MIMO-based Rate Adaptation to Enhance TCP Throughput over Wireless Fading Channels

Vishwanath Ramamurthi, Abu (Sayeem) Reaz, Dipak Ghosal, and Biswanath Mukherjee
University of California, Davis, CA, USA
Email: {rama, asreaz, dghosal, bmukherjee}@ucdavis.edu

Abstract— We study the performance of TCP (Transmission Control Protocol) over a wireless fading link using MIMO (Multiple Input Multiple Output) technology. MIMO technology using multiple transmit and receive antennas can potentially mitigate the effect of fading by providing diversity gain and/or increase the throughput by providing multiplexing gain. Given the number of transmit and receive antennas, there exists a tradeoff between the diversity and multiplexing gains. We study the effect of the so-called diversity-multiplexing tradeoff [1] on the throughput of TCP. We demonstrate by cross-layer simulations involving the physical and transport layers that diversity gain is more useful to enhance TCP throughput when the Signal-to-Noise Ratio (SNR) is low, and multiplexing gain is more useful when SNR is high. Our results indicate that there exist SNR levels at which switching from diversity-providing schemes to multiplexing schemes would provide enhanced TCP throughput. We propose a cross-layer rate-switching scheme to enhance TCP throughput over a wide range of SNR values.

Index Terms— MIMO, TCP, Diversity-Multiplexing Tradeoff, TCP Performance, Rate Adaptation, Wireless Fading channel.

I. INTRODUCTION

Multiple-Input Multiple-Output (MIMO) refers to multi-antenna communication systems wherein the transmitter is equipped with multiple antennas capable of transmitting independent signals and the receiver is similarly equipped with multiple receive antennas. Lately, MIMO has attracted a lot of attention in the area of wireless communications, since it provides significant improvement in throughput and range without any increase in the link bandwidth or transmission power [2-4]. MIMO technology increases the spectral efficiency of a wireless communication system by exploiting the space domain (using multiple antennas).

The performance of standard wireless communication techniques degrades in the presence of multipath environment which results in fading [5]. Fading causes the received signal to fluctuate widely within a distance of a few wavelengths due to constructive and destructive interference of the multipath signals at different points in space. This multipath nature of the wireless channel can be exploited to improve the performance of the wireless communication system if we use multiple antennas at the transmitter and the receiver. Multiple antennas at the transmitter and the receiver provide multiple independent paths from the transmitter to the receiver and thus can be used to obtain diverse copies of the same signal at the receiver, which can be suitably combined to improve the overall performance of the system. Space-Time Coding (STC) techniques work on this principle to provide a robust communication channel [6, 7].

There is another line of thought where the multipath channel between the transmitter and the receiver is used to transmit multiple streams of data from the transmitter to the receiver, thus increasing the capacity of the system. BLAST (Bell Labs LAyered Space-Time) architecture and its derivatives [2, 8] work on this principle to provide enhanced throughput.

The reliability that a MIMO system provides is quantified by the “diversity order” obtained, and the capacity gain that it provides is quantified by “multiplexing gain” obtained. These two approaches can be combined to obtain both reliability and high rate. But, due to the fundamental nature of the wireless channel, there exists a tradeoff between the diversity order and multiplexing gain that can be obtained [1].

The Transmission Control Protocol (TCP) is a very popular transport-layer protocol used by a myriad of applications to provide a reliable end-to-end communication. In [9], a TCP-aware rate adaptation for cellular networks has been proposed. We study the performance of TCP over a MIMO wireless channel using either STC or BLAST schemes. We address the issue of which scheme is better under given operating conditions to enhance TCP throughput. This is a cross-layer problem spanning the physical and the transport layers of the communication stack. We find that diversity-based schemes are better at low SNR levels, while multiplexing-based schemes are better at high SNR levels.

The rest of the paper is organized as follows. Section II describes the MIMO system model and the transmission schemes used. Section III reviews the performance of TCP over a lossy wireless link. In Section IV, we present numerical results from our cross-layer simulations of TCP over MIMO channel and then propose a cross-layer rate-switching scheme to enhance TCP throughput. Section V concludes the paper.

II. MIMO SYSTEM MODEL

Consider a Nt x Nr MIMO system with Nt transmit antennas and Nr receive antennas (see Fig. 1). The channel between the transmitter and the receiver is a Rayleigh fading channel. Transmission occurs in blocks of T transmission slots. The discrete-time transmission model for a MIMO system is given as:

\[ Y = HX + N \]  

(1)

where X is the Nt x T transmission matrix, Y is the Nt x T receiver matrix, H is the Nt x Nt channel matrix, and N is the Nt x T Additive White Gaussian Noise (AWGN) matrix with each element being distributed as \( N_c(0, E_s/\text{SNR}) \) where \( E_s \) is the signal energy and \( \text{SNR} \) is the Signal-to-Noise Ratio at each receive antenna element.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.
A. Space-Time Coding (STC) - The Alamouti Code

For a 2 x 2 MIMO system, the optimal diversity scheme is the “Alamouti Code” [6] which is given by:

\[
X = \begin{bmatrix}
  s_1 & -s_2^* \\
  s_2 & s_1^*
\end{bmatrix}
\]  

(2)

where \(s_1\) and \(s_2\) are symbols from the constellation set of the digital modulation used and \(s^*\) denotes the complex conjugate of symbol \(s\). Alamouti code achieves a maximum possible diversity order \(d = NtNr = 4\) by a 2 x 2 system and has a spatial rate \(r_s = 1\). Also, because of the orthogonal structure of \(X\), the optimal MIMO Maximum Likelihood (ML) decoding has a remarkably low complexity.

B. Spatial Multiplexing (SM) - BLAST

In SM, the incoming data is split into multiple streams and transmitted over the \(N_t\) transmit antennas. \(N_t\) independent symbols are transmitted per time slot, thus achieving a spatial rate \(r_s = N_t\). SM can be used in conjunction with either vertical encoding (resulting in Vertical BLAST or V-BLAST), or diagonal encoding (resulting in Diagonal BLAST or D-BLAST). In V-BLAST, the incoming bit stream undergoes temporal coding, interleaving, and symbol mapping before being demultiplexed into \(N_t\) streams that are transmitted over the \(N_t\) antennas (see Fig. 2). In D-BLAST, the incoming bit stream is first demultiplexed into \(N_t\) streams as in V-BLAST. Then, each of these streams is passed through a cyclic stream rotator that cycles the bit stream-antenna associations in a periodic fashion, resulting in a diagonal structure of transmission for each stream (see Fig. 2).

ML decoding of SM schemes is computationally intensive and not practical. The complexity of decoding can be reduced by use of linear receivers such as Zero-Forcing (ZF) receiver and Minimum Mean Square Error (MMSE) receiver in conjunction with Successive nulling and Interference Cancellation (SIC) [10]. SM schemes have higher data rates than the STC schemes, but they also possess higher BER than STC schemes. Figure 3 shows the comparative BER versus SNR performance of STC and SM (using MMSE-based SIC) for 2 x 2 MIMO systems with QPSK modulation obtained by Monte-Carlo simulations using MATLAB.

C. Diversity – Multiplexing Tradeoff

The multiplexing gain of a MIMO system is defined as [1]:

\[
r = \frac{R}{\log_{10} SNR}
\]

where \(R\) is spectral efficiency in bits per seconds per Hertz. The multiplexing gain \(r\) signifies the rate at which the data rate increases with SNR. The diversity gain \(d\) represents the rate at which the BER \(P_e\) decreases with SNR [1]:

\[
P_e \approx SNR^{-d}
\]

Due to the fundamental nature of the wireless channel, there is a tradeoff between the diversity order \(d\) and the multiplexing gain \(r\) that can be obtained with a given number of transmit and receive antennas. It has been shown in [1] that if \(r\) antennas are used for spatial multiplexing, then the maximum diversity order that can be achieved is \((N_t-r)(N_r-r)\). The optimal diversity-multiplexing tradeoff curve for a 2 x 2 MIMO system is shown in Fig. 4.

One extreme of the optimal tradeoff curve is the point (4, 0)

Fig. 1. MIMO communication model.

Fig. 2. V-BLAST and D-BLAST for \(N_t = 2\).

Fig. 3. Comparative BER performance of Alamouti STC and V-BLAST for 2 x 2 MIMO system using QPSK modulation.

Fig. 4. Diversity-multiplexing tradeoff for 2 x 2 MIMO system.
where the maximum diversity is achieved but the multiplexing gain is 0. The other extreme of the tradeoff is the point (0, 2) where the maximum possible multiplexing gain is obtained but the diversity order is 0.

III. PERFORMANCE OF TCP OVER WIRELESS LINKS

TCP was designed to provide a reliable end-to-end data pipe in traditional wired networks. TCP uses acknowledgements for transmitted packets. The application-layer data is fragmented into smaller packets called TCP segments which are sequentially numbered. The receiver acknowledges correctly-received packets in a cumulative manner.

TCP maintains a parameter $W$, called Congestion window, which indicates the amount of outstanding packets it can transmit. It also maintains a parameter $W_t$ called the slow-start threshold, which controls the increments in window size. Parameters $W$ and $W_t$ are updated in the Reno flavor of TCP as follows:

1. For every acknowledgement received,
   - If ($W < W_t$), Set $W = W + 1$ (Slow Start Phase)
   - Else Set $W = W + 1/W$ (Congestion Avoidance Phase)
2. When loss is indicated by $K$ duplicate acknowledgements,
   - Set $W_i = W/2$ and $W = W_i$ (Fast Retransmit)
   - Resume Congestion Avoidance Phase
3. When loss in indicated by timeout,
   - Set $W_i = W/2$ and $W = 1$
   - Start Slow Start phase

TCP gauges the capacity of the network by using the slow-start mechanism where $W$ is initially set to one packet and is increased by one for every acknowledgement received, resulting in exponential growth of the window size. Congestion-avoidance stage is started once the window size exceeds the slow-start threshold. Timeouts and duplicate acknowledgements are used to indicate packet loss. Whenever a packet is transmitted, a timer is started whose timeout value is the round-trip time of packets. Packets that are lost are then retransmitted. In the Reno and later flavors of TCP, the reception of $K$ ($K = 3$) duplicate acknowledgements is also taken as an indication of packet (or TCP segment) loss.

Packet loss was assumed only due to congestion at a bottleneck link and subsequent buffer overflow. So to deal with the situation, congestion-control mechanism is used where the transmitter reduces its window size by an amount dependent on the type of loss, thus effectively cutting on the rate of transmission of packets. Fast retransmit is used to decrease the recovery time due to a temporary lossy condition indicated by duplicate acknowledgements.

In [11], the problems arising when the TCP protocol is used over wireless links is discussed. TCP was designed for wired networks, but when it is retrofitted for the wireless domain, its performance degrades. TCP mistakes wireless errors to be congestion and unnecessarily invokes the congestion-avoidance algorithm. The performance of TCP over a lossy link has been analytically characterized in [12-15] under slightly varied assumptions. We assume that the RTT is dominated by the transmission time of TCP packets. Acknowledgements are small and contribute only negligibly to the RTT. If the total propagation delay is $\tau$, and the average packet processing rate is $\mu$, RTT is given by:

$$\text{RTT} = \tau + \frac{1}{\mu}$$

The average packet processing rate $\mu$ is related to the transmission bandwidth and also the packet size. If $B_w$ is the transmission bandwidth, and $l_p$ is the length of the transmitted packet, then the average packet processing rate is:

$$\mu = \frac{B_w}{l_p} = \frac{B_w \log \text{SNR}}{l_p}$$

The following approximate closed-form expression for TCP throughput $B(p)$ at a packet-error rate (PER) $p$ has been derived in [12, 15] and was found to be accurate over large range of loss rates and RTTs:

$$B(p) = \begin{cases} 
\frac{1-p}{p} + E[W] + Q(E[W]) \frac{1}{1-p} \\
\frac{1}{1-p} + \frac{W_{\text{max}} + Q(W_{\text{max}})}{1} - p \text{RTT} + \frac{f(p)}{1-p} \\
\frac{1}{1-p} + \frac{W_{\text{max}} + Q(W_{\text{max}})}{1} - p \text{RTT} + \frac{f(p)}{1-p} 
\end{cases}$$

where $W_{\text{max}}$ is the TCP maximum window size, and $T_0$ is the TCP timeout period. Expected window size $E[W]$ is given as:

$$E[W] = \frac{8(1-p)}{3bp} + \frac{(3b-1)^2}{3b} - \frac{3b-2}{3b}$$

The functions $Q(W)$ and $f(p)$ are given as follows:

$$Q(W) = \min \left\{ \frac{(1-(1-p)^{W-3})}{1-\frac{(1-p)}{W}} \right\}$$

$$f(p) = 1 + p + 2p^2 + 4p^3 + 8p^4 + 16p^5 + 32p^6$$

PER $p$ is related to BER $P_e$ obtained at the physical layer as:

$$p = 1 - (1 - P_e)^k$$

Using this analysis, Fig. 5 plots the TCP throughput as a function of loss probability with parameters $B_w = 1\text{MHz}$, $T_0 = 2.5\text{ sec}$, $l_p = 1400$ bytes, and $W_{\text{max}} = 6$. We can see that TCP throughput degrades with increasing packet-error probability.

IV. NUMERICAL RESULTS

TCP throughput degrades with increase in packet-error probability which is related to BER $P_e$ through Eqn. (11). BER is related to SNR through BER-SNR curves which, in turn, depend on the physical-layer technology used and the wireless environment. We saw in Section II that MIMO technology provides a tradeoff between rate and reliability. In this section, we evaluate the performance of TCP over MIMO channels.
We use MATLAB to characterize the physical-layer performance. We use the approximate TCP throughput analysis in Eqns. (5) – (11) and also OPNET simulations for evaluating the performance of TCP.

1. Fixed Rate

In this case, a fixed constellation size is used and the spectral efficiency $R$ does not increase with SNR. In Fig. 3, we plotted the comparative BER performance of Alamouti code and V-BLAST for a 2 x 2 MIMO system using QPSK constellation. We feed these as inputs to OPNET simulator to evaluate the comparative performance of these schemes in terms of TCP throughput. We used the model presented in [16] to simulate Rayleigh fading in OPNET and two different Wireless LAN rates (5.5 Mbps and 11 Mbps) to simulate the effect of different rates for the two schemes.

2. Variable Rates

Multiplexing gain of a MIMO system specifies the rate at which the data rate increases with SNR. This increase in data rate is accomplished by increasing the constellation size of the modulation technology. Diversity-Multiplexing tradeoff is a rate-reliability tradeoff that can be obtained with MIMO technology. It specifies a method of increasing rate and decreasing BER at the same time. If the rate of increase in data rate is higher, then the rate of decrease of BER is lower. This fact can be visualized in Fig. 7 which plots the theoretical BER versus SNR for various multiplexing gains for a 2 x 2 MIMO system.

The dashed curves in the background are for the maximum possible diversity order $d = 4$ and zero multiplexing gain. Each dashed curve corresponds to a different spectral efficiency (starting from 2 bits/sec/Hz up to 40 bits/sec/Hz). The solid curves in the foreground are each for a different multiplexing gain and the spectral efficiency increases with SNR along each curve at an exponential rate with exponent equal to the multiplexing gain $r$. As the value of $r$ increases, the diversity order decreases and so does the reliability of the system.

![Fig. 6. Comparative TCP throughput performance of Alamouti code and V-BLAST scheme.](image)

Fig. 6. Comparative TCP throughput performance of Alamouti code and V-BLAST scheme.

Figure 6 plots the comparative TCP throughput performance of Alamouti code and V-BLAST as a function of SNR. We observe that Alamouti code performs better when SNR is low and the V-BLAST scheme has a higher throughput when SNR is high. There is a threshold SNR below which STC would be better and above which SM would be better.

B. Variable Rates

Multiplexing gain of a MIMO system specifies the rate at which the data rate increases with SNR. This increase in data rate is accomplished by increasing the constellation size of the modulation technology. Diversity-Multiplexing tradeoff is a rate-reliability tradeoff that can be obtained with MIMO technology. It specifies a method of increasing rate and decreasing BER at the same time. If the rate of increase in data rate is higher, then the rate of decrease of BER is lower. This fact can be visualized in Fig. 7 which plots the theoretical BER versus SNR for various multiplexing gains for a 2 x 2 MIMO system.

The dashed curves in the background are for the maximum possible diversity order $d = 4$ and zero multiplexing gain. Each dashed curve corresponds to a different spectral efficiency (starting from 2 bits/sec/Hz up to 40 bits/sec/Hz). The solid curves in the foreground are each for a different multiplexing gain and the spectral efficiency increases with SNR along each curve at an exponential rate with exponent equal to the multiplexing gain $r$. As the value of $r$ increases, the diversity order decreases and so does the reliability of the system.

Fig. 6 plots the comparative TCP throughput performance of Alamouti code and V-BLAST as a function of SNR. We observe that, for high multiplexing gains ($r \geq 1$), TCP throughput is negligible. This is because of the high BER when the multiplexing gain is high (see Fig. 7). For $r < 1$ and at high SNR, the throughput

![Fig. 7. Visualizing Diversity-Multiplexing tradeoff in a 2 x 2 MIMO system.](image)

Fig. 7. Visualizing Diversity-Multiplexing tradeoff in a 2 x 2 MIMO system.

We calculate TCP throughputs along each of the solid curves in Fig. 7 using Eqns. (5)–(11), and we plot the corresponding curves in Fig. 8. We observe that, for high multiplexing gains ($r \geq 1$), TCP throughput is negligible. This is because of the high BER when the multiplexing gain is high (see Fig. 7). For $r < 1$ and at high SNR, the throughput

![Fig. 8. TCP throughput along diversity-multiplexing tradeoff.](image)

Fig. 8. TCP throughput along diversity-multiplexing tradeoff.
increases with SNR at a rate proportional to $r$. Throughput curves in Fig. 8 that have higher multiplexing gain have larger slope at high SNR than curves with lower multiplexing gain. For $r < 1$ and at low SNR, lower value of multiplexing gain results in higher absolute throughput than higher values of multiplexing gain. Therefore, at lower SNR, larger diversity order is beneficial; and at higher SNR, larger multiplexing gain is more beneficial.

C. A Cross-layer Rate-Switching Scheme

Based on the observations from Figs. 6 and 8, we propose a rate-switching scheme based on feedback from TCP acknowledgements. The rate-switching scheme chooses the best multiplexing gain that provides maximum TCP throughput under the prevalent wireless channel conditions.

In Fig. 8, there are four distinct regions separated by three SNR levels: A, B, and C. At low SNR it is best to use a scheme which provides full diversity order, i.e., zero multiplexing gain. As SNR increases, we switch to $r = 0.25$ at A, then to $r = 0.5$ at B and finally to $r = 0.75$ at C. The receiver has knowledge of the SNR that it is experiencing and it informs the TCP transmitter at the other end the best transmission scheme to use at the given SNR level, by piggybacking the information in the TCP acknowledgements. For example, in a fixed-rate implementation, one bit feedback can be used for switching between STC and SM schemes. A value of 0 would indicate that the transmitter should be using STC while a value of 1 would indicate that the transmitter should be using SM. TCP on the transmitter side observes this bit and informs its physical layer to switch to the appropriate scheme. In a variable-rate implementation, a separate field in TCP acknowledgement could be used to indicate the rate at which the transmitter should operate. Such a cross-layer rate-switching scheme would have the best TCP throughput over the whole range of SNRs.

V. CONCLUSION

This paper illustrated the interplay between TCP and MIMO technology which is a potential candidate for next-generation broadband wireless access systems. We evaluated the performance of TCP over MIMO channel using different transmission schemes. We found that diversity-based schemes (STC) are better at low SNR levels while multiplexing-based schemes (SM) are better at higher SNR levels. We then proposed a cross-layer rate-switching scheme which exploits this cross-layer interplay to enhance TCP throughput at all SNR levels.

REFERENCES