TAB: CSI Lossless Compression for MU-MIMO Network

Qigui Xu, Wei Xi, Lubing Han, and Kun Zhao(✉)

School of Computer Science and Technology, Xi'an Jiaotong University,
Xi'an 710049, China
`qiguixu@gmail.com`, `weixi.cs@gmail.com`, `254215143@qq.com`,
`pandazhao1982@gmail.com`

Abstract. Multi-user MIMO (MU-MIMO) is an important technology to improve data transmission efficiency for future network, such as 5G and WiFi 6, due to its ability of enabling multi-users' concurrent transmissions. To achieve concurrent diversity gains, MU-MIMO network resource allocation relies on the feedback of Channel State Information (CSI) from multiple clients. CSI feedbacks from large user population, however, heavily degrade the throughput of a MU-MIMO network. Pursuing smart CSI feedback, we present a CSI timeliness-aware balanced mechanism, named TAB. It is a novel MU-MIMO protocol to eliminate unnecessary feedback overhead and improve CSI utilization within channel coherence time. TAB is fully compatible with the WiFi 5/6 standard and most state-of-the-art CSI feedback strategies, and is easy to be deployed on existing WiFi systems. Our software-radio based implementation and testbed experimentation demonstrate that TAB substantially improves the throughput of both downlink and uplink MU-MIMO network by 1.5× at least.

Keywords: MU-MIMO network · Channel State Information · Channel overhead

1 Introduction

Wireless network has been facing the growing demand of higher speed and more efficiency for the mass real-time data transmission among interconnected and intercommunicated multi-users. To deal with more resource consumption and better spectrum coverage, Multi-user Multiple-Input Multiple-Output (MU-MIMO) technology is being employed in many standard wireless protocols and infrastructures for future network, such as WiFi 6 [1] and 5G [9]. By allowing concurrent transmissions between a multi-antenna access point (AP) and multi-users, MU-MIMO holds the enormous potential to substantially improve

---

This work was supported by National Key R & D Program of China 2019YFB2102201, NSFC Grant No. 61772413, 61802299, and 61672424.

H. Gao et al. (Eds.): CollaborateCom 2020, LNICST 349, pp. 95–114, 2021.
https://doi.org/10.1007/978-3-030-67537-0_7

spectrum efficiency and network throughput. In MU-MIMO, users can report the estimation of their corresponding Channel State Information (CSI) to AP, and the AP selects the concurrent users with strong channel orthogonality accordingly to maximize the total channel capacity. CSI is a comprehensive metric of the energy attenuation and phase delay of wireless signals in multi-path channel with frequency selective fading [6]. For the AP in MU-MIMO, its data pre-coding before sent and data decoding after received rely on the accurate CSI estimation.

However, CSI feedback introduces high superfluous channel overhead, which constrains the network performance gain from MU-MIMO technology [14]. First, CSI report frame contains so many channel matrix elements and defined parameters that multi-user CSI feedback can occupy much system time and spectrum resource originally available for data transmission. Second, channel state may often fluctuate over time due to the channel mobility, dynamic interference and noise, thus most standard protocols require CSI feedback before each data transmission, regardless of the validity period for real-time CSI. Even worse, the CSI feedback overhead will increase to an intolerable degree with the growing user population in MU-MIMO network, overwhelming the channel resource utilization of normal data transmission. Besides, to measure real-time CSI, calculate the beamforming matrix and select the best transmission strategies, MU-MIMO network has to sacrifice lots of superfluous computation overhead and energy consumption. Thus it is crucial to reduce the CSI feedback overhead in MU-MIMO network.

To address the performance degradation of MU-MIMO network, existing solutions try to reduce unnecessary CSI feedbacks. On the one hand, some CSI compression methods are proposed to reduce the volume of complicated CSI matrix [3,10,12,19]. Without enough specific elements, however, inaccurate CSI matrix can not profile multi-path channel feature efficiently. As a result, the data transmission and parsing relying on accurate CSI may suffer from more efficiency loss. On the other hand, many novel channel sounding mechanisms are designed to reduce CSI feedback times. Some approaches support CSI share among location-related users [7,21], some [2,11,20] reuse CSI according to the similarity of CSI samples before and after, and some utilize simple CSI or zero CSI feedback strategies [15,16,18,22,23] to access channel. Lacking timely discriminative CSI feedback, however, fuzzy channel estimation can neither resist the effect of multi-path channel fading, nor avoid the accumulative error drift problem over time. Not to mention whether the real-world channel quality can afford for high efficient data transmission in MU-MIMO network. To sum up, if not considering the variance and fluctuation of real-time CSI effectiveness, blindly reducing feedback is often counterproductive and even aggravate the network performance deterioration.

In this paper we propose a CSI timeliness-aware balanced mechanism (TAB) to compress CSI while keeping its quality at the same time. First, based on the significant difference between adjacent training symbol samples, user can reply a lightweight ACK to AP, indicating that the real-time CSI has no significant change. Second, based on that CSI can be utilized to sense the fluctuation of wireless channel, AP accumulates CSI packet samples to estimate the variant range of channel state so as to predict the current CSI valid period related to each user, *i.e.*, channel coherence time (CCT). Finally, according to the CCT magnitude, the AP groups the multi-users with diverse priority, and schedules them via Round Robin with the corresponding constraint of maximum service time slice. In this way, the unnecessary CSI feedback overhead and the consequent resource consumption could be reduced significantly meanwhile the channel efficiency and network performance get improved efficiently.

The contributions of this paper can be summarized as follows.

1. We use both theoretical and experimental results to show that CSI feedback overhead is the cause of the performance of MU-MIMO limitation. This is the first work to target on comprehensive CSI sample analysis to realize CSI lossless compression via timeliness-aware balanced mechanism.

2. We develops an adaptive CCT-based multi-user scheduling scheme at AP and a CSI report occasion estimation method at users to eliminates unnecessary CSI feedback by replying a lightweight ACK frame to reuse historical CSI, so that the system throughput can be significantly improved.

3. We conduct extensive real-world experiments under practical settings, with prototype implementation using software-radio devices. Results show that TAB significantly improves the throughput of both uplink (UL) and downlink (DL) MU-MIMO network and decrease superfluous resource consumption.

2 Related Work

MU-MIMO technology has been put around for some time to resolve the insatiable demand for wireless capacity. Many standard protocols also have been exploring the application of MU-MIMO technology, such as LTE [4], IEEE 802.11ac (WiFi 5) [5], as well as the ongoing 5G [9] and 802.11ax (WiFi 6) [1]. They all employ Beamforming technology to improve channel Signal Noise Ratio (SNR) by concentrating the transmitted energy on the target receiver. Meanwhile, CSI [6,17] is introduced in MU-MIMO network to resist the effects of multi-path and frequency selective fading. CSI is the key component of MU-MIMO communication, but frequent CSI feedback also produces much superfluous channel overhead, which constrains and even cancels out the network throughput gain using MU-MIMO technology. To maintain this type of channel overhead in MU-MIMO system, Current CSI feedback reduction methods mainly fall into two categories: channel feedback volume compression or CSI feedback times reduction.

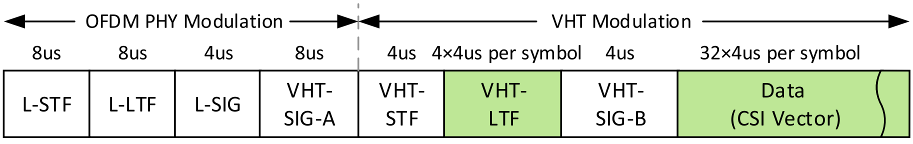
On the one hand, different compression methods are available to reduce the volume of CSI matrix. CSI-SF [3] uses the CSI value of a single data stream to predict the CSI value of multiple data streams, thereby reducing the CSI's oversampling. AFC [19] adaptively selects the compression level according to the Channel feature, quantizes or compresses CSI from 3 dimensions: time, frequency and numerical values. CSIFit [12] achieves high compression ratio by finding the sine signal of CSI and piecewise fitting it, based on Orthogonal Frequency Division Multiple (OFDM) technology. EliMO [10] adopt a two-way channel estimation to allow AP to accurately estimate DL CSI without explicit CSI feedback. Such compression methods sacrifice the CSI quality to save channel overhead, but inaccurate CSI may degrade the network performance in turn when faced with multi-path effect.

On the other hand, various CSI feedback mechanisms are designed to reduce CSI feedback times. Some schemes advise to share CSI in multi-user clusters. NEMOx [21] organizes a network into practical-size clusters with opportunistically synchronized APs in DL to balance their cooperation gain and spatial reuse. GCC [7] allows the location-related users to share a CSI matrix to limit the feedback quantity. Nevertheless, the premise of launching these schemes is that proper transmission mode and user groups are selected. Some strategies advocate reuse CSI to transmit more data packets according to CSI feature. Gabriel [2] reuses CSI by checking its validity period, which is obtained by F-distribution test of two CSI samples. RoFi [11] senses the rotation of device using Power Delay Profile (PDP) similarity and achieves rotation-aware CSI feedback while maintaining high throughput. QUICK [20] tries to reuse the CSI within its coherence time and schedule multi-users fairly. Unfortunately, these strategies can not prevent the real-time CSI from accumulative error drifting over time. There are also some approaches try to utilize simple CSI or zero CSI to access channel directly. OPUS [18] reduces CSI overhead via AP's iterative probing and users' competition report. Signpost [23] achieves scalable MU-MIMO signaling with zero CSI feedback. NURA [16] utilizes a lightweight UL user access mechanism with partial CSI feedback. TOUSE [15] employs Dynamic Time Warping (DTW) algorithm to evaluate the data rate and realizes the adaptive user selection with zero CSI feedback. Guidepost [22] combines the key ideas of NEMOx, Signpost and NURA, builds on a novel principle of indirection channel orthogonality evaluation to decouple and simplify the complicated computational and contention interaction among users. Although these methods can eliminate CSI feedback overhead significantly, they can not obtain a good Signal Interface Noise Ratio (SINR) without enough CSI in MU-MIMO transmission because of multi-path channel fading problem as well. In a word, these approaches have not consider the necessity for lossless CSI feedback or the role of historical CSI samples. Hence, this paper combines the current CSI sample analysis with the historical CSI sample, so as to reduce the unnecessary CSI feedback overhead when consider the effectiveness of the real-time CSI.

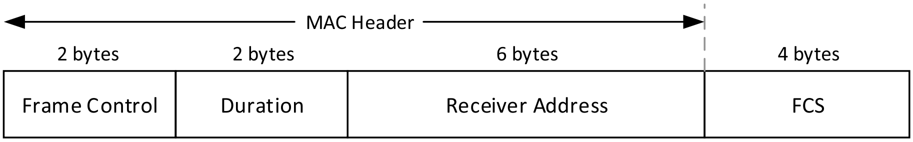
3 Channel Sounding Overhead

Channel State Information (CSI) is the key component for AP to allocate or parse data in MU-MIMO network. Considering the multi-path effect and frequency selective fading of wireless signals, CSI is a comprehensive measure of the energy attenuation and phase delay of signals from different paths, which reflects the diversities between different channels. In order to achieve accurate and efficient data transmission, it is essential for AP to obtain real-time CSI feedback via channel sounding. Nevertheless, CSI feedback will occupy a large number of network resources and bring much superfluous overhead to the MU-MIMO system. We must be aware that CSI reporting is time-consuming and frequency-spectrum-consuming while the concurrent capacity of AP can be maintained at the same time is limited. According to the channel sounding process, the influence of its channel overhead on system data throughput mainly depends on the size of CSI report frame and the feedback frequency of real-time CSI.

On the one hand, the CSI report frame in wireless protocol is very complicated, including many CSI matrix elements and specified preambles. For the scenario with a 4-antenna AP, the size of CSI matrix with 52 subcarriers estimated by each user with one antenna is $1 \times 4 \times 52 \times (32/8 \times 2) = 1664$ bytes, which requires 32 OFDM symbols including the Real and imaginary parts and it takes 4 µs to transmit each symbol. Then, the data of CSI vector is embedded in the data field of PHY Protocol Data Unit (PPDU) as shown in Fig. 1(a). In addition, the VHT-LTF (Long Training Field) in preamble is mainly applied to channel estimation and correction. Its length depends on the number of transmitted streams, and here it equals 4 symbols. As described above, reporting a CSI vector by a user may take up to 180 µs by sending the aggregation frame. If there are 20 users in the MU-MIMO network, it would take 3.6 ms to report all CSIs. Not to mention that during channel sounding, different users has various CSI valid periods which would gives rise to plenty of waiting and turnaround time for CSI request and report. According to the 802.11ac standard [5], however, the length of data field should not exceed 5.5 ms. That is, with the growing of user population, the CSI feedback overhead may approach or even overwhelm the channel resources available for data transmission. Even in 802.11ax standard [1], multi-user CSIs are fed back via UL MU-MIMO, the system time overhead is merely replaced with the channel spectrum resource consumption, rather than partially eliminated. Fortunately, it is found that a user may simply reply an ACK to indicate no change of the current CSI, which only costs 4 µs as shown in Fig. 1(b).



(a) The format of the PPDU type



(b) The format of the ACK frame

Fig. 1. CSI report frame analysis

On the other hand, real-time CSI on wireless channels has a certain period of validity, i.e., Channel Coherence Time (CCT). During this time period, the channel state remains almost constant, and all packets can be decoded using the same CSI. To analyze the CCT corresponding to different SNRs and user device moving speeds, the AP is allowed to continuously send packets to the user. The CCT is defined as the CSI update interval that leads to a 3 dB SNR loss for the decoding results of packets [13]. As Fig. 2, these stationary users have a relatively longer CCT exceeding 250 ms. Even at 1 m/s, their CCTs are much larger than the maximum packet transmission time 5.5 ms specified by 802.11ac. Only with extremely high moving speeds do the users need a much faster frequency to update CSI, which seems to be impractical because the human speeds on foot generally vary from 0.5 m/s to 2.0 m/s. It is evident that if the device in MU-MIMO network keeps stationary, the surrounding environments and system parameters of which remain the same during a long time, it is unnecessary to require the user to report CSI to AP before each transmission, like the 802.11ac protocol. Frequent CSI updates tend to many redundant operations, of which the resulting channel overhead undoubtedly restricts the throughput capacity of the MU-MIMO network.

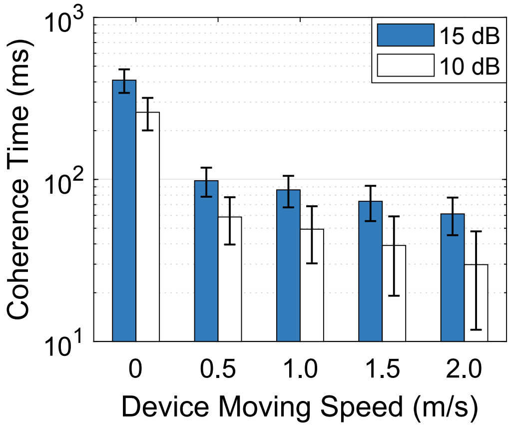


Fig. 2. Channel coherence time under different device moving speeds.

In addition, the request, estimation and utilization of real-time CSI also occupies quite a bit superfluous communication resource, computation overhead and energy consumption. Therefore, the purpose of TAB is to eliminate unnecessary channel sounding overhead as possible, including the superfluous channel overhead for CSI requests from AP, CSI report from users and the service scheduling among them. The final goal of TAB is to improve the channel efficiency for MU-MIMO data transmission.

4 TAB Protocol Design

In this section, we will introduce the details of TAB, which aims to resolve the performance degradation of MU-MIMO network. To reduce unnecessary channel sounding overhead, TAB combines CSI reporting occasion estimation method, channel coherence time (CCT) prediction method and multi-user balanced scheduling method. First, based on the significant difference between adjacent LTF symbols sample in time, unnecessary CSI report from users can be reduced. Second, according to the fluctuation state of CSI packets for a while, the CCT can be predicted to indicate the valid period of current CSI. Finally, with the priority grouping and time slice cycling, the multi-user balanced scheduling component determines which group of users should be scheduled and whether the current CSI needs updating.

4.1 Work Flow of TAB

TAB is compatible with 802.11ac/ax standard and QUICK protocol [20], except that TAB leverages a novel update operation on CSI Matrix List (CML). As Fig. 3, TAB includes both AP and client protocol adjustment, while the sampling form and window size are different, and the operation purpose is different.

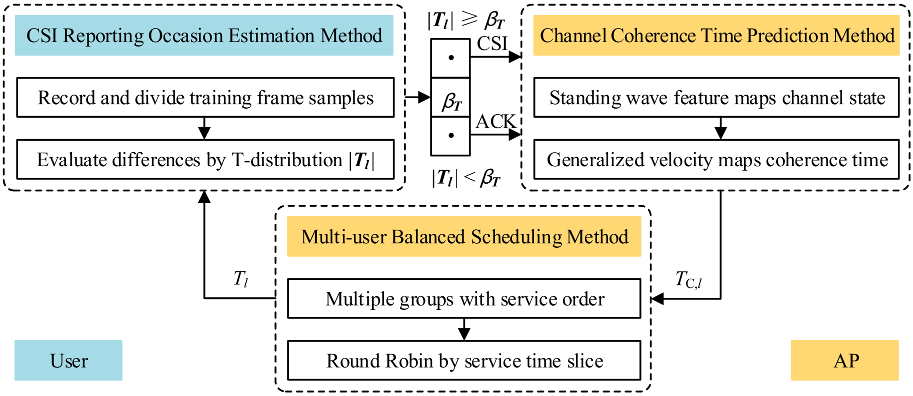


Fig. 3. The work flow of TAB

User conducts the short window sampling on the received training frame symbols, in order to determine whether to feedback the complete CSI estimation results. This process mainly consists of the following four steps:

(**i**) The user receives the Null Data Packet (NDP) frame including Long training Field (LTF) from the AP, and then analyzes and records the LTF as an evaluation sample for subsequent channel conditions.

(**ii**) The current user constructs a T-distribution hypothesis test to estimate the significant difference $|T_l|$ between the current LTF sample and the historical

LTF sample. If there is no significant difference between the samples before and after (*i.e.*, $|T_l| < \beta_T$), then go to step (**iii**); Otherwise, then go to step (**iv**).

(**iii**) The current user only reports a short ACK frame to inform the AP that the previous CSI can be reused, and accumulates the current LTF sample behind the historical LTF sample record as the new historical sample record.

(**iv**) The current user uses the current LTF sample to estimate the real-time CSI and immediately reports it to the AP, and then overwrites the cleared historical LTF sample with the current LTF sample.

AP conducts the long window sampling of effective CSI packets to estimate the corresponding channel coherence time of each user and schedule them fairly. This process mainly consists of the following three steps:

(**i**) AP records the effective CSI packet samples through QUICK protocol and starts analysis after reaching a certain length of time window, which is set to 30 ms here to ensure that the existing packets can effectively reflect the quasi-standing wave characteristic of CSI amplitude during device movement.

(**ii**) For cumulative CSI packet samples, pre-treatment of CSI, power distribution analysis, characteristic frequency estimation, device generalized speed estimation and CCT prediction via interpolation mapping are performed successively in order to achieve the CCT $T_{C,l}$ related to each user l.

(**iii**) According to the magnitude of predicted CCT, AP groups each user, sets the corresponding maximum service time slice T_{max} as system constraint and the suitable scheduling order as well as fair service time T_l during multi-user round robin scheduling.

4.2 CSI Reporting Occasion Estimation Method

CSI reporting occasion estimation method aims to eliminate unnecessary feedback overheads from users via just replying an ACK to AP to indicate no change of their CSIs. A straightforward criterion for CSI validation is to judge whether the historical CSI $H_{l,h}$ for the l-th user is equal to its current CSI $H_{l,c}$. Due to the impact of noise and device mobility during feedback interval, however, we cannot compare $H_{l,h}$ and $H_{l,c}$ directly by calculating the distance between their coordinates in the constellation diagram (c-diagram). Hence, we leverage statistical approach to verify whether there are significant differences between $H_{l,h}$ and $H_{l,c}$ by analyzing the received LTF samples in user.

We can reasonably assume that if the CSI changes, the mean of $H_{l,c}$ would be different from the mean of $H_{l,h}$. Whereupon, we can construct the following hypothesis:

$$\mathrm{H}_0 : \overline{H}_{l,c} = \overline{H}_{l,h}, \mathrm{H}_1 : \overline{H}_{l,c} \neq \overline{H}_{l,h}$$

where $\overline{H}_{l,c}$ and $\overline{H}_{l,h}$ is the expectation of $H_{l,c}$ and $H_{l,h}$ respectively. If H_0 holds, the historical CSI $H_{l,h}$ is still qualified in this turn. Considering different lengths of $H_{l,c}$ and $H_{l,h}$, TAB performs this hypothesis testing of H_0 via the *T-distribution test* and we can construct the tests statistic T as follows:

$$T_l = \frac{\left(\overline{H}_{l,c} - \overline{H}_{l,h}\right)}{S_w\sqrt{\frac{1}{N_{l,c}} + \frac{1}{N_{l,h}}}} \sim t\left(N_{l,c} + N_{l,h} - 2\right)$$

where

$$S_w = \frac{(N_{l,c} - 1)\, S_{l,c}^2 + (N_{l,h} - 1)\, S_{l,h}^2}{(N_{l,c} + N_{l,h} - 2)}$$

and $S_{l,c}^2$ and $N_{l,c}$ denotes the variance and sample size of $H_{l,c}$, respectively and similarly for $H_{l,h}$. Hence, with the significance level α, the rejection region is

$$|T_l| = \left| \frac{\left(\overline{H}_{l,c} - \overline{H}_{l,h}\right)}{S_w \sqrt{\frac{1}{N_{l,c}} + \frac{1}{N_{l,h}}}} \right| \geq t_{\alpha/2}\,(N_{l,c} + N_{l,h} - 2) \tag{1}$$

In TAB, we set $\alpha = 0.01$ in which $t_{\alpha/2} \geq 2.576$ when $N_{l,c} + N_{l,h} - 2 \leq \infty$. Thus the decision threshold is set as $\beta_T = 2.576$. As long as $|T_l| < \beta_T$, we can believe that H_0 holds 99 % confidence.

In this way, partial CSI report strategy can be adopted to reduced the length of CSI report frame. If most CSI vectors of all subcarriers have significant change, full CSI should be reported as usual. If partial CSI vectors of these subcarriers have significant change, the corresponding CSI vectors of continuous adjacent subcarriers should be reported. In contrast, the other CSI vectors making no evident difference can be replaced with some flag bits indicating sub-ACK. Once most CSI vectors of all subcarriers have no significant change, the user can reply an ACK frame to instead to indicate that its historical CSI can be reused for the AP. In addition, after a user report current CSI, the corresponding recorded LTF samples must be refreshed. Otherwise, the user should continue to record to generate longer historical samples, for avoiding CSI error drift over time.

4.3 Channel Coherence Time Prediction Method

Channel coherence time (CCT) is the time validity period T_C within which the data packets can be decoded by the same CSI without increased bit error ratio (BER). This feature can help us to constrain the operation of TAB system. For one, the CCT can guide AP reduce unnecessary channel sounding overhead including CSI request and feedback. For the channels within CCT, it can consider transmitting longer aggregate packets by reusing CSI. For another, CCT can provide effective basis for the AP to schedule users. The concurrent capacity of AP can be maintained at the same time is limited. Once faced with a large number of users, orderly user scheduling can reduce channel overhead caused by frequent channel sounding. Therefore, reasonable prediction of CCT will greatly maintain the effectiveness of real-time CSI and the fairness of user scheduling, and reduce unnecessary channel sounding overhead.

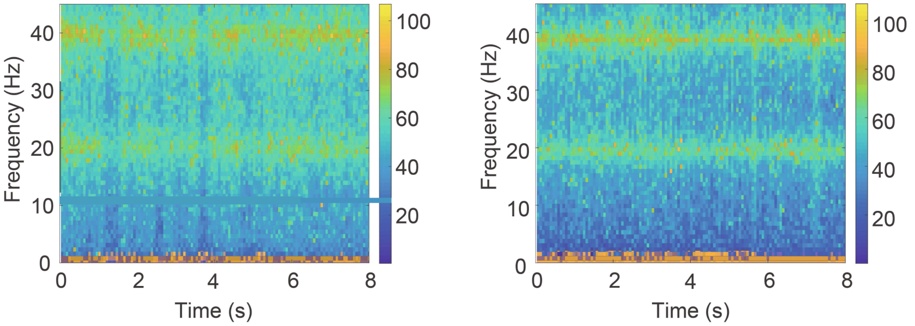
With the CSI-quality protection of QUICK, here CCT is mainly determined by both device motion state and environmental changes related to channel mobility. It has been proved that [8,20], when a wireless device is moving indoor with a low speed v, though the Doppler shift is negligible, some periodical ripples with frequency f_o in continuous CSI amplitude sample still appears. This quasi-standing wave phenomenon of wireless signals can satisfy

$$f_o = \frac{v}{\lambda/2} \tag{2}$$

where λ is the wavelength of original signal and $\lambda = 12.5\ cm$ in 2.4GHz WiFi. That is, for a mobile channel, its moving speed can be inferred by analyzing the characteristic frequency of the quasi standing wave taken on in CSI sample. Based on the above relationships, the CCT prediction method in TAB is a two-level mapping process, specific as follows.

1. *Pre-treatment of CSI sample.* The phase of CSI $\mathbf{H} = [H_1, H_2, \ldots, H_N]$ for N selected subcarriers in multi-path environment is uniformly distributed between $[0, 2\pi]$, which provides no discriminative information. Hence, we only use its amplitude $\mathbf{A} = |\mathbf{H}|$ to estimate the comprehensive channel state during sampling. Since the instantaneous reception rate of packets is unstable, $\mathbf{A}$ is resampled to a stable reception frequency f_w, denoted as $\mathbf{A_{re}}$.

2. *Power distribution Analysis via CSI.* We utilize the short-time Fourier transformation (STFT) with 50% overlapping Hanning window to obtain the power spectral density (PSD) of $\mathbf{A_{re}}$ in different subcarriers, in which the channel noise and motion state are both reflected significantly. Take device moving speed $v = 1.20\,\mathrm{m/s}$ for example as Fig. 4, specific moving speed will make the received samples present a certain quasi-standing wave characteristic frequency. Without loss of generality, the accumulative length of $\mathbf{A_{re}}$ is set to 30 ms.



(a) Low SNR (b) High SNR

Fig. 4. The STFT results for the CSI amplitude sample of the 15-th subcarrier under different SNRs: 10 dB and 15 dB.

3. *Characteristic Frequency Estimation.* We can extract the characteristic frequency of fundamental wave in the i-th subcarrier, of which the PSD is no less than 1/3 of the maximum PSD, as follows.

$$f_o^i = \frac{\sum_{f_{\min} < f_j < f_{\max}} f_j \times \omega_j}{\sum \omega_j} \tag{3}$$

where the weight ω_j represents the corresponding PSD value and the frequency range should be selected between $f_{\min} = 8$ Hz and $f_{\max} = 32$ Hz according to the Eq. (2). Without loss of generality, the final characteristic frequency f_o of the standing wave is estimated by the median of N selected subcarriers' characteristic frequency as follows

$$f_o = median\left(f_o^1, f_o^2, ..., f_o^N\right) \tag{4}$$

where $N = 30$ in our work.

4. *Device Generalized Speed Estimation.* To comprehensively measure the channel state fluctuation degree, we use the custom device generalized speed as the evaluation metric. Hence, the device generalized speed can be expressed as

$$\bar{\upsilon}=\frac{\lambda \times f_o}{2} \tag{5}$$

For $\upsilon = 1.2\,\text{m/s}$, the estimated results are $1.19\,\text{m/s}$ ($15\,\text{dB}$) and $1.24\,\text{m/s}$ ($10\,\text{dB}$) respectively, very close to the truth.

5. *CCT Prediction via Interpolation Mapping.* By employing Piecewise Cubic Hermite Interpolation (PCHIP), plenty of statistical results in $10\,\text{dB}$ can be used to fit the relationship between $\bar{\upsilon}$ and T_C, as Fig. 5. A simple lookup table is constructed at the AP, so that once the generalized speed is estimated, the corresponding CCT can be obtained immediately. TAB takes 0.05 as the step size, and constructs a lookup table containing 41 data pairs, which replaces a large amount of computing overhead with a little fixed memory.

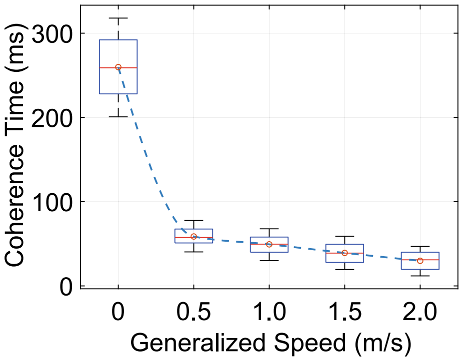


Fig. 5. PCHIP fitting between channel coherence time and Device Generalized Speed.

Note that the CCT prediction method is a process of estimating the overall CCT by partial cumulative CSI samples within the CCT. Under common human walking and channel noise, CSI samples would generally not suffer Cliff-like variations. The corresponding CCT is far greater than the specified maximum data packet length, as well as the sample time used to estimate the overall CCT. The historical CSI packets are recorded merely when satisfying QUICK protocol with affordable channel quality and often plays a auxiliary role. Only when there are extreme changes does the AP need to start over with CSI accumulating.

In addition, CCT only helps the AP select and schedule users, and limit CSI calibration timeout. Hence, a coarse-gained CCT prediction is enough efficient for TAB to work stably. Occasional inaccurate CCT values will only increase some redundant operations of feedback, record and refresh, but not obvious impacts on the whole MU-MIMO system.

4.4 Multi-user Balanced Scheduling Method

There is no doubt that the previous proposed method has reduced CSI feedback overhead before data transmission as much as possible. Nevertheless, the unfair service for multi-users will still lead to serious throughput imbalance problems between different users. Since AP prefers to schedule these users whose CSIs is still usable in CML, the user with a longer CSI lifecycle based on CCT would obtain a longer service time. Especially, the CCT of stationary user is much longer than that of mobile user. As a result, other users with short CCT will have to wait for service in a long time and even forever. Therefore, it is very necessary to group the users based on different priorities and schedule them with the maximum service time slice.

On the one hand, for the CCTs of vary length from different users, we adopt multi-group multi-priority service strategy. Based on the empirical results, we divides the users with different CCTs into 3 categories:

(**i**) $T_C \leq 48$ ms and $T_{max} = 24$ ms.

(**ii**) 48 ms $< T_C \leq 144$ ms and $T_{max} = 72$ ms.

(**iii**) $T_C > 144$ ms and $T_{max} = 216$ ms.

where T_{max} is the maximum time slice in each group, used to limit the maximum service time for each group of users. The priorities of different groups are divided by the urgency of the current transmission task. When the priority level of task urgency is similar, we adopt short run-time job first (SJF) algorithm to reduce the average waiting and turnaround time during user scheduling.

On the other hand, for each user in one group, we set the specific user service time according to the estimated CCT and the corresponding maximum time slice. To make a tradeoff between user throughput fairness and network global throughput, every user would be served in a continuous time of as close as possible to this time slice for each scheduling, according to the formulation

$$T_l = \left\lfloor \frac{T_{\max}}{T_{C,l}} + \frac{1}{2} \right\rfloor \times T_{C,l} \tag{6}$$

where T_l is the maximum continuous service time available to each user l, $T_{C,l}$ is the CCT related to the l-th user and $\lfloor x \rfloor$ is the floor function of x.

So far, the continuous service time of each user can be as close as possible to the maximum time slice in its group, which avoids that some individual users have been occupying channel resources. All the service time of users is integer times of their own CCTs, which can fully improve the utilization rate of real-time CSI and reduce unnecessary channel overhead. For a network connected with 20 users, each user can be served for twice or thrice at least in every second. Such service conditions will not affect the experience of users to use the Internet.

5 Implementation and Evaluation

In this section, we present and analyze the experimental results using a implemented TAB prototype.

5.1 Experimental Setup

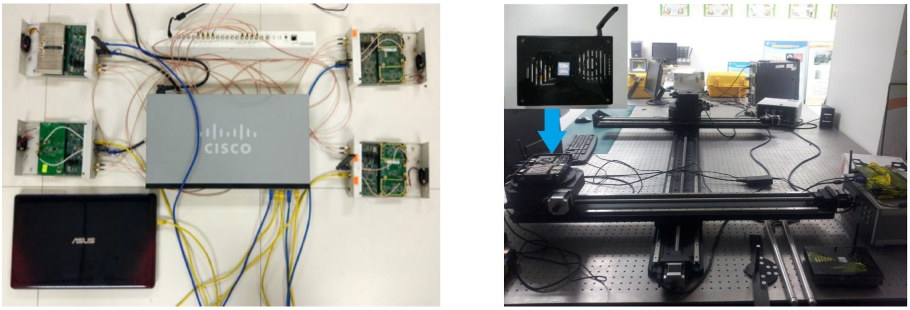
We have implemented TAB on USRP-N210 and USRP-X310 radio platforms with corresponding UHD software packages, as shown in Fig. 6(a). An AP with multiple antennas is built with one USRP-X310 plus multiple SBX daughterboards. Each concurrent user is a USRP-N210 equipped with a SBX daughterboard, providing 40 MHz bandwidth. To allow multiple users to transmit concurrently, we connect the USRP-N210 devices to a laptop and control their transmissions by an instruction script. A similar instruction script is also installed in the two SBX daughterboards of USRP-X310. For precise time synchronization, an external clock model is adopted as a common clock source to connect the AP and users.

The entire system is compatible with the 802.11ac/ax protocol, QUICK protocol and standard OFDM-MIMO specifications, including the modulations (16-QAM, 64-QAM, 256-QAM) and code rates. Hence, TAB can be easily implemented on Commercial Off-The-Shelf Network Interface Cards (COTS NICs) without hardware modification, such as Intel 5300 NIC. For the user, it's supported by the NICs that the device could calculate the CSI upon receiving NDP from AP, and place the CSI into data field of a frame and feed the frame back to the AP. For the AP, constellation diagram is supported for NICs to decode received symbols. Given essential information, we could implement TAB on the COTS NICs through programming its network card driver. In addition, we simulate the different SNR conditions by tuning the transmission power within the range of [5 dBm, 20 dBm].

TAB implements OFDM modulation, packet detection, channel estimation and symbol demodulation. We use LabVIEW to achieve OFDM and the channel estimation in the prototype system based on USRP. Besides, we have obtained some micro-benchmark results and the overall performance using Intel 5300NIC with 4-axis motion testbed in Fig. 6(b). Moreover, to verify our methods, we examined TAB in three different scenarios with general multi-path as shown in Fig. 7: (1) corridor, (2) office and (3) workshop.

5.2 CSI Reusability Within CCT

In this section, we evaluated the effectiveness of reusing CSI within CCT. To illustrate this point, we first verify the performance of TAB in reducing CSI feedback overhead. Figure 8 show the average number of CSI feedback times for continuous 100 rounds of UL and DL MU-MIMO transmission in TAB, respectively. Figure 8(a) shows that the average number of CSI feedback times in UL increases with the growth of the number of users and AP's antennas. Obviously, feedback overhead can be significantly reduced by leveraging CSI reuse, as they



(a) USRP with LabVIEW (b) 5300NIC on 4-axis testbed

Fig. 6. TAB Prototype system and experimental devices

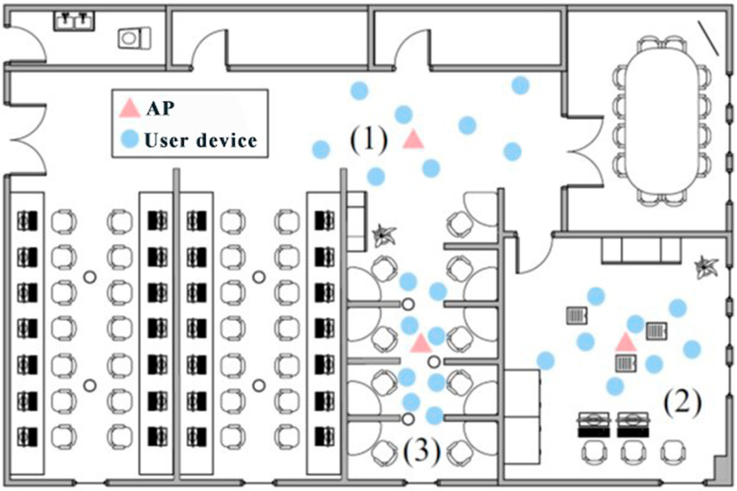
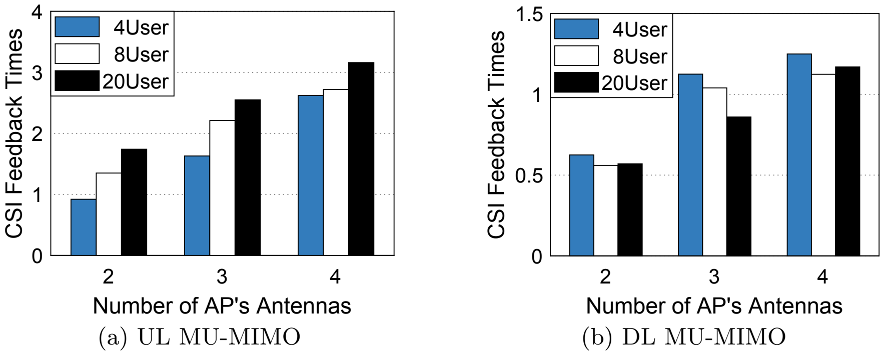


Fig. 7. Distribution of devices in different experimental scenarios

only need less than 4 times of feedback for 100 rounds. When the user population reached 5 times of the number of AP's antennas, the average CSI feedback times would be close to 70% of the number of AP's antennas, which means the performance of TAB begins to degrade to the traditional MU-MIMO network. Figure 8(b) shows that the average number of CSI feedback times in DL is even smaller. It slightly increases while the number of AP's antenna increases, finally fluctuating around 1. That means TAB has a great performance in reducing CSI feedback for both UL and DL MU-MIMO transmission.

Longer CCT does not mean higher reusability of CSI, which is also influenced by distributed user selection and scheduling. To measure CSI utilization, we introduce "Packet Length Per CSI" (PLPCSI) as the indicator to illustrate the average packets length decoded by different users' CSI in MU-MIMO net-



(a) UL MU-MIMO (b) DL MU-MIMO

Fig. 8. The count of CSI feedback for continuous 100 rounds of transmission in TAB.

work. Figure 9 presents the PLPCSI values for each user after 100 consecutive concurrent transmissions in a 8-user MU-MIMO network: (a) 2-AP scenario; (2) 4-AP scenario. Since the 802.11ac has specified the maximum data packet length, the PLPCSI of traditional network is not more than 5.5 ms. In TAB, all the PLPCSI values of different users have been improved to varying degrees. Moreover, when faced with the same scale of users, the PLPCSI values will increase significantly with the growing of AP's antenna, *i.e.*, CSI utilization is higher.

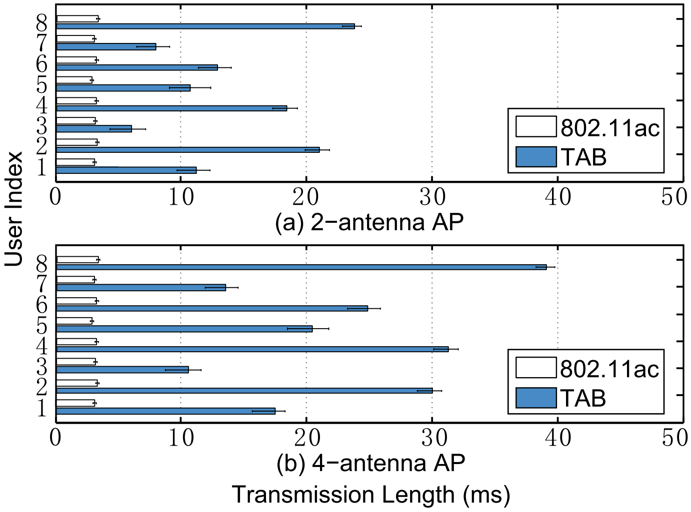
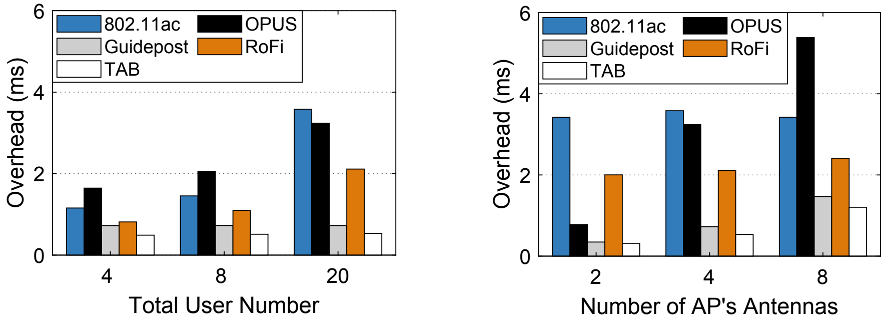


Fig. 9. PLPCSI values for each user after 100 consecutive concurrent transmissions.

5.3 Compare Channel Sounding Overhead

For Channel Sounding overhead comparison, we compare TAB with standard 802.11ac, OPUS [18], Guidepost [22] and RoFi [11]. Figure 10(a) illustrates the average Channel Sounding overhead during each round of transmission under different user population. We conduct these experiments with 4-antenna AP. While the overhead in standard 802.11ac and OPUS increases with the increase of total user number, the overhead in Guidepost and TAB is almost constant. In a topology with 20 users, it achieves almost 7.5×, 6.6×, 1.5× and 4.5× overhead decrease over 802.11ac, OPUS, Guidepost and RoFi, respectively. Obviously, TAB outperforms all other schemes. We run a benchmark scheme with 20 users to validate the effect of the number of AP's antenna on overhead, and plot the result in Fig. 10(b). It is clearly that the overhead increases when the number of AP's antenna grows in all these four MU-MIMO systems. However, the growth rate in TAB is still in a tolerable range. Hence, TAB can be scalable easily.



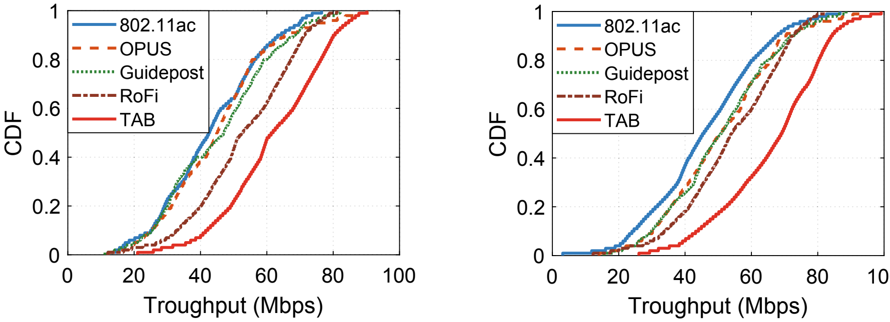
(a) Different scales of user population (b) Different numbers of AP antenna

Fig. 10. Average channel sounding overhead analysis

5.4 Compare Throughput Under Accurate CSI

In this micro-benchmark, we compare the throughput of different MU-MIMO systems, as shown in Fig. 11 and 12.

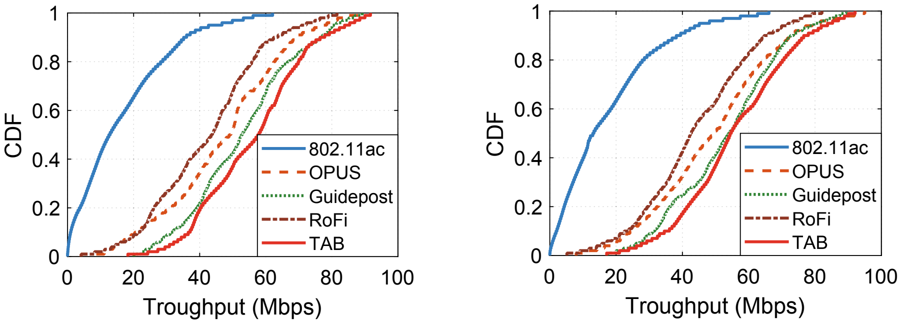
For DL MU-MIMO shown in Fig. 11, Standard 802.11ac, OPUS and Guidepost have almost identical throughput. Since there is no need of user selection, only concurrent users need to report their CSI. RoFi save the CSI feedback overhead by mobility-awareness, which also work with a little higher throughput in mobile scenarios. TAB achieves the highest throughput among all schemes. It is because that each concurrent users within CCT can continuously receive data from AP without CSI feedback. Moreover, the performance of TAB in static scenario is better than it in mobile scenario due to longer CCT.



(a) Mobile scenarios (b)Static scenarios

Fig. 11. The throughput comparison among different systems in DL MU-MIMO transmission under different device motion states.

For UL MU-MIMO shown in Fig. 12, RoFi, OPUS, Guidepost and TAB have too much higher throughput than 802.11ac. In order to select the orthogonal users, AP needs all the users to report their CSI in 802.11ac. The four novel MU-MIMO protocol have reduced much channel sounding overhead and user access time to different extent. Especially, TAB utilize multi-user balanced scheduling method to provide higher CSI utilization and stable data transmission availability, so that the throughput of TAB is the 1.5 5 times of 802.11ac.



(a) Mobile scenarios (b) Static scenarios

Fig. 12. The throughput comparison among different systems in UL MU-MIMO transmission under different device motion states.

5.5 Compare Energy Consumption

In this micro-benchmark, we approximately evaluate the energy consumption about CSI feedbacks of TAB compared with four existing MU-MIMO systems, including standard 802.11ac, OPUS, Guidepost and RoFi, which have tried to reduce CSI feedback overhead except standard 802.11ac.

The current metric for energy efficiency is energy consumption per data bit (ECPDB). The ECPDB combines the energy consumption per bit for transmitting data with that for receiving, respectively, *i.e.*, $et\,(m)$ and $er\,(m)$ as using MCS index m [11]. For the Intel 5300 WiFi NIC 40 Hz as bandwidth, 800 ns as Guard Interval(GI), 16QAM as modulation mode and 360 Mbps as date rate, $er\,(28)$ =11 nJ/bit and $et\,(0)$ = 90 nJ/bit, while the size of packets is $size\,(p_i) = 1500$ bytes and the size of CSI is $size\,(csi_i) = 1664$ bytes, which are much longer than the size of control frame and probing packets, *i.e.* $size\,(ctr_i)$ and $size\,(pro_i)$. Hence, the energy consumption of CSI feedback would almost cost $90\times1664\times8/(90\times1664\times8+11\times1500\times8) = 90.1\%$ of total energy consumption for transmission in standard 802.11ac. Fortunately, TAB has significantly reduced the number of CSI feedbacks $\sum_{i=1}^{N} size\,(csi_i)$ as much as possible.

Figure 13 shows that the results of energy consumption in different methods under both mobile and static scenarios. It is clearly that OPUS, Guidepost, RoFi and TAB can reduce much higher energy consumption than the standard 802.11ac. Besides, OPUS and Guidepost have reduced CSI feedback and even do not need CSI feedback, but they have to add some other information and steps to represent the CSI report. Moreover, zero CSI cannot profile the channel state accurately which degrades the efficiency of data transmission. Thereby RoFi and TAB have consumed quite less energy than them based on the direct reduction of CSI feedback times. In addition, the fluctuation of the energy consumption in TAB is lower than all other schemes, benefiting from the comprehensive and balanced guarantee for the channel efficiency.

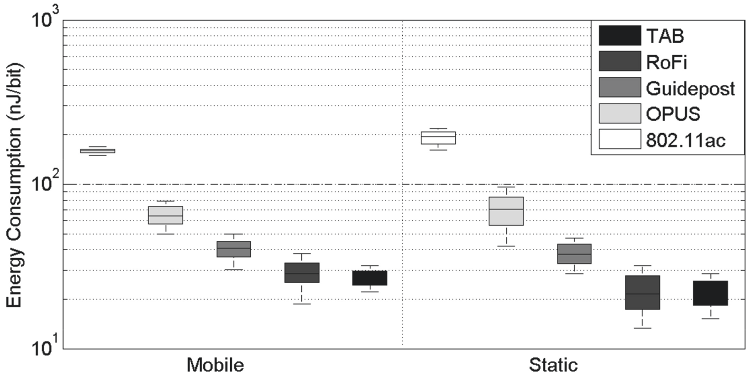


Fig. 13. Energy consumption analysis of different MU-MIMO systems.

6 Conclusion

Our work demonstrates that the unnecessary CSI feedback from concurrent users will lower the throughput in MU-MIMO network, due to the high superfluous overhead. Hence, we propose TAB to reduce unnecessary channel overheads for MU-MIMO WLANs. TAB designs a CSI timeliness-awareness balanced mechanism to determine whether a user's scheduling and CSI feedback is necessary, which significantly improve the CSI utilization and channel throughput. We implement the prototype TAB over software-radio devices. Extensive experiment results show that TAB can substantially improve the throughput for MU-MIMO networks and reduce superfluous resource consumption.

References

1. Afaqui, M.S., Garcia-Villegas, E., Lopez-Aguilera, E.: IEEE 802.11ax: challenges and requirements for future high efficiency WiFi. IEEE Wirel. Commun. **24**(3), 130–137 (2017)
2. Cai, Y., et al.: CSI feedback reduction by checking its validity period: poster. In: Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom 2016), pp. 469–470. ACM (2016)
3. Crepaldi, R., Lee, J., Etkin, R., Lee, S.J.: CSI-SF: estimating wireless channel state using CSI sampling & fusion. In: Proceedings of the 32nd IEEE Conference on Computer Communications (INFOCOM), pp. 154–162 (2012)
4. Duplicy, J., et al.: MU-MIMO in LTE systems. EURASIP J. Wirel. Commun. Netw. **2011**(1), 1–13 (2011). https://doi.org/10.1155/2011/496763
5. Gast, M.: 802.11ac: A Survival Guide. O'Reilly Media, Sebastopol (2013)
6. Halperin, D., Hu, W., Sheth, A., Wetherall, D.: Tool release: gathering 802.11n traces with channel state information. ACM SIGCOMM Comput. Commun. Rev. **41**(1), 53–53 (2011)
7. Fang, J., Wang, L., Qin, Z., Liu, J., Lu, B.: GCC: group-based CSI feedback compression for MU-MIMO networks. Mobile Netw. Appl. **23**(3), 407–418 (2018). https://doi.org/10.1007/s11036-018-1015-1
8. Jiang, Z.P., et al.: Communicating is crowdsourcing: Wi-Fi indoor localization with CSI-based speed estimation. J. Comput. Sci. Technol. **29**(4), 589–604 (2014). https://doi.org/10.1007/s11390-014-1452-7
9. Li, S., Li, D.X., Zhao, S.: 5G Internet of Things: a survey. J. Ind. Inf. Integr. **10**, 1–9 (2018)
10. Ma, Y., Gang, Z., Shan, L.: ELiMO: eliminating channel feedback from MIMO. In: 2017 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 1–8 (2017)
11. Ma, Y., Zhou, G., Lin, S., Chen, H.: RoFI: rotation-aware WiFi channel feedback. IEEE Internet Things J. **4**(5), 1684–1695 (2017)
12. Mukherjee, A., Zhang, Z.: Channel State Information Compression for MIMO systems based on curve fitting, pp. 1–9 (2016)
13. Shen, W., Lin, K.C., Chen, M., Tan, K.: SIEVE: scalable user grouping for large MU-MIMO systems, pp. 1975–1983 (2015)
14. Sur, S., Pefkianakis, I., Zhang, X., Kim, K.H.: Practical MU-MIMO user selection on 802.11 ac commodity networks. In: Proceedings of the 22nd ACM Annual International Conference on Mobile Computing & Networking (MobiCom 2016), pp. 122–134. ACM (2016)

15. Tang, Z., Qin, Z., Zhu, M., Fang, J., Ma, H.: TOUSE: a fair user selection mechanism based on dynamic time warping for MU-MIMO networks. KSII Trans. Internet Inf. Syst. **11**(9), 4398–4417 (2017)
16. Teng, W., Zhang, X.: Random access signaling for network MIMO uplink. In: IEEE Conference on Computer Communications (INFOCOM), pp. 1–9 (2016)
17. Xi, W., Ma, R., Cai, Y., Zhao, K.: Prevent CSI spoofing in uplink MU-MIMO transmission. In: Proceedings of the 1st Workshop on Context Sensing & Activity Recognition (CSAR 2015), pp. 13–18 (2015)
18. Xie, X., Zhang, X.: Scalable user selection for MU-MIMO networks. In: Proceedings of the 33rd IEEE Conference on Computer Communications (INFOCOM), pp. 808–816 (2014)
19. Xie, X., Zhang, X., Sundaresan, K.: Adaptive feedback compression for MIMO networks. In: Proceedings of the 19th Annual International Conference on Mobile Computing & Networking (MobiCom 2013), pp. 477–488 (2013)
20. Zhang, S., Xi, W., Xu, Q., Zhao, K., Cai, Y.: Accurate CSI estimation to eliminate unnecessary transmission for MU-MIMO networks. In: Proceedings of the 2019 International Conference on Embedded Wireless Systems and Networks (EWSN 2019), pp. 166–177. ACM, Junction Publishing, USA (2019)
21. Zhang, X., Sundaresan, K., Khojastepour, M.A.A., Rangarajan, S., Shin, K.G.: NEMOx: scalable network MIMO for wireless networks. In: Proceedings of the 19th Annual International Conference on Mobile Computing & Networking (MobiCom 2013), pp. 453–464. ACM (2013)
22. Zhou, A., Teng, W., Zhang, X., Ma, H.: Guidepost: scalable MU-MIMO user selection via indirect channel orthogonality evaluation. IEEE Trans. Mob. Comput. **18**(7), 1556–1570 (2018)
23. Zhou, A., Wei, T., Zhang, X., Liu, M., Li, Z.: Signpost: scalable MU-MIMO signaling with zero CSI feedback. In: Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2015), pp. 327–336 (2015)