p)\nu + \frac{p\nu - \sum_{m=0}^{M} m\Phi_{p\nu}(m)}{1 - \sum_{m=0}^{M} \Phi_{p\nu}(m)}.$ (16)

If the optimal transmission probability of $p^* = x_M^* / \nu$ is applied, we have

$$E[n|C] = \nu - x_M^* + \frac{x_M^* - \sum_{m=0}^M m \Phi_{x_M^*}(m)}{1 - \sum_{m=0}^M \Phi_{x_M^*}(m)}$$

= $\nu + C_M$, (17)

where $C_M = \frac{\sum_{m=0}^M (x_M^* - m)\Phi_{x_M^*}(m)}{1 - \sum_{m=0}^M \Phi_{x_M^*}(m)}$ and its value for each M is listed in Table I.

C. Algorithm Description

Algorithm 1 describes the overall procedure of our proposed algorithm. Note that the values of x_M^* and C_M are listed in Table I. Lines 3, 6, 11, 14 correspond to the estimation on the arrival rate where the previous arrival rate is given by multiplying the weight of θ . In Section VII, we set $\theta = 0.99$ in simulation. Lines 4, 7, 12, 15 correspond to the update on the mean backlog size whose mathematical background was introduced in the previous subsection. Line 17 corresponds to the calculation on the optimal transmission probability.

V. SIC RESOLVE ALGORITHM

The proposed transmission control algorithm, introduced in the previous Section IV, adapts the transmission probabilities for packet transmissions at each *normal* slot. For the design of the proposed algorithm, we applied the results of the average duration of SRP(m) which is initiated by $m (\in [2 : M])$ users' simultaneous transmissions. Once an SRP(m) starts, only the *m* users involved in the previous transmissions retransmit their packets under the *SIC resolve algorithm* in the subsequent slots. If all the *m* packets are successfully received by the AP, SRP(m) is terminated. Section V-A introduces the SIC resolve algorithm and Section V-B analyzes the expected duration of each SRP(m), i.e., $E[X_m]$.

Algorithm 2 SIC Resolve Algorithm

6 6		
1: Initialize \mathcal{V} as the set of users involved.		
2: Run SIC_RESOLVE(\mathcal{V}).		
3: function SIC_RESOLVE(\mathcal{V})		
4: Set $\mathcal{W} = \emptyset$ and $X_{ \mathcal{V} } = 0$.		
5: while $\mathcal{W} = \emptyset$ or $\mathcal{W} = \mathcal{V}$ do		
6: Each user in \mathcal{V} retransmits with probability $p_{ \mathcal{V} }$.		
7: Each transmitted user becomes a member of $\dot{\mathcal{W}}$.		
8: $X_{ \mathcal{V} } \leftarrow X_{ \mathcal{V} } + 1.$		
9: end while		
10: if $ \mathcal{W} = 1$ then		
11: The AP performs decoding and SIC as possible.		
12: end if		
13: if $ \mathcal{V} - \mathcal{W} \ge 2$ then		
14: Each non-transmitted user enters $\mathcal{V} \setminus \mathcal{W}$.		
15: end if		
16: if $ \mathcal{W} \ge 2$ then		
17: Run SIC_RESOLVE ($\mathcal{W}, X_{\mathcal{W}} $).		
18: $X_{ \mathcal{V} } \leftarrow X_{ \mathcal{V} } + X_{ \mathcal{W} }.$		
19: end if		
20: if $ \mathcal{V} - \mathcal{W} \ge 2$ then		
21: Run SIC_RESOLVE($\mathcal{V} \setminus \mathcal{W}, X_{ \mathcal{V} \setminus \mathcal{W} }$).		
22: $X_{ \mathcal{V} } \leftarrow X_{ \mathcal{V} } + X_{ \mathcal{V} \setminus \mathcal{W} }.$		
23: end if		
24: end function		

A. Algorithm Description

Let \mathcal{V} denote the set of the users that make the AP initiate an $SRP(|\mathcal{V}|)$ where $|\mathcal{V}|$ denotes the cardinality of the set \mathcal{V} . Then, the SRP is proceeded by invoking the function SIC_RESOLVE($\mathcal{V}, X_{|\mathcal{V}|}$) recursively, which is described in Algorithm 2. $X_{|\mathcal{V}|}$ is an output of the function SIC_RESOLVE($(\mathcal{V}, X_{|\mathcal{V}|})$) and it indicates the duration spent for the function.

At the first slot of SIC_RESOLVE($\mathcal{V}, X_{|\mathcal{V}|}$), the users in \mathcal{V} retransmit their packet with probability $p_{|\mathcal{V}|}$ as in Line 6. This retransmission probability could be minimized for each $|\mathcal{V}|$ as the AP knows the value of $|\mathcal{V}|$ when invoking the function. Each user who has transmitted in the first slot considers itself as a member of a new set \mathcal{W} as in Line 7. If $|\mathcal{W}| = 1$, the AP decodes the only transmitted user's packet and additionally performs SIC based on the previously received signals. If the number of non-transmitted users is larger than 2, those users become a member of the set $\mathcal{V} \setminus \mathcal{W}$ as in Line 14. If $|\mathcal{W}| \ge 2$, i.e., the number of transmitted users is larger than 1, the SIC_RESOLVE is invoked for \mathcal{W} as in Line 17. After the completion of SIC_RESOLVE($\mathcal{W}, |X_{\mathcal{W}}|$) if exists, the function SIC_RESOLVE runs again for $\mathcal{V} \setminus \mathcal{W}$ if $|\mathcal{V}| - |\mathcal{W}| \ge 2$. The time duration for SIC_RESOLVE($\mathcal{V}, X_{|\mathcal{V}|}$) is updated at Lines 8, 18, and 22.

B. Expected Duration of SIC Resolving Time

In this subsection, we analyze the expected SIC resolve time achievable by Algorithm 2.

1) Recursive relation for SIC resolving time: Let X_m be the number of time slots for SIC resolving when m users are initially involved. Obviously, $X_0 = X_1 = 0$. To derive the recursive relation between $\{X_m\}$, denote $\Delta_m \ge 1$ as the number of time slots required for starting the inherent SIC resolve phase like Lines 17 or 21 in Algorithm 2, and $M_m \in [1 : m - 1]$ as the number of users for the inherent SIC resolve phase. Notice that both Δ_m and M_m are random.

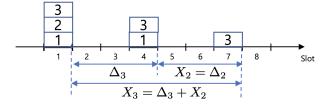


Fig. 2. Explanatory SIC resolving time when 3 users are involved.

From the proposed SIC resolve algorithm, we have $X_2 = \Delta_2$, $X_3 = \Delta_3 + X_2$, and for $m \ge 4$,

$$X_m = \begin{cases} \Delta_m + X_{m-1}, & \text{if } M_m = 1 \text{ or } M_m = m - 1, \\ \Delta_m + X_{M_m} + X_{m-M_m}, & \text{if } M_m \in [2:m-2], \end{cases}$$
(18)

For better understanding of the relation between Algorithm 2 and X_m , explanatory cases for resolving SIC started from three and four users are provided in Figs. 2 and 3, respectively. Specifically, in Fig. 2, three users (users 1 to 3) are initially involved and after random transmissions with probability p_3 , users 1 and 3 retransmit after three time slots, i.e., $X_3 = \Delta_3 + X_2$ and $\Delta_3 = 3$ in this case. Note that user 2 does not participate the remaining retransmission procedure because of the condition in Line 20 of Algorithm 2 is not satisfied. Then users 1 and 3 perform random transmissions with probability p_2 and user 3 retransmits alone after three time slots. Again, from the condition in Line 20 of Algorithm 2, user 1 does not participate the remaining retransmission procedure, which gives $X_2 = \Delta_2$ and $\Delta_2 = 3$ in this case.

In Fig. 3, four users (users 1 to 4) are initially involved and after random transmissions with probability p_4 , user 1 retransmits after two time slots, i.e., $X_4 = \Delta_4 + X_3$ and $\Delta_4 = 2$. Then users 2, 3, and 4 perform random transmissions with probability p_3 and users 2 and 4 retransmit after two time slots. In this case, $X_3 = \Delta_3 + X_2$ and $\Delta_3 = 2$. Finally, users 2 and 4 perform random transmissions with probability p_2 and user 2 alone retransmits after four time slots, i.e., $X_2 = \Delta_2$ and $\Delta_2 = 4$. Note that users 2 and 4 retransmit after two time slots, but the corresponding received signal is removed at the AP and users 1 and 4 perform the same random transmissions, see the condition in Line 5 of Algorithm 2.

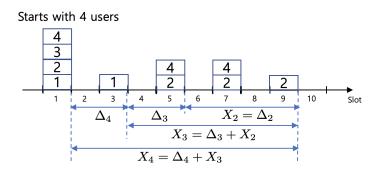


Fig. 3. Explanatory SIC resolving time when 4 users are involved.

2) Expected Duration Analysis: Now consider how to derive the expected time required for resolving SIC, i.e., $E[X_m]$ for $m \in [2:M]$.

From the definition of Δ_m , it follows the geometric distribution with success probability $1 - B_m^{p_m}(0) - B_m^{p_m}(m)$ and, as a result, we have

$$E[\Delta_m] = \frac{1}{1 - B_m^{p_m}(0) - B_m^{p_m}(m)}.$$
(19)

TABLE II MINIMUM EXPECTED NUMBER OF TIME SLOTS.

m	$E^*[X_m]$
2	2
3	3.333
4	4.761
5	6.210
10	13.426

Also, for $l \in [1:m-1]$, we have

$$\Pr(M_m = l) = \frac{B_m^{p_m}(l)}{1 - B_m^{p_m}(0) - B_m^{p_m}(m)},$$
(20)

which satisfies $\sum_{l=1}^{m-1} \Pr(M_m = l) = 1$. Therefore, $E[X_2] = E[\Delta_2]$, $E[X_3] = E[\Delta_3] + E[\Delta_2]$, and from (18),

$$E[X_m]$$

$$= \sum_{l=1}^{m-1} \Pr(M_m = l) E[X_m | M_m = l]$$

$$= (\Pr(M_m = 1) + \Pr(M_m = m - 1)) (E[\Delta_m] + E[X_{m-1}])$$

$$+ \sum_{l=2}^{m-2} \Pr(M_m = l) (E[\Delta_m] + E[X_l] + E[X_{m-l}])$$

$$= E[\Delta_m] + \sum_{l=2}^{m-1} (\Pr(M_m = l) + \Pr(M_m = m - l)) E[X_l]$$
(21)

for $m \ge 4$. Then, $E[X_m]$ can be recursively calculated from (21) by using (19) and (20).

Recall that p_m is the retransmission probability of each involved user when m users are involved for the SIC resolving, which can be optimized to minimize the expected SIC resolving time. Let $\{p_i^*\}_{i=1}^m$ be the set of optimal retransmission probabilities minimizing $E[X_m]$. Denote the minimum expected number of time slots for resolving SIC among m users by $E^*[X_m]$. Then, it is given by

$$E^*[X_m] = \min_{\{p_i \in (0,1)\}_{i=1}^m} E[X_m].$$
(22)

Table II summarizes $E^*[X_m]$ for m = 2, 3, 4, 5, 10. Through numerical calculations, we also observed that the setting of $p_m = 1/2$ for all m yields almost identical expected duration for resolving SIC as Table II. Therefore, from the engineering viewpoint, setting a constant probability of 1/2 is a reasonably good choice.

VI. IMPACT OF SIC FAILURE

Although introducing SIC capability in the physical layer decoding can provide substantial throughput improvement in random access systems, SIC is quite challenging in practice, especially as the number of users involving in the SIC procedure increases due to imperfect channel estimation, error propagation during SIC process, etc. In this section, we will address such practical issues and how to modify the proposed online transmission control algorithm in Section IV, minimizing the performance degradation due to SIC failure.

Let $p_e \in [0, 1)$ be the probability of SIC failure in each of SIC steps. For simplicity, we assume that an event of SIC failure independently occurs with probability p_e for each time slot involving more than one user are simultaneously transmitted.¹

¹The same analysis presented in Section VI is straightforwardly applicable for the case where the probability of SIC failure depends on the number of simultaneously transmitted users.

TABLE III $E^*[X_m^{(e)}]$ and p_m^* when $p_e = 0.5$.

m	$E^*[X_m^{(e)}]$	p_m^*
2	3	0.5
3	4.788	0.412
4	6.633	0.343
5	8.486	0.288
10	17.802	0.163

Obviously, the most efficient way to deal with such an SIC failure event is that the corresponding users involved in the event retransmit again except the users whose packets are already decoded, to construct the same received signal at the AP so that the AP can resume the rest of SIC procedure.

We can apply the same notion of the SIC resolving time to analyze the impact of SIC failure. Let $X_m^{(e)}$ be the number of time slots for SIC resolving when m users are initially involved counting the retransmission slots due to SIC failure. Note that the analytical method introduced in Section V-B is able to provide a simple universal tool for dealing with such probabilistic events. Define Δ_e as the number of time slots required to have the received signal with no SIC failure at the AP excluding the originally received signal. Then, its mean is given by $E[\Delta_e] = \frac{1}{1-p_e} - 1$ from the geometric distribution. Now define $X'_m = X^{(e)}_m - \Delta_e$, which corresponds to the number of time slots for SIC resolving when *m* users are initially involved assuming that there is no SIC failure for the received signal involving m users.

Then, similar methods in Section V-B are applicable to derive recursive relations between $\{X_m^{(e)}\}$ and $\{X_m'\}$. In particular, $X_2' = \Delta_2$ and $X_2^{(e)} = X_2' + \Delta_e$. Also,

$$X'_{3} = \begin{cases} \Delta_{3} + X'_{2}, & \text{if } M_{3} = 1, \\ \Delta_{3} + X^{(e)}_{2}, & \text{if } M_{3} = 2, \end{cases}$$

and $X_{3}^{(e)} = X_{3}' + \Delta_{e}$. For $m \ge 4$,

$$X'_{m} = \begin{cases} \Delta_{m} + X'_{m-1} & \text{if } M_{m} = 1, \\ \Delta_{m} + X^{(e)}_{M_{m}} + X'_{m-M_{m}} & \text{if } M_{m} \in [2:m-2], \\ \Delta_{m} + X^{(e)}_{m-1} & \text{if } M_{m} = m-1, \end{cases}$$

and $X_m^{(e)} = X'_m + \Delta_e$. Then, from (19) and (20), we have $E[X'_2] = E[\Delta_2]$, $E[X_2^{(e)}] = E[X'_2] + E[\Delta_e]$. Also,

$$E[X'_3] = E[\Delta_3] + E[X'_2] + \Pr(M_3 = 2)E[\Delta_e]$$
(23)

and $E[X^{(e)}] = E[X'_3] + E[\Delta_e]$. For $m \ge 4$, we have

$$E[X'_{m}] = E[\Delta_{m}] + \sum_{l=2}^{m-1} (\Pr(M_{m} = l) + \Pr(M_{m} = m - l))E[X'_{l}] + \sum_{l=2}^{m-1} \Pr(M_{m} = l)E[\Delta_{e}]$$
(24)

and $E[X_m^{(e)}] = E[X_m'] + E[\Delta_e]$. Similarly, let $E^*[X_m^{(e)}]$ be the minimum expected number of time slots for resolving SIC among *m* users. Unlike the case of no SIC failure, $p_m = 1/2$ is not optimal if $p_e > 0$. For instance, Table III summarizes $E^*[X_m^{(e)}]$ and the corresponding $\{p_i^*\}_{i=1}^m$ for m = 2, 3, 4, 5, 10 when $p_e = 0.5$.

Now, from the derived $E^*[X_m^{(e)}]$, we can obtain the maximum service rate and the corresponding parameters required for applying the proposed online transmission control algorithm as the same manner in Section IV. For instance, Table IV summarizes \hat{S}_M^* and the corresponding parameters for online estimation when $p_e = 0.5$.

M \hat{S}_M^* x_M^* C_M 0.3678 1.3922 1 1 1.2580 2 0.4821 2.1220 3 0.5155 1.4700 2.8892 4 0.5264 1.6380 3.6960 5 0.5300 1.760 4.5461 10 9.2964 0.5316 1.8840 0.7 Lessessesses 0.6 0.5 Throughput 0.3 Leeeeeeeeeeeeeeeeeeeeee + M=1. Online Estimation M=1. Ideal Case 0 M=2. Online Estimation M=2, Ideal Case 0.2 * M=3. Online Estimation M=3, Ideal Case 0.16 M=5. Online Estimation M=5, Ideal Case 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 Arrival Rate $[\lambda]$

TABLE IV PARAMETERS FOR ONLINE ESTIMATION WHEN $p_e = 0.5$.

Fig. 4. Average throughput with respect to λ .

VII. NUMERICAL EVALUATION

In this section, we perform numerical evaluation to demonstrate the performance of the proposed online estimation and transmission control algorithm for random access with SIC. Recall the throughput improvement in Table I requires the backlog size in real-time, which is impossible for most random access systems. Furthermore, in practice, SIC is quite challenging especially as the number of users involving in the SIC procedure increases due to imperfect channel estimation, error propagation during SIC process, etc. In the following, we will address these practical issues and demonstrate that the proposed online estimation and adaptation framework can still provide such throughput improvement under more practical environment. Unless otherwise stated, we assume Poisson packet arrivals and the update weight θ in Algorithm 1 is set as $\theta = 0.99$ for the online estimation.

A. Online Estimation and Adaptation

In order to investigate the performance of the proposed algorithm in Section IV, Fig. 4 compares the average throughput achievable by the proposed algorithm with that of the ideal case where the backlog size is known. As seen in the figure, the maximum service rate supportable by the proposed algorithm is the same for that of the ideal case. In order to investigate the impact of the online estimation and transmission control closely, Fig. 5 plots the average access delay with respect to λ . The average delay rapidly increases as the arrival rate converges to the maximum service rate so that the system becomes unstable for this regime. Note that the average delay gap between the online estimation and the ideal case is less than four time slots for the stable regime of λ . The results demonstrate that the proposed online estimation and transmission control performs well by only learning from the historical information.

Not only for Poisson user arrivals, the proposed algoirthm is universally applicable for a board class of arrival distributions. As an explanatory case, we assume that for every 100 time slots the arrival rate is changed either the Poisson distribution of mean 2λ or zero with equal probability. Hence, the mean arrival rate is equal to λ . Figs. 6 and 7 plot the average throughput and delay of this non-Poisson arrival model, respectively. For comparison, we

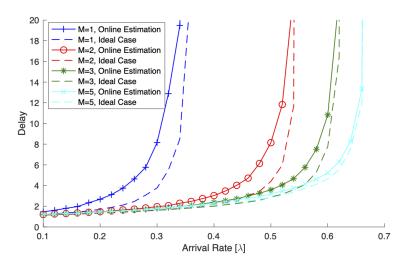


Fig. 5. Average delay with respect to λ .

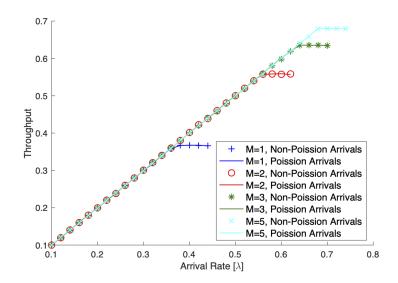


Fig. 6. Average throughput with respect to λ for non-Poisson arrivals.

also plot the performances of the poisson arrival model. Because the considered non-Poisson model corresponds to on-off user arrivals, the delay performance is expected to worse that that of the Poisson model. As seen in the figures, the proposed algorithm still performs well even for non-Poisson user arrivals.

To investigate the real-time estimation performance of the proposed algorithm, Fig 8 compares the estimated backlog size (the parameter ν in Algorithm 1) with the actual backlog size by observing 10^5 time slots when M = 2, where 100 time episodes are averaged out to plot the figure. For each time episode, we set $\lambda = 0.4$ for $t \in [1:3 \times 10^4]$ and $t \in [7 \times 10^4 + 1:10^5]$ and $\lambda = 0.5$ for $t \in [3 \times 10^4 + 1:7 \times 10^4]$. The result shows that Algorithm 1 can acurately estimate the actual backlog size in real-time, providing an efficient transmission control for random access systems reflecting network dynamics.

B. Impact of SIC Failure

To evaluation the performance degradation due to SIC failure, Figs. 9 and 10 plot the average throughput and delay performance of the proposed algorithm when the probability of SIC failure is given by $p_e = 0.5$. It is worthwhile to mention that the maximum arrival rate supported by the proposed algorithm coincides with the analytical results of \hat{S}_M^* in Table IV. In spite of the performance degradation due to SIC failure, the results show

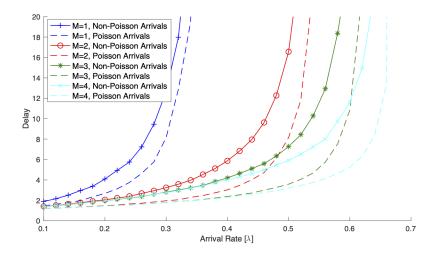


Fig. 7. Average delay with respect to λ for non-Poisson arrivals

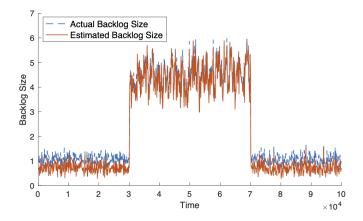


Fig. 8. Estimated backlog size of Algorithm 1 when M = 2.

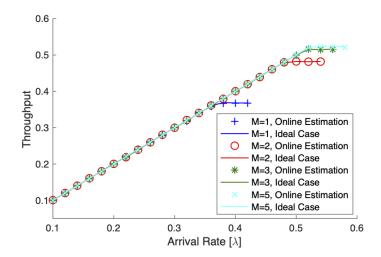


Fig. 9. Average throughput with respect to λ when $p_e = 0.5$.

that SIC is still beneficial for improving throughput compared to the known best achievable throughput of 0.487 without applying SIC [6] even for the very restrictive environment of M = 3 and $p_e = 0.5$. The overall delay

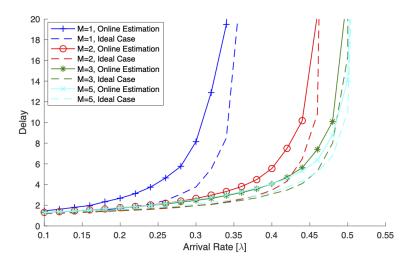


Fig. 10. Average delay with respect to λ when $p_e = 0.5$.

performance shows similar tendency as the case of no SIC failure as in Fig. 5. Because of retransmissions of SIC-failed packets, increasing the access delay is inevitable. Nonetheless, the proposed algorithm optimally adjusts the transmission probabilities during the SIC-enabled collision resolve procedure to minimize the retransmissions due to SIC failure in probabilistic sense.

VIII. CONCLUSION

In this paper, we proposed the online estimation and adaptive transmission algorithm for slotted random access systems. The proposed algorithm has been designed to fully utilize the SIC capability of the AP by carefully adjusting its transmission via online estimated information. Thanks to such online estimation and adaptation features, it is universally applicable for general non-Poisson arrivals and also robust to limited SIC capability and SIC failure.

REFERENCES

- E. de Carvalho, E. Bjornson, J. H. Sorensen, P. Popovski, and E. G. Larsson, "Random access protocols for massive MIMO," *IEEE Commun. Mag*, vol. 55, no. 5, pp. 216-222, May 2017.
- [2] R. Harwahyu, R. G. Cheng, C. H. Wei, and R. F. Sari, "Optimization of random access channel in NB-IoT," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 391-402, Feb. 2018.
- [3] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo, "Contention-based access for ultra-reliable low latency uplink transmissions," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 182-185, Apr. 2018.
- [4] K. Kannan, and M. Devaraju, "QoS supported adaptive and multichannel MAC protocol in vehicular ad-hoc network," *Cluster Comput.*, pp. 1-13, Mar. 2018.
- [5] J. J. Capetanakis, "Tree algorithm for packet broadcast channel," IEEE Trans. Inf. Theory, vol. IT-25, no. 5, pp. 505-515, Sep. 1979.
- [6] D. Bertsekas and R. Gallager, Data Networks, Prentice-Hall, 1987.
- [7] M. Sidi and I. Cidon, "Splitting protocols in presence of capture," IEEE Trans. Inf. Theory, vol. 31, no. 2, pp. 295-301, Mar. 1985.
- [8] G. Fayolle, P. Flajolet, M. Hofri, and P. Jacquet, "Analysis of a stack algorithm for random multiple-access communication," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 244-254, Mar. 1985.
- [9] H. Jin and A. Acampora, "Performance of a tree-based collision resolution algorithm in cellular systems with smart antennas," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 1143-1151, Mar. 2007.
- [10] Y. Yu and G. B. Giannakis, "High-throughput random access using successive interference cancellation in a tree algorithm," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4628-4639, Dec. 2007.
- [11] X. Wang, Y. Yu, and G. B. Giannakis, "Design and analysis of cross-layer tree algorithms for wireless random access," *IEEE Trans. Commun.*, vol. 55, no. 12, pp. 2253-2256, Dec. 2007.
- [12] X. Wang, Y. Yu, and G. B. Giannakis, "A robust high-throughput tree algorithm using successive interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 909-919, Mar. 2008.
- [13] X. Qin and R. Berry, "Opportunistic splitting algorithms for wireless networks," in *Proc. IEEE INFOCOM*, pp. 1662-1672, Mar. 2004.
- [14] D. Y. Kim, H. Nam, and H. Jin, "Pseudo-Bayesian broadcasting algorithm for opportunistic splitting scheduling systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5451-5455. Jun. 2017.
- [15] D. Xu and Y. Yao, "Splitting tree algorithm for decentralized detection in sensor networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6024-6033, Dec. 2013.

- [16] A. D. Gore and A. Karandikar, "Power-controlled FCFS splitting algorithm for wireless networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 842-856, Feb. 2010.
- [17] Y. Hu, W. Yang, and Y. Cai, "Throughput analysis of slotted ALOHA with cooperative transmission using successive interference cancellation," *Science in China Series F: Information Sciences*, vol. 52, no. 12, pp. 2354-2359, Dec. 2009.
- [18] M. Mollanoori and M. Ghaderi, "Uplink scheduling in wireless networks with successive interference cancellation," *IEEE Trans. Mobile Comput.*, vol. 13, no. 5, pp. 1132-1144, May 2014.
- [19] E. Paolini, G. Liva, and M. Chiani, "Graph-based random access for the collision channel without feedback: Capacity bound," in *Proc. IEEE ICC*, Dec. 2011.
- [20] S.-H. Lee, B. C. Jung, and S.-W. Jeon, "Successive interference cancellation with feedback for random access networks," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 825-828, Apr. 2017.
- [21] C. Xu, L. Ping, P. Wang, S. Chan, and X. Lin, "Decentralized power control for random access with successive interference cancellation," *IEEE J. Select. Areas Commun.*, vol. 31, no. 11, pp. 2387-2396, Nov. 2013.
- [22] H. Lin, K. Ishibashi, W.-Y. Shin, and T. Fujii, "A simple random access scheme with multilevel power allocation," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2118-2121, Dec. 2015.
- [23] Y. Li and Lin Dai, "Maximum sum rate of slotted aloha with successive interference cancellation," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5385-5400, Nov. 2018.
- [24] E. Paolini, G. Liva, and M. Chiani, "Coded slotted ALOHA: A graph-based method for uncoordinated multiple access," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6815-6832, Dec. 2015.
- [25] C. Stefanovic and P. Popovski, "ALOHA random access that operates as a rateless code," *IEEE Trans. Wireless Commun.*, vol. 61, no. 11, pp. 4653-4662, Nov. 2013.
- [26] L. Lu, L. You, and S. C. Liew "Network-coded multiple access," *IEEE Trans. Mobile Comput.*, vol. 13, no. 12, pp. 2853-2869, Dec. 2014.
- [27] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations," *IEEE Commun. Mag*, vol. 55, no. 9, pp. 55-61, Sep. 2017.
- [28] H. Zhang, D.-K. Zhang, W.-X. Meng, and C. Li, "User pairing algorithm with SIC in non-orthogonal multiple access system," in *Proc. IEEE ICC*, May 2016.
- [29] B. Ling, C. Dong, J. Dai, and J. Lin, "Multiple decision aided successive interference cancellation receiver for NOMA systems," *IEEE Wireless Commun. Lett.* vol. 6, no. 4, pp. 498-501, Aug. 2017.
- [30] K. Hicuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. E98?B, no. 3, pp. 403-414, Mar. 2015.
- [31] B. S. Tsybakov and V. A. Mikhailov, "Random multiple packet access: Part-and-try algorithm, (in Russian) Probl. Pered. Inform., vol. 16, no. 4, pp. 6579, Oct./Dec. 1980.
- [32] R. L. Rivest, "Network control by Bayesian broadcast," IEEE Trans. Inf. Theory, vol. 33, no. 3, pp. 323-328, Mar. 1987.