TRINITY: A Practical Transmitter Cooperation Framework to Handle Heterogeneous User Profiles in Wireless Networks

Shailendra Singh
University of California, Riverside
singhs@cs.ucr.edu

Karthikeyan Sundaresan
NEC Labs America, Inc.
karthiks@nec-labs.com

Srikanth V. Krishnamurthy
University of California, Riverside
krish@cs.ucr.edu

Xinyu Zhang
University of Wisconsin-Madison
xyzhang@ece.wisc.edu

Amir Khojastepour
NEC Labs America, Inc.
amir@nec-labs.com

Sampath Rangarajan
NEC Labs America, Inc.
sampath@nec-labs.com

ABSTRACT

To handle increased capacity demands, sophisticated MIMO-based transmission strategies, based on transmitter cooperation, have emerged. However, different types of users' channels (e.g., static vs mobile, stable vs dynamic channels) that make up today's enterprises, require different MIMO transmission strategies. With the wrong strategy, a user could even see a degradation in performance. Our overarching goal is to design and implement a framework, TRINITY, that can simultaneously cater to a heterogeneous mix of users, by intelligently combining a plurality of MIMO transmission strategies wherein the transmitters at different nodes can cooperate to deliver significant performance gains. Three key challenges that we address in building TRINITY are: (i) how to categorize users into channel profiles such that a single transmission strategy caters to the users of a profile, (ii) how to combine strategies to communicate with users of different profiles simultaneously, and (iii) what is the granularity of transmitter cooperation needed to balance efficiency with complexity. We implement and evaluate TRINITY on our WARP radio testbed. Our extensive experiments show that TRINITY's intelligent combining of transmission strategies improves the total network rate by 50%-150%, satisfies the QoS requirements of thrice as many users, and improves PSNR for video traffic by 10 dB compared to individual transmission strategies.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication

General Terms

Algorithms, Experimentation, Measurement, Performance

MobiHoc'15, June 22–25, 2015, Hangzhou, China.

http://dx.doi.org/10.1145/2746285.2746321.

Keywords

Network MIMO, Distributed Antenna System (DAS), Transmitter Cooperation, Heterogeneous user profiles

1. INTRODUCTION

Enterprises previously dominated by wireless static users with devices like laptops are increasingly seeing mobile users, thanks to smartphones and Bring-Your-Own-Device (BYOD) initiatives. At the same time, wireless transmission strategies have evolved to cope with increased capacity demands. Specifically, the past decade has seen the emergence of sophisticated MIMO-based strategies.

While standards have emerged for both single user (802.11n) and multi-user (802.11ac) MIMO, they still only dictate transmissions within a cell. Recent research has led to networked MIMO systems, wherein the APs of different cells can cooperate and transmit simultaneously. With *network multiplexing*, such cooperation can control interference to deliver a scalable, network-wide multiplexing gain (scales with the number of transmitters [16]). With *network diversity*, the cooperation could lead to higher robustness to fading and mobility due to a diversity gain. Transmission cooperation could lead to significant throughput gains over conventional systems that simply reuse the spectrum (e.g., using CSMA).

Different types of users need different strategies: However, harnessing the potential gains from transmission cooperation in a heterogeneous setting is not easy. While network multiplexing (e.g., using netMIMO [16]; Fig. 1c) supports concurrent data flows from transmitters in close proximity (e.g., APs of neighboring cells) and could thus, maximize throughput, it can only apply to static users. This is because, such multiplexing can only be realized if the entire process (eg. for netMIMO) of estimating channel state information (CSI) at a receiver, feeding back the CSI to the transmitter and subsequently, transmitting the data are completed within the coherence time (the time after which the CSI can be expected to change) of a channel. Applying netMIMO for mobile users or for users with short coherence times, can even hurt performance by as much as 70% for a 3-user network as shown later. For mobile users, one can conceivably apply network diversity, where there is no need for CSI feedback; here, multiple versions of the same data stream are sent from the different transmitters (e.g., distributed antenna system or DAS technology or space-time codes; Fig. 1b). The diversity gain increases the reliability of the transmission as the channel state changes rapidly. It also provides a much larger transmission footprint (at the expense of spatial reuse) that allows mobile users to receive data reliably as they move around without handoffs.

Note here that static clients would obtain much lower throughputs with network diversity, than that with netMIMO.

With both network multiplexing and network diversity we veer from traditional WiFi access (e.g., CSMA) by having the base stations (each possibly with multiple antennas) in close proximity coordinate their transmissions. However, there could be static clients that experience low coherence times due to the dynamics of the environment they are in (e.g., human motion, closing/opening of doors). For such clients, using network diversity that limits spatial reuse is an overkill since they do not move from the footprint of one transmitter to that of another. On the other hand, one cannot apply network multiplexing since the CSI changes rapidly. For such clients, the use of traditional CSMA type spectrum reuse policies seems to be the best way to go as shown via experiments in Section 2.2.

Classification of users and strategies: To cater to a mix of users with diverse channel and mobility characteristics (referred to as user/channel *profiles*), our thesis is that one must tailor transmission strategies to user profiles towards maximizing performance and improving user quality of experience. The users in an enterprise network can be categorized as follows based on the above discussion. First, they can be categorized into those with large and small channel coherence times. Further, among those with small coherence times, they can be classified based on the contributing factor - user mobility or environment dynamics. Similarly, the gamut of wireless transmission strategies can be grouped into three distinct categories based on transmitter cooperation and CSI availability. At the top level, we have those that allow transmitter cooperation and those that do not. We refer to the latter as simply Reuse strategies (eg. CSMA). Schemes that enable transmitter cooperation can be further classified into (a) *network multiplexing* requiring CSI feedback (e.g., netMIMO, MU-MIMO), and (b) *network diversity* without requiring CSI feedback (e.g., space-time block codes, DAS etc.). Note that this classification subsumes conventional diversity-multiplexing strategies (e.g., 802.11n) in single transmitter settings (the multiple antennas are treated as separate transmitters). With real-world experiments, we show that applying the right strategy for each user profile is important not just for improving performance but more importantly to prevent any performance degradation (from applying the wrong strategy).

Challenges: Many challenges arise in ensuring that the right strategy is applied to each user profile in practice: (i) What is the overhead in obtaining timely CSI ? (ii) Can CSI by itself suffice in categorizing users into the different profiles or are other forms of feedback needed ? (iii) Given the high CSI overhead of netMIMO, how can one scale transmission cooperation to a large network ? (iv) How can a transmitter simultaneously apply the different transmission strategies in the presence of a heterogeneous mix of users ?

Contributions: In this paper, we design and implement a practical framework, TRINITY, that can (i) accurately categorize users into the various profiles, (ii) combine multiple strategies at transmitters to effectively cater to users of various profiles simultaneously, and (iii) determine the appropriate granularity of transmitter cooperation to balance performance against overhead and complexity. Briefly, TRINITY is deployed at a central controller (managing enterprise networks) and incorporates three key design elements.

1. It enables simultaneous operation of all strategies by intelligently combining them in either frequency or time. We focus on the frequency domain since it allows for power pooling [9] benefits not available with time domain operations. Moreover, most of today's networks (802.11n, WiMAX, LTE) are based on OFDM; the available OFDM spectrum units (called *sub-carriers*) are split between different strategies. However, note that the design can be trivially modified to combine the strategies in the time domain.

2. The resources (e.g., # sub-carriers) allocated to a strategy depends on the traffic load of the corresponding user profile and is integrated with a user categorization process. A measurement (CSI and SNR) driven approach coupled with sensor hints (e.g., accelerometer [17]), is used to accurately categorize users into profiles.

3. To optimize network performance, we propose a novel scheme for clustering transmitters that addresses the performance-complexity trade off with network multiplexing and the coverage-capacity trade off with network diversity, in large networks.

We implement and evaluate TRINITY via extensive experiments on our 13 node WARP [4] testbed and with large-scale simulations. We see that TRINITY categorizes users with 90-95% accuracy. Tailoring transmission strategies to heterogeneous user profiles at finer time-scales compared to application layer dynamics, TRINITY can benefit both throughput and delay sensitive applications. It improves network rate by 50%-150%, satisfies QoS requirements for *thrice* as many users, and improves PSNR for video traffic by 10 dB, compared to individual transmission strategies.

2. APPLYING THE RIGHT STRATEGY

We first provide a primer on transmission cooperation strategies. Later, we determine what strategy to use for each user profile.

2.1 Classification of Strategies

Network Multiplexing: With network multiplexing, multiple independent data streams are transmitted concurrently from multiple cooperating APs to different users by converting interference into a multiplexing gain; an example of this is network (distributed) MIMO [16]. The data streams for different users are shared at all APs, which are tightly synchronized (at the carrier signal level). From the PHY layer perspective, this can be realized using a data encoding (called precoding) scheme called zero-forcing beamforming (ZFBF) [23]; this applies a precoding matrix (computed from inverse of the channel matrix between APs and the clients) to send a linear combination of the data streams through each AP, such that unwanted streams (interference) cancel out at each client, leaving only the desired stream. A simple scenario is shown in Fig. 1c. Network MIMO (netMIMO) allows the capacity to scale *linearly* with the number of cooperating APs. However, this comes at the cost of tight phase synchronization across APs and overhead caused by clients' channel state information (CSI) feedback to the APs. We refer to variants of closed-loop (requiring CSI) precoding schemes (e.g., netMIMO, adaptive beamforming, MU-MIMO) both across and within APs collectively, as *network multiplexing* schemes.

Network Diversity: These schemes send multiple, coherent versions of a data stream via multiple transmitters to provide transmit diversity. An example is the use of distributed Alamouti space-time (ST) codes [23]; when APs 1 and 2 employ the 2×1 Alamouti code to jointly transmit to client 1, the resulting diversity gain scales the SNR at the client by the factor $|h_{11}|^2 + |h_{21}|^2$, where h_{ij} is the complex channel gain between AP i and user j. A simpler form of transmit diversity is to transmit the same data (symbols) from multiple transmitters as shown in Fig. 1b, wherein client 1 receives a combination of the streams transmitted from all APs over a composite channel $\tilde{h} = h_{11} + h_{21} + h_{31} + h_{41}$. The power pooled from the multiple transmitters contributes to a combining (SNR) gain on average. This form of network diversity is similar in principle to broadcast and is referred to as a distributed antenna system (DAS). Unlike ST diversity, DAS does not require a receiver to *estimate* the state with respect to the individual channels from the different transmitters (no associated pilot overhead). This has made it highly popular for deployments in both WiFi and cellular frameworks [1].

These schemes not only improve coverage and robustness - factors critical for handling mobility, but also provide a diversity (SNR)

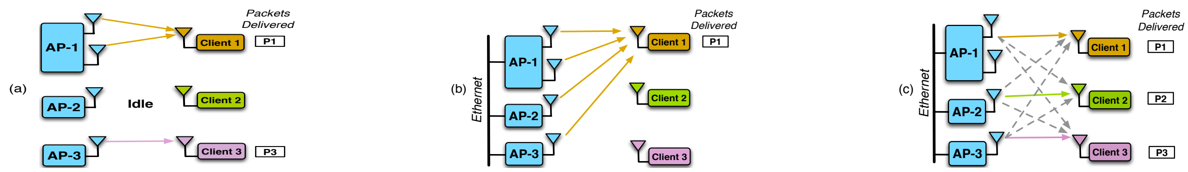


Figure 1 (a) Reuse: Two slot schedule: AP1, AP3 active simultaneously in slot 1, AP2 in slot 2. (b) DAS: 3 slot schedule - 3 APs sending same data to one client in each slot. (c) netMIMO: 3 APs sending 3 data streams to 3 clients in each slot.

gain that translates to a *logarithmic* increase in capacity. Further, they do not need timely CSI feedback (from the clients to the APs). We refer to such schemes broadly as *network diversity* schemes.

Spectrum reuse: In addition to network multiplexing and diversity, it is also important to consider *non-cooperating* transmitters. This includes conventional time-divisioned access (e.g., CSMA) wherein, APs defer transmissions in the presence of interference (a busy channel) while spatial reuse is automatically leveraged otherwise. In the example in Fig. 1a , APs 1 and 3 can transmit in tandem on the same channel, while AP 2 time shares the medium with 1 and 3. Such schemes (e.g., distributed CSMA or centralized scheduling [19]) will collectively be referred to as *Reuse* schemes.

2.2 Guidelines for Strategy Selection

Users in enterprise networks can be categorized into one of the following three categories based on their mobility and channel coherence characteristics: **(i)** *Mobile users with Short Coherence Times (Mobile)*: Coherence time (Tc) varies based on the speed of a client which can range from walking speeds of 3-4 Kmph (Tc = 10 ms) to vehicular speeds of 75 Kmph[1] (Tc = 1.1 ms). **(ii)** *Static Users with Short Coherence Time (Short-Tc)*: Static users can also experience short Tc (10-20 ms or less) due to dynamic settings where objects or other users move. **(iii)** *Static Users with Long Coherence Time (Static)*: Clients in static settings in the absence of mobility, experience a more stable channel (longer Tc, 100 ms or more).

Experimental Study Next, we perform experiments to determine the appropriate transmission strategy for each user category; our studies here form the basis for TRINITY's design. We will use CSMA (a Reuse strategy), DAS and netMIMO as the representative transmission strategies. Our experimental set-up (more details in Section 5) shown in Figures 1a, 1b and 1c is deployed in an indoor lab with three transmitters and three clients using a WARP testbed. For all mobile experiments, we placed the WARP radios on a cart and moved it on the same path at the same walking speed. Aggregate network rate (throughput normalized to unit bandwidth, bits/s/Hz) is used as the metric and depends on both client SNR received as well as the schedules indicated in Figs. 1a, 1b and 1c. The reported results are averaged over multiple runs.

Static Clients: From Fig. 2a, we see that *when all the clients are static, netMIMO is the most appropriate strategy*, outperforming Reuse and DAS by up to 69.8%. It achieves high network rate by multiplexing three data streams to three clients at the same time. Since the channel coherence time in our static environment is large, all clients can decode their data with high reliability. Further, *since the benefit of spatial reuse (absent in DAS) outweighs the diversity gain in static environments, Reuse performs better than DAS.*

Mobile Clients: However, as we vary the number of mobile clients in the network, the performance of both netMIMO and Reuse start to degrade; the degradation for netMIMO is especially severe. On the other hand, *owing to coverage and combining gain in SNR from three transmitters without reliance on any CSI, DAS is unaffected by client mobility*. Thus, DAS outperforms netMIMO and Reuse by up to 96.7% depending on the number of mobile clients.

[1]Vehicular speeds are considered as our schemes are also applicable to outdoor small cell (eg. LTE) mobile networks.

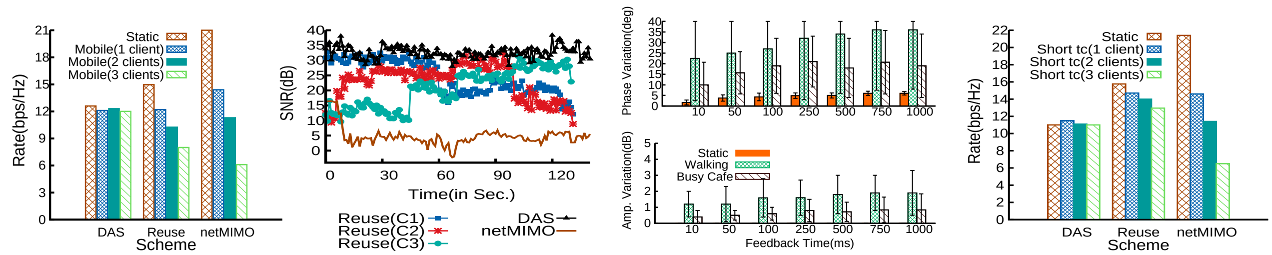
To better understand the behavior of these transmission schemes, we also recorded the received SNR of the symbols from each transmitter. Fig. 2b shows the average SNR of the mobile clients under Reuse, DAS and netMIMO over multiple runs. For DAS and netMIMO, we only report the SNR values of a single client; other clients also exhibit a similar trend. In the Reuse scheme, client-1(C1) is associated with AP-1, client-2(C2) with AP-2, client-3(C3) with AP-3. During experiments, client-1 is moved from left to right while client-3 is moved from right to left. Client-2 is moved in both directions. It can be inferred from Fig. 2b that each client in Reuse, experiences a high SNR only when they are near their respective APs. In DAS, a mobile client experiences a high SNR throughout the experiment due to the coverage and the signal combining effect from the three transmitters, with about 5-6 dB of SNR gain over the highest SNR possible with Reuse (with ideal handoffs).

Why not Reuse or netMIMO for Mobile Clients? *In addition to link degradation, mobility also impacts the benefits of spatial reuse in Reuse*. Even for centralized schemes [19] to leverage reuse, interference conflicts have to be determined (from transmitters to various clients) to establish appropriate reuse schedules - a process that is executed only once every several packets to reduce overhead/complexity. In the presence of client mobility, tracking interference conflicts of mobile clients accurately is not practical; this in turn affects the Reuse performance with respect to such clients.

In netMIMO, the SNR drops if the client becomes mobile. We see in Fig. 2c that unlike for a static client, there is a large variation in channel phase (20°-40°) and magnitude (1-2 dB) for a walking client even with a CSI feedback rate of once every 10 ms. Since precoding depends on CSI to remove interference between concurrent streams, *stale CSI from such variations from mobility has a bigger impact on netMIMO performance*. Further, if there are N clients/AP, M APs and S sub-carriers in a OFDM transmission, to apply netMIMO, one needs to measure channel qualities from each of the M APs (sequentially, M transmissions) to all the $M \cdot N$ clients on all the S sub-carriers ($S \cdot M \cdot N$ channels). This has to be then fed back from each of the clients ($M \cdot N$ transmissions) to construct the channel matrix required by the precoding algorithm. Further, a few measurements per client are typically needed for stable channel estimates. Thus, it is hard to accomplish this entire process within the channel coherence times of even walking clients.

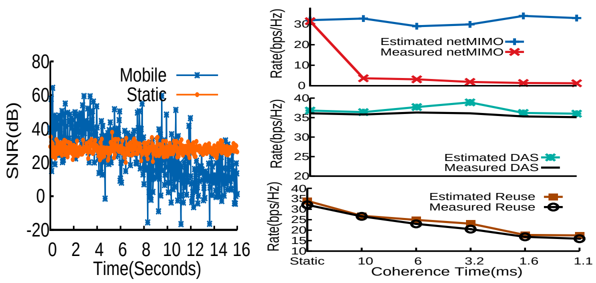
The expanded coverage along with combining gain in SNR, without reliance on any CSI, makes DAS the strategy that is least impacted by (and thus, best suited for dealing with) client mobility.

Short-Tc Clients: A static client can also experience fluctuating channel conditions due to environment dynamics etc. This can result in small channel coherence times ($\approx$ 10 ms) even if the client is static as seen in Fig. 2c, where a static client is in a busy indoor cafe. Fig. 2d shows the aggregate network rate achieved by each scheme with an increasing number of such static clients. With phase variations as high as 20°, even with a CSI feedback period of 10 ms, netMIMO is highly susceptible to stale CSI and the achieved rate degrades in a manner similar to that in the mobile scenario. However the degradation in Reuse is now less pronounced. While link quality is impacted, since the topology (interference pattern) does not change, spatial reuse is not impacted and thus, Reuse outperforms DAS (reuse gains will increase with topology size).



(a) (b) (c) (d)

Figure 2 : (a) Static vs. Mobile. (b) Impact of Mobility. (c) Channel variations. (d) Static with Short Tc.



(a) (b)

Figure 3 : (a) SNR: Mobile Vs. Static. (b) Estimated vs. Measured.

Hence, interestingly, *a simple Reuse scheme is the best for static clients that experience small coherence times*.

Our work in perspective: Our strategy classification is based on core features related to transmitter cooperation and CSI availability and is generic enough to accommodate most open-loop and closed-loop transmission schemes. A few important remarks are in order. **(A)** Our classification of multiplexing, diversity and reuse, is from the *network* perspective, accounting for transmitter cooperation; it subsumes conventional multiplexing-diversity schemes applied at a single transmitter. When transmitters and clients have multiple antennas themselves, each transmitter in our network-level open-loop (no CSI) schemes (Reuse and network diversity) can in addition apply 802.11n's (MIMO) open loop multiplexing/diversity. **(B)** The focus of our work is to exploit transmission cooperation to maximize the performance of the network as a whole, and not of specific individual clients; however, the latter also improves as a by-product. For example, for a static client with poor links to the multiple spatially dispersed transmitters, DAS may yield its highest individual throughput; however, DAS prevents the transmitters from multiplexing other clients and thus hurts aggregate network performance. **(C)** Network multiplexing only refers to the ability to multiplex multiple streams across different transmitters and does not preclude diversity. Indeed, by inherently accounting for client link qualities (based on CSI) [8], it determines the appropriate number of streams that can be sent to each client. If clients with poor but stable channels are chosen for the multiplexing session, then more antennas would automatically be used to generate fewer streams (i.e. providing diversity) to these clients, such that it maximizes the aggregate throughput performance of the session.

3. DESIGN OF TRINITY

In applying the right strategy to each user profile based on the aforementioned inferences, two key challenges arise in practice: (i) How to categorize users into various profiles? and (ii) How to intelligently combine various strategies to cater to a heterogeneous set of users simultaneously and manage resources effectively between strategies? We now present the key elements in TRINITY designed to address these challenges. TRINITY is implemented at a central controller that manages a set of distributed transmitters.

3.1 Categorization

While device-based categorization (smartphone ⇒ mobile, laptop ⇒ static) is simple, it does not capture a user's true channel state (e.g., laptop in a busy cafe). However, measuring the channel coherence time of a user directly is hard. Hence, TRINITY uses indirect indicators such as mobility cues (sensor/SNR hints) along with rate degradation (from CSI estimates) *jointly* for categorizing users.

3.1.1 Mobility Triggers

Sensor hints: Most mobile devices (smartphones/tablets) have GPS and accelerometers. While GPS provides a good estimate of user speeds outdoors (especially for vehicular users), accelerometer readings have been successfully used (e.g., [17]) to decipher user mobility indoors with high accuracy. When accelerometer data is available, TRINITY employs the approach in [17]. Briefly, for each new accelerometer sample, the standard deviation of the sample's magnitude is computed over a sliding window (w) of samples. If this standard deviation exceeds a threshold (a), movement is detected. If it is within the threshold for n successive windows, the device is deemed stationary. We use the values $w = 5$, $a = 0.15$ m/s^2, and $n = 10$; in [17] it was experimentally shown that with these values, device movement can be detected with an accuracy over 95%.

SNR hints: While sensor hints provide high accuracy, they may not be always available. In such cases, TRINITY uses SNR hints to decipher user mobility using an approach similar to that in [11]. Specifically, the variation in the average SNR of a client is used for this purpose. Consider a simple experiment where a mobile client is moving away from the AP at walking speed; we record the SNR variation that it experiences with time. We perform 20 such trials and the SNR trace from one such trial is compared against that of a static client in Fig. 3a. As expected, the SNR variation for a mobile client is high due to the dynamic environment. We find that this variation is consistently more than 10 dB in all 20 trials. Similar experimental results are also reported in [11]. Using this value as a threshold, we consider a user to be mobile (mobility hint) as follows: we compute the standard deviation (σ) over a moving window (2 seconds) of 10 measured samples of average SNR over a 200 ms (t_o) interval. If $\sigma > 10$ dB, the user is considered mobile.

Compared to sensor hints, SNR hints only provide an accuracy of just over 80%. Hence, TRINITY employs sensor hints for detecting user mobility whenever available and resorts to SNR hints only when the former is not present.

3.1.2 Sensitivity of netMIMO

Finally, netMIMO's sensitivity to fluctuating channels can also be effectively used for categorization. When channel states are measured from the transmitters to users for netMIMO, one can estimate the SINR and thus, the rate *expected* at users when netMIMO is applied (calculated using standard techniques [22]). When net-

MIMO is executed, the resulting SINR or rate can be *measured* and then compared against the estimated value. A significant drop in measured rate compared to the expected rate will result if netMIMO can no longer be supported by the user's channel. To understand the validity of our claim, we simulate a scenario with different channel coherence times to create mobility. Instead of transmitting the symbols over the air we pass them through a flat AWGN channel, so as to keep it constant over the coherence time (channel feedback rate is 20ms), after which it changes independently to a new realization. We find that (Fig. 3b) netMIMO rate drastically degrades for a client when it changes its state from being static to mobile even at walking speed. Unlike DAS or Reuse, the deviation between the estimated and measured rates is very large for netMIMO (more than 50% consistently) due to its strong reliance on CSI.

CSI alone is insufficient: Instead of using sensor/SNR hints, one might wonder if CSI estimates alone are sufficient for user categorization. Such an approach has two major shortcomings. First, it requires CSI estimation/feedback overhead for all users, regardless of whether or not they can support netMIMO; this will be prohibitive in terms of overhead. Second, CSI is limited to characterizing coherence times, but cannot attribute reasons to why coherence times are short; this is important to distinguish between DAS and Reuse.

3.1.3 Proposed Algorithm

TRINITY leverages the above two indicators *jointly* to categorize users with high accuracy as follows.

Initial Step 1: When a user joins the network, with sensor and/or SNR hints, it first determines if the user is mobile and if so categorizes it to be a DAS user. Otherwise, it aggressively assumes the user to be a netMIMO user.

Initial Step 2: During netMIMO operation, if there is a significant gap between the estimated (from CSI feedback) and measured netMIMO rates (by more than 50%) for the user, then the user is removed from the netMIMO category. If the user cannot support netMIMO and if its sensor and/or SNR hints do not indicate it to be a DAS user, only then the user is classified as a Reuse user.

Recurring Step: Since a user's profile can change temporally, categorization cannot be a one-time process. If the user is a netMIMO or Reuse user and if its profile degrades to a lower multiplexing (rate) category (netMIMO$\rightarrow$reuse$\rightarrow$DAS), then this would automatically be reflected in its rate (for netMIMO) and sensor/SNR hints (for Reuse). Using these triggers, the user can then be appropriately assigned to DAS or Reuse. To account for temporal changes, categorization is implicitly integrated within TRINITY.

3.2 Joint Application of Strategies

In reality, users of different profiles are inter-twined in various regions of the network. Hence, it is inevitable that we combine different strategies (at each transmitter) either in the time or frequency domain to serve such a mix of users.

TRINITY combines the strategies in the frequency domain. Thus, it can leverage *power pooling* benefits at the transmitters [9]. The sub-carriers in an OFDM system (say N in total) are split orthogonally (e.g., N_m to netMIMO, N_d to DAS, $N - N_m - N_d$ to Reuse; see Fig. 6) between the various strategies. Each transmitter then transmits with the appropriate subset of sub-carriers carrying data for users of the corresponding profile. On the downlink (AP$\rightarrow$users), when an AP (with fixed transmit power) has users that do not span *all* profiles, the power on the sub-carrriers assigned to the unused strategie(s) will be pooled to the sub-carriers assigned to the remaining strategies in operation. This results in a higher SNR on the sub-carriers in operation, a gain termed as *power pooling gain*. Note that when strategies are combined in the time-domain (which TRINITY can do), all sub-carriers are used for a given strategy at a time and thus there is no room for power pooling.

3.3 Resource Management

Once users are categorized, an estimate of the traffic loads from each category are collected by the central controller (from the APs). The different user profiles (i) are then weighted (v_i) based on the traffic load, and the allocation of number of sub-carriers to each of the strategies across the entire network is made proportional to their weights (e.g., $N_i = \frac{v_i N}{\sum_i v_i}$). Note that other weight choices based on priority/QoS can also be employed. Once the v_i's are determined, the sub-carrier allocation is done by the central controller, and is updated periodically at coarse time-scales to track traffic load variations. Among the sub-carriers allocated to a strategy, users in the respective profile are then scheduled based on a desired fairness model (proportional fairness in our case, details are in [7]). Since resource management is tightly coupled with user categorization, it continuously adapts to dynamics in user profiles. Once resources have been allocated to different strategies (user profiles), the next step is to determine how the transmitters serving users in different categories access these resources.

4. MEDIUM ACCESS IN TRINITY

While the Reuse strategy can be realized in a completely distributed manner, cooperation strategies like netMIMO and DAS require a central controller to manage synchronization, processing, etc. Thus, in TRINITY downlink transmissions are *scheduled* by its central controller. However, media access in TRINITY can also be easily realized in a quasi-distributed manner as discussed later in Section 4.2. Further, since the resources allocated to the strategies are orthogonalized in the frequency domain, it is sufficient to consider scheduling within each strategy in isolation. We later discuss (Section 4.4) how the various strategies are integrated at a transmitter.

4.1 Clustering for netMIMO and DAS

Need for Clustering: For netMIMO, the ideal operation would be to execute one large netMIMO between all the transmitters in the network and the users in the netMIMO category. While this would provide the maximum number of concurrent streams (scaling with the # of transmitters), this would also incur the overhead of synchronizing all the transmitters, measuring (and feeding back) CSI from all transmitters to all netMIMO users and sharing of all netMIMO users' data streams across all transmitters. Clearly this is not scalable beyond a few transmitters (as in [12, 16]).

Similarly with DAS, grouping all transmitters in one big DAS set-up eliminates the need to identify appropriate transmitters for various clients. However, the cost incurred is that the data on each sub-carrier is broadcast by all the transmitters, thereby significantly limiting reuse in the network. Given that users' mobility may be restricted to regions of the network for a given period of time, this could cause a gross under-utilization of the spectrum.

Approach: To address these issues, TRINITY groups the transmitters into smaller, contiguous clusters (e.g., Fig. 5), and netMIMO and DAS are applied only within each cluster. Interference between clusters is avoided either in the time or frequency domain. Applying cooperation strategies at cluster granularity ensures feasibility and scalability for netMIMO, and allows for spectrum reuse across clusters in DAS. The maximum cluster size (Q, decided based on practical considerations), and the clusters themselves can be different for netMIMO and DAS.

Clustering Algorithm: To establish clusters, we design a novel, topology-aware clustering algorithm for TRINITY. The underlying principle is that transmitters within each cluster should have strong mutual interference, in order to maximize the cooperation or com-

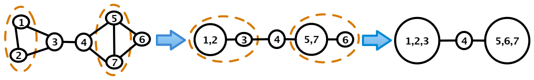


Figure 4 : Graph coarsening with $Q = 3$

bining gain. Moreover, transmitters across different clusters should be weakly coupled, so as to maximize frequency reuse opportunities. TRINITY adopts a graph coarsening approach to clustering (executed by the central controller), whereby neighboring vertices are consolidated (merged) recursively based on a desired metric.

Step 1: For each strategy, the entire network can be represented by a conflict graph $G(V, E)$, where transmitters constitute the vertices V and an edge $(u, v) \in E$ indicates interference (to clients) between transmitters u and v operating under the given strategy.

Step 2: For each randomly selected vertex v, the algorithm finds the neighbor u that with v, suspends the locally maximum *clique* (members interfere with each other). Then, it merges u, v and their edge into a single vertex u', which is added to the conflict graph.

Step 3: This process is applied recursively on the new conflict graph, until no more vertices can be merged or the cluster size constraint Q is reached. Each vertex in the resultant conflict graph contains multiple transmitters in the original network, which then form the desired clusters (Fig. 4).

TRINITY employs *maximum clique* as its coarsening metric as it creates clusters with good cooperation gain, while also reducing the interference across clusters. Note that the above conflict graph is determined at coarse time scales as it depends only on the AP layouts (which do not change) and aggregate client distributions around each AP (which only change gradually).

4.2 Interference Avoidance across Clusters

An inevitable feature of clustering in netMIMO and DAS is that transmitters on the edge of the clusters will receive interference from neighboring clusters, albeit only from transmitters using the same strategy. To eliminate such inter-cluster interference, TRINITY takes the following approach through resource allocation.

Step 1: It considers a conflict graph with clusters themselves as vertices (an edge exists between adjoining clusters) and performs a simple, (degree-based) greedy multi-coloring on the set of sub-carriers allocated to the strategy.

Step 2: The orthogonal sets of sub-carriers assigned to the clusters are then used by their cluster-edge (interfering) transmitters *only* to remove inter-cluster interference. *All* sub-carriers allocated to the strategy are used by the cluster-interior (non-interfering) transmitters, thereby allowing for efficient reuse of resources across all the clusters. In the example in Fig. 5, APs 3 and 4 are cluster-edge, interfering transmitters and hence operate on orthogonal sub-carriers (first 24 sub-carriers), while the other APs operate on all sub-carriers (see Fig. 6) for their netMIMO clients.

Note that with *Reuse*, there is no clustering (the cluster size is one with every transmitter being an edge transmitter).

Random access approach: While TRINITY adopts a scheduled approach, it can also use a distributed random access MAC to handle interference between clusters. Such a quasi-distributed approach would retain coordination only within the clusters (for applying cooperative strategies), while decoupling operations across clusters.

4.3 Framing Transmissions

TRINITY combines multiple strategies in the frequency domain at a transmitter, which amounts to transmitting to multiple clients in a single frame. This is the essence of OFDMA, which is popular in cellular networks (LTE, WiMAX). Like a conventional OFDMA frame, TRINITY's transmission frames are two-dimensional in time and frequency (time symbols and sub-carriers), and consists of a preamble followed by a control region. The control region contains information on which sub-carriers carry data for which user, and the MCS (modulation and coding rate) to be used for decoding the data. The controller ensures that data for users on a specific strategy are scheduled only on sub-carriers assigned to that strategy. Training or reference signals/pilots are embedded in the frame (similar to WiMAX/LTE, 802.11n/ac) to enable frequency synchronization, channel measurement and CSI reporting for netMIMO.

4.4 TRINITY: Summary of Operations

We summarize the operations in TRINITY with the simple example in Fig. 5; there are 6 APs and 9 clients, six of which (1,2,3,5,6,7) are static, two (4,8) are short-Tc and one (9) is mobile.

1. With sensor and/or SNR hints, an uncategorized user is first checked to see if it is a DAS user (client 9). Otherwise, it is assumed to be a netMIMO user. Channels are estimated for that user and netMIMO is then applied. Based on the estimated and observed rates for netMIMO operation, and coupled with sensor/SNR hints, the user is appropriately categorized as either netMIMO or a Reuse user (clients 1,2,3,5,6,7 are netMIMO; 4,8 are Reuse users).

2. Sub-carriers in the frequency domain are partitioned between strategies based on traffic loads of users in the associated categories, priority and fairness. (Assuming identical traffic load for all clients, netMIMO, Reuse and DAS strategies receive 24, 8 and 4 sub-carriers respectively out of a total of 36 sub-carriers.)

3. For a given cluster size constraint ($Q_{netMIMO} = 3$, $Q_{DAS} = 6$), transmitters are clustered for netMIMO and DAS, using the heavy clique graph coarsening approach (two clusters for netMIMO and one for DAS). Transmitters in the cluster interior (APs 1,2,5,6 for netMIMO; all APs for DAS) use all sub-carriers assigned to the strategy; those at the cluster edge operate on a subset determined by multi-coloring to avoid interference (for netMIMO, APs 3 and 4 operate on orthogonal sets of 12 sub-carriers out of the allocated 24). For Reuse, transmitters use multi-coloring to determine their frequency resources (APs 2 and 5 do not interfere and use all 8 sub-carriers assigned for Reuse). Clients are then scheduled at each transmitter for each strategy on the appropriate set of sub-carriers.

4. Data for multiple clients belonging to different profiles are combined in the frequency domain and transmitted simultaneously from each AP (frame transmissions at each AP are shown in Fig. 6).

5. While netMIMO users are re-categorized automatically when their channel state changes to one with low multiplexing gain, Reuse and DAS users are explicitly moved to netMIMO periodically to see if they have moved to a state with a higher multiplexing gain.

5. IMPLEMENTATION AND TEST-BED

5.1 Implementation

We prototype TRINITY on the WARP (FPGA-based) SDR platform using the WARPLab framework. Multiple WARP boards can be connected to a host PC via an Ethernet switch. The host PC performs the (de-) construction and processing of baseband waveforms transmitted (received) to (from) the FPGA board with the RF front-end. TRINITY and all its components are executed on the Host PC, which also serves as the central controller. Single antenna WARP boards are used for both the AP and clients.

Transmission Strategies: Our PHY layer implementation for Reuse is a centralized version of 802.11g with standard (de-) modulation techniques with 64 sub-carriers per OFDM symbol. When transmitters apply netMIMO, the central controller performs a linear combination of the BPSK symbols (zero-forcing beamforming precoding [23]) to be sent by the different transmitters so as to precancel their mutual interference. The channel matrix, whose inverse provides the precoding matrix, is obtained before netMIMO

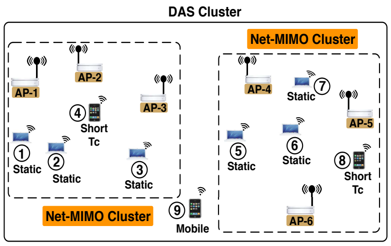


Figure 5 : Sample network topology

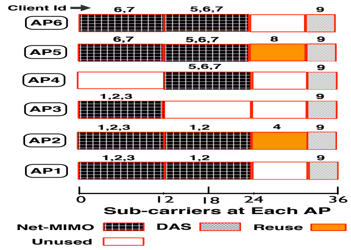


Figure 6 : Execution of Trinity

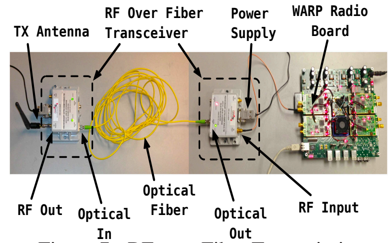


Figure 7 : RF over Fiber Transmission

Frequency	2.48GHz	Bandwidth	10MHz
Feedback Rate	50ms	Symbols/Pkt	1024
Symbol Time	$20\mu s$	Modulation	BPSK

Table 1 : Physical layer parameters

operation from the CSI piggy-backed (by clients) on the most recently received ACK packets. Each receiver estimate its CSI based on pilots from the last few OFDM symbols received in the previous frame. Due to the nature of WARPLab framework, the minimum feedback delay for CSI is 50 ms; this is still sufficient for enabling netMIMO for static clients. The DAS mode is an extension to the Reuse PHY. The same data symbols are modulated and sent over the air by different transmitters simultaneously without precoding.

Categorization: The central controller collects and keeps track of rate information (for netMIMO clients), and sensor and SNR hints needed for categorization. SNR hints are realized from measurements from the sent and received data bits. However, WARP devices do not provide sensor hints unlike modern mobile devices. Several WARP-based PHY layer designs [4] have employed WARP's User I/O push buttons for events such as switching antenna mode, changing modulation and coding etc. Similarly, in TRINITY we program one of the push buttons to generate the sensor hint to the central controller that the client is mobile once it starts walking.

5.2 Deployment and Synchronization

In our implementation, we make the transmitters lightweight (serve as just RF boards that transmit signals). All the baseband processing is pushed to the central controller. This allows for an efficient and practical realization of netMIMO, especially for synchronizing, estimating channels and sharing of data streams across APs over a large geographic area. With each WARP mother board serving 4 daughter RF boards, we move the mother boards to the central controller, while the daughter boards serve as the RF transmitters. The RF signals processed at the mother boards are carried to the RF daughter boards through optical fibers (as shown in Fig. 7) that have very low latency ($\approx 5\ \mu$s/Km), thereby preserving the symbol-level synchronization needed for netMIMO. The RF signal is modulated to optical through an off-the-shelf commercial RoF (radio over fiber) transceiver, carried over the fiber and then converted back to RF signal at the other end. Co-locating multiple WARP mother boards also enables easier clock and phase synchronization between them (using approaches recently proposed in [18]) - a feature imperative for the functioning of netMIMO. Note that the above distributed, *thin* AP deployment model has recently become a very cost-effective and popular option for cellular and WiFi coverage in convention centers, hospitals, etc. [5].

5.3 Experimental Set-up

Our WARP testbed, consisting of 6 transmitters and 7 clients, is deployed in an indoor office setting as shown in Fig. 8. Different scenarios are created by moving the transmitters and clients to different locations. For mobility, WARP clients placed on a cart are moved at (timed and repeatable) fixed walking speeds on the same path across all experiments. Since a controlled environment is not possible for short-Tc clients (e.g., busy cafe), for repeatability, we emulate their channel variations by moving only their antennas in the proximity of their original location. Fig. 9 shows that such emulation produces channel variations very similar to a busy cafe, while allowing for repeatability.

To avoid external interference, we use channel 14 (unused by other devices) in the 2.4 Ghz band. Back to back packets are sent for 6 minutes using appropriate strategies in each experiment and the results are averages over multiple runs unless otherwise stated.

Metrics: The main metrics of evaluation are aggregate network rate (bits/s/Hz or bits/s as appropriate) and utility. A client's rate is obtained from its achievable Shannon rate ($log(1+\rho)$) and depends on the average received SINR (ρ, per sub-carrier) per transmitted OFDM symbol, and the number of sub-carriers allocated to the client. For netMIMO, transmission rates are implicitly determined as part of the pre-coding phase once CSI is known. For Reuse and DAS, where CSI is not employed, any of the conventional SNR based rate adaptation schemes can be used. However, due to the inherent feedback latency in WARP, rate adaptation is currently not supported for Reuse and DAS in our test-bed. In a way, measuring the rate directly from SINR of received symbols assumes ideal rate adaptation and hence prevents the vagaries of rate adaptation algorithms from influencing the inferences on strategies. Utility is used to characterize fairness among clients in the same category. The utility of a user is the logarithm of its rate ($\log(R_i)$) [3]. We seek to maximize the sum of the utilities of all users (network wide utility); this results in proportional fairness [3].

To evaluate QoS, we consider video delivery. Specifically, we consider typical video file [6] (encoding rate = 1Mbps, frames per second = 30) downloads and estimate the achieved Peak-Signal-to-Noise-Ratios (PSNR) for quantifying the delivered QoS. PSNR is defined as a function of the mean squared error (MSE) between the pixels of the decoded video and that of the original version. It is computed as follows: $PSNR = 10 \log_{10} \frac{2^L - 1}{MSE}\ [dB]$, where L is the number of bits used to encode pixel luminance (8 bits).

6. EVALUATION

6.1 Categorization

We use a sample network topology containing 3 APs and 6 clients as shown in Figure 10. Two scenarios are considered: (1) categorizing a single client (client 4) in isolation, and (2) categorizing two clients (2 and 4) simultaneously. To emulate a realistic scenario, where newly joined clients in the network need to be categorized, we assume that the pre-existing clients in the network are already categorized accurately. Given that all devices may not provide sensor hints, we also consider two variants - one with sensor hints and one without sensor hints (only SNR hints and CSI).

Accuracy with and without sensor hints: Client 4's channel state is set to be in one of 3 states: static, mobile or short-Tc. 20

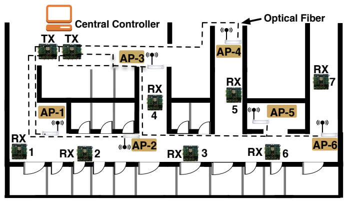


Figure 8 : Testbed deployment

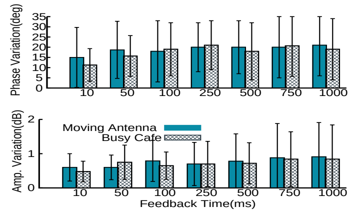


Figure 9 : Channel Variation

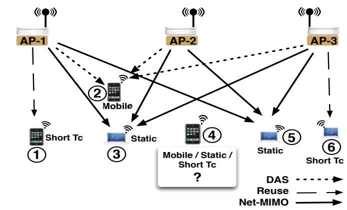


Figure 10 : Sample categorization topology

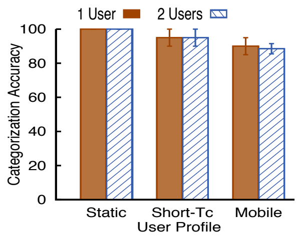


Figure 11 : Categorization Accuracy

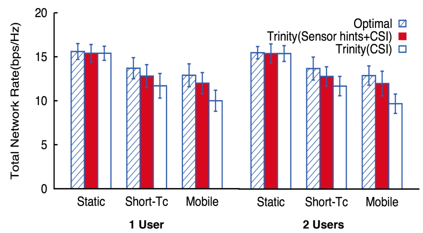


Figure 12 : Impact on Network Rate

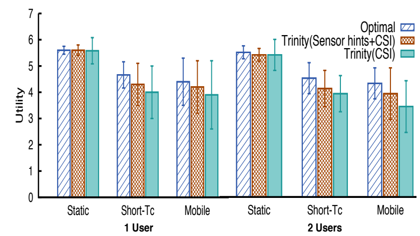


Figure 13 : Impact on Network Utility

trials are performed with each state. In the case of two clients (4 and 2), we experiment over all possible combinations of their profiles. Categorization accuracy is evaluated by comparing the profile categorized by TRINITY against *ground truth*, which is the profile that yields the highest aggregate network rate. To determine a client's optimal profile, we perform an exhaustive search over all possible combinations of strategies (DAS, Reuse and netMIMO) for each client, to find the combination which maximizes the overall network rate.

As seen in Fig. 11, TRINITY achieves a high accuracy of 95% for netMIMO users, 90-95% for Reuse users and 80-85% for DAS users. The accuracy is highest for netMIMO, where rate discrepancies provide a clear indicator, while channel dynamics for DAS users cause some classification errors. The accuracy with sensor hints is about 11% higher on average, than without it, indicating that SNR hints may be sufficient in providing an accurate classification of DAS and Reuse users, although at the expense of network performance. Further, introducing multiple clients for simultaneous categorization does not impact the accuracy of categorization. This is expected as categorization is specific to individual clients.

Impact on network rate and utility: Next, we examine the implication of the accuracy on on the aggregate network rate. Here, after categorization, the apt transmission schemes are used to transfer data to clients and the aggregate network rate is measured. The results in Fig. 12 clearly indicate that the impact of making local categorization decisions for users, although sub-optimal compared to the exhaustive global approach (e.g., some slow moving users may be under Reuse), does not incur appreciable loss (compared to optimal), while providing a large reduction in complexity. Given the high accuracy of categorizing netMIMO users, their loss in network rate is negligible. For Reuse and DAS users, the loss is small with sensor hints; in its absence it can get to 10% and 15%, respectively, but still remains within reasonable limits. Similar results are seen when utility is used to capture the fairness in Fig. 13.

The above experiments demonstrate *TRINITY's ability to categorize users appropriately with high accuracy using a combination of standard CSI/SNR and sensor hints, thereby yielding close-to optimal network performance*.

6.2 Combining Strategies in Frequency

For all the strategy-combining experiments, a topology where 2 APs jointly apply netMIMO to 2 netMIMO clients and DAS to a single DAS client, is used. In addition, one of the APs serves a single Reuse client. Since different number of sub-carriers may be employed for each strategy when combining them in a frame, we use rate in bit/s (accounting for # sub-carriers) as our metric. In the interest of space, we defer results on feasibility of frequency-domain combining, power pooling and flexible resource allocation policies in TRINITY to [7].

To understand the implication of power pooling gains with frequency domain combining (FDC), we consider its impact on network rate compared to time domain combining (TDC). In our experimental topology, when the Reuse client is associated with AP1, the sub-carriers allocated for Reuse at AP2 are un-used and their power is redirected to those that serve netMIMO and DAS users. Considering an equal split of sub-carriers between strategies, this results in about a 10% gain in the aggregate network rate compared to TDC (Fig. 14a). When the DAS user is removed, its sub-carriers are re-allocated between netMIMO and Reuse users. This results in additional un-used sub-carriers (and their power) from the Reuse strategy at AP2, which are used to serve the netMIMO users. This further increases the rate gain to $\approx 15\%$.

Increasing the number of APs, increases the possibility of more sub-carriers being left un-used at APs, since all APs may not be serving users from all profiles. To see if this contributes to increased network rate, we increase the number of APs within the interference range of the current two APs in our topology. Indeed, Fig. 14b shows that increasing the number of APs can harness additional power pooling gains from FDC and contribute to $\approx 20\%$ increase in network rate compared to TDC. Note that, while in practice the amount of power that can be pooled is limited by FCC regulations [2], it can still deliver appreciable gains [9].

6.2.1 Case study in a dynamic network

We consider a sample case study to highlight TRINITY's adaptation to network dynamics (more cases in [7]). The experiment consists of 6 APs and 6 clients as shown in Figure 14c. APs 1 and 6 do not interfere with other APs, whereas APs 2, 3, 4 and 5 are in a cluster, where they interfere with each other and can invoke either netMIMO or DAS. There are 3 static and 3 short-Tc clients at the start of the experiment. We depict TRINITY's operations and reactions on a time line in Fig. 15. We see that at the start of the experiment, TRINITY categorizes the clients accurately into different profiles - static clients are assigned to netMIMO, whereas short-Tc clients are assigned to Reuse. TRINITY then runs its strategy combining algorithm to apply these strategies on orthogonal frequencies (see Fig. 15) and begins transmissions. APs 2, 3, 4 and 5 split their sub-carriers between netMIMO (served by APs 2,3,4,5) and

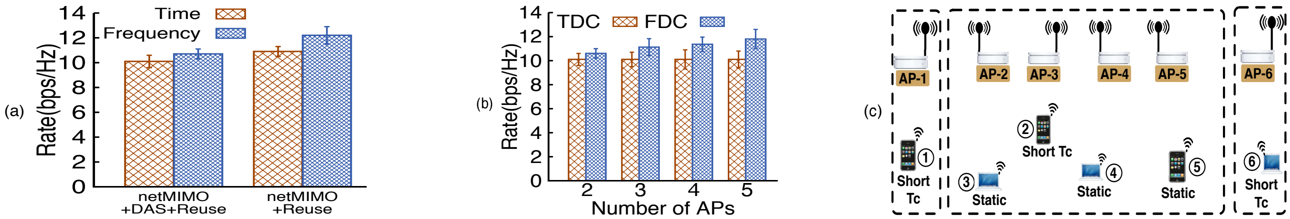


Figure 14 :(a) TDC Vs. FDC. (b) Impact of # APs on FDC. (c) Sample network topology for case study.

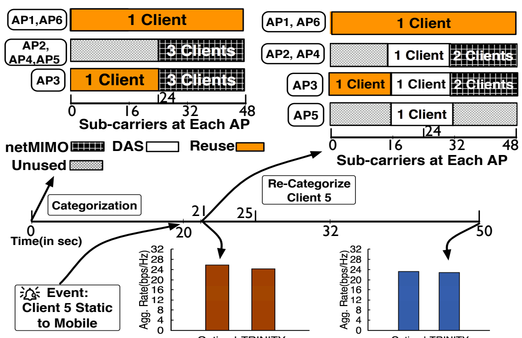


Figure 15 : Execution of TRINITY on sample topology.

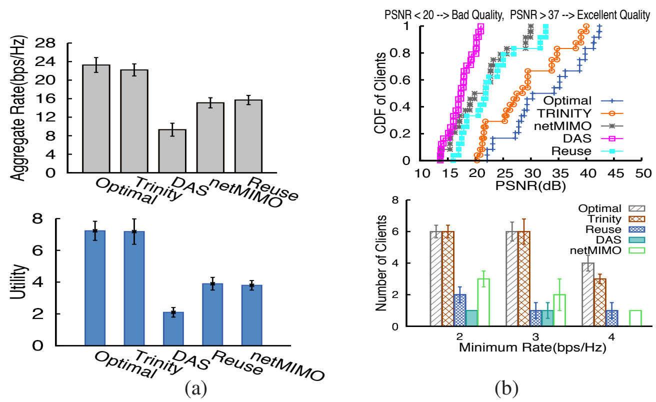


Figure 16 : (a) Performance of TRINITY (b) QoS

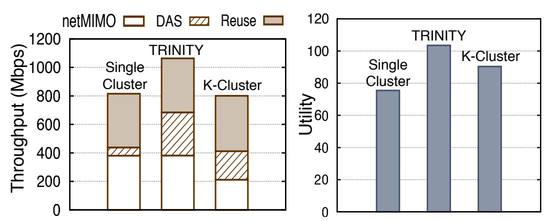


Figure 17 : Trace-driven simulation results.

reuse (served by AP 3) in the cluster. Hence, APs 2, 4 and 5 benefit from the power pooled from their un-used Reuse sub-carriers.

At the 20th second, we change client 5's state from static to mobile; we move it at walking speed in the vicinity of APs 2, 3, 4 and 5 until the experiment ends. The change in profile to a lower multiplexing one (from netMIMO) is identified automatically.

TRINITY re-categorizes client 5 within a second, into the DAS profile and obtains the new sub-carrier mapping for the APs as shown in Fig. 15. To understand the impact on network rate during network dynamics, we compare TRINITY's rate against that of the optimal (exhaustive search) solution which is also executed immediately after re-categorization as well as after 50s. It can be seen that the loss in network performance during the re-categorization process (latency to find the new profile) is small and gets further amortized when measured over a longer time window. Note that the time line experiment is part of a longer 6 min. experiment but only the first 50 seconds are shown here.

6.3 Integrating TRINITY Components

Impact on Network Rate/Utility: To understand TRINITY's performance (network-level), we compare it with four other schemes: optimal (exhaustive search), netMIMO-only, DAS-only and Reuse-only. The latter three are baseline strategies that apply only one strategy to all users irrespective of their profile and is hence representative of state of the art schemes that do not differentiate between user profiles. Our evaluations include various network topologies with 6 APs and 6 clients in our testbed, similar to those in Figs. 5 & 14c but with a more comprehensive set of clustering patterns for the APs, namely topologies with (i) linear placement of APs (only neighboring APs interfere with each other's clients), (ii) two clusters of 3 APs each, (iii) three clusters of 2 APs each, (iv) clusters of size 1, 2 and 3 APs, and (v) a single cluster of 6 APs. In each topology, clients' locations and profiles are randomly configured, and the APs transmit back-to-back data for 180 seconds.

Fig. 16a shows the average of the aggregate network rate and utility respectively, achieved over all scenarios for each scheme. We see that *TRINITY's performance is very close to that of optimal and outperforms other schemes by almost 50-150% in network rate and by two folds in utility*. This exemplifies the merit of catering transmission strategies to user profiles with TRINITY.

6.3.1 Impact on User Quality of Experience

Since we consider clients of various profiles, a given rate or throughput may not be valued the same by clients with different profiles. Hence, we also evaluate TRINITY in terms of client QoS satisfaction (using same set of topologies as previous experiment). To this end, we first consider a simple metric, wherein each client requires a minimum rate for its application for satisfactory experience and the goal is to see how many clients' requirements can be satisfied. Fig. 16b (bottom) shows the number of clients satisfied by a given scheme as a function of their increasing minimum rate requirement. We see that TRINITY is capable of satisfying as many clients as the optimal scheme in most of the cases. Further, *it can cater to thrice as many clients as the other individual strategies*. Next, we consider a standard encoded video of rate 1 Mbps with 30 frames per second for each client to understand the direct impact on application performance in Fig. 16b (top). Here, we see that *TRINITY is able to improve clients' video quality significantly by providing a large (median) gain of 10 dB in PSNR*.

6.3.2 Large-scale Trace-driven Simulations

Since it is hard to scale a testbed running netMIMO on WARP to larger topologies, we now evaluate the *complete* TRINITY solution using trace-driven simulations. To create traces for a realistic network topology, we move the WARP nodes to various locations of an office building, and collect the per-subcarrier CSI (both magnitude and phase distortion) between nearby transceiver pairs. These WARP nodes together form a network of 20 APs and 30 clients. Each AP has 2 to 6 neighboring APs within its sensing range, and serves 1 to 3 clients within its communication range. Among the clients, there are 8 Reuse, 10 netMIMO and 12 DAS users. The CSI and interference relations between neighboring nodes are then fed into a simulator executing TRINITY. We evaluate TRINITY against its variants, where its clustering component is replaced by one of the following schemes: (i) Single Cluster, which aggregates all netMIMO or DAS APs into a single cluster; (ii) K-Cluster, which greedily clusters nearby K transmitters (K is fixed to 4 in our experiments). In TRINITY, the maximum cluster size is 4.

Fig. 17(a) shows that Single Cluster achieves the highest throughput for netMIMO users owing to its global synchronization and data sharing across all netMIMO APs. However, TRINITY suffers only a small loss in performance due to splitting the APs into small clusters of practical size. This is because not all APs interfere with each other in a large network, and TRINITY can leverage spatial reuse between remote APs in different clusters. The K-cluster scheme does not account for the topology factor in its clustering process, *i.e.*, weak/strong interference between nearby APs, and thus it loses 26.1% throughput compared to TRINITY.

For DAS users, Single Cluster achieves the lowest performance, as it completely sacrifices spatial reuse for diversity, causing substantial throughput loss when there are multiple DAS users spread over a large area. TRINITY and K-cluster strike a good balance between diversity (needed for mobility) and spatial reuse. For Reuse clients, all the schemes achieve a comparable level of performance (Similar inferences can be made for utility results in Fig. 17(b)).

7. RELATED WORK

Experimental software radio platforms have allowed network practitioners to take sophisticated PHY layer multiplexing schemes like beamforming (e.g., [13]), multi-user MIMO (e.g.,[10]), interference alignment [12, 13], network MIMO (e.g., [16]),from theory to practice. Similar efforts exist with respect to Reuse [19] and diversity [15] strategies as well. While these works have made great strides in highlighting the potential and practical limitations of such schemes, their focus has understandably been on static clients with stable channels (for reliable CSI estimation) as a first step. On the theory front, works [14] are now exploring how to effectively leverage outdated CSI. However, their use in netMIMO is still in its early stages and is hence not considered here.

Given the heterogeneity of user profiles in both enterprises and outdoor cellular (e.g., LTE/WiMAX small cells) networks, it is important to understand which *strategies* are apt for which user profiles and how to combine them. While [20] articulated the need for using different strategies for different user profiles, no system was designed or implemented towards realizing this vision.

8. DISCUSSIONS AND CONCLUSIONS

We design, implement and evaluate a practical framework TRINITY, that identifies and optimizes transmission (cooperation) strategies to cater to a heterogeneous mix of users effectively, thereby improving their quality of experience in wireless networks.

Complexity and Overhead: TRINITY requires only as much overhead as any closed-loop transmission scheme like beamforming or netMIMO. By identifying the right set of netMIMO clients, it saves on the complexity and feedback overhead incurred in applying netMIMO generically to all clients. Further, by restricting netMIMO to clusters, its complexity and overhead are further reduced.

Uplink: Realizing OFDMA on the uplink requires sub-carrier level synchronization between clients being multiplexed, which is challenging although doable [21]. Thus, TRINITY is currently limited to FDC on the downlink, while using TDC on the uplink.

9. ACKNOWLEDGEMENTS

We thank the MobiHoc'15 reviewers for their feedback on this paper. This work was supported

10. REFERENCES

[1] DAS in Action : Atlanta. `http://bit.ly/ACRt4V`.

[2] Fcc rules for unlicensed wireless equipment operating in the ism bands. `http://www.afar.net/tutorials/fcc-rules/`.

[3] Network utility maximization. `http://bit.ly/1el4QAb`.

[4] Rice University WARP project. `http://warp.rice.edu`.

[5] Wired is the new wireless. `http://bit.ly/UaD2AC`.

[6] Xiph.org video media. http://media.xiph.org/video/derf/.

[7] Tailoring wireless transmission strategies to user profiles in wireless networks: Tech. report. `http://bit.ly/1suD5xO`, 2013.

[8] A.H., M., AND A., N. Diversity of MIMO linear precoding. In *IEEE Trans. on Info. Theory* (2014).

[9] ARSLAN, M. Y., SUNDARESAN, K., KRISHNAMURTHY, S. V., AND RANGARAJAN, S. Design and implementation of an integrated beamformer and uplink scheduler for ofdma femtocells. In *ACM MobiHoc* (2012).

[10] ARYAFAR, E., AN, N., SALONIDIS, T., AND KNIGHTLY, E. W. Design and Experimental Evaluation of Multi-user Beamforming in Wireless LANs. In *ACM MobiCom* (2010).

[11] FARKAS, K., HOSSMANN, T., LEGENDRE, F., PLATTNER, B., AND DAS, S. K. Link quality prediction in mesh networks. *Elsevier Comput. Commun.* (2008).

[12] GOLLAKOTA, S., PERLI, S. D., AND KATABI, D. Interference alignment and cancellation. In *ACM SIGCOMM* (2009).

[13] KUMAR, S., CIFUENTES, D., GOLLAKOTA, S., AND KATABI, D. Bringing cross-layer mimo to today's wireless lans. In *ACM SIGCOMM* (2013).

[14] MADDAH-ALI, M. A., AND TSE, D. Completely stale transmitter channel state information is still very useful. In *Allerton* (2010).

[15] RAHUL, H., HASSANIEH, H., AND KATABI, D. Sourcesync: a distributed wireless architecture for exploiting sender diversity. In *ACM SIGCOMM* (2010).

[16] RAHUL, H. S., KUMAR, S., AND KATABI, D. JMB: Scaling Wireless Capacity with User Demand. In *ACM SIGCOMM* (2012).

[17] RAVINDRANATH, L., NEWPORT, C., BALAKRISHNAN, H., AND MADDEN, S. Improving Wireless Network Performance Using Sensor Hints. In *USENIX NSDI* (2011).

[18] SHEPARD, C., YU, H., ANAND, N., LI, E., MARZETTA, T., YANG, R., AND ZHONG, L. Argos: practical many-antenna base stations. In *ACM Mobicom* (2012).

[19] SHRIVASTAVA, V., AHMED, N., RAYANCHU, S., BANERJEE, S., KESHAV, S., PAPAGIANNAKI, K., AND MISHRA, A. Centaur: realizing the full potential of centralized wlans through a hybrid data path. In *ACM Mobicom* (2009).

[20] SINGH, S., SUNDARESAN, K., KHOJASTEPOUR, A., RANGARAJAN, S., AND KRISHNAMURTHY, S. V. One strategy does not serve all: Tailoring wireless transmission strategies to user profiles. In *ACM HotNets* (2012).

[21] TAN, K., FANG, J., ZHANG, Y., CHEN, S., SHI, L., ZHANG, J., AND ZHANG, Y. Fine-grained channel access in wireless lan. In *ACM SIGCOMM* (2010).

[22] TARIGHAT, A., SADEK, M., AND SAYED, A. A multi user beamforming scheme for downlink mimo channels based on maximizing signal-to-leakage ratios. In *IEEE ICASSP* (2005).

[23] TSE, D., AND VISWANATH, P. Fundamentals of wireless communication. Cambridge University Press, 2005.