Scientific Research

International Journal of Communications, Network and System Sciences

ISSN: 1913-3715
Volume 3, Number 7, July 2010

ISSN: 1913-3715

www.scirp.org/journal/ijcns/

JOURNAL EDITORIAL BOARD

ISSN 1913-3715 (Print) ISSN 1913-3723 (Online)
http://www.scirp.org/journal/ijcns/

Int. J. Communications, Network and System Sciences, **2010, 3, 579-637**
Published Online July 2010 in SciRes (http://www.SciRP.org/journal/ijcns/)

TABLE OF CONTENTS

Volume 3 Number 7 July 2010

International Journal of Communications, Network and System Sciences (IJCNS)

Journal Information

SUBSCRIPTIONS

The *International Journal of Communications, Network and System Sciences* (Online at Scientific Research Publishing, www.SciRP.org) is published monthly by Scientific Research Publishing, Inc., USA.

Subscription rates:

Print: $50 per issue.

To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements

Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)

Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.

E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:

E-mail: ijcns@scirp.org

Int. J. Communications, Network and System Sciences, **2010, 3, 579-584**
doi:10.4236/ijcns.2010.37077 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Scientific Research

Hybrid Authentication Cybersystem Based on Discrete Logarithm, Factorization and Array Entanglements

Boris S. Verkhovsky
Computer Science Department, New Jersey Institute of Technology, Newark, USA
E-mail: *verb@njit.edu*
Received July 13, 2009; *revised February* 28, 2010; *accepted May* 10, 2010

Abstract

A hybrid cryptographic system providing digital authentication is described and analyzed in this paper. The proposed cryptosystem incorporates three features: complexity of the discrete logarithm problem, complexity of integer factorization of a product of two large primes and a combination of symmetric and asymmetric keys. In order to make the cryptosystem less vulnerable to cryptanalytic attacks a concept of digital *entanglements* is introduced. As a result, the proposed cryptographic system has four layers (entanglement-encryption-decryption-disentanglement). It is shown that in certain instances the proposed communication cryptocol is many times faster than the RSA cryptosystem. Examples provided in the paper illustrate details of the proposed authentication cryptocol.

Keywords: Crypto-Immunity, Cybersecurity, Digital Authentication, Array Entanglements, Multi-Layer Cryptographic Protection, Hybrid Cryptocol

1. Introduction and Basic Definitions

In this paper a hybrid digital signature cyber-secure communication system is described and analyzed. In order to make this cryptosystem faster and less vulnerable to cryptanalytic attacks a concept of *entanglements* is introduced [1,2]. Furthermore, in this cryptographic protocol there are *four* layers (entanglement-encryption-decryption-disentanglement). Since there is no one-to-one mapping between a plaintext block and the corresponding ciphertext block, this system of communication is less vulnerable to plaintext attacks. The overall cryptographic algorithm is a hybrid protocol that incorporates three features: discrete logarithm problem modulo large prime [3], factorization of a product of two large primes [4] and a combination of symmetric and asymmetric keys.

To describe the proposed cryptosystem, let's consider

A1. An array $m = (a_1, a_2, .., a_r)$ (1)

consisting of r blocks of a digitized plaintext that is to be transmitted from a sender (Alice) to a receiver (Bob);

B1. A square $r \times r$ non-singular matrix E with

$$|E| \neq 0 \quad \text{and} \quad h = Em. \tag{2}$$

$$\text{In the paper} \quad h = (h_1, .., h_r) \tag{3}$$

and E are respectively called a vector and matrix of *entanglements* [1].

C1. A sufficiently strong cryptographic protocol L that is used for encryption of one of the entanglements, for example, h_1, with corresponding ciphertext c_1.

In order to speed up the encryption/decryption procedure and as a result to minimize the entire communication time it is necessary to minimize the amount of computations. For that reason there is no necessity to encrypt all other entanglements $h_{j \neq i}$, where j = 1, 2,.., i – 1, i + 1,.., r and h_i is the encrypted entanglement. Indeed, if h_i is not known to a potential intruder, then he or she must solve a system of r equations, where only r – 1 components of vector h are publicly known. In the cryptosystem described below the size r of the array m is a trade-off between crypto-immunity and acceleration of the decryption: the larger the value of r, the faster the overall communication protocol. On the other hand, the larger r is, the less time is required to cryptanalyze the entire message.

To ***avoid confusions***, it is important to indicate the following distinctions:

- The matrix of entanglement E (and non-linear mappings) discussed below are not secret keys as in an affine cryptographic algorithm; all elements of matrix E are *publicly known*;
- In contrast with the RSA and Rabin algorithms, n_k

is a *private* key of the *k*-th user, not a public key.

2. Digital Signature Scheme

2.1. System Design Module (Users Establish their Private and Public Key):

A2. All users agree on a large prime p and a generator g, where

$$2 \le g \le p-2 \tag{4}$$

***Remark* 1**: selection of a generator for a large prime p is a non-deterministic procedure. However, if both p and $(p-1)/2$ are primes, then

$$g := (3p - 1)/4 \tag{5}$$

is a generator [5].

B2. Each user selects large primes p_k and q_k, such that

$$p_k \equiv q_k \equiv 2 (\mathrm{mod}\, 3), \tag{6}$$

and that their product n_k satisfies two constraints:

$$\alpha p < n_k < p \tag{7}$$

C2. Each user computes her/his public key e_k (encryption key) and private key d_k, where e_k is co-prime with z_k, *i.e.*,

$$\gcd(z_k, e_k) = 1 \tag{8}$$

D2. Every user computes a multiplicative inverse d_k of e_k modulo

$$z_k := (p_k - 1)(q_k - 1) \quad [4]$$

i.e., d_k satisfies the equation

$$d_k e_k \equiv 1 (\mathrm{mod}\, z_k) \tag{9}$$

E2. If Alice and Bob intend to secretly exchange authenticated information, they establish a secret key $w_{AB} := g^{ab} \bmod p$ by using the Diffie-Hellman key exchange [3].

2.2. Encryption/Decryption Module (Alice Sends to Bob a Plaintext Array *m*):

F2. Using an open channel Alice asks Bob to secretly send to her Bob's secret key n_B;

G2. Bob computes $x := n_B w_{AB} \bmod p$ and sends x to Alice;

she recovers $n_B := x w_{AB}^{-1} \bmod p$;

H2. Using the RSA protocol Alice encrypts h_i {see (2)}:

$$c_i := h_i^{e_B} \bmod n_B; \quad [4]; \tag{10}$$

I2. Alice transmits the array $\{h_1, .., h_{i-1}, c_i, h_{i+1}, .., h_r\}$ to Bob;

J2. Bob decrypts c_i:

$$v := c_i^{d_B} \bmod n_B; \{= h_i\}; \tag{11}$$

K2. Using $h = (h_1, .., h_r)$ Bob recovers all plaintext blocks $m = (a_1, a_2, .., a_r)$;

L2. If the original array m is intelligible, but the recovered text is not, then Bob realizes that it was forged by an intruder; otherwise Bob accepts authenticity of the text.

2.3. Selection of Block Size and Matrix of Entanglements

To make sure that the entanglements are smaller than every n_i, {otherwise the entire array $m = (a_1, a_2, .., a_r)$ is not recoverable}, select the matrix of entanglement E and such division of a plaintext onto blocks that the maximal value of the i-th entanglement h_i does not exceed αp, (7)}.

Example 1: Let $m := (a, b, c, d, e)$ and

$$h = (h_1, h_2, h_3, h_4, h_5), \tag{12}$$

$$\begin{aligned} &\text{where } h_1 := d + 2e; h_2 := a - 2b; \\ &h_3 := 2a - b + c; \\ &h_4 := c - d + 2e; \\ &h_5 := a + 2b + c + d. \end{aligned} \tag{13}$$

$$\begin{aligned} &\text{Then } b = h_1 + 3h_3 - (4h_2 + h_4 + 2h_5); \\ &a = h_2 + 2b; \quad c = h_3 - 2a + b; \\ &d = h_5 - a - 2b - c; e = (h_1 - d)/2. \end{aligned} \tag{14}$$

Let's specify that every block in m must satisfy a threshold $a_k < t$.

Then (13) implies that

$$\max h_i = 5t < \alpha p < n_i < p. \tag{15}$$

Therefore, for every $k = 1, \ldots, r$ must hold

$$a_k \le t \le \alpha p / 5;$$

and if $\alpha = 2/3$, then $a_k \le t \le 2p/15$.

From the recovery procedure (14) it is clear that we can compute all initial blocks *a*, *b*, *c*, *d* and *e* **only if** we know all numeric values h_1, h_2, h_3, h_4, h_5 from (13). Henceforth, this fact implies that it is sufficient to encrypt at least one of these entanglements to securely protect all five plaintext blocks.

Furthermore, it is necessary to notice that entanglements themselves do not provide secure protection. In the proposed cryptographic scheme instead of employing just one layer (plaintext/encryption/ciphertext) we propose **two layers** (plaintext/entanglement/encryption/ci-

phertext) between the plaintext array (*a, b, c, d, e,*…) and ciphertext ($c_1, h_2, h_3, h_4, h_5, \ldots$).

Remark 2: The RSA algorithm discussed below is just an example of how h_i can be encrypted. Any strong cryptocol based on the complexity of factorization of $n = pq$ can also be used. The Rabin algorithm [6] or (hyper) elliptic-curve cryptography [7-10] based on modulo composite *n* are other possible applications.

2.4. Essence of RSA Digital Signature Algorithm

In order to demonstrate the advantages of the proposed digital signature algorithm, let's recall the RSA digital signature algorithm [4,11]. Suppose Alice wants to send to Bob a message $m = (a_1, a_2, .., a_r)$ with a digital signature. Then for ***every*** k = 1, 2,…, r Alice signs a_k $f_k := a_k^{\,d_A} \mod n_A$ with her private key, then encrypts it with Bob's public key

$$c_k := f_k^{\,e_B} \bmod n_B; \tag{16}$$

and transmits the ciphertext c_k to Bob over an open communication channel. Bob decrypts $x := c^{d_B} \bmod n_B$ and then verifies the signature:

$$y := x^{e_A} \bmod n_A; \quad \{y = m\}; \tag{17}$$

If y is intelligible, then Bob accepts it as an authenticated message from Alice.

3. Examples of Entanglements

3.1. Linear Transformations

Example 2: Let

$$\begin{aligned} &h_0 := a_1 + \cdots + a_{r-1} + a_r;\ h_1 := -a_1 + a_2 + \cdots + a_r; \\ &h_2 := a_1 - a_2 + \cdots + a_r;\ .., h_{r-1} := a_1 + \cdots - a_{r-1} + a_r. \end{aligned} \tag{18}$$

Proposition 1: If all entanglements $h_0, h_1, h_2, .., h_{r-1}$ are known and integer, then for every k = 1,.., r – 1

$$a_k = (h_0 - h_k)/2 \tag{19}$$

$$a_r = h_0 - (a_1 + a_2 + \cdots + a_{r-1}) \tag{20}$$

and all a_k are integers as well.

Proof follows from two observations:

- for every k = 1,.., r – 1 $h_k = h_0 - 2a_k$; (21)
- all $h_0, h_1, h_2, .., h_{r-1}$ have the same parity which implies that their pair-wise differences are *even*. Therefore, every $a_1, a_2, ..$ and a_r is an integer. Q. E. D.

Complexity of recovery: It requires r – 1 subtractions and divisions by 2 (binary shifts) to recover the first r – 1 blocks in (19) and r – 1 subtractions to recover the last block in (20).

If a sender (Alice) encrypts only s of all entanglements, where $0 < s < r$, then the intruder will not be able to deduce any blocks (provided that the matrix E is properly selected and a portion of entanglements is encrypted with a sufficiently strong PKC protocol). In an extreme case, if s = 1, then the intruder must solve a system of equations $Ea = g$, where the matrix E is **known** but only $r - s$ elements of vector g are known. However, this is impossible, because to find the blocks $a_1, a_2, .., a_r$ the intruder must know all r elements of vector g.

3.2. Non-Linear Transformations

In the more general case, the entanglements can be non-linear, *i.e.* $h := E(a)$, and/or some components of the transformation $E(a)$ can be also encrypted. For example, if $h := Ea$, then we can encrypt several elements of matrix E. This approach is beyond the scope of this short paper. It is important to bear in mind that the selection of the transformation $E(a)$ affects the computational complexity of the recovery process.

The choice of the mapping E is important. If E is a matrix, then it must be non-singular and selected in such a way that the recovery will not become a computationally formidable.

Example 3: Let's consider an array of r plaintext blocks $h_1, h_2, .., h_{r-1}, h_r$ and the following r entanglements:

$$\begin{aligned} &h_1 := a_1^2 - a_2^2;\ h_2 := a_2^2 - a_3^2; \cdots \\ &h_{r-1} := a_{r-1}^2 - a_r^2;\ h_r := a_1 + a_r. \end{aligned} \tag{22}$$

It is obviously sufficient to encrypt only one of the entanglements. Then, after the decryption, we proceed as follows:

$$w := h_1 + h_2 + \cdots + h_{r-1} = a_1^2 - a_r^2. \tag{23}$$

Therefore,

$$a_1 = (h_r + w/h_r);\ a_r = (h_r - w/h_r); \tag{24}$$

and for k from 2 to r – 1

$$a_k = \sqrt{a_{k-1}^2 - h_{k-1}}. \tag{25}$$

Combined with encryption these non-linear entanglements provide secure protection and recovery for every transmitted array. Yet, they require divisions of integers and extraction of square roots, which are computationally more complicated procedures.

3.3. Improper Entanglements

Example 4: Let

$$\begin{aligned} &h_1 := 2a_1 + a_2;\ h_2 := a_1 + a_2; \\ &h_3 := a_1 + a_2 + a_3;\ .., h_r := a_1 + a_2 + \cdots + a_r. \end{aligned} \tag{26}$$

If $r > 2$, it is insufficient to encrypt only one of these entanglements.

Indeed, if h_1 is encrypted, then for all $3 \le k \le r$

$$a_k = h_k - h_{k-1}. \quad (27)$$

In general, if i is fixed, $i \ge 2$, and only h_i is encrypted, then

$$a_1 = h_1 - h_2; \quad (28)$$

and for all $3 \le k \le i-1$ and $i+2 \le k \le r$

$$a_k = h_k - h_{k-1}. \quad (29)$$

Therefore, $r - 2$ blocks are cryptographically unprotected in every array.

4. Trade-off Analysis

Every block in (16)-(17) requires *four* exponentiations for encryption and decryption. In contrast, in the protocol A.2-L.2 described above, (4)-(11), the *array* of r blocks requires only one exponentiation for its encryption and decryption. Therefore, the larger the transmitted array r is, the more efficient the speed-up of A.2-L.2 is. If r = 100, then A.2-L.2 is *four hundred* times faster than the RSA algorithm.

Furtermore, if $n_B \le m \le n_A$, then the RSA digital signature algorithm (16)-(17) fails to recover the original plaintext m unless special measures are taken [4,11]. The application of entanglements (linear or non-linear transformations) is a tool that is proposed to accelerate the encryption-decryption process. Although the entanglements themselves do not provide protection, yet, when used in combination with other measures, they decrease the amount of computations necessary for the entire encryption/decryption process.

It is necessary to mention that any detailed and credible *quantification* of the trade-off between the size r of the array and cryptoimmunity requires analysis of all strategies potentially available to the intruder. Yet to *qualitatively* illustrate this point of view, let's consider an asymptotic case, where the size r of the transmitted array of plaintext blocks is very large. From one point of view, the larger r is, the more advantageous the proposed cryptosystem is. Indeed, only one entanglement is encrypted/decrypted instead of all r entanglements as it is done in the RSA, ElGamal, Rabin [6], ECC [7-9] and other PKC cryptosystems [10]. On the other hand, if the size r of the array is very large, then the intruder can invest the required time and computing resources to cryptanalyze the encrypted entanglement.

Let's consider an extreme case, where the entire message M consists of N blocks. Let's select a square non-singular $N \times N$ matrix E, compute N entanglements $h_1, h_2, .., h_N$ using (18) and encrypt only one of them, say, h_1. For instance, if the sender transmits information regarding highly-sensitive issues of long-term national policy or the details of a major corporate policy, the intruder will invest all available resources to break the encrypted entanglement h_1 [12-18]. Therefore, for security purposes, it is safer to divide the entire file M onto several parts/arrays and securely protect each array.

5. Decryption: Reduction of Complexity

The most serious computational bottleneck of the present public-key cryptographic protocols is that they are notoriously slow and therefore cannot be used in the real-time exchange of sensitive information.

Although we are far away from completely eliminating this bottleneck, the proposed cryptosystem is a systemic tool that accelerates secure communication via open channels of the Internet or within corporate networks.

Eliminating the bottleneck mentioned above is one of major research areas today and will likely occupy hundreds of communication specialists and system designers for years ahead. Various PKC algorithms were introduced in the last thirty years. Elliptic-curve cryptography and its hyper-elliptic extension are vivid examples of this research: to accelerate the encryption/decryption process. The proposed cryptosystem is another illustration of how we can accelerate the PKC protocols if the entangled arrays rather than individual blocks are encrypted.

6. Illustrative Numeric Example

The steps A4-H4 describe a system *design* stage and the steps I4-L4 describe its implementation for signed encryption and authenticated decryption of arrays

$$m = (a_1, .., a_r):$$

A4. Let Alice and Bob select p = 1907, a generator g = 1430, (5), and $\alpha = 2/3$, (13-15);

B4. Let each Alice and Bob select two pairs of primes: $\{p_A, q_A\} = \{29, 47\}$ and $\{p_B, q_B\} = \{17, 89\}$ where

$$p_A \equiv q_A \equiv p_B \equiv q_B \equiv 2 (\text{mod } 3), \quad (30)$$

and compute their products [1]:

$$n_A := p_A q_A = 1363 \text{ and } n_B := p_B q_B = 1513; \quad (31)$$

then $\{p_A, q_A, n_A\}$ is a triad of Alice's *private* keys and $\{p_B, q_B, n_B\}$ is a triad of Bob's *private* keys;

C4. {Establishment of a secret key w}: w must satisfy the inequality $w < \alpha p$; Alice and Bob randomly select secret integers a = 7 and b = 10 respectively and compute $u := g^a \text{ mod } p = 1601$;

and $y := g^b \text{ mod } p = 1733$;

D4: Alice transmits u to Bob, who transmits y to Alice;

E4. Alice and Bob compute respectively

$$w_A := y^a \bmod p \quad \text{and} \quad w_B := u^b \bmod p. \qquad (32)$$

As a result,

$$w_{AB} = w_A = w_B = g^{ab} \bmod p = 1118 \qquad (33)$$

is their secret key;

F4. Alice and Bob compute a multiplicative inverse w_{AB}^{-1} of their secret key $w_{AB} : w_{AB}^{-1} = 1281$ [11];

G4. {Alice requests Bob to send his private key n_B to her}:

Bob computes *v* and sends it to Alice:

$$v := n_B w_{AB} \bmod p = 25;$$

H4. Alice recovers Bob's private key:

$$n_B = v w_{AB}^{-1} \bmod 1907 = 1513;$$

I4. Suppose that Alice and Bob select their public keys $e_A = e_B = 3$.

Consequently, $d_A e_A \bmod z_A = 1$;

and $d_B e_B \bmod z_B = 1$;

imply that $d_A = 909$; and $d_B = 1009$.

J4. Suppose Alice intends to transmit to Bob over the Internet an encrypted array

$$m := \{324, 241, 332, 108, 412\}$$

with her digital signature.

If she selects the entanglements (13), then

$$h = \{1234, 500, 568, 1350, 1588\}.$$

If $\alpha = 2/3$, then h_1 satisfies the requirement (15);

K4. Alice encrypts h_1:

$$c_1 := h_1^{e_B} \bmod n_B = 1476,$$

and transmits $(c_1, h_2, h_3, h_4, h_5)$ to Bob;

L4. Bob decrypts the ciphertext c_1:

$$x := c_1^{d_B} \bmod n_B = 1234 \{= h_1\};$$

M4. Using (14), Bob recovers $h = (h_1, .., h_5)$. Because nobody except Bob knows his private key n_B, only he can recover the correct values of all plaintext blocks. If the recovered message is intelligible, Bob accepts it as the authentic message from Alice.

Preliminary results of this paper are published in [19].

7. Conclusions

This paper describes a novel concept for the PKC that employs a combination of DLP, factorization and entanglements, which facilitates otherwise computationally difficult problem [12,14,19].

Let's summarize the most important issues that were described and briefly discussed in this paper:

A: In contrast with RSA, n_k is a private key of the *k*-th user, not the public key [19];

B: In another contrast, the encryption/decryption is applied not to every block of the plaintext, but to every *array* of blocks; in other words, the ***unit of protection*** is not a block, but an array of several blocks [20];

C: Within each array prior to encryption all blocks are entangled [1];

D: The advantage of entanglements is that they are interdependent; the disadvantage is that if one entanglement is corrupted, it affects the entire array. Namely, that array cannot be recovered by the receiver [2];

E: If the information is transmitted in an aggressive media and subject to networking failures or errors, the proposed cryptosystem cannot be used unless additional measures of information assurance are applied (see [21,22]).

F: As a by-product of interdependence, there is no necessity to encrypt and decrypt each block or each entanglement. Instead it is sufficient to encrypt only one of *r* entanglements [23]. This is the first advantage of the proposed protocol.

G: The application of cryptography based on a cubic-root provides the second advantage. The encryption requires only two multiplications [1];

H: The overhead of the entanglements is on the stage of information recovery: it is necessary to solve a system of *r* equations with *r* unknowns. Yet, there are many ways how to select matrix *E* that will make these computations easier. Several linear and non-linear examples of entanglements are provided above for illustration. Additional examples of entanglements are described in [20]. The proposed cryptosystem also provides a digital signature protocol.

8. Acknowledgements

The author would like to express his appreciation to I. V. Semushin for assistance and to P. Fay for comments that improved the style of this paper.

9. References

[1] B. Verkhovsky, "Entanglements of Plaintext Streams and Cubic Roots of Integers for Network Security," *Advances in Decision Technology and Intelligent Information Systems*, Vol. IX, 2008, pp. 90-93.

[2] B. Verkhovsky, "Information Assurance Protocols: Efficiency Analysis and Implementation for Secure Communication," *Journal of Information Assurance and Security*, Vol. 3, No. 4, 2008, pp. 263-269.

[3] W. Diffie and M. E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory*, Vol. IT-22, No. 6, 1976, pp. 644-654.

[4] R. Rivest, A. Shamir and L. Adleman, "A Method of Obtaining Digital Signature and Public-Key Cryptosystems," *Communication of ACM*, Vol. 21, No. 2, 1978, pp. 120-126.

[5] B. Verkhovsky, "Deterministic Algorithm for Generators of Strong Primes," CS-06 Research Report, NJIT, 2006.

[6] M. O. Rabin, "Digitized Signatures and Public Key Functions as Intractable as Factorization," MIT/LCS Technical Report, TR-212, Cambridge, 1979.

[7] V. S. Miller, "Use of Elliptic Curves in Cryptography," *Lecture Notes in Computer Science*, Vol. 218, No. 85, 1985, pp. 417-426.

[8] N. Koblitz, "Elliptic Curve Cryptosystems," *Mathematics of Computation*, Vol. 48, No. 177, 1987, pp. 203-209.

[9] N. Koblitz, A. Menezes and S. Vanstone, "The State of Elliptic Curve Cryptography," *Designs, Codes and Cryptography*, Vol. 19, No. 2-3, 2000, pp. 173-193.

[10] N. Koblitz, "Hyperelliptic Cryptosystems," *Journal of Cryptology*, Vol. 1, No. 3, 1989, pp. 139-150.

[11] B. Verkhovsky, "Overpass-Crossing Scheme for Digital Signature," *International Conference on Systems Research, Informatics and Cybernetics*, Baden-Baden, Germany, July 29-31, 2001.

[12] J. M. Pollard, "Monte Carlo Methods for Index Computation Mod P," *Mathematics of Computation*, Vol. 32, No. 143, 1978, pp. 918-924.

[13] V. I. Nechaev, "Complexity of a Deterministic Algorithm for the Discrete Logarithm," *Mathematical Notes*, Vol. 55, No. 2, 1994, pp. 165-172.

[14] A. M. Odlyzko, "Discrete Logarithms: The Past and the Future," *Designs, Codes and Cryptography*, Vol. 19, No. 2-3, 2000, pp. 129-145.

[15] J. M. Pollard, "Kangaroos, Monopoly and Discrete Logarithms," *Journal of Cryptology*, Vol. 13, No. 4, 2000, pp. 437-447.

[16] D. R. Stinson, "Some Baby-Step Giant-Step Algorithms for the Low Hamming Weight Discrete Logarithm Problem," *Mathematics of Computation*, Vol. 71, No. 237, 2002, pp. 379-391.

[17] M. Chateauneuf, A. Ling and D. R. Stinson, "Slope Packings and Coverings, and Generic Algorithms for the Discrete Logarithm Problem," *Journal of Combinatorial Designs*, Vol. 11, No. 1, 2003, pp. 36-50.

[18] J. Coron, D. Lefranc and G. Poupard, "A New Baby-Step Giant-Step Algorithm and Some Applications to Cryptanalysis," *Lecture Notes in Computer Science*, Vol. 3659, 2005, pp. 47-60.

[19] B. Verkhovsky, "Fast Digital Signature Hybrid Algorithm Based on Discrete Logarithm, Entanglements of Plaintext Arrays and Factorization," 7^{th} *International Conference Mathematics Modeling in Physics, Technology, Socio-Economic Systems and Processes*, Ulyanovsk, Russia, 2009, pp. 13-16.

[20] B. Verkhovsky, "Information Assurance and Secure Streaming Algorithms Based on Cubic Roots of Integers," *In the Fifth International Conference on Information Technology*: *New Generations* (*ITNG*-08), Las Vegas, USA, 2008, pp. 910-916.

[21] B. Verkhovsky, "Control Protocols Providing Information Assurance in Telecommunication Networks," *Journal of Telecommunications Management*, Vol. 2, No. 1, 2009, pp. 59-68.

[22] B. Verkhovsky, "Selection of Entanglements in Information Assurance Protocols and Optimal Retrieval of Original Blocks," *Journal of Telecommunications Management*, Vol. 2, No. 2, 2009, pp. 186-194.

[23] B. Verkhovsky, "Accelerated Cybersecure Communication Based on Reduced Encryption/Decryption and Information Assurance Protocols," *Journal of Telecommunications Management*, Vol. 2, No. 3, 2009, pp. 284-293.

Int. J. Communications, Network and System Sciences, **2010, 3, 585-592**
doi:10.4236/ijcns.2010.37078 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Scientific Research

Analysing TCP for Bursty Traffic

Israfil Biswas, Arjuna Sathiaseelan, Raffaello Secchi, Gorry Fairhurst
Electronics Research Group (*ERG*), *University of Aberdeen, AB*24 3*UE, King's College, Aberdeen, UK*
E-mail: {*israfil, arjuna, raffaello, gorry*}*@erg.abdn.ac.uk*
Received April 8, 2010; *revised May* 15, 2010; *accepted June* 19, 2010

Abstract

The Transmission Control Protocol (TCP) has been designed to support interactive and bulk applications, with performance tuned to support bulk applications that desire to continuously send data. In contrast, this paper analyses TCP performance for a class of applications that do not wish to send continuous data, but instead generate bursts of data separated by application-limited periods in which little or no data is sent. In this context, the paper evaluates an experimental method, Congestion Window Validation (CWV), proposed to mitigate the network impact of bursty TCP applications. Simulation results show that TCP-CWV exhibits a conservative behaviour during application-limited periods. The results also show that TCP-CWV is able to use the available capacity after an idle period over a shared path and that this can have benefit, especially over long delay paths, when compared to slow-start restart specified by standard TCP. The paper recommends the development of CWV-like algorithms to improve the performance for bursty applications while also providing an incentive for application designers to use congestion control.

Keywords: Congestion Window Validation, Slow Start, TCP, Congestion Control

1. Introduction

TCP [1] provides an Internet transport protocol that has been designed to support a range of applications. A TCP sender encapsulates data to from TCP segments, which are sent as packets over the Internet. TCP also incorporates congestion control to limit the impact of each flow on other flows that share the network.

The standardized TCP congestion control [2] techniques maintain a record of the currently available capacity along a network path in a variable, known as Congestion Window (cwnd). Senders increase the number of TCP segments in flight each time positive feedback is received indicating that the current rate is not contributing to congestion, and reduce cwnd when it is perceived that the network may be congested as indicated by loss or congestion marking. This regulates the number of packets an application can inject into the network (*i.e.*, the transmission rate). In this method, received ACKs may be thought of as clocking-out new data [1] based on the concrete evidence that recent path capacity was available.

Recent years have seen a change in way many applications use TCP. There has been a significant growth in applications that are "bursty", that is applications that alternate periods of data transmission at a rate with limited by the application with periods where there is no, or little data to be sent. We call a class of applications that send at a rate lower than the actual available rate or at a rate controlled by the application, "application-limited" [3]. Applications that result in such traffic include online games, video conferencing, etc. This behaviour can also result when already deployed applications, such as when the hyper-text transfer protocol (HTTP) [4] is used with persistent connections. Accompanying the growth of bursty application there has been increased deployment of residential Internet [5].

VoIP and video streaming have become popular real-time applications. However, conventional perception is that TCP may be inappropriate for such applications because of congestion controlled reliable delivery may lead to excessive end-to-end delays. More than 50% of the commercial streaming traffic is carried over TCP [6]. As wide-deployment of Network Address Translators, NATs and firewalls often prevent popular media applications over UDP traffic, bursty applications such as Skype [7] and Windows Media Services [6] use TCP. Researchers [7,8] have evaluated the feasibility of constant bit rate (CBR) over TCP and motivated us to evaluate bursty applications performance using CBR traffic over TCP.

Any application-limited TCP flow sends fewer packet probes along the path than allowed by the cwnd. In this case the TCP sender cannot validate that the current value of the cwnd is appropriate by the reception of

ACKs. Therefore, standard TCP reduces the cwnd to the Restart Window (RW) when the TCP sender leaves an idle period, resetting the window to min (IW,cwnd) [9]. This results in poor performance for bursty applications. It also could result in under-utilised capacity for several round trip times (RTTs).

Standard TCP also unnecessarily increases the cwnd during an application-limited period, extending this beyond the size confirmed by reception of ACKs.

To support the congestion control [2] for bursty flows over networks with variable characteristics, the traditional congestion control methods need to be revisited. This includes selection of an appropriate TCP-friendly transmission rate [10] inter-flow and intra-flow fairness, multimedia congestion control. We suggest TCP should allow a flow to return, after an idle period, to a recent previously permitted transmission rate, providing there is no indication that the capacity has changed.

Congestion Window Validation (CWV) [3] is an experimental modification to the TCP congestion control that was proposed to partially solve the problem of an inappropriate cwnd value. CWV modifies TCP congestion control to affect behaviour in two circumstances: when a connection needs to resume transmission after an idle period, and when the flow sending rate is limited by the rate that the application generates data (*i.e.*, application-limited). In both cases, the current value of the cwnd cannot be validated by reception of positive feedback at the sender, since the number of packet probes along the transmission path is lower than the congestion window itself. In other words, the reception of an ACK does not provide evidence that the network path is able to sustain the transmission rate recorded in the cwnd.

CWV also modifies the TCP congestion control procedure by updating cwnd during application-limited to match the application rate. It saves the latest cwnd for use when the flow resumes after an idle or application-limited period.

Many operating systems adopt a conservative approach where TCP resumes transmission in the slow-start phase from a RW of only one packet immediately following an idle period [11]. This indicates that implementers may prefer the safe approach of RFC 5681 which obsoletes RFC 2581, rather than the more recent CWV proposed in RFC 2861. The remainder of this paper explores the limited success of CWV. The next section contains analysis on CWV. Section III analyses the performance of CWV using simulation. Section IV discusses CWV issues. Finally, the last section provides a conclusion.

2. Congestion Window Validation

To understand the response of standard TCP after an idle period, this section describes three application behaviours (this forms the basis of the tests performed in [11] and [12]).

We refer to the case where an application stops sending for a period less than the TCP Retransmission Time Out (RTO) as a "short idle" period [3]. In this case, standard TCP allows a sender to resume transmission at a rate constrained by the cwnd (*i.e.*, at the same rate of increase of sequence number with time). Therefore, TCP can potentially send a cwnd-size line-rate burst into the network after such an idle period. The hypothesis here is that the previously determined cwnd is still valid when the application resumes transmission.

An application that is idle for a period greater than the RTO using standard TCP must restart with slow-start [2]. This resets the cwnd to no more than the Restart Window (RW) and results in exponential growth of the sequence number with time up to the stored ssthresh value. Hence, TCP restarts the ACK clock as at the beginning of a transfer.

An application that stops sending for a period greater several RTOs should not make assumptions about the previous congestion state of the path that it was using, nor that it is necessarily using the same path. Hence, standard TCP recommends a sender that is idle over several RTOs should continue from the RW by also resetting the cwnd.

In [3], a technique was proposed to provide a conservative estimate of cwnd. If the TCP connection has been inactive (*i.e.*, no packets in flight) for a period larger than one retransmission timeout (RTO), cwnd is reduced by a half as many times as the number of RTTs the TCP sender had been idle. This is equivalent to exponentially decaying cwnd during the idle period. Since TCP halves cwnd each time a negative feedback is received (that is, at most once per RTT), CWV provides a safe value for cwnd, which is then expected to be validated by the reception of ACKs during the first round of transmission following the idle period.

TCP also resets the slow-start threshold (ssthresh) to max (ssthresh, 3 × cwnd/4), which keeps previous information on the available path capacity. Since TCP resumes with a cwnd larger than the restart window (RW), the TCP sender can quickly recover its previous transmission rate. CWV also suggests that an application that stops sending for a period greater several RTOs should make no assumptions about the previous congestion state of the path it is using. Hence, a sender using TCP-CWV will exhibit a performance resembling standard TCP.

When the TCP sender detects that the cwnd has not been fully used for a period larger than an RTO (e.g. observing that each packet transmission does not use a full cwnd with an empty transmission buffer for more than RTO seconds), cwnd is reduced to (cwnd + w_used)/2. Here, w_used is an estimate of the portion of cwnd that was effectively used by the application. Hence, this mechanism avoids growth of cwnd to an arbitrary larger

value than the window-size actually used (because of slow-start or congestion-avoidance cwnd increase) and allows cwnd to be validated by reception of ACKs by the sender.

Figure 1 shows the dynamics of cwnd for standard TCP and TCP-CWV. In this case, we consider an application that generates data at an application-defined rate, interspersed by idle times when the application is inactive. The figure shows that at point 'a' both standard TCP and TCP-CWV start using slow start. Point 'b' denotes the point where the maximum rate permitted by cwnd is less than the application rate. Standard TCP continues to grow the cwnd, while CWV does not increase cend above that corresponding to the used rate. At point 'd', there is an idle period greater than a RTO, CWV reduces cwnd by a half, but standard TCP resumes from the RW (point 'c').

The discussion so far has focused on stable paths, and it may be argued that path conditions often remain relatively stable, at least for periods of several minutes [13]. However there are also cases where the Internet path characteristics can change (e.g., routing topology changes [14], mobility changes [15], intermittent wireless links) or a change in traffic scenarios could invalidate the congestion window. This may mean that a previously safe rate may become unsuitable, if too a long a time has passed since the network the path was last used. This may also be a cause of an inappropriate cwnd.

To explore this, we examine a highly dynamic network scenario where, not only significant variations in traffic rate, but also changes in the transmission path (due for instance to terminal mobility) modifying capacity or delay characteristics of a path may happen. The problem is that any path change experienced while the sender was idle could result in a significant increase of drop rate. It is therefore important to assess whether it is reasonable to allow an application to send faster after idle. Hence,

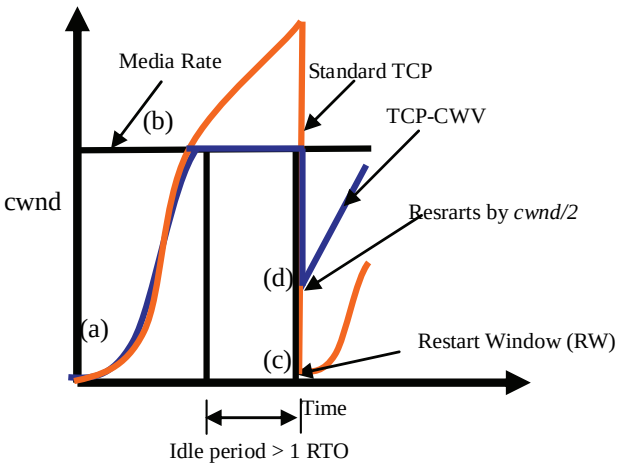


Figure 1. Illustration of cwnd dynamics for standard TCP and TCP-CWV.

the assessment could contribute to limited transient congestion in times of change, in return for improving application responsiveness. Next section analyses the impact of these methods that send faster after idle period in transient conditions using TCP-CWV.

The section explores this hypothesis for a set of simultaneous bursty flows that share a single bottleneck path. We analyse performance in a severely congested scenario for both idle period and application-limited period case.

3. Simulation Analysis of CWV

This section compares the performance of TCP-CWV and standard TCP following a significant change of the path characteristics.

The section considers two cases for analysis: 1) several idle TCP connections restart simultaneously (idle-period case), 2) several application-limited TCP flows simultaneously subjected to a sudden variation of application sending rate (application-limited case).

3.1. Idle Period Case

We analyse the performance of CWV after an idle period in two cases:

1) when the bottleneck is shared between equal numbers of standard TCP and CWV flows (heterogeneous-flow scenario). Hence, heterogeneous flows are mixed flows that are competing with flows using one alternate algorithm at a time.

2) when CWV or standard TCP only flows are present (homogeneous-flow scenario). Hence, all flows use one congestion control algorithm. Here a path is occupied by one kind or same type of flows and not sharing with other types of flow.

All simulations used a large advertised receiver window (1.5 MB) so that the TCP sender could send constrained only by the cwnd or CWV. Hence, flows at steady-state are interrupted and restarted after a short period of time. The interrupt and restart time of each flow was chosen from a uniform random distribution. Varying the duration of the idle period allowed investigation of CWV behaviour compared to standard TCP. **Table 1** summarises the simulation parameters.

Figure 2 illustrates the simulation topology where multiple TCP flows [S_1, S_2 …S_n, S_{n+1}] contribute traffic at the node n_0 and destinations are [R_1, R_2 …R_n, R_{n+1}]. These flows used a rate appropriate to medium quality video [16] over IP with an encoding rate of 512 kbps (packet size of 1500 bytes). To reach the destination via node n_3, one path is n_4-n_5 (capacity of 100 Mbps) and the alternate path is n_1-n_2.

Assuming a scenario where there is a path break on the path n_4-n_5 and n_0 chooses n_1-n_2 with a currently

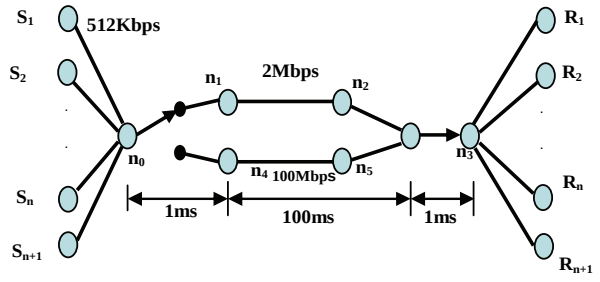


Figure 2. Single bottleneck topology.

Table 1. Configuration parameters for idle period simulations.

Bottleneck Link	
Bandwidth (Mb/s)	100
One-way Link Delay (ms)	100
Router Buffer Size	BDP
Access Links	
Bandwidth (Mb/s)	100
One-way Link Delay (ms)	1
TCP Configurations	
Maximum Segment Size (B)	1460
Maximum Advertised Window Size (kB)	1500
Minimum Retransmission Timeout (sec)	1
Simulation duration (sec)	40
CBR traffic Parameters	
Size (bytes)	1460
Rate (Kbps)	512
Idle period	
Duration of periods (sec)	0.5 to 5
Changed Bottleneck Bandwidth (Mb/s)	2
Flow/Drop monitor duration (RTT)	5

available capacity of 2 Mb/s. We assume both paths have the same path delay of 100 ms.

While TCP flows are idle, the common path changes from 100 Mb/s to 2 Mb/s. The rate of 2 Mb/s was chosen because if each TCP connection carries a 512 Kb/s constant bit-rate flow. Increasing the number of connections to 32 allows evaluation of the TCP performance under high congestion. When TCP is used for bursty applications, it is expected to match the application transmission rate or media rate. We measure the 'Average received rate', this is the arrival rate at the receiver over a short period of time. The performance is measured over 5 RTT following an idle period (as an indication of the goodness of restart response).

The less conservative approach of CWV after an idle period can result in more packet losses compared to standard TCP.

The packet drop rate was evaluated at the bottleneck as an indication of protocol aggressiveness. This allowed investigation of the trade-off between user-perceived performance, reflected by TCP received packet rate, and network performance (*i.e.*, the loss rate). This drop-rate is only relevant in the homogeneous-flows scenario, where the cause of buffer overflows can be attributed to the investigated protocol.

3.1.1. Heterogeneous-Flows Scenario

When the idle period is less than one RTO (about 0.5 sec in the simulations), cwnd remained unchanged using both TCP-CWV and standard TCP. These simulations confirm that the two protocols achieve the same performance. However, when the idle period is larger than one RTO, TCP-CWV achieves better performance in terms of packet arrival rate at the receiver.

Figure 3 shows performance for an idle period of 1.5 sec with the heterogeneous-flows. In this case, standard TCP resumes from the RW (one packet in this experiment) and performs slow-start, whereas CWV restarts from a level significantly larger than RW. Finally, if the idle period is larger than several RTOs, the simulations show that TCP-CWV achieves the same performance as standard TCP and that CWV reduces the cwnd to the RW as in standard TCP.

3.1.2. Homogeneous-Flows Scenario

When TCP flows compete with flows of the same type (*i.e.*, using the same congestion control algorithm), similar behaviour to the heterogeneous-flow case is observed. That is, when the idle period is smaller than an RTO or sufficiently large (e.g., 5.0 sec), standard TCP and TCP-CWV achieves the same performance in received rate and drop rate. Whereas for an idle period of few RTOs (e.g., 1.5 sec) TCP-CWV outperforms standard TCP in terms of achieved bit rate (**Figure 4**). However, this also produces higher congestion at the bottleneck as illustrated by the drop rate graph (**Figure 5**).

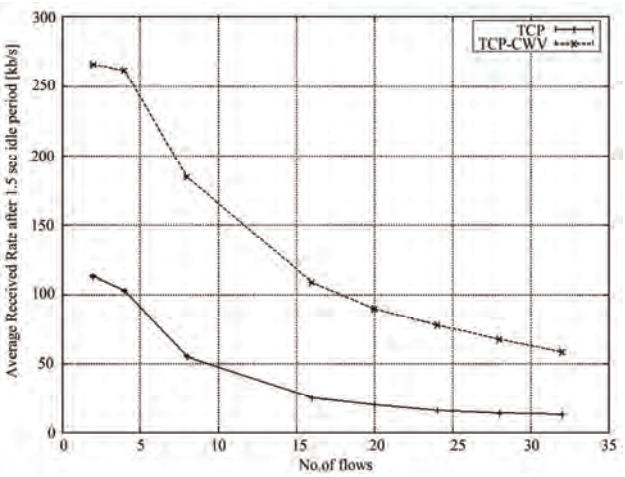


Figure 3. Average received rate vs. number of flows for the heterogeneous flows after 1.5 sec idle period.

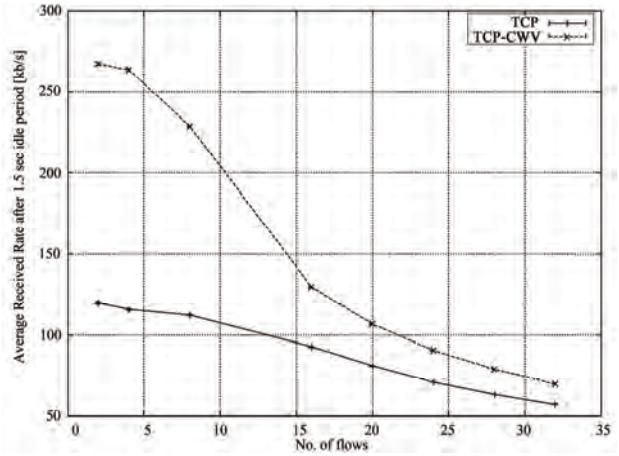


Figure 4. Average received rate vs. number of flows after 1.5 sec idle period for homogeneous-flows.

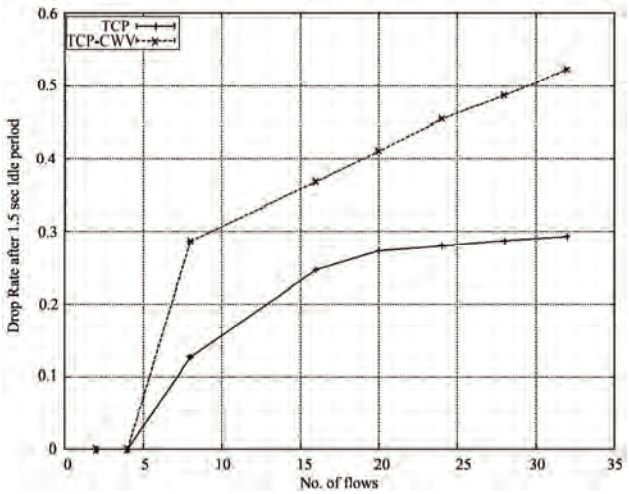


Figure 5. Drop Rate after 1.5 sec idle period for homogeneous-flows.

In conclusion, simulations show CWV restarts quickly and hence higher received rate compare to standard TCP. CWV shows best performance over a less congested path by dropping fewer packets. The heterogeneous case shows a higher received rate after an idle period CWV utilises more bottleneck capacity than standard TCP. A quick restart helps CWV to be fair to itself, and also shows fairness to standard TCP by reducing the rate similar to standard TCP rate in the longer period.

3.2. Impact over Long Network Path

We also considered the performance of standard TCP and TCP-CWV over a Long Fat Network (LFN) [17], specifically one with a long network path. Hence, in this experiment, the one-way delay is increased to 300 ms (*i.e.*, more than 600 ms RTT). CWV has an advantage of quick restart after an idle period. **Figure 6** shows the average received rate and **Figure 7** shows the loss rate for the

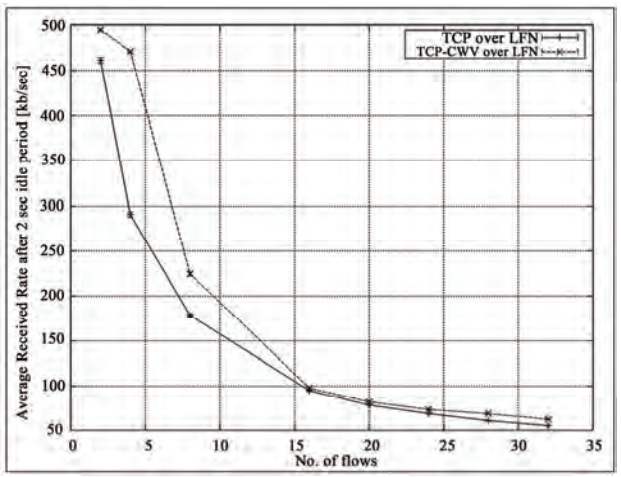


Figure 6. Average received rate vs. number of flows after a 2 sec idle period with homogeneous-flows over a long network path.

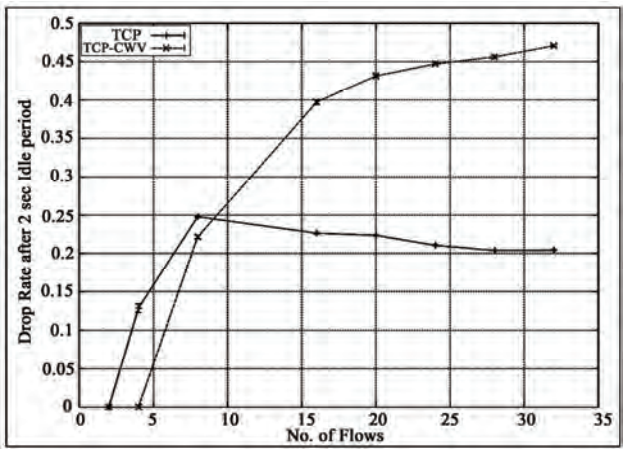


Figure 7. Drop Rate 2 sec idle period in homogeneous-flows over a long network path.

scenario with homogeneous-flows.

This shows that TCP-CWV is able to achieve a higher received rate at moderate congestion without severe router buffer overflow. The graphs also show that TCP-CWV allows a sender to maintain a cwnd sufficiently large to utilise the capacity after an idle period.

3.3. Delay Jitter

Figure 8 shows the one-way delay measured as the application-generated packet time at the sender and reception time at the receiver socket buffer by the application against the packet sequence number. The queuing delay induced by TCP-CWV is substantially decreased with respect to standard TCP for a restart time after an idle period longer than one RTO. This benefit can be observed providing that cwnd is not reduced to RW, which

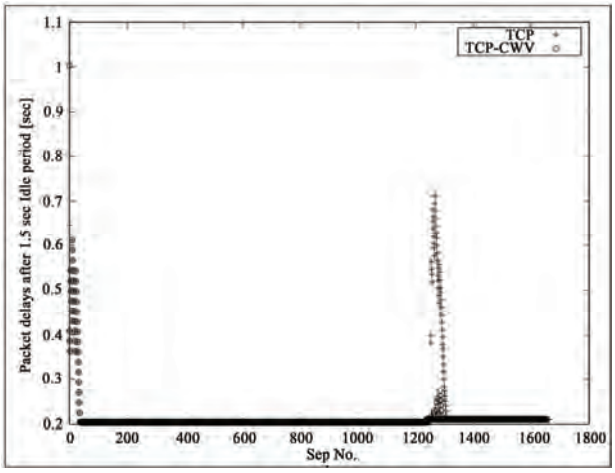


Figure 8. Average one-way delay over 25 standard TCP and TCP-CWV connections in the homogeneous-flows scenario (1.5 sec idle period and 200 ms propagation delay).

is for idle periods of up to several RTOs. This result illustrates that the TCP-CWV flows suffer high delay only at the beginning of a connection because of the three-way handshake and the initial slow-start phase.

The maximum delay variation is strongly dependent on the duration of the idle period. In this simulation, the best performance improvements are observed between five to twenty-two RTTs of idle period duration. For less than five or shorter RTT, CWV performs the same as standard TCP (*i.e.*, there is no reduction of cwnd) and after a longer RTT (e.g., more than 20 RTTs) CWV decays the cwnd to a value almost equal to that of the initial window of standard TCP.

3.4. Application-Limited Period

The increase of cwnd allowed by standard TCP during an application-limited period can be inappropriate (*i.e.*, allowing the cwnd to grow unnecessarily while transmitting at a rate lower than the available capacity). If there is a change in application transmission rate, this behaviour could lead to many packet drops. This is because the value of cwnd no longer reflects the current path capacity. CWV was proposed to enhance congestion control by continuously updating cwnd during application-limited periods to avoid unnecessary increase during application-limited periods.

To create an application-limited scenario, we arrange for the sending rate of a CBR application to be lowered to a steady-state from 512 kb/s to 12 kb/s (for instance, when a high bit rate media flow switches to a low bit rate media flow). After an interval of several RTOs, the application rate is restored to its original rate.

At the same time, the shared link capacity is changed from 100 Mb/s to 2 Mb/s, as in the previous set of simulations. The average rate of packet arrival at the receiver was measured (over a several RTOs starting from the capacity discontinuity) as a measure of the response time, and the packet drop rate at the bottleneck as a measure of capacity sharing of the congestion control protocol. **Table 2** shows the configuration parameters and we used the same topology of **Figure 2**.

3.4.1. Heterogeneous-Flow Scenario

Results for an application-limited scenario show that TCP-CWV is conservative and cannot respond quickly to a change in the application rate, whereas standard TCP allows TCP to immediately send additional packets (see **Figure 9**). Hence, application-limited period responses are opposite to the idle period responses for standard TCP and TCP-CWV.

Table 2. Configuration parameters for simulations with an application-limited period.

Bottleneck Link	
Bandwidth (Mb/s)	100
One-way Link Delay (ms)	100
Router Buffer Size	BDP
Access Links	
Bandwidth (Mb/s)	100
One-way Link Delay (ms)	1
TCP Configurations	
Maximum Segment Size (B)	1460
Maximum Advertised Window Size (kB)	1500
Minimum Retransmission Timeout (sec)	1
Simulation duration (sec)	40
CBR traffic Parameters	
Size (bytes)	1460
Rate (Kbps)	512
Change of Rate-application-limited period (Kbps)	12
Application-limited period	
Duration of periods (sec)	0.5 to 5
Changed Bottleneck Bandwidth (Mb/s)	2
Flow/Drop monitor duration (RTT)	10

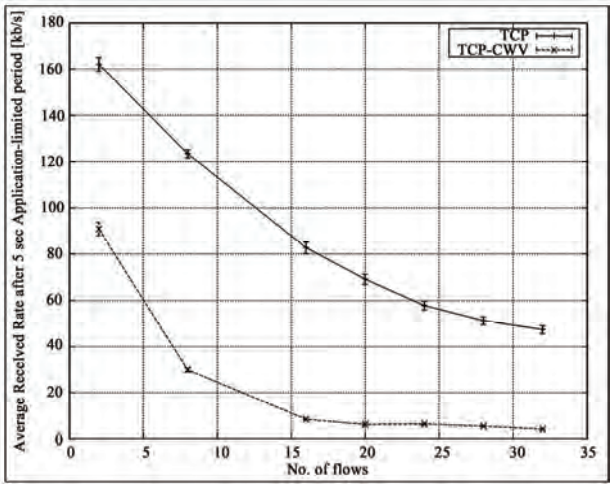


Figure 9. Average packet arrival rate at the receiver for standard TCP and TCP-CWV connections in the heterogeneous-flows scenario after 5 sec application-limited period and 100 ms delay path.

3.4.2. Homogeneous-Flow Scenario

Figure 10 confirms that in the case of homogeneous-flows standard TCP provides higher received rate than TCP CWV.

On the other hand, the inefficient cwnd growth offered by standard TCP results in a much larger packet drop rate (**Figure 11** highlights) with respect to CWV. Changing the size of the interval the application rate remains at low rate, we observed that the longer the interval, the lower the correlation between the cwnd and the bandwidth-delay product.

Finally, CWV has no benefit to the application because it offers a lower received rate compared to standard TCP. CWV also is less desirable for the shared bottleneck compared to current TCP. However, it is observed that the CWV application-limited period approach is safe for these transient events.

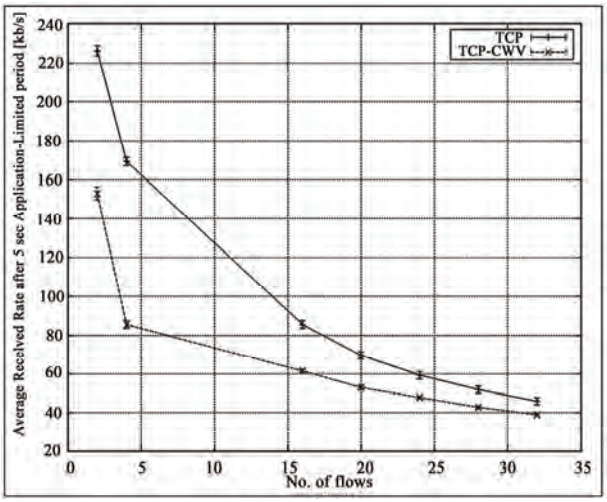


Figure 10. Average arrival rate at the receiver after 5 sec idle period in homogeneous-flows scenario for CWV and standard TCP connections.

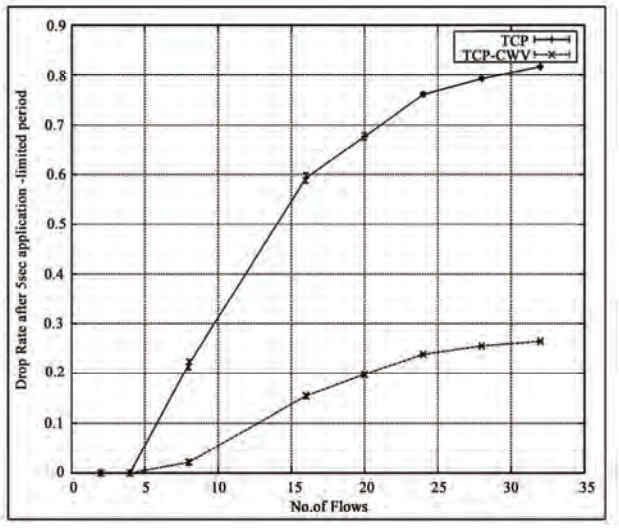


Figure 11. Drop rate after 5 sec application-limited period in homogeneous-flows scenario for CWV and standard TCP connections.

4. Discussion

Standard TCP takes a conservative approach during an idle period that lasts more than one RTO. It reduces the cwnd to the RW and then slow-starting back to the application rate. This approach is beneficial to the network, but does not benefit the application (as shown by the reduced average received rates for the case of Heterogeneous, Homogeneous flows and the packet drop rates in **Figures 3-5** respectively).

TCP-CWV mitigates the poor network performance of standard TCP by reducing the cwnd by one half for every RTO period that the application is idle. This benefits the application allowing an application to send packets much faster after a restart from an idle period. However, CWV impacts the network more if there is a transient event that changes the network path characteristics while the application was idle (shown by the increased average receive rates and packet drop rates in **Figures 4** & **5**).

During an application-limited period of more than one RTO, standard TCP behaves more aggressively compared to TCP-CWV. Standard TCP benefits an application-limited application by the faster sending, whereas TCP-CWV behaves more conservatively. The conservative approach of TCP-CWV ensures that in transient conditions, the impact caused by TCP-CWV flows restarting from a application-limited period is less compared to standard TCP (as shown by the lower average received rates and packet drop rates in **Figures 10** & **11**). **Table 3** shows the comparison as response after an idle or application-limited period.

Table 3. Comparison of response after idle or application-limited period: Standard TCP and TCP-CWV.

Approaches	Period	Standard TCP [RFC 5681]	TCP-CWV [RFC 2861]
1. Fairness to Application	Idle period	Conservative probing but losses application fairness.	Aggressive probing. Fair to the application.
	Appliction-limited period	Unnecessarily increases the cwnd, good for the application	Decay cwnd to utilise the capacity. Conservative approach, not fair to the application
2. Fairness to Path	Idle period	Safe probing. Good for the Internet path	Probing is not safe So, no benefit for the path.
	Appliction-limited period	Higher drops in transient events, bad for the path	Less drops to transient events So fair to the path

Our analysis of TCP-CWV poses a question: What is best for application designers that develop bursty applications? TCP-CWV would benefit an application if it exhibits regular idleness. However TCP-CWV would be of benefit only if the idle period was several RTOs. Applications exhibiting very large idle periods (tens of seconds) would experience no benefit from using TCP-CWV, since the behaviour would be the same as for standard TCP. Although TCP-CWV benefits the network in an application-limited scenario, the conservative approach of TCP-CWV does not provide an incentive to application to use this.

5. Conclusions

The current TCP specification defines a conservative slow start algorithm that can penalise an application which restarts from an idle period, making it undesirable for interactive bursty applications. TCP-CWV suggests a remedy to this problem, allowing a faster restart after an application was idle. This is seen to be beneficial to the application, and suggests the need for appropriate methods can be found, that balance the threat of network collapse against application performance. TCP-CWV exhibits a much more aggressive faster restart behavior after idle, however when an application is limited by the application rate, TCP-CWV has a much more conservative approach. Standard TCP has a more aggressive approach for application-limited flows. This non uniform approach of TCP-CWV has been a deterrent for it being deployed widely in the Internet, with TCP-CWV only deployed in the Linux OS (enabled by default).

Our future work will propose a new method(s) applicable to both idle and application-limited periods. We hope our work would lead to standards paving a way for application designers. The availability of methods that effectively support burst applications will provide an incentive for application designers to change to use a standard method to share the network resources in a more efficient and friendly manner.

6. References

[1] J. Postel, "Transmission Control Protocol," STD 7, RFC 793, September 1981.

[2] V. Jacobson, "Congestion Avoidance and Control," *ACM SIGCOMM Computer Communication Review*, Vol. 25, No. 1, 1995, pp. 157-187.

[3] M. Handley, J. Padhye and S. Floyd, "TCP Congestion Window Validation," RFC 2861, June 2000.

[4] H. F. Nielsen, J. Gettys, A. Baird-Smith, E. Prud'hommeaux, H. Lie and C. Lilley, "Network Performance Effects of HTTP/1.1, CSS1, and PNG," *Proceedings of the ACM SIGCOMM*'97 *Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, Cannes, France, September 1997, pp. 155-166.

[5] J. Heidemann, K. Obraczka and J. Touch, "Modeling the Performance of HTTP over Several Transport Protocols," *IEEE/ACM Transactions on Networking*, Vol. 5, No. 5, 1997, pp. 616-630.

[6] L. Guo, E. Tan, S. Chen, Z. Xiao, O. Spatscheck and X. Zhang, "Delving into Internet Streaming Media Delivery: A Quality and Resource Utilization Perspective," *Proceedings of the* 6*th ACM SIGCOMM Conference on Internet Measurement*, Rio de Janeriro, Brazil, 2006, pp. 217-230.

[7] B. Eli, B. S. Abdul, R. Dan and S. Henning, "The Delay-Friendliness of TCP," *Proceedings of the* 2008 *ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, Annapolis, MD, USA, 2008, pp. 49-60.

[8] S. Baset, E. Brosh, V. Misra, D. Rubenstein and H. Schulzrinne, "Understanding the Behavior of TCP for Real-Time CBR Workloads," *Proceedings of the* 2006 *ACM CoNEXT Conference*, Lisboa, Portugal, 2006.

[9] M. Allman, V. Paxson and E. Blanton, "TCP Congestion Control," RFC 5681, September 2009.

[10] M. Handley, S. Floyd, J. Padhye and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification," RFC 3448, January 2003.

[11] Md. I. Biswas and G. Fairhurst, "A Practical Evaluation of Congestion Window Validation Behaviour," 9*th Annual Postgraduate Symposium in the Convergence of Telecommunications, Networking and Broadcasting PGNet*, Liverpool, UK, 2008.

[12] Md. I. Biswas and G. Fairhurst, "An Investigation of TCP Congestion Window Validation over Satellite Paths," 4*th Advanced Satellite Mobile Systems Conference*, Bologna, Italy, 2008, pp. 37-42.

[13] H. Balakrishnan, S. Seshan, M. Stemm and R. Katz, "Analysing Stability in Wide-Area Network Performance," *ACM Sigmetrics Performance Evaluation Review*, Vol. 25, No. 1, 1997, pp. 2-12.

[14] J. Ni, H. Xie, S. Tatikonda and Y. R. Yang, "Network Routing Topology Inference from End-to-End Measurements," Technical Report, Yale University, 2007.

[15] A. C. Snoeren and H. Balakrishnan, "An End-to-End Approach to Host Mobility," 6*th ACM/IEEE International Conference on Mobile Computing and Networking*, Boston, Massachusetts, August 2000, pp. 155-166.

[16] S. A. Baset and H. Schulzrinne, "An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol," *IEEE INFOCOM*, Barcelona, Spain, 2006, pp. 1-12.

[17] V. Jacobson and R. Braden, "TCP Extensions for Long-Delay Paths," RFC 1072, October 1988.

Int. J. Communications, Network and System Sciences, **2010, 3, 593-601**
doi:10.4236/ijcns.2010.37079 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Scientific Research

Design and Analysis of a Multiple-Input Receiver for Mimo Wireless Applications

Constantinos Votis, Panos Kostarakis
Physics Department, University of Ioannina, Panepistimioupolis, Ioannina, Greece
E-mail: *kvotis@grads.uoi.gr, kostarakis@uoi.gr*
Received April 21, 2010; *revised May* 28, 2010; *accepted July* 1, 2010

Abstract

In this article, we present multiple-input receiver architecture for (Multiple-Input Multiple-Output) MIMO wireless communication applications. The proposed implementation is provided by a defined number of identical receiver units that are fed by a RF modulated signal on specific carrier frequency, power strength and initial phase. These units carry out the corresponding amplification, filtering and demodulation procedures. Details on design and implementation of this Printed-Circuit-Board are introduced and further discussed. Experimental results are also presented, allowing the validation of investigation on the performance of the current receiver architecture. Besides, these measurements indicate that the proposed device, combining with a suitable antenna array, provides a versatile receiver platform for baseband signal processing. The incoming RF modulated signals have frequencies on the range of 2.4 GHz, several phases, magnitudes and modulation modes. From these, it seems that the proposed receiver implementation supports MIMO communication and multiple port channel characterization applications at 2.4 GHz ISM (Industrial, Scientific and Medical) band.

Keywords: Channel Sounder, MIMO Systems, Baseband Processing

1. Introduction

Modern wireless communication systems continue to push for wider bandwidth capabilities, higher data rates and better quality of services. Scientific and engineering community provides a number of novel techniques and methods to meet these requirements. One of them is called Multiple-Input Multiple-Output (MIMO) architecture that could exploit the capacity of a wireless communication channel [1-3]. Using multiple antenna elements on both the transmitter and receiver ends offers significant capacity enhancement on radio propagation applications. In order to achieve this benefit, appropriate design aspects on such systems have to be taken into account. It is obvious that a receiver device with multiple input ports is mainly required. Furthermore, appropriate synchronization and data acquisition procedures have to be supported by this device in order to collect and record the data transmission streams from each sub-channel at any scattering radio propagation environment.

The efficiency of such systems depends on several performance and channel parameters. One of them is referred to the profound knowledge of the time-variant radio channel in various indoor or outdoor environments. Devices that could provide knowledge of the wireless channel status are referred as channel sounding systems. Furthermore, several multiplexing techniques are applied to these systems for channel estimation purposes. Time, frequency, code division multiplexing and hybrid methods are mainly used in these applications [4,5]. Generally, these devices improve MIMO system performance and offer crucial assumptions that provide a resource for channel model developments.

In addition, the hardware is crucial for the performance of such MIMO systems. Resolution accuracy and capability are dominated by the corresponding strategy adopted for the channel sounder and communication applications. In particular, the choice of the receiver architecture indicates the method of channel acquisition and estimation, as well as the efficiency of the MIMO communication system. More precisely, fully switched, semi-switched and parallel transmission are the main techniques that supports channel characterization applications, using one or a combination of the multiplexing methods (TDM, FDM, CDM), each with different advantages and drawbacks.

These methods also support MIMO communication systems, providing transmit and receive diversity and

channel capacity enhancements. Increment on data rates, decrement on bit error rates, independent sub-channels existence and quite excellent quality of services are the main advantages that are provided by them.

In order to design and construct a RF platform for MIMO channel sounder and communication applications, we propose this multiple-input receiver architecture. Existing experimental measurements set-ups are usually not able to probe a number of parallel streams, simultaneously. This feature is crucial for MIMO channel characterization and communication applications. In any case, the proposed receiver device corresponds to an experimental testbed that enables multiple channel applications, simultaneously. Due to this fact, an antenna array configuration is required for feeding the inputs of the proposed receiver implementation. Design and implementation aspects of this device and its complementary circuitry, including radio-frequency amplifying and down-conversion as well as calibration procedure are included in Section 2 of the current paper. In Section 3, the multi-channel receiver performance is also simulated on MIMO applications and further baseband signal processing techniques are also included, as well as the corresponding results are discussed; the paper concludes in Section 4.

2. Multiple-Input Receiver Design and Implementation

The proposed implementation employs broadband quadrature demodulator architecture and comprises of several identical units, which operate synchronously and are fed by a matched multiple-element antenna array. A schematic block diagram of this receiver architecture is depicted in **Figure 1**. Besides, **Figure 2** shows the corre-

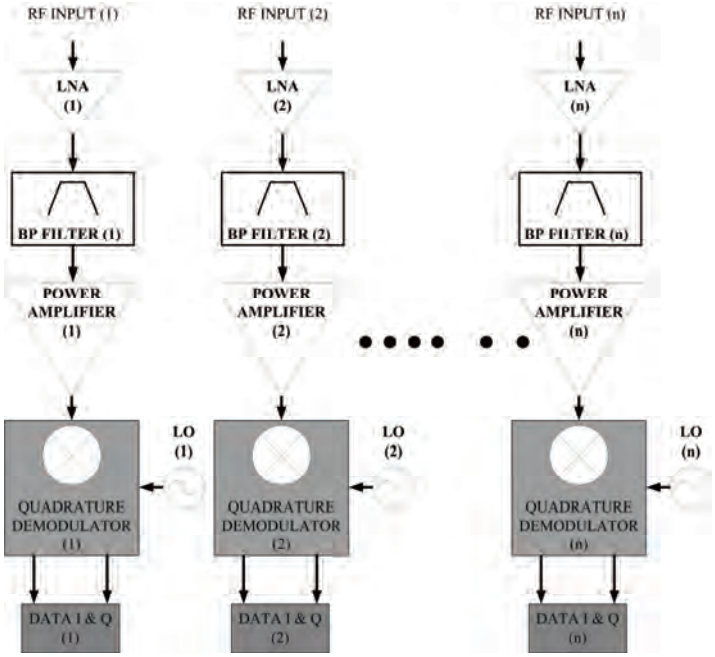


Figure 1. Generic receiver architecture.

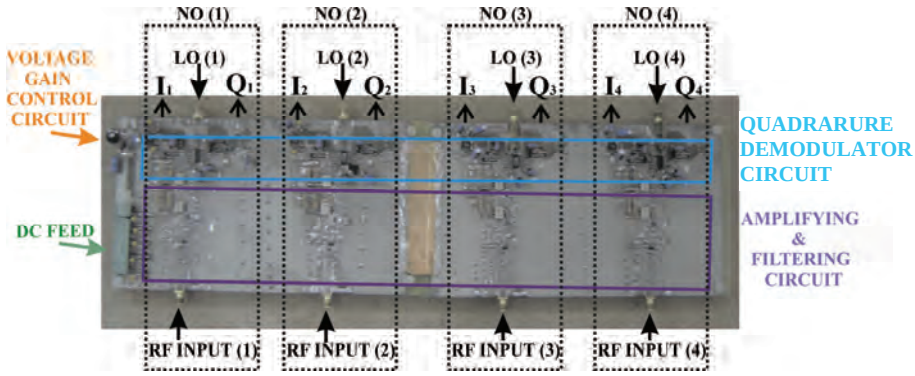


Figure 2. Receiver hardware implementation.

sponding Printed-Circuit-Board (PCB).

In **Figures 1** and **2**, it is obvious that the first stage, at each of the receiver units, is represented by an appropriate amplifying and filtering circuitry. This provides a quite great power gain, increasing the received signal strength and eliminating noise enhancements as the noise figures of these amplifying devices are quite low [6,7]. The corresponding measured SNR at the output of this stage approximates to 18.6 dB. In addition, the filtering circuit corresponds to a 2.45 GHz bandpass filter that decreases significantly the level of the signals that are out of the frequency range of interest. For further analysis, this amplifying and filtering circuitry was investigated in terms of the S-parameters. **Figure 3** shows the corresponding S21 parameter measurements. These results were provided by a Vector Network Analyzer and indicate the non-linear behavior of the current circuitry, giving a measure of unwanted phase and amplitude distortion that may be occured.

It is obvious that the offering power gain approximates to 42 dB for any value of the input (received) power. The corresponding small variations are quite negligible. Moreover, the AM to PM conversion is close to 0.2 deg/dB for the mean value of the input power (–62.5 dBm). The reverse transmission coefficient is also a significant parameter that corresponds to the isolation characteristics of the current circuitry. **Figure 4** depicts the measured results.

These results introduce a –57 dB isolation value that provides quite effective performance on the amplifying and filtering circuitry. Furthermore, **Figure 5** represents the input return loss.

From these measurements, it is obvious that a quite small amount of the input power is reflected. This fact is

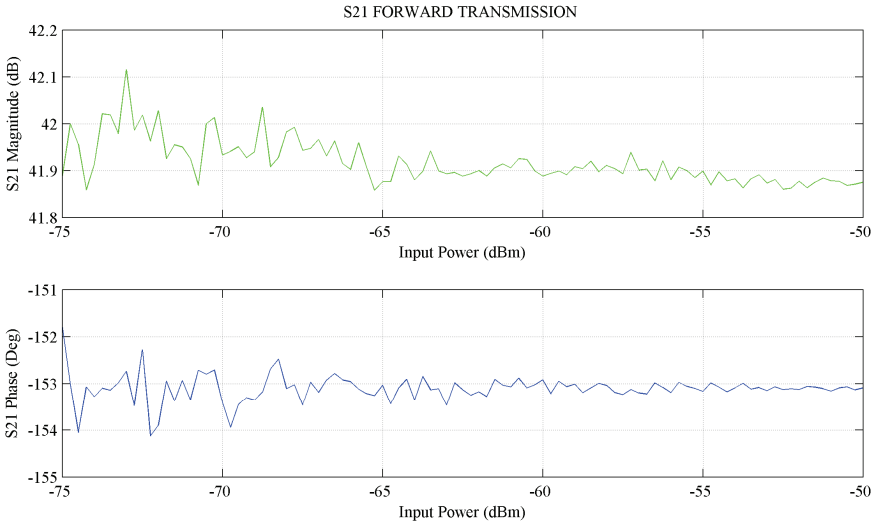


Figure 3. Transmission Gain of the first stage circuitry.

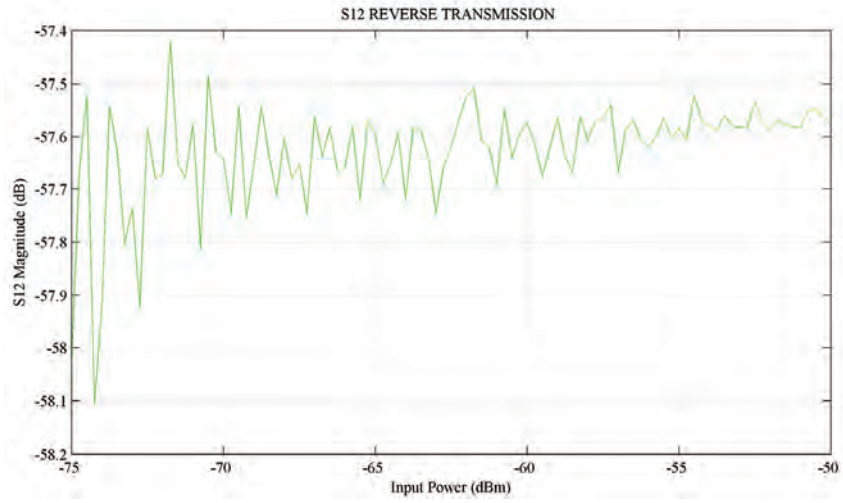


Figure 4. Isolation of the first stage circuitry.

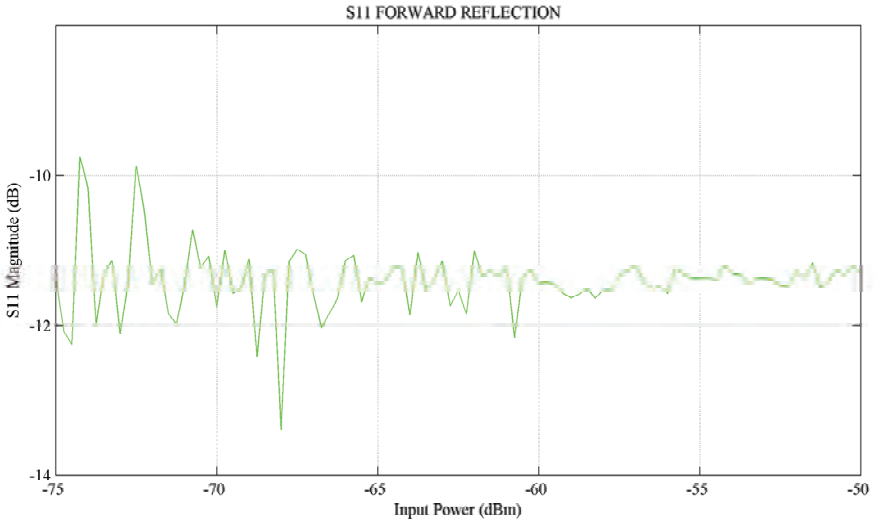


Figure 5. Input Return Loss of the first stage circuitry.

provided by the –11 dB return loss value on the dynamic input power range of each receiver unit.

These characteristics indicate that the received RF modulated signal is prepared to drive the corresponding quadrature demodulator input at each of the four identical, albeit independent units. Demodulation procedure necessitates four local-oscillator signals that are provided via by a 1-to-4 power divider by a frequency generator at the frequency range of 2.4 GHz. Besides, the demodulator circuits offer quite 69.5 dB gain control adjustment on the RF signal strength at their corresponding outputs. For this, a precision control circuit sets the linear-in-dB gain response to the gain control voltage. Furthermore, these demodulator integrated circuits employ polyphase filters to achieve high quadrature accuracy and amplitude balance over the entire operation frequency range [8]. Each of the units provides *Ix* and *Qx* output signals that correspond to the RF input signal, where *x* is an index which ranges from 1 to 4.

Test experimental measurements of the proposed implementation indicates constant amplitude and phase declinations presented on the signal outputs (*Ix, Qx*) due primarily to the different paths (transmission line lengths) from the LO source to the quadrature demodulator, as well as the demodulator inputs to the antenna array elements interface. These also include amplitude and phase errors from the coaxial lines that provide interconnection between RF inputs of the proposed receiver implementation with the elements of the antenna array [9]. These errors were measured and it is going to be taken into account in dynamic control on acquisition and data collection procedure.

3. Multiple-Input Receiver Performance

As noted above, the proposed multiple-input receiver implementation provides MIMO wireless communication and channel sounder applications in the frequency range of 2.4 GHz. For better analysis on the performance of this receiver device, a course of test measurements was made. For this purpose, signal generator platform was used to provide the RF inputs of the proposed implementation in order to simulate the signal reception of the corresponding antenna array. This equipment operates in transmit mode and provide four independent RF signals with several amplitudes, initial phases, frequencies and modulation modes. These signals are synchronized and drive the proposed receiver device for calibration and initialization purposes. Furthermore, a digital oscilloscope was also used to collect and store the *Ix* and *Qx* output signals, for further analysis.

At first, signal amplitude variations were investigated by an appropriate local-oscillator and RF input signaling. Each receiver unit was fed by a single-tone signal at frequency f_{RF} that differs from the LO signal frequency f_{LO} by the parameter *df*. The corresponding value ranges from 2 kHz to 10 kHz, with step 2 kHz, providing a single tone IF signal at the corresponding *Ix* and *Qx* output at frequency *df*. Collection and acquisition of them were achieved via the digital oscilloscope equipment.

In addition, voltage gain control unit supports power gain adjustments on the RF signal at each demodulator input. A constant mean level was chosen at the corresponding dynamic range. The measured results are depicted in **Figures 6-9** for each receiver unit, respectively. In these figures, the corresponding incoming RF power ranges from –80 dBm to –52 dBm.

From these figures, it is obvious that the signal strength of the output *Ix* is quite constant at the frequency range of 10 kHz for each value of *x* parameter. Besides, the quadrature nature of the current demodulator ensures that the signal strength of the output *Qx* var-

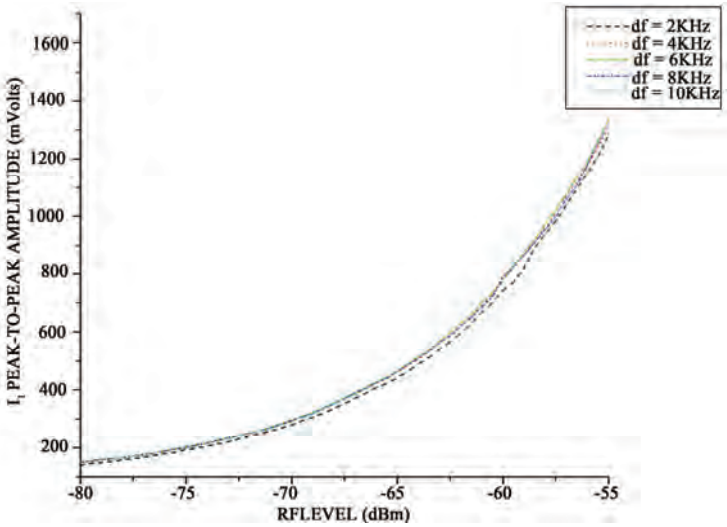


Figure 6. Output *I* signal strength at Receiver Unit 1.

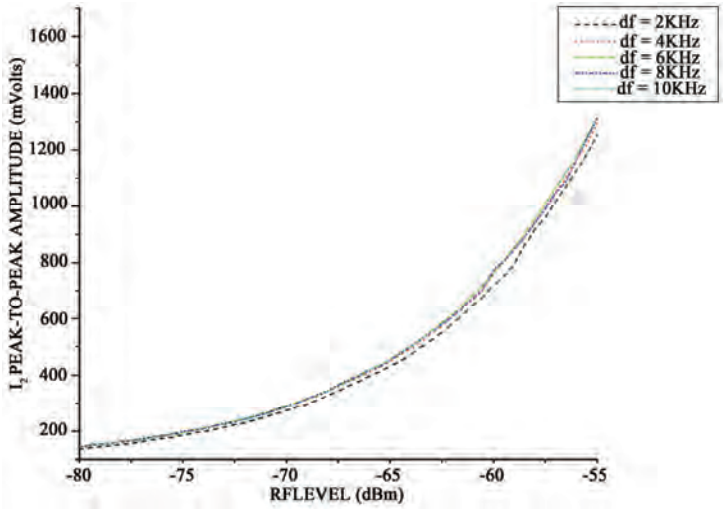


Figure 7. Output *I* signal strength at Receiver Unit 2.

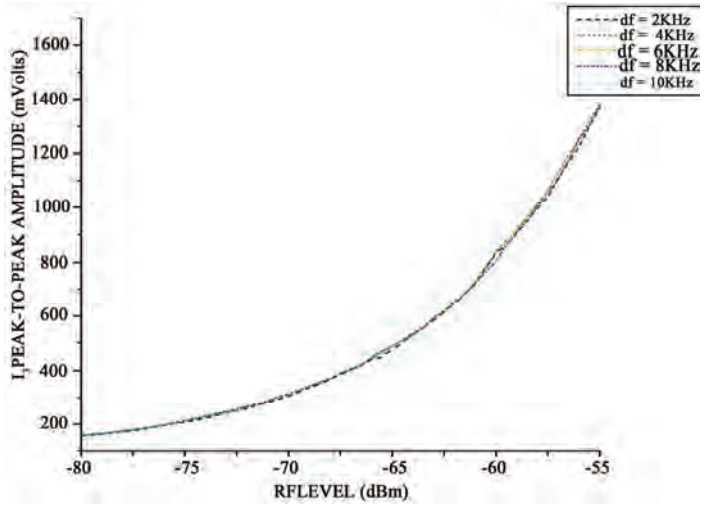


Figure 8. Output *I* signal strength at Receiver Unit 3.

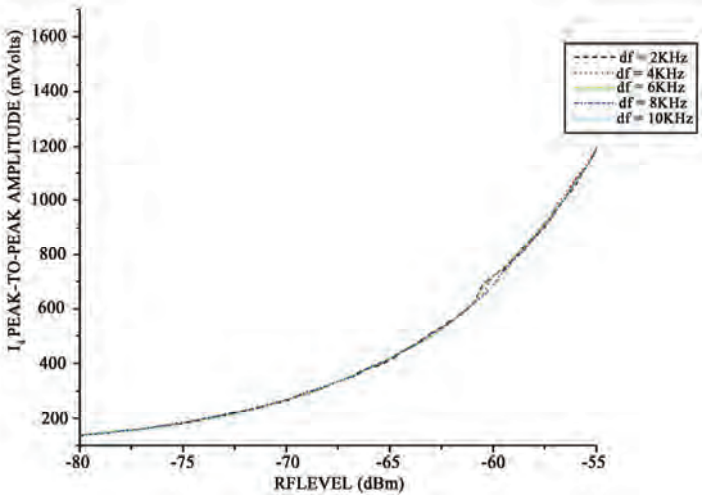


Figure 9. Output *I* signal strength at Receiver Unit 4.

ies with the same way as the corresponding *Ix*. There is only a 90 degrees phase shift amongst each pair *Ix* and *Qx* at each receiver unit. As these curves have quite identical forms at the frequency range of 10 kHz, we chose a frequency point of 6 kHz as the quite mean value on this bandwidth and we calculated the mathematical expressions that associates the output signal strengths with the RF received signal levels. **Figures 10-13** depict these variations at each receiver unit, in order to provide the corresponding mathematical calculations. The corresponding results are included in **Table 1**.

These mathematical calculations introduce an effective way to define the received signal power in dBm, when the value of *Ix* signal amplitude, expressed in voltage peak-to-peak, is obtained at each of the receiver units.

Furthermore, using the RF equipment that feed the proposed four channel receiver implementation, a course of calibration steps was made in order to investigate the phase difference on the receiving RF single tone signals. As mentioned above, these phase declinations are provided by the current receiver structure. For this, we used a pair of calibrated coaxial lines to connect the receiver RF inputs with the RF generator equipment. Two of the receiver units were fed by a single tone RF signal at frequency range of 2.4 GHz (f_{RF}), simultaneously. In each case, the third receiver unit was used as reference. The corresponding LO inputs were also fed by a single tone signal at frequency $f_{LO} = f_{RF} + df$, where *df* ranges from 1 kHz to 10 kHz with step 1 kHz. With this measured platform, we stored the phase difference *dphi* between the *Ix* output signals for several *df* values by a digital oscilloscope acquisition equipment. The corresponding results are presented in **Figure 14**.

As mentioned above, these *dphi* phase difference variations correspond to the frequency range of 2.4 GHz with 10 kHz bandwidth. The forms of these curves indicate that the parameter *dphi* is quite constant at this frequency range for each receiver unit. Moreover, receiver units 2 and 4 introduce quite identical phase shifts (24 degrees) with respect to the reference receiver unit 3. Instead, the receiver unit 1 provides phase shifting on the order of 60 degrees. These declinations on phase shift values are caused by transmission line differences on the PCB layout and by declinations on linearity at the particular amplifying, filtering and demodulator circuitry.

In order to investigate the performance of the proposed receiver architecture on full MIMO channel characterization, we used this device with an appropriate four-element antenna array. In particular, these experimental results were provided by a single-element antenna and a single-channel RF platform at the transmitter end, as well as the four element antenna array and the proposed receiver implementation at the receiver end. **Figures 15** and **16** depict the *Ix* output signals for two azimuth angle orientations of the antenna array at the receiver. These angles were 0 and 180 degrees and correspond to A and B antenna array orientations, respect-

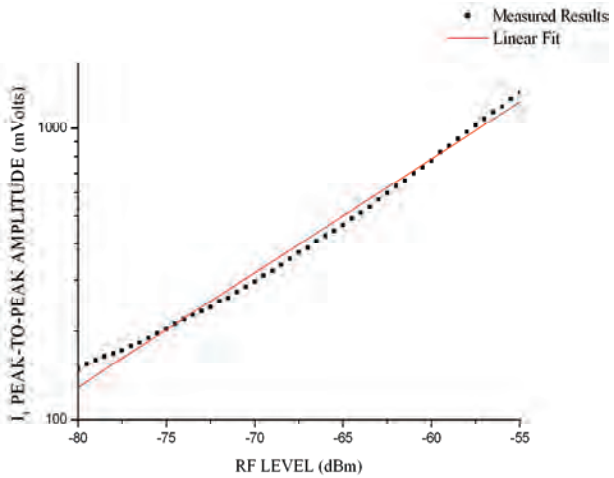


Figure 10. Linear fitting of the Output *I* signal strength at Receiver Unit 1 for *df* = 6 kHz.

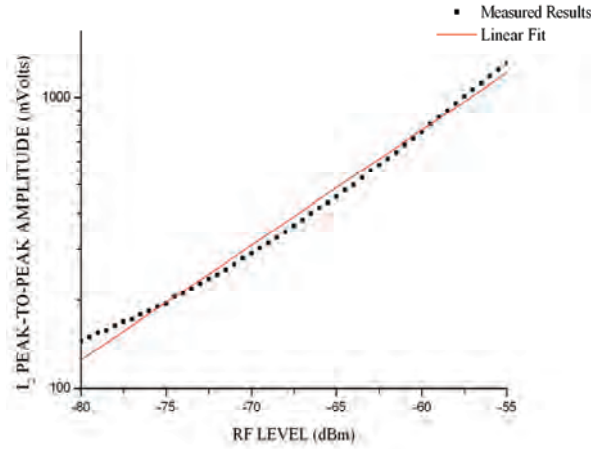


Figure 11. Linear fitting of the output *I* signal strength at receiver unit 2 for *df* = 6 kHz.

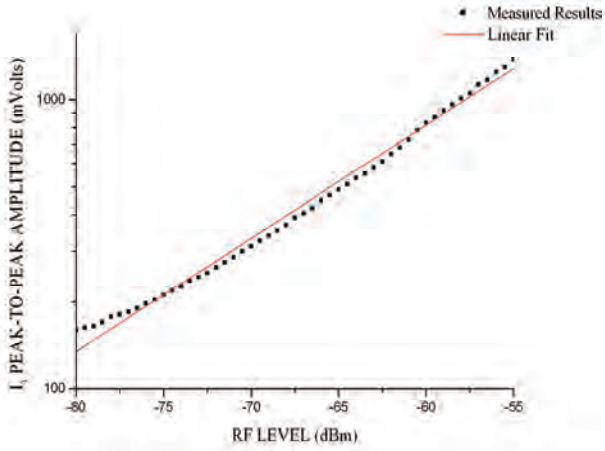


Figure 12. Linear fitting of the output *I* signal strength at receiver unit 3 for *df* = 6 kHz.

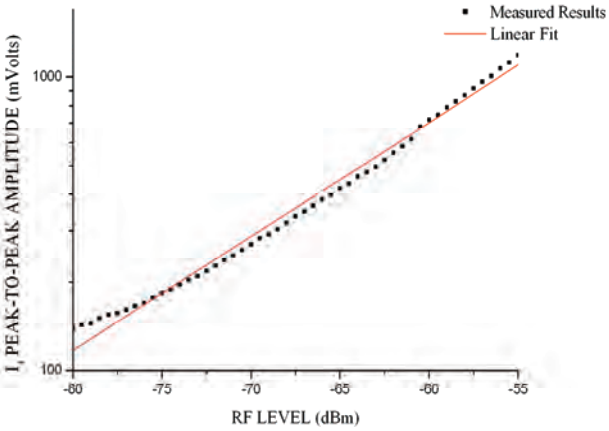


Figure 13. Linear fitting of the output *I* signal strength at Receiver Unit 4 for *df* = 6 kHz.

tively. In each case, the locations of the transmitter and receiver antenna, as well as the radio propagation environment were constant.

Table 1. Linear fit calculated results.

Expression **Y = A + B * X**				
Receiver unit	**A**	**Error of A**	**B**	**Error of B**
First	5.25	0.03	0.0393	0.0005
Second	5.26	0.03	0.0396	0.0005
Third	5.26	0.03	0.0392	0.0005
Forth	5.18	0.03	0.0389	0.0005

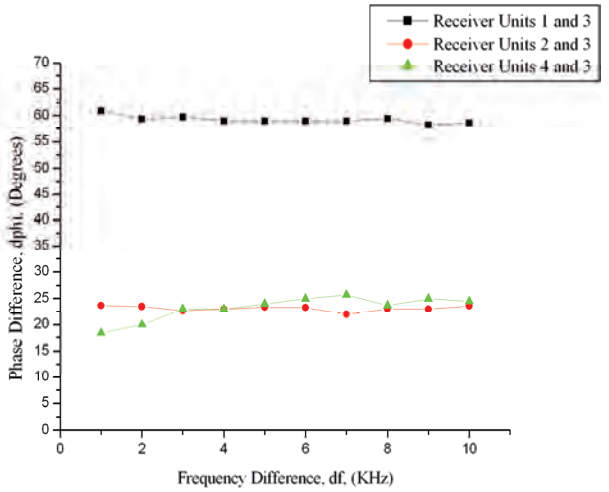


Figure 14. Phase declinations on the proposed receiver units.

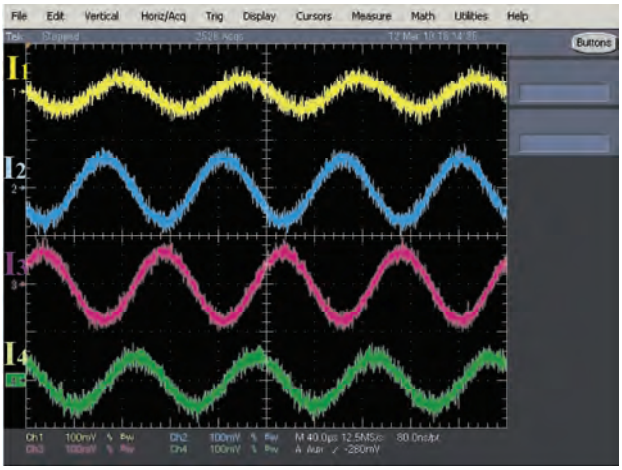


Figure 15. Output *Ix* signal strengths at Receiver for *df* = 10 kHz and orientation A.

From these representative measured results, it is obvious that the *Ix* output signals introduce amplitude and phase variations that correspond to the RF receiver signals. In fact, a single tone RF signal was transmitted and propagated via the multipath environment. As we used a four-element receiver antenna array, four copies of this RF signal were received. Each of them experiences different fading environment, decreasing the corresponding

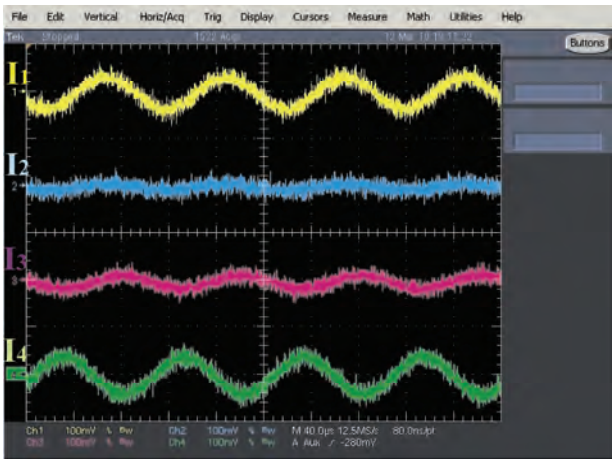


Figure 16. Output *Ix* signal strengths at Receiver for *df* = 10 kHz and orientation B.

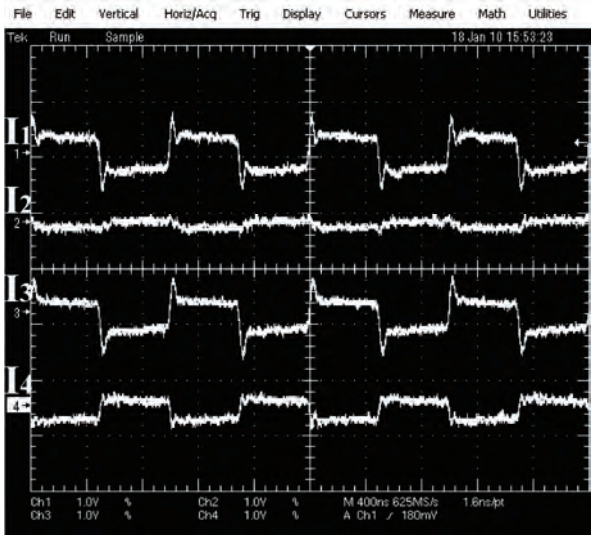


Figure 17. Data acquisition at BPSK modulation mode.

amplitude and changing its initial phase. Both these amplitude and phase variations are depicted in the corresponding *Ix* output signals of the proposed receiver device. These considerations indicate an interesting method of channel matrix calculations at several time snapshots of the radio propagation environment. As the proposed receiver implementation provides multiple-port applications and the *Ix* output signals are collected and recorded by a digital oscilloscope, simultaneously, we could exploit the experiment above, in order to study the full complex channel response between a multiple element antenna array at the transmitter and at the receiver ends, too. This procedure is known as MIMO channel sounder application.

Another issue is introduced by the quadrature demodulator devices that support the proposed receiver architecture. These integrated circuits provide direct demodulation, recovering the transmitted data stream. In particular, we used the transmitter RF platform in order to provide a BPSK signal at the frequency range of 2.4 GHz. This modulated signal was transmitted by a single element antenna and received by a four element antenna array that was connected with the corresponding inputs of the proposed receiver structure. These results are included in **Figure 17**.

From this figure, it seems that the same data stream could be recovered in the receiver end at each of the four units in the proposed implementation. The corresponding *Ix* output signals depend on the common data codeword and the radio propagation environment at each channel between the transmitter and receiver ends. In particular, I_1, I_3 and I_4 output signals provide the data codeword in desirable form but the I_2 output signal has quite negligible amplitude. Besides, there is a time delay between I_4 and both I_1 and I_3. All these observations are provided by the corresponding channel propagation and multipath fading environment.

4. Conclusions

The design and construction of a multiple-input receiver, using state-of the art quadrature demodulation technique for MIMO wireless communication and channel sounder applications have been presented and investigated. The performance of the receiver in terms of frequency, phase and amplitude accuracy, as well as modulation mode and synchronization has been further discussed. Experimental measurements introduce the performance characteristics of the proposed multi-channel implementation according to the MIMO application requirements. In conclusion, the receiver design represents a versatile and efficient implementation for modern wireless applications. This design and an appropriate antenna array structure provide a RF platform for MIMO communications and channel characterization applications.

5. Acknowledgements

This research project (PENED) is co-financed by E.U.-European Social Fund (80%) and the Greek Ministry of Development-GSRT (20%).

6. References

[1] H. Winters, "On the Capacity of Radio Communications Systems with Diversity in a Rayleigh Fading Envroment," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-5, No. 5, 1987, pp. 871-878.

[2] G. J. Foschini, "Layered Space-Time Architecture for Wireless Communications in a Fading Environment When Using Multiple Antennas," *Bell Labs Technical Journal*, Vol. 1, No. 2, 1996, pp. 41-59.

[3] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment When Using

Multiple Antennas," *Wireless Personal Communications*, Vol. 6, No. 3, 1998, pp. 311-335.

[4] M. H. Ullah and A. U. Priantoro, "A Review on Multiplexing Schemes for MIMO Channel Sounding," *International Journal of Computer Science and Netwotk Security*, Vol. 9, No. 6, 2009, pp. 294-300.

[5] S. A. Charles, E. A. Ball, T. H. Whittaker and J. K. Pollard, "A 5.5 GHz Channel Sounder for Fixed Wireless Channels," *Communications IEE Proceedings*, Vol. 150, No. 4, 2003, pp. 253-258.

[6] "Analog Devices Preliminary Datasheets of MAX2640."

[7] "Mini-Circuits Preliminary datasheets of ERA-5."

[8] "Maxim Preliminary Datasheets of AD8347."

[9] D. M. Pozar, "Microwave Engineering," Wiley, New York, 1998.

Int. J. Communications, Network and System Sciences, **2010, 3, 602-607**
doi:10.4236/ijcns.2010.37080 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Using Bayesian Game Model for Intrusion Detection in Wireless Ad Hoc Networks

Hua Wei, Hao Sun
Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an, China
E-mail: *wh*860127@163.*com*, *hsun@nwpu.edu.cn*
Received April 12, 2010; *revised May* 15, 2010; *accepted June* 22, 2010

Abstract

Wireless ad hoc network is becoming a new research fronter, in which security is an important issue. Usually some nodes act maliciously and they are able to do different kinds of Denial of Service (Dos). Because of the limited resource, intrusion detection system (IDS) runs all the time to detect intrusion of the attacker which is a costly overhead. We use game theory to model the interactions between the intrusion detection system and the attacker, and a realistic model is given by using Bayesian game. We solve the game by finding the Bayesian Nash equilibrium. The results of our analysis show that the IDS could work intermittently without compromising its effectiveness. At the end of this paper, we provide an experiment to verify the rationality and effectiveness of the proposed model.

Keywords: Wireless Ad Hoc Networks, Game Theory, Intrusion Detection System, Bayesian Nash Equilibrium

1. Introduction

A wireless ad hoc network (WANET) is a collection of mobile nodes in which the nodes communicate with each other without the help of any fixed infrastructure [1]. Nodes within each other's radio range communicate directly via wireless links, while those that are far apart use other nodes as relays. Because of the limited resource, some nodes may act selfishness. Ad hoc network misbehavior maybe inflicted by malicious nodes, each of which aims at harming the network operation; consequently, mechanisms that enforce security present a particular challenge. In order to avoid the harm of malicious nodes, one way is the use of an intrusion detection system, which watches out for any intrusion and sets out an alarm when an intrusion is detected. The intrusion detection and response mechanism is described in [2].

In recent years, we have seen researchers using game theory in the area of ad hoc networks. It is a powerful tool in that it can be used to model any system which exhibits the characteristics of a game. In WANET, mobile nodes typically have selfish motivations, lack of cooperation among themselves, and have conflicting interests with each other. These characteristics make game theory (GT) a promising tool to model, analyze, and design various aspects of WANET. We have given a two-player game to model the interactions between an intrusion detection system and an attacker in wireless ad hoc network. Each defender is equipped with an intrusion detection system (IDS) in order to monitor the activeness of an attacker.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 describes the one-stage game and multi-stage game, and Bayesian Nash equilibrium solutions are investigated. Section 4 presents numerical examples to verify the effectiveness of the proposed game. The conclusion of the paper is in section 5.

2. Related Work

Game theory has been successfully applied to many disciplines including economics, political science, and computer science. Game theory usually considers a multiplayer decision problem where multiple players with different objectives can compete and interact with each other. In the context of intrusion detection, several game theoretic approaches have been proposed to wired networks, sensor networks, and ad hoc networks.

Yenumula B. Reddy [3] discuss currently available intrusion detection techniques, attack models using game theory, and then propose a new framework to detect malicious nodes in wireless sensor networks using zero sum game approach for nodes in the forward data path. The first part of the research provides the game model with probability of energy required for transferring the data packets. The second part derives the model to detect the malicious nodes using probability of acknowledgement at source. Yuhan Moon, Violet R. Syrotiuk [4] present CCM-MAC, a cooperative CDMA-based multi-channel

medium access control (MAC) protocol for mobile ad hoc networks (MANET) in which each node has one half- duplex transceiver. They provide an analysis of the maximum throughput of CCM-MAC and validate it through simulation in MATLAB, and also compare the throughput it achieves to IEEE 802.11, a multi-channel MAC protocol, and a CDMA-based MAC protocol.

In [4] Hadi Otrok *et al*. address the problem of increasing the effectiveness of an intrusion detection system (IDS) for a cluster of nodes in ad hoc networks, and formulate a zero-sum non-cooperative game between the leader and intruder. They solve the game by finding the Bayesian Nash equilibrium where the leader's optimal detection strategy is determined. Finally, empirical results are provided to support their solutions.

Yu Liu, Cristina Comaniciu and Hong Man [5] have used static Bayesian game and dynamic Bayesian game to model the interactions between attacker and defender in ad hoc networks. They have shown that the static game leads to a mixed-strategy Bayesian nash equilibrium when the defender's belief of the attacker being malicious is high, and the dynamic game has a mixed-strategy Perfect bayesian equilibrium. In [6], they have used game theory for developing efficient defense strategies for a network with multiple IDSs. They have formulated a non-zero-sum, noncooperative attacker/defender game where the payoffs of players are non-strictly competitive. They have showed that the game achieves at least a Nash equilibrium that leads to a defense strategy for the defender.

A two-player, non-cooperative, non-zero-sum game has also been studied by Agah *et al.* [7] and Alpan and Basar [8] to address attack-defense problems in sensor networks. In their models, each player's optimal strategy depends only on the payoff function of the opponent and the game is assumed to have complete information. [9-11] have given the similar model, but the game is assumed to have incomplete information.

Our model is similar to the ones mentioned in the aforementioned works in that it is a two-player, non-zero-sum and noncooperative game. However, our work is not aimed at giving the best strategy of the defender. In this paper, we have given a one-stage game and multi-stage game. In the proposed works, the IDS of defender runs all the time, which is a costly overhead for a battery-powered mobile device since nodes have limited resource. The results of our model show that the IDS could work intermittently.

3. Bayesian Game

3.1. Game Model

In this section we present our game model. An IDS attempts to detect intrusion from an attacker. Hence, we may look at this as a game between two players, the IDS and the attacker. The attacker is denoted by i and IDS is denoted by j. The player i's intent is to attack the network without getting caught, whereas that of the player j is to detect intrusion when the attacker attacks. There is no cooperation whatsoever between the two players.

Player i has two types, regular that is denoted by $\theta_i = 0$ and malicious is denoted by $\theta_i = 1$. Node's type is his private information and IDS is uncertain about its opponent's type. IDS has only one type, that is regular or $\theta_j = 0$ and it is common knowledge for both players.

To present our model, we make the following assumptions. An IDS needs not be running all the time during which the wireless ad hoc network is up. The pure strategy space of this player is denoted by S_j = (Monitor t of the time, Not monitor), $t \in [0,1]$. The first strategy of player j depicts the situation when the IDS is active for some percentage (denoted by t). For example, if the IDS detects by monitoring the traffic, the IDS periodically monitors the traffic and the rest of the time, it sits idle. Likewise, an attacker need not be trying to attack 100% of the time. The malicious type of player i has two pure strategies: Attack s of the time and Not attack, $s \in [0,1]$. The regular type of player i has one pure strategy: Not attack. The two players choose their strategies simultaneously at the beginning of the game, assuming common knowledge about the game (costs and beliefs).

We first consider the scenario of the IDS. **Tables 1-2** illustrate the payoff matrix of the game in strategic form. In the matrix, a represents the detection rate of the IDS, b represents the false alarm rate of the IDS, and $a, b \in [0,1]$. In the **Table 1(a)**, the payoff matrix for the

Table 1. The type of player i is malicious.

(a) Payoff matrix of IDS.

$i \setminus j$	$S_j(1)$	$S_j(1)$
$S_i(1)$	$(2a-1)tsm-(1-t)sl-tc_d$	$-sl$
$S_i(2)$	$-btn-tc_d$	0

(b) Payoff matrix of attacker.

$i \setminus j$	$S_j(1)$	$S_j(1)$
$S_i(1)$	$(1-2a)tsm+(1-t)sl-sc_a$	$sl-sc_a$
$S_i(2)$	0	0

Table 2. The type of player i is regular.

$i \setminus j$	$S_j(1)$	$S_j(1)$
$S_i(2)$	$(0,-btn-tc_d)$	$(0,0)$

player j when player i is malicious is given. m denotes the overall gain of the player i for detecting the attack, and l is the overall loss for not detecting the attack during the whole lifetime. Costs of attacking and monitoring are denoted by c_a and c_d during the whole period. In our model, we assume that $m \geq l$ and $l \geq c_a$, c_d is reasonable since otherwise the player i does not have incentive to attack and the player j does not have incentive to monitor. The player j monitors t of the time, the player i attacks s of the time. The probability of the player j monitoring when the attack is on is ts, during which the player j gets a gain of tsm. Similarly, the probability of the player j not monitoring when the attack occurs is $(1-t)s$ because of which the player j loses an amount of $(1-t)sl$. tc_d is the cost incurred due to monitoring. The expected payoff of detecting the attack depends on the value of a, which is $(2a-1)tsm-(1-t)sl-tc_d$. When the player j is not active and there is an attack, so the payoff of the player j is $-sl$. The entry at position (row 2, column 1) is $-btn-tc_d$. n is the overall loss incurred by the player j for the false detection. The rest of the entry of the matrix is zero as the player i plays Not attack.

The payoff matrix for the player i when the player i is malicious is defined as shown in **Table 1(b)**. In contrast, the gain of player i is the loss of player j, which is $(1 - 2a)tsm + (1 - t)sl$. The entry at (row 1. column 2) is the same as in previous scenario. For the other entries, when the player i plays $S_i(2)$ (Not attack), his payoff is always 0.

The payoff matrix for the player i when it is regular is given in **Table 2**. The player i has only one strategy when it is regular. The payoff of player i is always 0. If player i decides not to monitor, his payoff is 0; if he decides to play $S_j(1)$, he has the monitoring cost tc_d and an expected loss $-btn$ due to the false alarm, so his payoff is $-btn-tc_d$.

3.2. One-Stage Game

The intent of both players is to maximize their own payoff. This implies that we assume that both players are rational. Suppose player j assigns a prior probability μ_0 to player i is malicious. In the following, we use Bayesian Nash equilibrium (BNE) to analyze the game model, based on the assumption that is a common prior.

If player i plays his pure strategy pair (Attack s of the time if malicious, Not attack if regular), then the expected payoff of player j is

$$E_j(S_j(1)) = \mu_0(atsm-(1-a)tsm-(1-t)sl-tc_d) - (1-\mu_0)(btn+tc_d)$$

$$E_j(S_j(2)) = -\mu_0 sl$$

So if $\mu_0 > \dfrac{bn+c_d}{2asm-sm+sl+bn}$, $E_j(S_j(1)) > E_j(S_j(2))$, then the best strategy of player j is to play Monitor t of the time. However, if player j plays this strategy, Attack s of the time will not be the best strategy if player i is malicious, and he will transfer to play Not attack instead. Hence, ((Attack s of the time if malicious, Not attack if regular), Monitor t of the time, μ_0) is not a BNE. If $\mu_0 < \dfrac{bn+c_d}{2asm-sm+sl+bn}$, ((Attack s of the time if malicious, Not attack if regular), Not monitor, μ_0) is a BNE. Similarly, ((Not attack s of the time if malicious, Not attack if regular), Not monitor, μ_0) is not a BNE.

THEOREM 1: In the described game-theoretic model, there is no pure-strategy BNE when μ_0 satisfies the inequality

$$\mu_0 > \frac{bn+c_d}{2asm-sm+sl+bn}.$$

We previously showed that no pure-strategy BNE exists for the game when $\mu_0 > \dfrac{bn+c_d}{2asm-sm+sl+bn}$. But there is a mixed-strategy BNE.

Let p be the probability with which the player i plays its first strategy. Hence, $(1-p)$ is the probability with which it plays the second strategy. Similarly, let q be the probability with which the player j plays its first strategy. Hence, $(1-q)$ is the probability with which it plays the second strategy. Then the expected payoff of player j is

$$E_j(S_j(1)) = p\mu_0(atsm-(1-a)tsm-(1-t)sl-tc_d) - (1-p)\mu_0(btn+tc_d)-(1-\mu_0)(btn+tc_d)$$

$$E_j(S_j(2)) = -p\mu_0 sl$$

From $E_j(S_j(1)) = E_j(S_j(2))$, we get that the malicious type of player i's equilibrium strategy is to play first strategy with probability

$$p^* = \frac{bn+c_d}{\mu_0(2asm-sm+sl+bn)}.$$

and the expected payoff of player i is

$$E_i(S_i(1)) = q(atsm+(1-a)tsm+(1-t)sl-tc_a) + (1-q)(sl-sc_a)$$

$$E_i(S_i(2)) = 0$$

From $E_i(S_i(1)) = E_i(S_i(2))$, we get that the equilibrium strategy of player j is to play first strategy with probability

$$q^* = \frac{l - c_a}{2atm - tm + tl}.$$

THEOREM 2: In the described game-theoretic model, the strategy pair ((Attack *s* of the time with probability p^* if malicious, Not attack if regular), Monitor *t* of the time with probability q^*, μ_0) is a mixed-strategy BNE.

The above described game is a static game, for which the players maximize their utilities based on the payoff matrix for the game. Due to the difficulty of assigning accurate prior probabilities for player *i*'s type, we extend the static to dynamic game, where the player *j* can update his beliefs according to the Bayes' rule.

3.3. Multi-Stage Game

The aforesaid one-stage game is static Bayesian game, for which the player j maximizes his payoff based on a fixed prior about the maliciousness of his opponent. The lifetime of the network could be broken down into intervals of the time and our game could be used as a repeated game over these intervals. So, we extend the one-stage game to multi-stage game.

We assume that the one-stage game is repeatedly played in each time period t_k, where k = 0, 1, … An interval of T seconds maybe selected for each stage game. In order to get a simple model, we assume that T = 1. The payoffs of the players in each stage game are the same as in the proceeding one-stage game, and we assume that there is no discount factor with respect to the payoffs of the players. The extensive form of each stage game can be represented in a similar manner as for the static one-stage game.

In our model, the player j's type is known to all the player while the player i's type is selected from the type set Θ ={malicious, regular}. Knowing that the player i's type is a private information. Bayesian equilibrium [12] dictates that the player i's action depends on his type θ. By observing the behavior of the player i, the player j can calculate the posterior belief evaluation function $\mu_{t_{k+1}}(\theta_i \,|\, a_i(t_k))$ using the following Bayes' rule

$$\mu_{t_k}(\theta_i \,|\, a_i(t_k)) = \frac{\mu_{t_k}(\theta_i \,|\, a_i(t_k)) P(a_i(t_k) \,|\, \theta_i)}{\sum_{\theta_i \in \Theta} \mu_{t_k}(\theta_i \,|\, a_i(t_k)) P(a_i(t_k) \,|\, \theta_i)} \quad (1)$$

where $\mu_{t_k}(\theta_i \,|\, a_i(t_k)) > 0$ and $P(a_i(t_k) \,|\, \theta_i)$ is the probability that strategy $a_i(t_k)$ is observed at this stage of the game given the type θ of the player i. From the assumption of described game, we know that

$$P(a_i(t_k) = Attack \,|\, \theta_i = 1) = ap + b(1-p)$$

$$P(a_i(t_k) = Not\ Attack \,|\, \theta_i = 1) = (1-a)p + (1-b)(1-p)$$

$$P(a_i(t_k) = Attack \,|\, \theta_i = 0) = a$$

$$P(a_i(t_k) = Not\ Attack \,|\, \theta_i = 0) = 1-b$$

LEMMA 1: the multi-stage game satisfies the four Bayesian conditions (1)-(4).

1) Posterior beliefs are independent, and all types of player j have the same beliefs, and even unexpected events will not change the independence assumption for the type of the opponents.

2) Bayes'rule is used to update beliefs from $\mu_{t_k}(\theta_i \,|\, a_i(t_k))$ to $\mu_{t_{k+1}}(\theta_i \,|\, a_i(t_{k+1}))$ whenever possible.

3) The players do not signal what they do not know.

4) All players must have the same belief about the type of another player.

Proof: condition (1) is trivially satisfied because player j has only one type. We can see that the multi-stage game satisfies (2) from Equation (1). In our multi-stage game context, player i's signal is part of attack actions, thus (3) is satisfied. Because there are only two players in the game at any stage, the condition (4) is satisfied.

THEOREM 3: The multi-stage game has a perfect Bayesian equilibrium (PBE).

At stage game t_k, duo to the updated belief $\mu(\cdot)$, the probability p^* is also updated continuously. From the previous analysis of section 3.2, the malicious type of player i's equilibrium strategy is to play his first strategy with probability

$$p^* = \frac{bn + c_d}{\mu(\cdot)(2asm - sm + sl + bn)} \quad (2)$$

the equilibrium strategy of player j is to play his first strategy with probability

$$q^* = \frac{l - c_a}{2atm - tm + tl} \quad (3)$$

So the PBE of the game is given as $(p^*, q^*, \mu(\cdot))$, with $(p^*, q^*, \mu(\cdot))$ given by Equations (1)-(3).

4. Example

For each experiment, we assume that $m = l = 1000$, $n = 100$. **Figures 1** and **2** assume $s = 0.85$, $t = 0.85$, $c_d = 5$, **Figure 3** assumes $t = 0.85$, $c_d = 5$, $a = 0.9$, $b = 0.02$, and **Figure 4** assumes $s = 0.9$, $t = 0.9$, $a = 0.95$, $b = 0.14$. **Figure 5** assumes $s = 0.9$, $t = 0.5$,

$a = 0.95$, $b = 0.02$, $c_d = 5$. For all four scenarios player j's prior probability $\mu_0 = 0.5$.

From **Figure 1**, we see that the higher a is, the faster posterior belief converges to 1. By contrast, **Figure 2** shows that the lower b is, the faster posterior belief converges to 1. In other words, the detection accuracy of the IDS affects the convergence speed of player j's posterior belief. From **Figure 3**, we see that the lower time of attacking, the faster posterior belief converges to 1. From **Figure 4**, we see that the higher c_d, the faster the convergence speed of player j's posterior belief will be.

Figure 5 shows the posterior belief of the player j for these two scenarios. The belief for the first scenario converges to 1 faster than the second scenario. This is because in the first scenario the player i starts to attack earlier compared to the second scenario. Once the belief reaches 1, it does not go down even if the player i is not attacking since the type has already been identified.

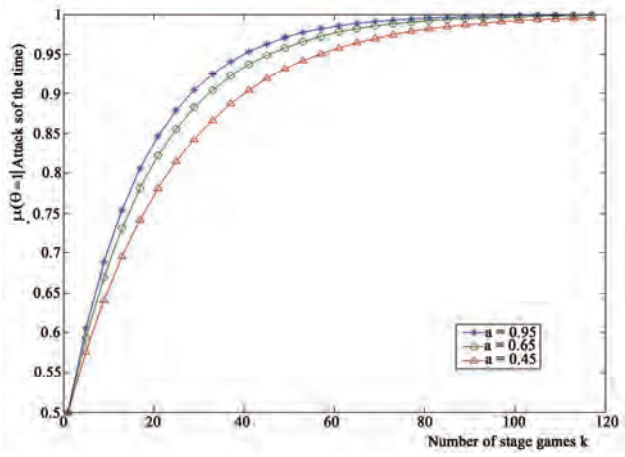


Figure 1. Convergence of player j's posterior beliefs given the observations of a sequence of a sequence of consecutive Attack actions under various *a*.

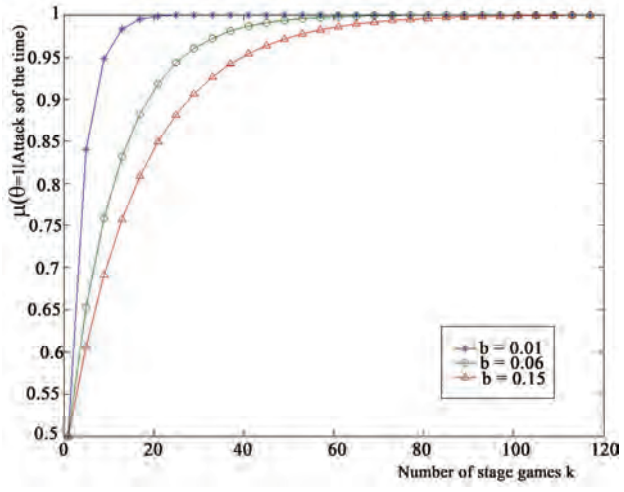


Figure 2. Convergence of player j's posterior beliefs given the observations of a sequence of a sequence of consecutive Attack actions under various *b*.

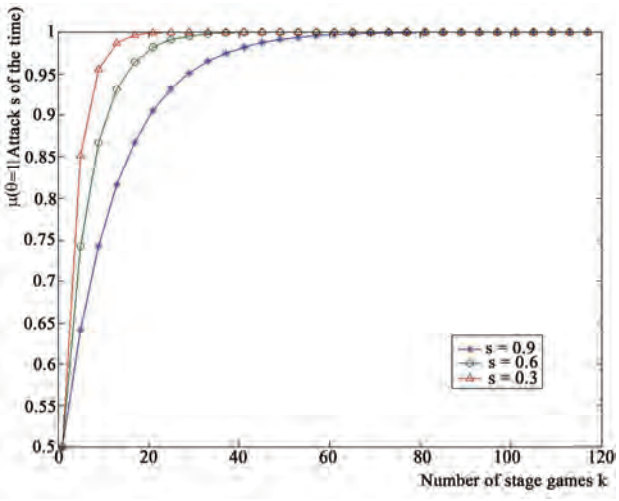


Figure 3. Convergence of player j's posterior beliefs given the observations of a sequence of a sequence of consecutive Attack actions under various *s*.

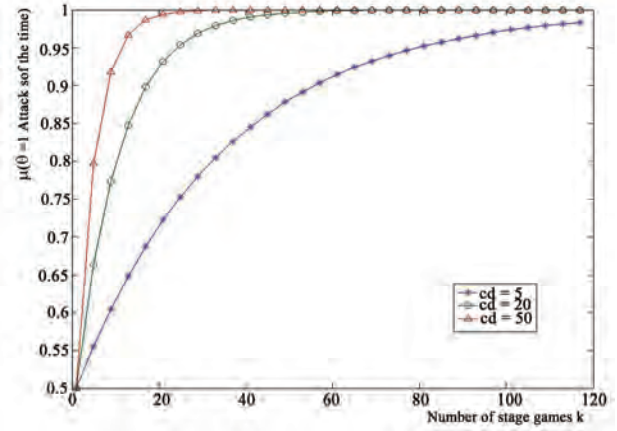


Figure 4. Convergence of player j's posterior beliefs given the observations of a sequence of a sequence of consecutive Attack actions under various c_d.

5. Conclusions

In this paper, our goal is to determine whether it is essential to always keep the IDS running without compromising on its effectiveness. First of all, we assume that the IDS works intermittently. Then, we model the interaction between intrusion detection system and an attacker as a one-stage game, and show that this game has two Bayes ian Nash equilibriums. Second, we model this game as a multi-stage game, where IDS does not have fixed prior

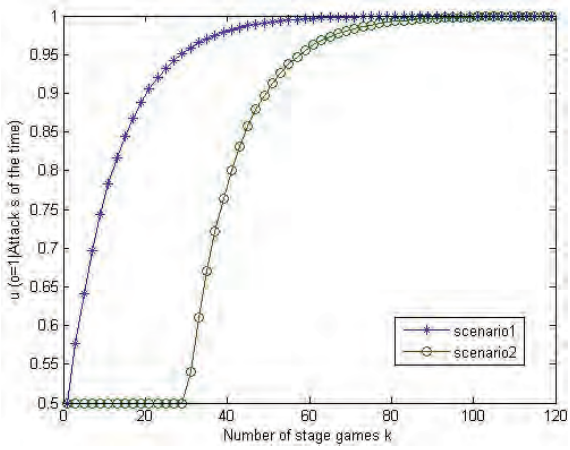


Figure 5. Posterior belief.

probabilities about the type of its opponent and can update its belief at the end of each stage of the game, and show that this game has a mixed-strategy perfect Bayesian equilibrium. The results of the proposed two games show that IDS could work intermittently while getting the same effectiveness.

6. Acknowledgements

The paper is supported by the National Natural Science Foundation of China under Grant Nos.70871098 and 70901063.

7. References

[1] R. Ramanathan and J. Redi, "A Brief Overview of Ad Hoc Networks: Challenges and Directions," *IEEE Communications Magzine*, Vol. 40, No. 5, 2002, pp. 20-22.

[2] Y. G. Zhang and W. Lee, "Intrusion Detection in Wireless Ad-Hoc Networks," *Proceedings of the* 6*th Annual International Conference on Mobile Computing and Networking*, Boston, 2000, pp. 275-283.

[3] Y. B. Reddy, "A Game Theory Approach to Detection Mmalicious Nodes in Wireless Sensor Networks," *Proceedings of the* 2009 *Third International Conference on Sensor Technologies and Applications*, Athens, June 18-23, 2009, pp. 462-468.

[4] H. Otrok, N. Mohammed, L. Y. Wang, M. Debbabi and P. Bhattacharya, "A Game-Theoretical Intrusion Detection Model for Mobile Ad Hoc Networks," *Computer Communications*, Vol. 31, No. 4, 2008, pp. 708-721.

[5] Y. Liu, C. Comaniciu and H. Man, "A Bayesian Game Approach for Intrusion Detetion in Wireless Ad Hoc Networks," *Proceedings from the* 2006 *Workshop on Game Theory for Communications and Networks*, Pisa, Italy, October 14, 2006, pp. 1-12.

[6] Y. Liu, H. Man and C. Comaniciu, "A Game Theoretic Approach to Efficient Mixed Strategies for Intrusion Detection," *IEEE International Conference on Communications*, Istanbul, 2006, pp. 2201-2206.

[7] A. Agah, S. K. Das, K. Basu and M. Asadi, "Intrusion Detection in Sensor Networks: A Non-Cooperative Game Approach," *Proceedings of the Third IEEE International Symposium on Network Computing and Applications*, Boston, August-September 2004, pp. 343-346.

[8] T. Alpcan and T. Basar, "A Game Theoretic Approach to Decision and Analysis in Network Intrusion Detection," *Proceedings of the* 42*nd IEEE Conference on Decision and Control*, Maui, Hawaii, December 2003, pp. 2595-2600.

[9] N. Marchang and R. Tripathi, "A Game Theoretical Approach for Efficient Deployment of Intrusion Detection System in Mobile Ad Hoc Networks," *Proceedings of the* 15*th International Conference on Advanced Computing and Communications*, Guwahati, 2007, pp. 460-464.

[10] T. Poongothai and K. Jayara, "A Noncooperative Game Approach for Intrusion Detection in Mobile Ad Hoc Networks," *Proceedings of the* 2008 *International Conference on Computing Communication and Networking*, St. Thomas, VI, December 18-20, 2008, pp. 1-4.

[11] A. Patcha and J.-M. Park, "A Game Theoretic Approach to Modeling Intrusion Detection in Mobile Ad Hoc Networks," *Proceedings of the* 2004 *IEEE Workshop on Information Assurance and Security*, United States Military Academy, West Point, NY, June 10-11, 2004, pp. 30-34.

[12] M. Willem, "Minimax Theorem," Birkhauser, Boston, 1996.

Int. J. Communications, Network and System Sciences, **2010, 3, 608-611**
doi:10.4236/ijcns.2010.37081 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Routing Strategy Selection for Zigbee Mesh Networks

Ramanathan Karthikeyan
Department of Electronics & Communication Engineering, Kumaraguru College of Technology, Coimbatore, India
E-mail: *karthikeyanrece@gmail.com*
Received April 29, 2010; *revised May* 31, 2010; *accepted July* 3, 2010

Abstract

Based on IEEE 802.15.4 Low Rate-Wireless Personal Area Network (LR-WPAN) standard, the Zigbee standard has been proposed to interconnect simple, low rate and battery powered wireless devices. The deployment of Zigbee networks is expected to facilitate numerous applications such as Home-appliance networks, home healthcare, medical monitoring and environmental sensors. An effective routing scheme is more important for Zigbee mesh networks. In order to achieve effective routing in Zigbee Mesh networks, a Zigbee protocol module is realized using NS-2. The suitable routing for different data services in the Zigbee application layer and a best routing strategy for Zigbee mesh network are proposed. The simulation shows the selection of suitable routing for continuous data services and for bursting data services in the Zigbee application layer and the comparison of three routing strategies namely ERD (All packets Enable Route Discovery), SRD (All packets Suppress Route Discovery) and BOS (routing Based on Data Services) with respect to efficiency and overhead.

Keywords: LR-WPAN, NS-2, ERD, SRD, BOS

1. Introduction

Zigbee is an emerging worldwide standard for Wireless Personal Area Networks (WPAN). Under the main goal to provide low-power, cost effective, flexible, reliable and scalable wireless products Zigbee Alliance has been developing and standardizing the Zigbee network. Based on IEEE 802.15.4 [1], Zigbee defines three types of devices. They are Zigbee Coordinator, Zigbee Router and Zigbee End device. Zigbee networks support star, tree and mesh topologies, self-forming and self-healing as well as more than 65000 address spaces; thus the network can be easily extended in terms of size and coverage area. The star topology of Zigbee is mainly designed for the simple communication from one node to several nodes. The tree network uses a hierarchical tree routing mechanism. The mesh network uses the mixed routing method combined with Z-AODV and hierarchical tree routing.

2. Zigbee Routing Algorithms

There are two routing algorithms in Zigbee network layer. They are modified Ad Hoc on Demand Distance Vector (Z-AODV) and Hierarchical Routing algorithms [2].

2.1. Z-AODV

Currently AODV [3] is the easiest and most widely implemented MANET protocol. Z-AODV is one of the earliest AODV simplified versions. Z-AODV removes the following items from the AODV specification such as Sequence number, gratuitous RREP, hop count, Hello message, precursor limits. In Z-AODV if the communications are unidirectional, the destination sends connect message to the source. If data traffic is bidirectional, no additional messages are used. In any case, a source detects a link break in a route when it doesn't receive messages from the destination.

2.2. Hierarchical Routing Algorithm

The hierarchical routing algorithm depends on the topology and a distributed addressing scheme of Zigbee networks. There are three types of devices in Zigbee networks. They are Coordinator, Router and End device. A Zigbee coordinator is responsible for initializing, maintaining and controlling the network. A star network has a coordinator with end devices directly connecting to the coordinator. For tree and mesh networks, Zigbee devices can communicate with each other in multihop

fashion. The network is formed by one Zigbee coordinator and multiple Zigbee routers. A device can join a network as an end device by associating with a coordinator or a router [4].

Before forming a network, the coordinator determines the maximum number of children of a router (C_m), the maximum number of child routers of a router (R_m), and the depth of the network (L_m) [5]. A child of a router can be a router or an end device, so ($C_m \geq R_m$). Zigbee specifies a distributed address assignment using parameters C_m, R_m and L_m to calculate the nodes network addresses. In Zigbee if a device joins a network successfully, it can obtain a network address from the coordinator or a router. The basic idea of the assignment is that for the coordinator and the routers in every layer, the whole address space is logically partitioned into $R_m + 1$ block. The first R_m blocks are to be assigned to the router child devices and the last block is reserved for the ($C_m - R_m$) child end devices. In order to make the assignment easily, a function *Cskip* can be computed by C_m, R_m and L_m. The value of this function is the size of address sub-block being distributed by each parent at the depth of its router child devices for a given network depth *d*.

$$Cskip(d) = \begin{cases} 1 + C_m(L_m - d - 1), & if\ \ R_m = 1 \\ \dfrac{1 + C_m - R_m - C_m R_m^{L_m - d - 1}}{1 - R_m}, & otherwise \end{cases} \quad (1)$$

A_n is computed by the following formula.

$$A_n = A_{parent} + Cskip(d-1)R_m + n \quad (2)$$

Here A_{parent} denotes the address of the parents, n denotes the n^{th} end device, and $1 \leq n \leq (C_m - R_m)$. When Zigbee adopted the hierarchical routing algorithm and a device called X with address A and depth d received a packet, the device extracted the destination address called D.

$$N = \begin{cases} D, & if\ \ D > A + R_m \times Cskip(d) \\ A + 1 + \left\lfloor \dfrac{D - (A+1)}{Cskip(d)} \right\rfloor \times Cskip(d), & Otherwise \end{cases} \quad (3)$$

If $D > A + R_m \times Cskip(d)$, the destination is the direct descendent of X, and X forward the packet to this direct descendent. If not, the destination is the indirect descendents of X, so X forward the packet to one of its child with address computed by

$$A + 1 + \left\lfloor \frac{D - (A+1)}{Cskip(d)} \right\rfloor \times Cskip(d). \quad (4)$$

In this way the network address is assigned to the network elements in the Zigbee network using hierarchical routing.

3. Simulated Results

3.1. Performance of Tree Routing and Z-Aodvrouting over Zigbee Networks

A Zigbee protocol module is developed using NS-2 with the following specifications. The simulation area is 50 × 50 m^2, the number of nodes is 21, Transmission range is 12 m, Packet error ratio is 0.2%, Data rate is 250 Kbps, Packet size is 70 bytes and Simulation time is 150 sec.

From **Figure 1** we observe that, the tree routing has faster response in forwarding the data packets since it doesn't need to initiate the routing tables. The flow starts at 25th second. At 27th second, 7 tree routed data packets are arrived at the destination node. The Z-Aodv routed data packets arrived at 27th second only because Z-Aodv must initiate the routing discovery. Z-Aodv after establishing its routing table the number of data packets in two routing methods will tend to be the same. Z-Aodv always chooses the route with less number of hops and the tree routing usually won't obtain shorter route. So the data frames transmitted in tree routing network are always 1.2 to1.4 times more than the data frames transmitted in Z-Aodv network. So the tree routing of Zigbee is suitable for bursting data transmissions and Z-Aodv is suitable for continuous data transmissions.

3.2. Comparison of Three Routing Strategies for Zigbee Mesh Network

A data flow consists of mixture of continuous data and bursting data is used in this simulation. The Zigbee mesh network uses a mixed routing mechanism combined with tree routing and Z-Aodv routing [6].

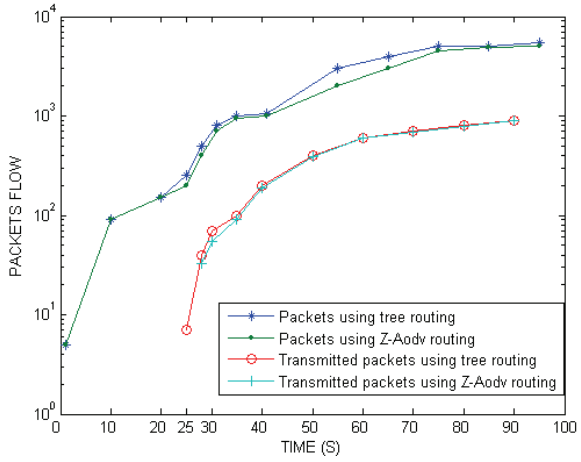


Figure 1. Performance comparison of tree routing and Z-AODV routing.

The Discover Route Field in the header of the data frames decides the routing approach for the data frames. If it has the value of Suppress Route Discovery, it uses the routing tables that exist already. When there is no corresponding address of the destination node the network will use tree routing.

If the Discover Route Field has the value of Enable Route Discovery and when the routing address is there in the routing table, the routing will follow this routing table. Otherwise the router will initiate the routing discovery. When the node has no ability to initiate the routing discovery, it will use tree routing. Based on the previous section simulation results, we can choose the binding data services in the Zigbee application layer will always use Enable Route Discovery routing method and the bursting data services in the Zigbee application layer will always use the Suppress Route Discovery routing method. This kind of routing method is called as routing based on data services. Efficiency is defined as the ratio between the transmitted data bytes and the total transmitted bytes. **Figure 2** shows the comparison of three routing strategies with respect to efficiency.

The efficiency of SRD routing method is the highest. The efficiency of ERD routing method is the lowest. BOS has to initiate the routing discovery for continuous data flow alone. ERD has to initiate the routing discovery for both continuous and bursting data flow. The increase of control overhead makes the efficiency of ERD low.

Figure 3 shows the average number of frames required to transmit a single data packet in all the three routing strategies for different bursting data conditions. SRD requires more frames to transmit a single data packet, since it uses the defined tree path. But ERD elects the shortest path and in turn it requires fewer

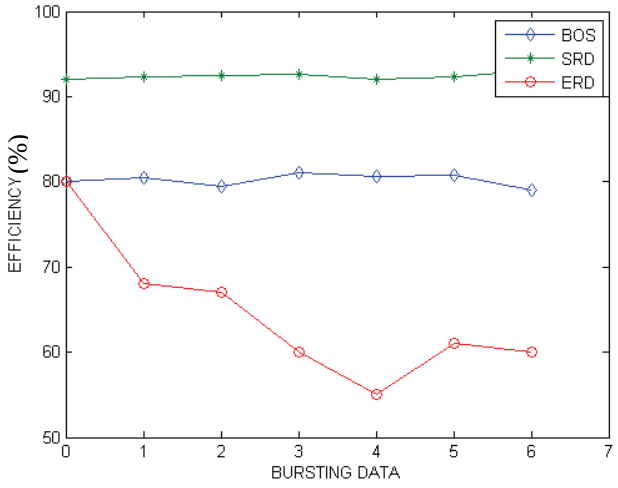


Figure 2. Comparison of ERD, SRD & BOS in the aspect of Efficiency.

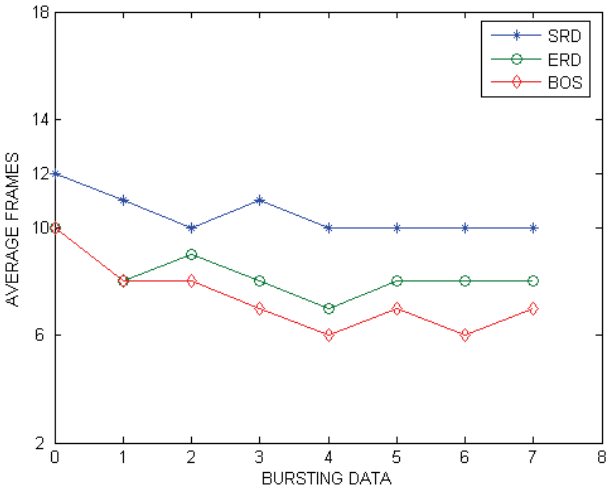


Figure 3. Average number of frames required to transmit a data packet in ERD, SRD & BOS.

frames. The BOS routing method cuts down the consumption of routing discovery for bursting data since it uses tree routing. So it has the least overhead. Comparatively BOS has the least overhead than ERD and SRD, which accordingly reduce the power consumption.

Thus in turn it is more suitable and much beneficial for Low Power IEEE 802.15.4 & Zigbee.

4. Conclusions

In this paper, the selection of suitable routing for continuous data services and for bursting data services in the Zigbee application layer is proposed. The three routing strategies namely Enable Route Discovery(ERD), Suppress Route Discovery(SRD), and routing Based On data Services(BOS) are compared in the aspects of Efficiency, overhead and BOS is proposed as the suitable routing strategy for Zigbee mesh networks.

5. References

[1] IEEE. Std. 802.15.4, "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks," 2003.

[2] J. Sun, Z. X. Wang and H. Wang, "Research on Routing Protocols Based on Zigbee Network," *Proceedings of the Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Kaohsiung, Taiwan, November 26-28, 2007, pp. 639-642.

[3] C. Perkins and E. B. Royer, "Ad Hoc On-demand Distance Vector(AODV) Routing," RFC 3561, July 2003.

[4] T. Kim, D. Kim and S. Yoo, "Shortcut Tree Routing in Zigbee Networks," *In 2nd International Symposium on Wireless Pervasive Computing*, San Juan, Puerto Rico, February 5-7, 2007, pp. 42-47.

[5] Zigbee Alliance, "Zigbee Specification Version 1.0," December 14th, 2004. http://www.zigbee.org

[6] X. H. Li and K. L. Fang, "An Improved Zigbee Routing Strategy for Monitoring Systems," *IEEE Proceedings of First International Conference on Intelligent Networks and Intelligent Systems*, Wuhan, 2008, pp. 255-258.

Int. J. Communications, Network and System Sciences, **2010, 3, 612-619**
doi:10.4236/ijcns.2010.37082 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Scientific Research

Techniques of Transmitting Beamforming to Control the Generated Weights

Imen Sfaihi[1,2], Noureddine Hamdi[2], Ammar Bouallegue[2]
[1]*The National school of engineering of Tunis* (*ENIT*), *El-Manar University*, *Tunis*, *Tunisia*
[2]*The communication System Laboratory* (*SysCom Lab*) *in ENIT*, *Tunis*, *Tunisia*
E-mail: *imene_sfaihi@yahoo.fr*, {*Noureddine.Hamdi*, *ammar.bouallegue*}*@enit.rnu.tn*
Received April 16, 2010; *revised May* 25, 2010; *accepted June* 29, 2010

Abstract

In this paper, we consider the limited feedback Transmitting Beamforming for (multiple in single out) MISO systems. In conventional techniques, all vectors of a large codebook (CB), used for the feedback of the quantized channel state information (CSI), are broadcasted to all users, in a guard period which is followed by data burst periods. Instead of transmitting a large number of codevectors, we thought to divide the CB into several sub-codebooks (SC) and the broadcast would be based on the switch between them. Accordingly, a good performance can be provided while minimizing the required feedback channel capacity applying some proposed techniques such as "the switched Sub Codebook (SSC)" and "the Fairness SSC (FSSC)". To minimize the quantization error, we propose two other techniques. The first is based on making Transmit SSC vectors controlled by a rotation weight (RW) to obtain almost a zero correlation between the SSC vectors used for the selected spatial channels. The second is based on introducing "the Schmidt algorithm" to construct an orthonormal weights using the generated weights. These two proposed techniques increase the probability of the selection of the worst case user on his best codevector to make zero the angle between his couple codevector and channel response. To analyze and validate the performance of these proposed techniques, simulation results are presented.

Keywords: MISO, Beamforming, Limited Feedback, CSI, SSC, FSSC, RW-FSSC, SSC-Schmidt Algorithm

1. Introduction

In the last few years, Multiuser MISO systems also named as (Space Division Multiple Access) SDMA based on Limited Feedback Transmitting Opportunistic Beamforming (OBF) have been a lot of interests in recent research studies. The goal is to provide high system spectral efficiency [1] while reducing the complexity. Due to the constraint of narrowband of the feedback channel, transmitting OBF on the broadcast channel with limited feedback has been widely studied in the literature as [2-4] and references therein.

Moreover, transmitting OBF is provided in the literature as a more practical design that ameliorates the performance of SDMA [5] and [6]. Each user selects the correspondent beamformer from the Beamforming CB. Therefore, we found many techniques to formulate the Beamforming CB vectors, for example: random orthonormal beamforming CB as proposed in [7-9] or transmitting beamforming based on grassmannian line packing as described in [10] and [11].

This SDMA design can be combined with multiuser scheduling [5] and [6]. Therefore BS uses an algorithm to select the best pair index of user-CB vector that increases the system capacity such as: Max-rate [4] or sub-optimal algorithms as proposed in [12] where a selection is based on the best pair of user-beam vectors. In [13], an algorithm is proposed, known as semi-orthogonal user selection scheme. This algorithm is based on upper bounded techniques where the value of the SINR and the value of the error quantization are compared to predefined thresholds which are defined in [2-4].

Up to the moment, the generated CB vectors are transmitted to all users and then select the best pair CB vector-user. After investigating these studies, we thought to reduce the complexity by introducing new techniques based on a score that measure frequency of access by using vectors in SC. Moreover, the second formulation of OBF CB vectors named Grassmanian method is the most prac-

Identify applicable sponsor/s here. (*sponsors*)

tical and near to the reality but the components of the OBF CB vectors are with non zero correlation. Consequently, the correlation between OBF CB vectors and the selected user channel is not null. Then, the interference level would be increased and then the system throughput decreases. Therefore, we thought to minimize the probability to make errors.

In this paper, we propose four new techniques of transmitting OBF at which the first technique is based on SSC applied to the Max-Rate scheduler with limited feedback system. The components of CB composed by $N = 2^B$ vectors would be divided into n_D SC. In [2-4], the N components of the CB vectors are transmitted to all users at each time slot. This number is reduced in this proposal to N/n_D components of the CB that would be transmitted to all users. Then, we minimize the complexity and respect the bandwidth of the feedback channel.

Besides, for the goal of providing fairness among users, we propose the second technique that is based on fairness SSC (FSSC). This proposal is the continuity of the SSC technique when we investigate and improve the SSC by introducing the proportional fair principle (PF) to switch the specific SC.

Moreover, we intend to meet the performance of the FSSC and to reduce the interference level at user receivers. Whereby, the correlation between OBF CB vectors is compared to predefined thresholds to make the specific rotation that minimize this correlation. Then, we put the threshold to a given value and at each transmission the value of correlation is controlled and compared to this threshold.

After this, we thought to use from the beginning an orthonormal weights in transmitting OBF in order to reduce the interference level at user receivers. Whereby, the correlation between OBF CB vectors is converge to zeros by applying the Schmidt algorithm and construct the new orthonormal OBF weights that give a zeros correlation between the generated OBF CB vectors.

The remainder of this paper is organized as follows. Section 2 describes the system and channel models. In Section 3, we present an overview of the design of different CB proposed in the literature. In section 4, we present an overview of the CSI quantization. In section 5, we describe the different steps of the proposed techniques of transmitting OBF: SSC, FSSC, RW-FSSC and SSC-Schmidt algorithm. In section 6, we present the Max-Rate scheduler. In section 7, we analyze the system capacity of the proposed techniques and give the closed form of capacity. In section 8, we present a selection of simulation results.

2. System Model

We consider a MISO system with M_t antennas at the base station (BS) and K users when each is equipped with one receive antenna. Each user has her own rate β_k. It is assumed that slow power control is employed to equally share the total transmitted power P_t on all transmit antennas at the BS. Users symbols are loaded on transmit antennas using Beamforming, *i.e.* the BS assigns a Beamforming vector to each of up to M_t selected active users.

The Beamforming vectors $\{W_i\}_{i=1}^{M_t}$ are obtained using a generated orthogonal unitary beamforming vectors as defined in [7-9] or using grassmannian line packing codebook as described in [2-4]. To solve the problem of limited resources allocated to the feedback channel, users estimate their CSI and feedback them in a quantized form on B bits to the BS through an uplink limited capacity feedback (LCFB) channel. We denote by $N = 2^B$ the number of CB vectors which is defined by $CB = \{W_1, W_2, ..., W_N\}$. At each time slot, the number of components that would be used in optimal side of the transmission is equal to the number of transmit antennas M_t.

It is assumed that transmit signals experience path loss, log-normal shadow fading, and multi-path fading. The CSI is measured by the vectors which represent the short term fading CSI on all branches from the BS to the k^{th} user assumed to be constant during a time slot. According to the slow power control, 1) each entry of the vector h_k is an independent and identically distributed complex Gaussian random variable $CN(0; 1)$ representing the short term fading; 2) the CSI experiences flat fading during each time slot, and varies independently over time slots.

We denote by $h_k(t)$ the $M_t \times 1$ channel vectors, $W_i(t)$ the $M_t \times 1$ CB, $s(t)$ the $M_t \times 1$ transmitted symbol, $n_k(t)$ the additive white Gaussian noise (AWGN) vector with distribution $CN(0; N_0/2)$ for each element, and $y_k(t)$ the received signal. Then, the received signal for the considered multi-user MISO system in the time slot t is represented by

$$y_k(t) = \sqrt{\frac{P_t}{M_t}} \sum_{i=1}^{M_t} h_k^H(t) W_i(t) s_i(t) + n_k(t) \qquad (1)$$

According to Equation (1), the received signal for user k when using the W_i CB vector can be:

$$y_{k,i}(t) = \sqrt{\frac{P_t}{M_t}} h_k^H W_i s_i + \sqrt{\frac{P_t}{M_t}} \sum_{j=1..M_t, j\neq i} h_k^H W_j s_j + n_k \qquad (2)$$

Hence, the corresponding expression of the signal to interference plus noise ratio $SINR$ for the k^{th} user and the i^{th} CB vector is expressed as follows:

$$SINR_{k,i} = \frac{\left| h_k^H w_i \right|^2}{\sum_{j=1..M_t, j\neq i} \left| h_k^H w_j \right|^2 + \frac{M_t}{P_t}} \qquad (3)$$

To evaluate the sum capacity, we need the statistical distribution of the $SINR_{k,i}$.

2.1. The Statistical Distribution of the Signal to Interference Plus Noise Ratio

To simplify the Equation (3), we let $a_i = \left|h_k^H w_i\right|^2$; i = 1, …, M_t. The *SINR* on the i^{th} CB vector for user k given in the Equation (3) can be rewritten as

$$SINR_{k,i} = \frac{a_i}{\sum_{j=1..M_t, j\neq i} a_j + \frac{M_t}{P_t}} \quad (4)$$

Consequently,

$$SINR_{k,i} = \frac{a_i}{\sum_{m=1}^{i-1} a_m + \sum_{m=i+1}^{M_t} a_m + \frac{M_t}{P_t}} \quad (5)$$

Then,

$$SINR_{k,i} = \frac{a_i}{b_i + c_i + \frac{M_t}{P_t}} \quad (6)$$

where $b_i = \sum_{m=1}^{i-1} a_m$ and $c_i = \sum_{m=i+1}^{M_t} a_m$.

Although, the random variables a_i are of independent χ^2 distribution with two degrees of freedom. Note that the $SINR_{k,i}$ with different k are independent. Then the cumulative distribution function (CDF) of the largest SINR for user k denoted $SINR_{k,i}$ can be calculated in terms of the joint probability density function (PDF) of the largest one of M_t i.i.d χ^2 random variables, denoted by $f_{b_i,a_i,c_i}(x,y,z)$, as [14]

$$F_{SINR_{k,i}}(x) = \int_0^{\infty}\int_0^{(M_t-i)w/(i-1)}\int_0^{x\left(M_t/P_t+z+w\right)} f_{b_i,a_i,c_i}(w,y,z)\,dydzdw \quad (7)$$

After taking derivative with respect to x, the PDF of $SINR_{k,i}$ is given by

$$F_{SINR_{k,i}}(x) = \int_0^{\infty}\int_0^{(M_t-i)w/(i-1)} f_{b_i,a_i,c_i}\left(w, x\left(M_t/P_t + z + w\right), z\right)dzdw \quad (8)$$

It was further shown in [14] that the joint PDF $f_{b_i,a_i,c_i}(x,y,z)$ is available in closed form, after some modifications as given by

$$f_{b_i,a_i,c_i}(x,y,z) = M_t\binom{M_t-1}{i-1}\frac{\left[w-(i-1)y\right]^{i-2}}{(i-2)!(M_t-i-1)!}\times e^{-w-y-z}U\left(w-(i-1)y\right)\times \sum_{j=0}^{M_t-i}\binom{M_t-i}{j}(-1)^j\left(z-jy\right)^{M_t-i-1}U\left(z-jy\right), \quad y\succ 0; w\succ (i-1)y; z\prec (M_t-i)y. \quad (9)$$

where U(.) denoted the Heaviside unit step function.

Remark:

If the largest *SINR* of the user k is the first component of the CB vectors then i = 1 and the correspondent CDF is as given in [15]

$$F_{SINR_{k,1}}(x) = \int_0^{\infty}\int_0^{x\left(M_t/P_t+z\right)} f_{a_1,b}(y,z)\,dydz \quad (10)$$

And

$$f_{SINR_{k,1}}(x) = \int_0^{\infty}\left(M_t/P_t + z\right)\int_0^{\infty} f_{a_1,b}\left(x\left(M_t/P_t+z\right), z\right)dz \quad (11)$$

It was further shown in [15] that the joint PDF is available in closed form, as given by

$$f_{a_1,b}(y,z) = \frac{M_t}{(M_t-2)!}e^{-y-z}\times \sum_{j=0}^{M_t-1}\binom{M_t-1}{j}(-1)^j\left(z-jy\right)^{M_t-2}U\left(z-jy\right) \quad (12)$$

After replacing this expression in (24), the PDF of the largest *SINR* of the user k ($SINR_{k,1}$) can be written as

$$f_{SINR_{k,1}}(x) = \sum_{j=0}^{M_t-1}\frac{(M_t-1)M_t}{(M_t-1-j)!\,j!}\int_0^{\infty}\left(\frac{M_t}{P_t}+z\right)\times \left((1-jx)z - jx\frac{M_t}{P_t}\right)^{M_t-2}\times e^{-(1+x)z-x\frac{M_t}{P_t}} U\left((1-jx)z - jx\frac{M_t}{P_t}\right)dz \quad (13)$$

3. Conventional Codebook Design

According to the considered system model, we can give a description of the CB vectors design. At each time slot, M_t symbols of users are multiplied by M_t random orthonormal vectors w_i $M_t \times 1$ for i = 1, ..., M_t. Where w_i's are generated according to an isotropic distribution [9] and [10] or are computed according to formulation described in [7]. This random orthonormal vectors are used to define a random CB. This CB vectors are generated with zero correlation which is near to the reality since it require that the chanal is known at both transmitter and receiver.

Moreover, we found in some prior work such as in [2] and [3] the term of CB vectors with none zero correlation such as grassmannian CB vectors defined in [8]. Relying to previous approaches a good beamforming is specifically using a grassmannian. Therefore, in this work, we can use this technique to generate CB vectors with none zero correlation.

4. CSI Quantization

Due to constraint of the bandwidth of feedback channel, each user just feeds back its maximum *SINR* quantized in *B* bits and the index of the corresponding codebook vector. In [2] and [3], the random vector quantization (RVQ) method is applied for quantizing the CSI. The CB vectors $\{W_i\}_{i=1}^{M_t}$ are obtained using a generated orthogonal unitary beamforming vectors as defined in [7-9] or using grassmannian line packing codebook as described in [2-4,12]. At each time slot, each user identified by *k* select his 'best' vector from the CB. For that, a quantization CB vector is selected as:

$$i_s = \arg\max_{\{W_j\}_{j=1}^{M_t}} \left| h_k^H \times W_j \right| \tag{14}$$

The quantization error is expressed as:

$$\delta = \sin^2\left(\angle\left(W_{i_s}, H\right)\right) \tag{15}$$

In the literature, it is often assumed for simplicity that the feedback is without errors. Then, the quantization error δ should be converges to zero.

5. The Proposed Techniques of Transmitting OBF

5.1. Design of SSC

For transmitting OBF, the CB vectors are used randomly such as in [2-4]. At each time slot, all of *N* components of CB are transmitted to all users. This is clearer in the previous step of quantization of CSI. Therefore, we propose an idea that based on dividing the *N* components of CB vectors into n_D SC to minimize the number of components to be transmitted to all users at each time slot.

Then, we define how to switch to a given SC at each time slot to increase the system capacity and give equal opportunity among users to the channel accesses. The switching is based by investigating user scores based on the historic use of each SC at a number of previous time slots. This would provide fairness among users.

The switching step is defined as follow:

1) The initialized matrix ($Score = zeros(K, n_D)$) is updated at each time slot and would be used in the following time slot:

$$Score(k_s, i_s) = Score(k_s, i_s) + 1 \tag{16}$$

where k_s is the index of each user among the M_t served users and i_s is the index of each CB vector among the M_t selected CB vectors of the SSC satisfying Equation (14).

2) At each time slot, we search the index of the user that has the worst capacity to give fairness among users

$$k^* = \min_k (C) \tag{17}$$

3) After, we search the index of the switched SC vector that satisfies the following expression

$$i^* = \max_i \left(Score\left(k^*, i\right)\right) \tag{18}$$

4) Finally, we update the matrix of Scores and compute the system capacity.

The most obvious benefit is to take consideration of the historic use of the CB vectors that will be represented by a score based on the user access frequency.

5.2. Design of FSSC

5.2.1. Fairness Criteria

In SSC, the idea to give fairness is derived and the selected score is that has the user with the worst capacity $\left(\min_{k\in K}(C)\right)$. If we apply this, we can assume that the number of users to have the worst capacity can be large. Then, the selected user is chosen randomly and the same user can be chosen many times. Therefore, in our proposal design, the most obvious goal is to give fairness among users. And accordingly, we thought to an idea in basis on max-min schedulers introduced in previous work such as proportional fairness in [7] when all users have the equal chance to be served.

Now, we can suppose that each user has her own rate β_k and we propose in this technique to select the index of the SC that has the minimum of the user's capacity's divided by the proportional component α_k when α_k is expressed in the following sub section in (20). Then, we can assume that to select user with using $\left(\min_{k\in K}(C_k/\alpha_k)\right)$ should be given most chance to users who can be served at the following time slot. Accordingly, the probability to choose the same user many times is minimized and this probability converges to zero. Moreover and according to the most aim of our proposal, we can talk about the index of Jain for fairness defined in [16] and expressed as follows

$$j = \frac{\left(\sum_{k=1}^{K} C_k(t)\right)^2}{K\sum_{k=1}^{K}\left(C_k(t)\right)^2} \tag{19}$$

where $C_k(t)$ is the system capacity of k^{th} user and *K* is the total number of users. We are going to present and to discuss this term in the simulation results to validate our scheme and their results.

5.2.2. FSSC Algorithm

The fairness switching algorithm is described as follows:

1) Initialization:

a) Let β_k the rate of user k.

b) Let the proportional components α_k as

$$\alpha_k = \beta_k \Big/ \sum_{i=1}^{K} \beta_i \quad (20)$$

c) The score matrix is defined in the previous subsection.

2) The Score matrix is updated at each time slot and should be used in the following time slot (the same in (16)).

3) At each time slot, we search the index of the user that has the worst of the capacity divided by α_k to give fairness among users

$$k^* = \left(\min_{k \in K} \left(\frac{C_k}{\alpha_k} \right) \right) \quad (21)$$

4) After, we search the index of the fairness switched SC vector that satisfies the following expression

$$i^* = \max_i \left(Score\left(k^*, i\right) \right) \quad (22)$$

5) Finally, we update the matrix of Scores and compute the system capacity.

The most obvious benefit is the use of an access control based on PF constraint to provide the fairness access channel among users.

5.3. Design of RW-FSSC

In the literature, it is often assumed for simplicity that the feedback is without errors. Then, the quantization error expressed in (15) should converge to zero i-e the probability to make errors should be converge to zero. But in reality, the use of grassmanian method to generate the OBF CB vectors when the components of OBF CB vectors are with not zero correlation should be make errors. Consequently, the correlation between OBF CB vectors using to select the user channel is not null. Therefore, we thought to control this value, compared with the value of threshold and make the rotation weight (RW).

The RW is consisting of the applying a rotation on the best codevector to make zero the angle between the couple codevector and channel response. Then, the RW increases the probability of the selection of the worst case user and if this user is selected he would be assigned the best possible symbol rate.

The RW is consisting of three cases at which the value of the angle between the couple codevector and channel response defined as θ is compared with a specific values of thresholds that described as follows:

1) Let θ_{max} the threshold: it is the maximum angle between the couple codevector and channel response and W_r the OBF codebook vectors after rotation.

2) Let $\theta = \angle\left(W_i, H\right)$

3) if $|\theta| \le \frac{\theta_{\max}}{4}$ Then $W_r = W_i$

4) else if $-\theta_{\max} \le \theta \le -\frac{\theta_{\max}}{4}$ Then

$$W_r = W_i^{'} \times \exp\left(-j \times \frac{\theta_{\max}}{2} \right)$$

5) else if $\frac{\theta_{\max}}{4} \le \theta \le \theta_{\max}$ Then

$$W_r = W_i^{'} \times \exp\left(j \times \frac{\theta_{\max}}{2} \right)$$

6) else if $\theta > \theta_{\max}$ Then SNR = 0

5.4. Design of the Method that Introduce the Schmidt Algorithm

5.4.1. Schmidt Algorithm

***Theorem:* (*Process of Gram-Schmidt*)**

Let $\{a_1, ..., a_N\}$ a family of vectors linearly independent.

Then it exists a family of orthonormal vectors $\{q_1, ..., q_N\}$ when for all $i = \{1, ..., N\}$, we have

$V\,ect\,\{a_1, ..., a_i\} = V\,ect\,\{q_1, ..., q_i\}$.

According to the process of Gram-Schmidt and to the SSC technique, we thought to introduce the Schmidt algorithm to construct a new orthonormal OBF CB vectors using the generated OBF CB vectors using one of the conventional CB designs and that used in the following step to select the user channel.

5.4.2. Steps of this Proposal Technique

On the first hand, we use one of the conventional CB design such as random orthonormal CB that have an isotropic distribution or the grassmanian method.

1) We denote by $\{W_i\}_{i=1}^{N}$ the generated OBF CB vectors

2) We apply the Schmidt algorithm and we obtain the new orthonormal OBF CB vectors W_1

And after, we use this new orthonormal OBF CB vectors W_1 to quantize the CSI and use the SSC transmit technique.

6. Max-Rate Scheduling

To maximize the system capacity, the technique of scheduling to share resources among active users is studied and applied. In this section, we describe the main idea of the Max-rate scheduling to select users. Accordingly, the selected M_t users to be served at each time slot experiences peak level signal to interference plus noise ratio (*SINR*) expressed in (3). This can be expressed as:

$$k_s = \arg\max_k \left(SINR_{k,i_s} \right) \quad (23)$$

where i_s is the index of the CB vectors selecting with the proposed technique of transmitting beamforming at the

correspondent time slot. We use the Max-rate scheduling because it gives the optimal performance and the aim is to investigate the resources with the most efficiently.

7. Capacity Analysis

According to the step of [14], the exact sum capacity expression for such scheme under consideration can be written as

$$C = M_t \int_0^{\infty} \log_2(1+x)\, f_{SINR}(x)dx \qquad (24)$$

Based on the mode of assignment of one of the proposed techniques, we can assume that the largest *SINR* of different users are i.i.d and the correspondent PDF is given by

$$f_{SINR}(x) = KF_{SINR_{k,1}}(x)^{K-1} f_{SINR_{k,1}}(x) \qquad (25)$$

where $F_{SINR_{k,1}}$ and $f_{SINR_{k,1}}$ are the CDF and the PDF of the largest *SINR* for a particular user, given in (11) and (24) respectively. Then, the capacity can be written as

$$C = M_t \int_0^{\infty} \log_2(1+x) KF_{SINR_{k,1}}(x)^{K-1} f_{SINR_{k,1}}(x)dx \qquad (26)$$

8. Simulation Results

In this section, the performances of the Max-rate scheduler with limited feedback are evaluated using the SSC, FSSC, the RW-FSSC and SSC-Schmidt algorithm to control the generated CB (for grassmannian CB vectors) [8] in terms of system capacity. The number of active users K used for these simulations varies from 1 to 30; the number of SC n_D is equal to 2, the time moving window T is of 200 and the feedback bits B is of 3.

In **Figure 1**, we compare the sum capacity performances of the "SSC" and the RBF techniques (random Max-rate scheme with LF such as in [2-5]).

Figure 2 illustrates the performance of the FSSC design in terms of system capacity. These results are compared to the SSC technique.

Figure 3 plots the index of jain for fairness of these two techniques versus the number of active users K when the value of SNR = 20 dB. This figure shows the fairness degree of the capacity in FSSC and SSC techniques. We can conclude that FSSC and SSC provide respectively a quasi optimal (~1) and near optimal fair degree. This is explained by the number of users that have the same worst capacity can increase with the number of users.

Figure 4 shows the performance of the RW-FSSC in terms of system capacity. These results are compared to the FSSC technique. **Figure 5** shows the simulation results of the system capacity of the SSC applying the Schmidt algorithm and SSC versus the number of active users.

In **Figures 1, 2, 4** and **5**, the simulation results of the

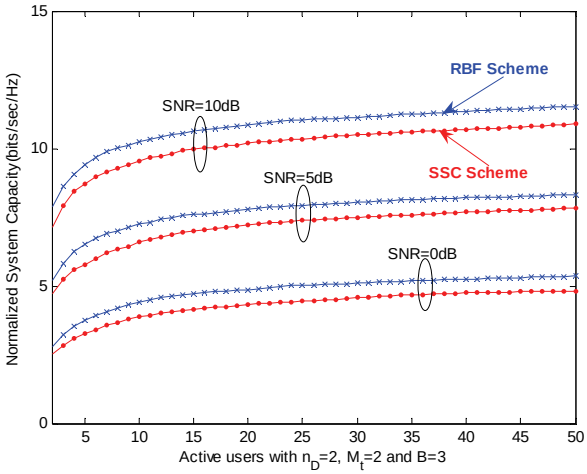


Figure 1. System capacity of the SSC and OBF techniques vs. number of active users.

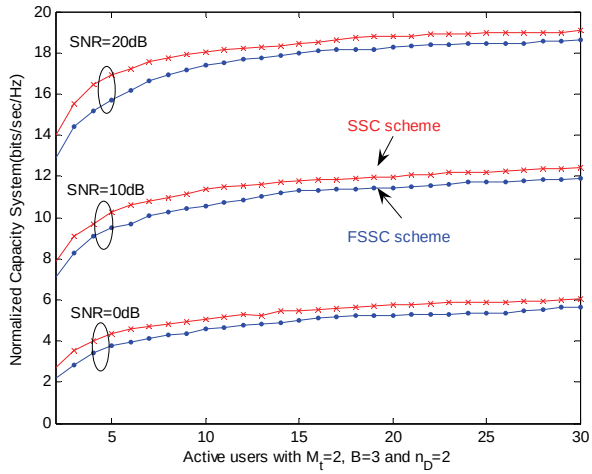


Figure 2. System capacity of the FSSC technique vs. number of active users.

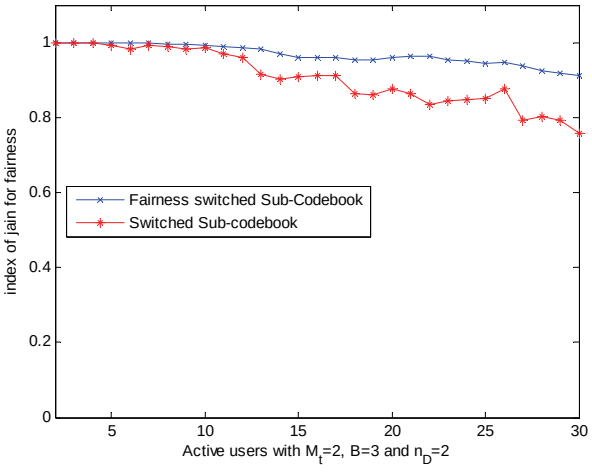


Figure 3. The index of jain for fairness.

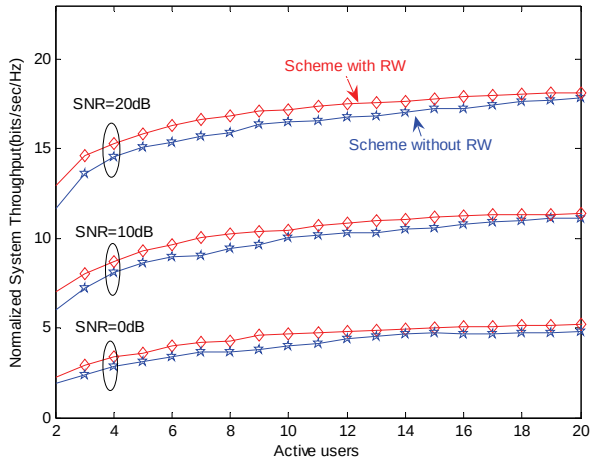


Figure 4. System capacity of RW-FSSC vs. number of active users with variation of the average SNR.

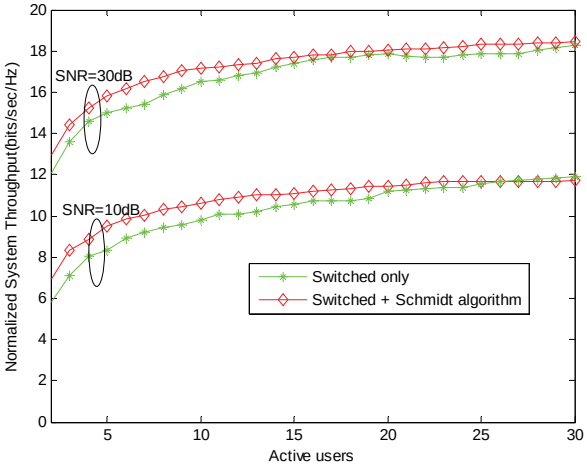


Figure 5. System capacity of SSC-Schmidt algorithm vs. number of active users with variation of the average SNR.

system capacity of the proposed techniques are plotted with different values of average SNR. Since, we can see that the system capacity applying the proposed techniques is nearly independent of the number of active users *K*. As can be seen from these figures, the difference between the curves of the transmit techniques is very small. In addition to that, the results of the proposed techniques are in good concordance and the system capacity grows as the average SNR increases.

9. Conclusions

The system capacity of limited feedback using OBF CB vectors and applying one of the new proposed techniques for Max-rate technique has been analyzed in this paper to deal with the performance of the coherent transmitting OBF. The transmit antenna are assigned to different up to M_t users at each time slot to increase system capacity.

According to the simulation results, we can conclude that the FSSC technique for Max-rate give fairness among users for a lot of number of users and a good performance. Moreover, we can conclude that the techniques to control the generated weights applied the FSSC for Max-rate minimize the probability to make error and give fairness among a number of users and a good performance while reducing the complexity of generating the OBF CB vectors as SSC technique. Accordingly, we can see that the system capacity can be improved using our proposed techniques.

10. References

[1] A. F. Molish, M. Z. Win and J. H. Winters, "Capacity of Mimo Systems with Antenna Selection," *IEEE Transaction on Wirelless Communications*, Vol. 4, No. 4, 2005, pp. 1752-1772.

[2] K. Huang, R. W. Heath and J. G. Andrews, "Performance of Orthogonal Beamforming for Sdma Systems with Limited Feedback," *IEEE Transaction on Vehicular Technology*, Vol. 58, No. 1, 2009, pp. 152-164.

[3] S. Zhou, Z. Wang and G. B. Giannakis, "Quantifying the Power Loss When Transmit Beamforming Relies on Finite Rate Feedback," *IEEE Transaction on Wirelless Communications*, Vol. 4, No. 4, 2005, pp. 1948-1957.

[4] R. Agarwal, C.-S. Hwang and J. Cioffi, "Opportunistic Feedback Protocol for Achieving Sum-Capacity of the Mimo Broadcast Channel," *Proceedings IEEE* 66*th Vehicular Technology Conference*, Baltimore, MD, 2007, pp. 606-610.

[5] T. Kim, R. W. Heath and S. Choi, "Multiuser MIMO Downlink with Limited Feedback Using Transmit-Beam Matching," *IEEE International Conference on Communication*, Beijing, May 19-23, 2008, pp. 3506-3510.

[6] J. L. Vicario, R. Bosisio, C. Anton-Haro and U. Spagnolini, "Beam Selection Strategies for Orthogonal Random Beamforming in Sparse Networks," *IEEE Transactions on Wireless Communications*, Vol. 7, No. 9, 2008, pp. 3385-3396.

[7] P. Viswanath, D. N. C. Tse and R. Laroia, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Transaction on Information Theory*, Vol. 48, No. 6, 2002, pp. 1277-1294.

[8] B. Hassibi and T. L. Marzetta, "Multiple Antennas and Isotropically Random Unitary Inputs: The Received Signal Density in Closed Form," *IEEE Transaction on Information Theory*, Vol. 48, No. 6, 2002, pp. 1473-1484.

[9] M. Sharif and B. Hassibi, "On the Capacity of Mimo Broadcast Channels with Partial Side Information," *IEEE Transaction on Information Theory*, Vol. 51, No. 2, 2005, pp. 506-522.

[10] D. J. Love, R. W. Heath and T. Strohmer, "Grassmannian

Beamforming for Multiple-Input Multiple-Output Wireless Systems," *IEEE Transaction on Information Theory*, Vol. 49, No. 10, 2003, pp. 2735-2747.

[11] B. Clercks, Y. Zhou and S. Kim, "Practical Codebook Design for Limited Feedback Spatial Multiplexing," *IEEE International Conference on Communications*, Beijing, May 19-23, 2008, pp. 3982-3987.

[12] M. Trivallato, F. Boccardi and F. Tosato, "User Selection Schemes for MIMO Broadcast Channels with Limited Feedback," *IEEE Transaction on Vehicular Technology Conference*, Dublin, April 2007, pp. 2089-2093.

[13] T. Yoo, N. Jindal and A. Goldsmith, "Multi-Antenna Broadcast Channels with Limited Feedback and User Selection," *IEEE Journal on Selected Areas in Communications*, Vol. 25, No. 7, 2007, pp. 1478-1491.

[14] H.-C. Yang, P. Lu, H.-K. Sung and Y.-C. Ko, "Exact Sum-Rate Analysis of MIMO Broadcast Channels with Random Unitary Beamforming Based on Quantized SINR Feedback," *IEEE International Conference on Communications*, Beijing, China, May 19-23, 2008, pp. 3669-3673.

[15] Y.-C. Ko, H.-C. Yang, S.-S. Eom and M.-S. Alouini, "Adaptive Modulation with Diversity Combining Based on Output-Threshold MRC," *IEEE Transactions on Wireless Communications*, Vol. 6, No. 10, 2007, pp. 3728-3737.

[16] R. Jain, "The Art of Computer Systems Performance Analysis," JohnWiley and Sons, New York, 1991.

Int. J. Communications, Network and System Sciences, **2010, 3, 620-624**
doi:10.4236/ijcns.2010.37083 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Design of Rectangular Dielectric Resonator Antenna Fed by Dielectric Image Line with a Finite Ground Plane

Fatemeh Kazemi[1], Mohammad Hassan Neshati[2], Farahnaz Mohanna[3*]
[1]*Electrical Department, University of Sistan and Baluchestan, Zahedan, Iran*
[2]*Electrical Department, Ferdowsi University of Mashhad, Mashhad, Iran*
[3]*Electrical Department, University of Sistan and Baluchestan, Zahedan, Iran*
E-mail: *fatemeh.kazemi.ms@gmail.com, neshat@ieee.org, f_mohanna@hamoon.usb.ac.ir*
Received April 1, 2010; *revised May* 8, 2010; *accepted June* 16, 2010

Abstract

A Rectangular Dielectric Resonator Antenna (RDRA) fed by Dielectric Image Line (DIL) through a narrow slot placed on a finite ground plane is numerically investigated. The effects of ground plane size on the radiation performance of the antenna are analyzed. To increase the antenna gain, four sidewalls are placed around the corners of the ground. Also, a reflector is placed at the back side of the structures to reduce backward radiation. Results show that 7.7 dB gain is obtained at 10 GHz with a broadside radiation pattern. For the DRA with four sidewalls maximum gain of 10.4 dB at 10.4 GHz is achieved which is 2.7 dB higher than the gain of the structure without them. The effect of air gap between dielectric resonator and ground plane is also investigated. The results show that with increasing distance between the DR and ground, antenna gain is decreased.

Keywords: Dielectric Image Line (DIL), Dielectric Resonator Antenna (DRA)

1. Introduction

Microstrip lines are used to excite slot-coupled patch and DR antennas, while their transmission loss is high especially at microwave and millimeter frequencies. To avoid conductor loss and to increase radiation efficiency, dielectric transmission line such as DIL could be used to excite a patch or a DRA through a narrow slot. The slot-coupled microstrip patch antenna and its array fed by the DIL were designed and investigated in [1], and a good gain, low return loss and low backward radiation were obtained. However, DRAs have been proposed as an efficient antenna at microwave and millimeter frequency, offering several advantages over the conventional microstrip patch antennas such as smaller in size, wider in bandwidth and no excitation of surface waves [2-5]. Moreover, due to no inherent conductor loss in dielectric materials, DRAs provide high radiation efficiency.

In this paper an RDRA fed by DIL, excited through slot on the ground plane is studied based on the Finite Element Method (FEM). The effects of the ground plane width are studied on the radiation performance of the DRA. Results show that the best width of the ground is nearly 100 mm for maximum gain and broadside radiation pattern. The structure provides a good return loss with peak gain of 7.7 dB at 10 GHz. The slot length and width are 3.7 mm and 0.144 mm respectively.

To increase the DRA gain, four sidewalls are placed around the corners of the ground plane. Results show that maximum gain of 10.4 dB is achieved at 10.4 GHz which is 2.7 dB higher than the gain of structure without sidewalls. Moreover, to reduce the backward radiation, a reflector is placed at the back of the structure under the waveguide tapers. Results show that backward radiation is decreased nearly 10 dB in E-plane.

2. Antenna Structures

The geometry of the RDRA is shown in **Figure 1(a)**. A rectangular DR of length a, width b, height c with the relative permittivity of ε_{rd} is placed on the ground plane with width W_a. A slot of length L and width W is etched at the center of the metal plane to excite the resonator. DIL as the transmission media consist of a rectangular dielectric slab of relative permittivity ε_r is placed under the ground plane. All dimensions of the structure are summarized in **Table 1**. Also, antenna structure adding four sidewalls is shown in **Figure 1(b)**.

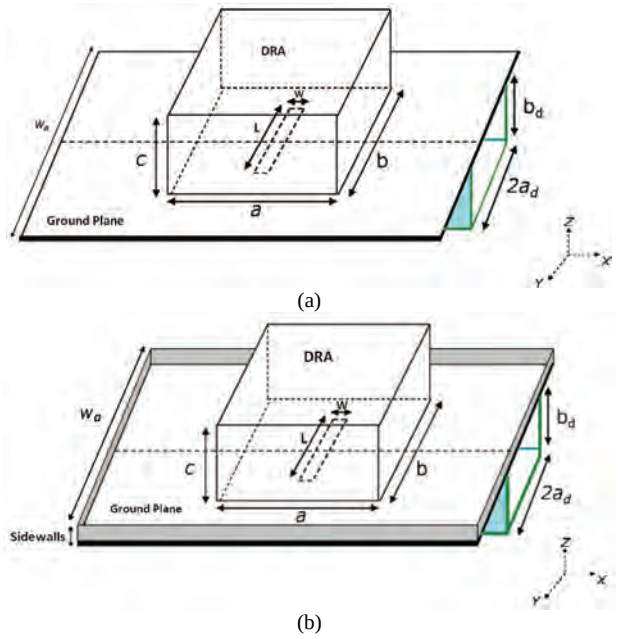


Figure 1. The geometry of the DRA fed by DIL. a) single antenna; b) antenna with four sidewalls.

Table 1. Antenna dimensions.

DRA		DIL	
a	6.2 mm	a_d	4.25 mm
b	6 mm	b_d	4.03 mm
c	6.1 mm	ε_r	10.2
ε_{rd}	10.2		

3. Antenna Simulation

The structures are numerically investigated using HFSS based on the Finite Element Method (FEM) which calculates full 3-D electromagnetic field inside and outside (far field) of the structures [6]. The detailed structure of the RDRA defined in HFSS is shown in **Figure 2(a)**. A standard metal waveguide, WR90 is used to excite the DIL, at the input and output of the transmission media. Three sections of waveguide using a proper tapering provide transition from TE_{10} mode of the metal rectangular waveguide to dominant mode of the DIL [7]. The DRA structure has two ports. Port one is defined as the input to excite the TE_{10} mode of the metal waveguide. The second port at the output is terminated to a matched load so; a traveling wave is propagated in DIL which efficiently excites the RDRA at the resonance frequency. The slot on the ground plane upon which the RDRA is located determines the amount of power coupled from the DIL to the resonator. The slot operates as a magnetic current in parallel to the resonator length exciting the RDRA at the principal TE^z_{111} mode of the operation [8, 9]. **Figure 2(b)** shows the detailed structure of the antenna with sidewalls and PEC reflector. Height of sidewalls is $0.25\lambda_o$, where λ_o is the wavelength in free space.

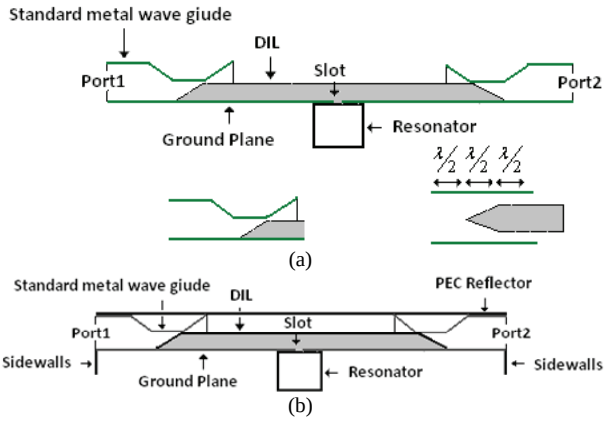


Figure 2. Detailed feed structure of RDRA. a) antenna structure; b) antenna with sidewalls and reflector plane.

4. Result and Discussion

4.1. Effect of Ground Plane Width

Figure 3(a) shows the effect of ground plane width on the RDRA peak gain. It can be seen that for low values of width antenna gain is very low. However, with increasing W_a backward radiation would decrease and hence, antenna gain is increased. For W_a = 100 mm, while the antenna structure is not too big, maximum gain is obtained. The antenna gain versus frequency for three values of W_a is shown in **Figure 3(b)**, which shows that for 100 mm width 7.7 dB gain is obtained at 10 GHz. Return loss and radiation pattern for this width size are also shown in **Figures 4(a)** and **4(b)** respectively. DRA structure provides broadside radiation pattern perpendicular to the ground plane and has good return loss.

4.2. Effect of Sidewalls

For increasing DRA gain, four sidewalls are placed around the corners of the ground plane. The height of sidewalls is chosen around $0.25\lambda_o$, while λ_o is the wavelength in free space. **Figure 5(a)** shows the effect of sidewalls on RDRA peak gain. It can be seen that antenna gain is increased to 10.4 dB, while backward radiation is not reduced significantly and the resonance frequency is shifted about 0.4 GHz. **Figure 5(b)** shows the simulated radiation patterns at 10.4 GHz. It can also be concluded that backward radiation is high. It is said that this is due to radiation from slot and feed line.

4.3. Effect of Reflector

To reduce the backward radiation, another ground plane

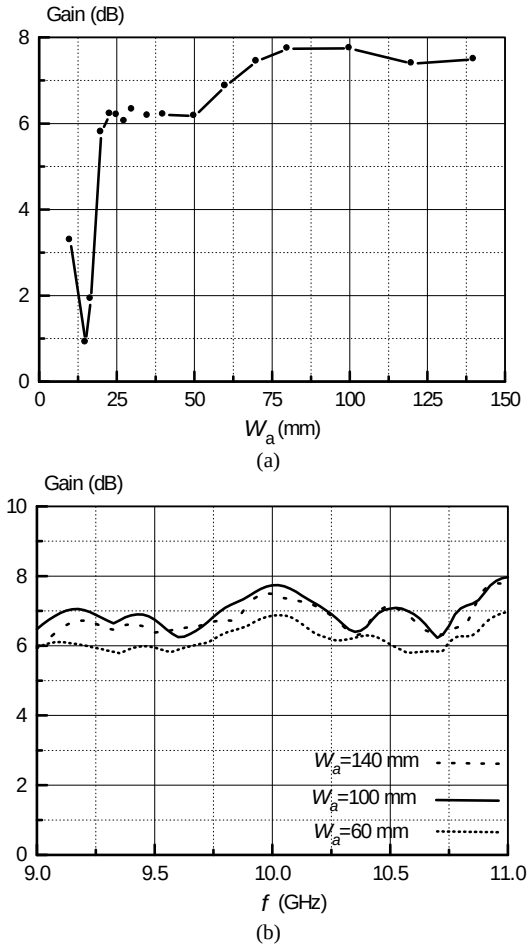
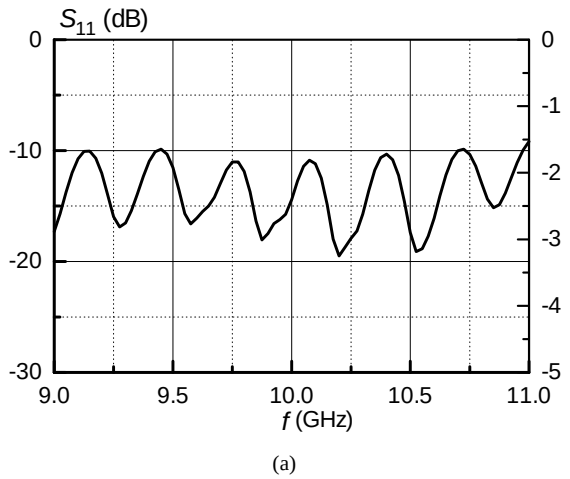


Figure 3. RDRA Peak gain. a) versus ground plane width at 10 GH; b) versus frequency and different values of W_a.

as a reflector is placed at the back side of the structures. Its size is chosen same as the main ground. In this case, the radiation patterns are shown in **Figure 6(a)**, which shows the backward radiation is significantly reduced. Also, the results show that maximum obtained gain of the RDRA with sidewalls and reflector is 10.4 dB at 10.4 GHz



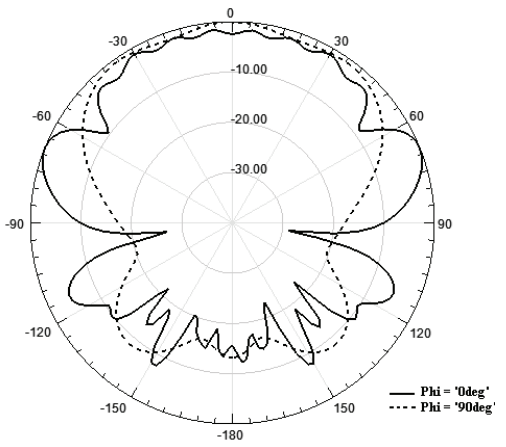


Figure 4. RDRA with 100 mm of ground plane width. a) return loss; b) radiation patterns at 10 GHz.

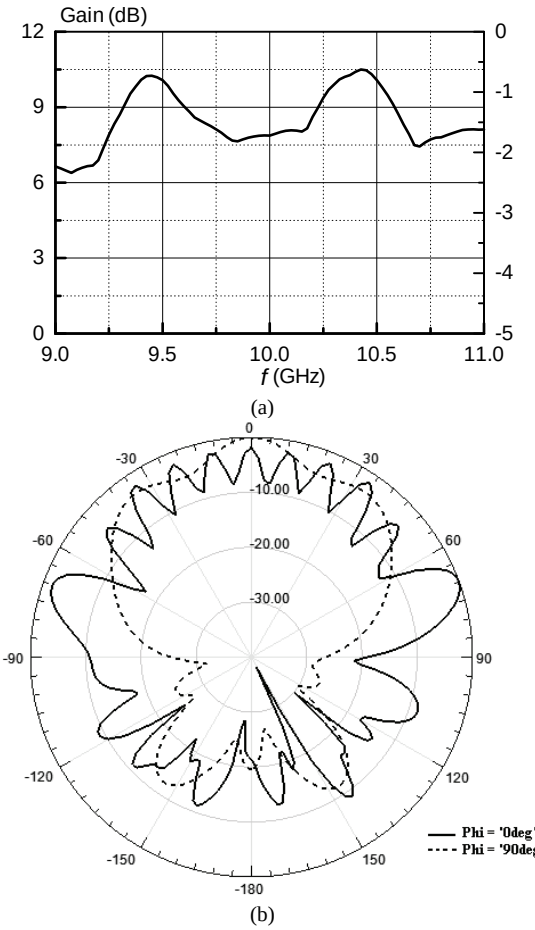
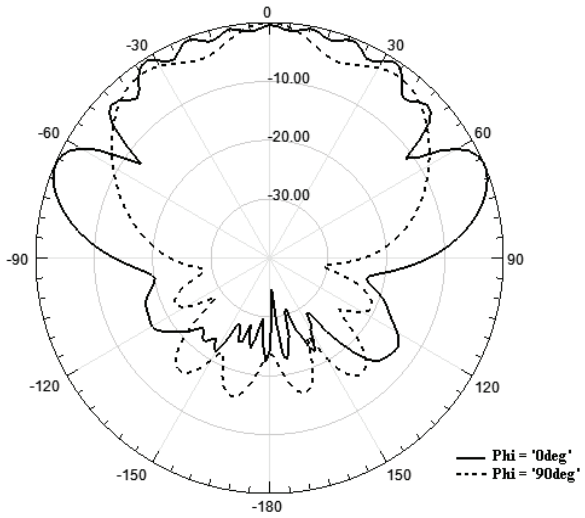
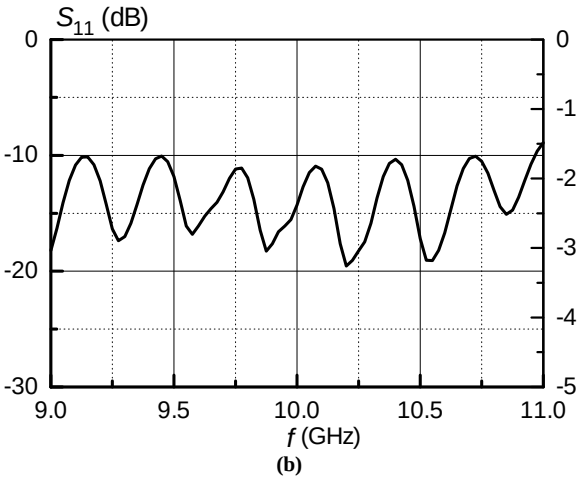


Figure 5. RDRA with sidewalls. a) peak gain versus frequency; b)radiation patterns at 10.4 GHz.

which is 2.7 dB higher than the gain of the structure without them. Therefore, the reflector decreases the backward radiation of the feed slot and the DIL. **Figure 6(b)**



(a)



(b)

Figure 6. RDRA with sidewalls and reflector plane. a) radiation patterns at 10.4 GHz; b) return loss.

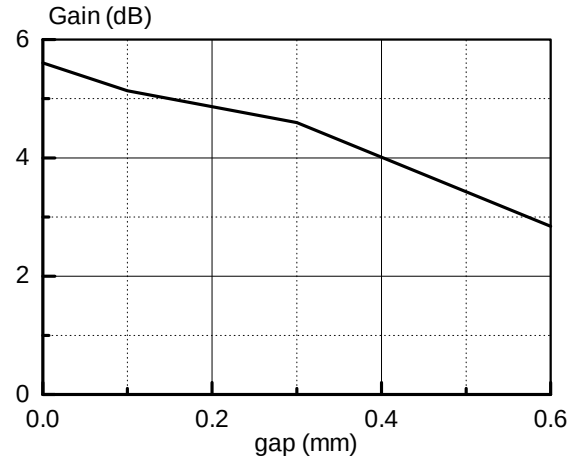
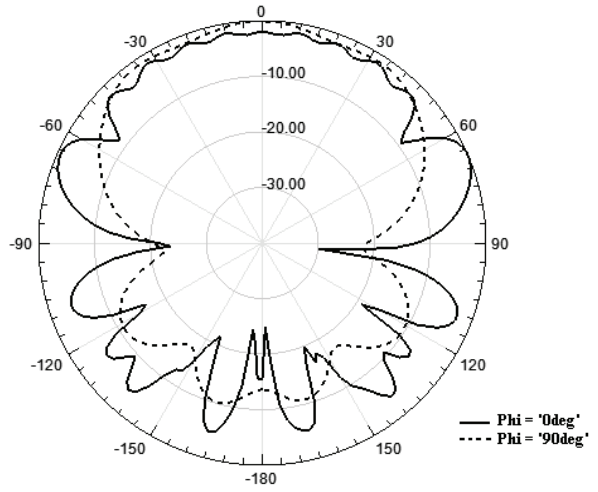
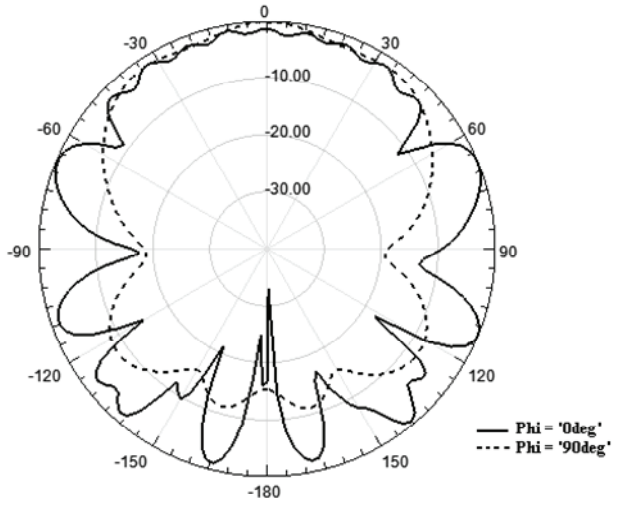


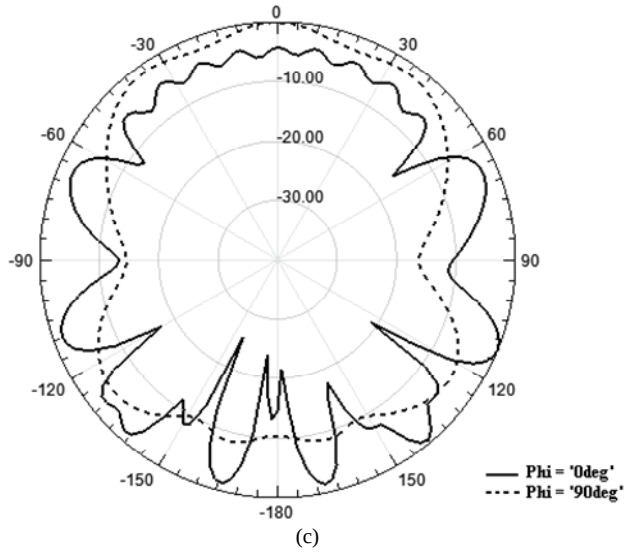
Figure 7. RDRA Gain versus air gap at 10.4 GHz.



(a)



(b)



(c)

Figure 8. RDRA radiation pattern at 10.4 GHz for: a) gap = 0.1 mm; b) gap = 0.3 mm; c) gap = 0.6 mm.

shows the return loss of antenna structure.

4.4. Effect of Air Gap between Dielectric Resonator Antenna and Ground

Introducing a thin air gap, due to the roughness of the ground surface or failure to ensure complete contact between the DR and conducting parts of the RDRA structure, may significantly affects the radiation performance of a DRA. When an air gap exists between the resonator and the ground, the electric field component normal to the metallic part of the structure is much stronger in air gap than the field component inside the resonator, especially, when it is composed of a material of high dielectric constant.

To investigate the effect of air gap, a few simulation processes was carried out for different value of air gaps. The effect of gap on antenna gain is shown in **Figure 7**, which shows that for low values of distance between the resonator and ground, gain is high. However, with increasing air gap, antenna gain would decrease. **Figure 8** shows the effect of air gap on radiation patterns of the antenna. It confirms that increasing the gap, would decrease antenna gain.

5. Conclusions

In this paper a single RDRA excited by a DIL through a slot was numerically investigated by HFSS. The best ground plane width for maximum gain with a broadside radiation pattern was obtained. Results show that 7.7 dB gain at 10 GHz was obtained for 100 mm of ground plane width. Moreover, to increase antenna gain four sidewalls were added and maximum gain of 10.4 dB at 10.4 GHz was obtained which is 2.7 dB higher than the gain in case of structure without them. To reduce backward radiation, a reflector was placed at the back of the antenna structure. The results show that adding the reflector lead to reduce the backward radiation around 10 dB in E-plane, especially. Also, the effects air gap between dielectric resonator and ground plane on the radiation performance of the antenna was investigated and it was concluded that antenna gain decreased with increasing air gap.

6. References

[1] S. Kanamaluru, M. Y. Li and K. Chang, "Analysis and Design of Aperture Coupled Microstrip Patch Antennas and Arrays Fed by Dielectric Image Line," *IEEE Transactions on Antennas and Propagation*, Vol. 44, No. 7, 1996, pp. 964-974.

[2] A. Petosa, "Dielectric Resonator Antennas Handbook," Artech House Inc., 2007.

[3] A. A. Kishk, "Dielectric Resonator Antenna, a Candidate for Radar Applications," *Proceedings of* 2003 *IEEE Radar Conference*, Huntsville, May 5-8, 2003, pp. 258-264.

[4] J. Shin, A. A. Kishk and A. W. Glisson, "Analysis of Rectangular Dielectric Resonator Antennas Excited through a Slot over a Finite Ground Plane," *IEEE Antennas and Propagation Society International Symposium*, Salt Lake City, July 2000, pp. 2076-2079.

[5] Y. Coulibaly and T. A. Denidni, "Design of a Broadband Hybrid Dielectric Resonator Antenna for X-Band Applications," *Journal of Electromagnetic Waves and Applications*, Vol. 20, No. 12, 2006, pp. 1629-1642.

[6] Ansoft Corporation, "HFSS: High Frequency Structure Simulator Based on Finite Element Method," Vol. 11, Ansoft Corporation, 2003.

[7] H. Dashti, M. H. Neshati and F. Mohanna, "Numerical Investigation of Rectangular Dielectric Resonator Antennas (DRAs) Fed by Dielectric Image Line," *Progress in Electromagnetic Research Symposium Proceedings*, Moscow, Russia, August 18-21, 2009, pp. 1164-1168.

[8] A. S. Al-Zoubi, A. A. Kishk and A. W. Glisson, "Analysis and Design of a Rectangular Dielectric Resonator Antenna Fed by Dielectric Image Line through Narrow Slots," *Progress in Electromagnetics Research*, Vol. 77, 2007, pp. 379-390.

[9] P. Bhartia and I. J. Bahl, "Millimeter Wave Engineering and Applications," John Wiley, New York, 1984.

Int. J. Communications, Network and System Sciences, **2010, 3, 625-630**
doi:10.4236/ijcns.2010.37084 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Scientific Research

A QoS-Based Multichannel MAC Protocol for Two-Tiered Wireless Multimedia Sensor Networks

GholamHossein EkbataniFard, Mohammad H. Yaghmaee, Reza Monsefi
Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
E-mail: {*Ekbatanifard, Yaghmaee*}*@ieee.org, Monsefi@um.ac.ir*
Received March 8, 2010; *revised April* 24, 2010; *accepted May* 27, 2010

Abstract

Rapid penetration of small customized wireless devices and enormous growth of wireless communication technologies have already set the stage for large-scale deployment of wireless sensor networks. Offering precise quality of service (QoS) for multimedia transmission over sensor networks has not received significant attention. However offering some better QoS for wireless multimedia over sensor networks raises significant challenges. In this paper, we propose an adaptive Cross-Layer multi-channel QoS-MAC protocol to support energy-efficient, high throughput, and reliable data transmission in Wireless Multimedia Sensor Network (WMSNs). Our proposed protocol uses benefit of TDMA and CSMA/CA to adaptively assign channels and timeslots to active multimedia sensor nodes in clusters. Simulations show that the proposed system achieves the performance objectives of WMSNs with increased network throughput at the cost of a small control and energy overhead.

Keywords: Wireless Multimedia Sensor Networks, MAC, Multichannel, Cross-Layer, Cluster, Adaptive

1. Introduction

The main component of wireless (multimedia) sensor network, are the sensor nodes, which are small in size, capable of self-organizing, sensing, processing data and communicating with other nodes. The availability of inexpensive hardware such as CMOS cameras and microphones that can ubiquitously capture multimedia content from the environment has fostered the development of Wireless Multimedia Sensor Networks [1], *i.e.*, networks of wirelessly interconnected devices that can retrieve video and audio streams, images, and scalar sensor data.

The major objectives behind the research and deployment of sensor networks [2] lie in the following two broad aspects:

1) Event detection and possible data acquisition by sensing, data processing and communication through node coordination and data transmission [3,4] to the sink or to the interested user.

2) Conservation of energy [5] to maximize the post deployment, active lifetime of individual sensor nodes and the overall network. The reason is that replenishing the energy of sensor nodes by battery-replacement is clearly not feasible for a large network consisting of hundreds of nodes. Moreover, wireless sensors are often deployed in an area which is in-approachable to humans and away from any sustained power-supply.

On the other hand, today's wireless communication is a gradually changing paradigm from its existing voice-alone services to a new world of real-time audio-visual applications.

This ever-increasing popularity of multimedia applications has already started penetrating the domain of wireless sensor networks—thereby giving birth to the new terminology wireless multimedia sensor networks [6].

Video surveillance, telemedicine and traffic-control are going to be the killer-applications of these emerging WMSNs. While the need to minimize the energy consumption has driven most of the existing research in wireless sensor networks, these new applications require the sensor network paradigm to be re-investigated in view of application-specific quality of service (QoS).

A quick look into the existing MAC protocols for sensor networks reveals that lack of standardization and application-specific diverse requirements has deprived wireless sensor networks from having a single de-facto standard MAC protocol. Most of the existing MAC protocols for wireless sensor networks can be divided into two categories: 1) time division multiple access (TDMA)-based and 2) carrier sense multiple access (CSMA) based with (possible) collision avoidance (CA) [7].

TDMA protocols have a natural advantage of colli-

sion-free medium access; CSMA–CA protocols have a lower delay at varying traffic loads. However, transmitting multimedia applications with QoS offers significant new challenges over these energy-constrained sensor networks. Design of an efficient sensory MAC protocol, satisfying QoS requirements, is one major step in end to end QoS provisioning over WMSNs.

Current sensor nodes, such as MICAz and WINS, already support multiple channels for communication, for example, 40 channels in WINS [1].

Thus, by developing a multichannel MAC protocol, which can effectively utilize the available channel capacity through the cooperative work from other sensor nodes, we can achieve a better support for multimedia applications which demand for high data rates [8].

This motivates us to look for QoS-based, yet energy-aware, MAC protocols for WMSNs. The objective of this work is to develop a new QoS-based, energy-aware MAC protocol for WMSNs. In this paper, we propose an adaptive cross-layer multichannel protocol for MAC layer in WMSNs. This protocol use benefits of TDMA and CSMA/CA techniques in one MAC protocol.

The rest of the paper is organized as follows. Section II reviews existing works in sensory MAC protocols. Subsequently, in Section III we explain our proposed MAC protocol at some sub sections. Simulation results in Section IV corroborate the efficiency of the protocol in achieving the desired throughput and delay. Section V concludes the paper.

2. Related Work

A good survey of major MAC protocols for wireless sensor networks is provided in [9]. Provisioning QoS in MAC layers for wireless cellular and local area networks [10] is an active research area, QoS-based MAC protocol for wireless sensor networks have received relatively less attention. While both TMAC [11] and DSMAC [12] attempt to reduce the latency, little of the other MAC protocols are developed with an objective to optimize (or improve) some application-specific quality of service (like delay, throughput etc).

Protocols, like SPEED [13], cluster-QoS [14] and delay-constrained least cost routing [15] discuss the QoS-routing issues in wireless sensor networks. Unfortunately, all these works attempt to optimize QoS in the sensor-routing from higher layers only. However, end-to-end QoS in WMSNs cannot be satisfied without designing an efficient QoS-aware MAC protocol. Unfortunately only a handful of works exist for QoS-MAC in wireless sensor networks. These include Q-MAC [16], PQ-MAC [17], and RL-MAC [18]. To the best of our knowledge COM-MAC [8] is the first one that use multi channel techniques for MAC protocol in WMSNs. But it uses static time slots at control channel and doesn't propose any mechanism for nodes that do not have data for sending at start of intervals or for nodes that start sending data between an interval that these could increase delay and degrades throughput of the network.

This motivates us to develop a new cross-layer multi-channel QoS-aware MAC protocol for clustered WMSNs that adaptively changes the intervals and use dynamic nature at channel and time slot assigning, therefore promoting the throughput of the network and exploit the high energy efficiency.

3. Design of the Proposed Protocol

3.1. Network Architecture

As shown in **Figure 1**, a WMSN consists of several more powerful nodes as cluster heads that located at the center of different monitoring area, a number of identical and stationary multimedia sensor nodes surrounding each cluster head and a remote data sink which stores the multimedia content locally for later retrieval. Each sensor node can communicate directly with its cluster head and cluster head can communication directly with the data sink using an out-of-band channel. But, if direct communication is not available, multi-hop routing is also employed.

3.2. Our Assumptions

We make some assumptions with relation to the configuration of the network. These assumptions are:

Topology of network is cluster based.

There are N different channels available for use and all channels have the same bandwidth except one that has lesser bandwidth than others and use as reserved channel, namely channel-R.

Cluster heads can transmit or receive on N channels at the same time.

Cluster heads will have sufficient power supply and more processing capacity than other sensor nodes.

All multimedia sensor nodes in a cluster can transmit or

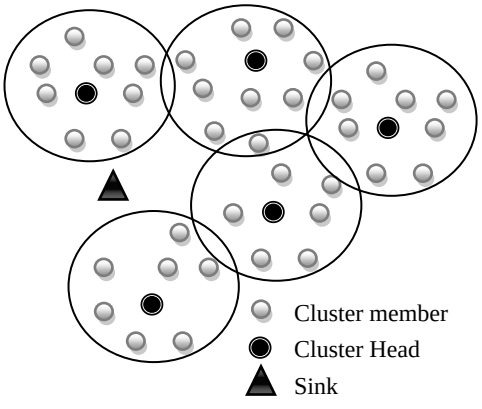


Figure 1. Network architecture.

receive on three channels, namely channel-1 and channel-2 and channel-R. Channel-1 is a contention based channel that assigned by cluster head at first phase of network deployment. Channel-2 is contention free and dynamically assigned by cluster head. Channel-R is a contention based reserved channel that is share between a cluster members. Sensor nodes are able to switch among channels dynamically. The channel switching time is less than 224 μs according to [1].

The working of a cluster of sensor nodes is synchronized to the cluster head and each sensor node can communicate directly with its cluster head.

3.3. Proposed QoS-MAC Protocol

We assume that the clustering process has been completed by performing a clustering protocol, the assignment of sensor nodes to clusters can be handled by existing clustering techniques [2]. Within each cluster, all tasks are done in time intervals (ΔT), which are dynamically changed. ΔT as Time interval can be varying depending on application and traffic load of the network.

We suppose three type nodes in a cluster as illustrated in **Figure 2**. These nodes are: cluster head, active nodes which are nodes that have data for sending, and passive nodes which have not data for sending at present. First of all, when the network is initially deployed, channel-1 allocation phase begin in each cluster and doing only one time. Channel-1 will be used as control channel for sending request message from multimedia sensor nodes to cluster head. As mentioned before the number of channels at cluster head is limited. So, it may happen that a channel assigned to more than one sensor nodes in a cluster. N-1 channels of cluster head could be assigned at this phase as channel-1 of multimedia sensor nodes. One remained channel from N channels of cluster head will be used as channel-R. That is share between all of nodes in a cluster. The usage of channel-R will be expressed later.

The operations of a cluster on ΔT are organized in three sequential phases: request phase, scheduling phase and data transmission phase. We now explain our MAC protocol details in three phases.

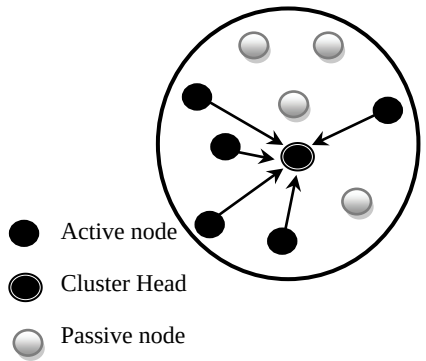


Figure 2. Node types in a cluster.

3.3.1. Request Phase

After channel-1 assignment phase, that runs only one time, the network operations begin at ΔT intervals. At the start of each ΔT, the network layer of each multimedia sensor node determines that whether information exists for sending or not in a cross layer manner. Then nodes that have information for sending, active nodes, start request phase on channel-1, and send a request message (REQ) to the cluster head.

The REQ message includes QoS requirements, such as amount of multimedia data to be transmitted, maximum delay, priority information and traffic class (streaming video, Non-Real Time (NRT), Best Effort (BE)), and Packet Error Rate (PER).

Because channel-1 may be assigned to more than one node in a cluster, so, adaptive contention window protocol [19] can be used for better performance on this channel.

When active nodes received acknowledgement of its REQ message, they go to standby mode and waiting for scheduling message from cluster head.

Request phase time (ΔTr) determined dynamically, based on event rate, traffic load and average of previous ΔTr periods.

3.3.2. Scheduling Phase

After request phase, cluster heads gather REQ messages and then start scheduling phase. Each cluster head calculate an appropriate schedule, based on priority and other QoS requirements that specified in REQ messages, to coordinate the data transmissions of active nodes. Then cluster heads broadcast scheduling messages through all N-1 channels.

In scheduling message, cluster head assign a channel as channel-2 to each active node for data transmission. If the number of active nodes in a cluster is more than N – 1, then a channel should assign to more than one active node. In fact, a time slot in a channel may be assign to an active node as channel-2.

Therefore, the scheduling message includes a channel and probably a time slot in it as channel-2 for each active node. Moreover new ΔT that calculated based on REQs is included in scheduling message. New ΔT specifies end of this interval and start of next interval indeed. A cluster time intervals have been illustrated in **Figure 3** for three intervals.

Lengths of time slots that assign to active nodes are depending on amount of data that specified in REQ me-

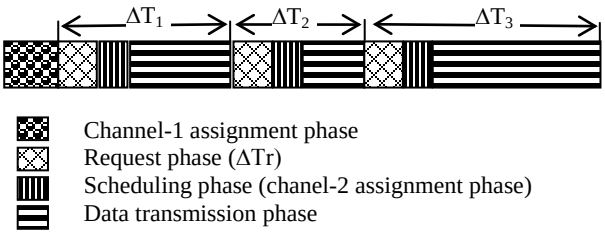


Figure 3. A cluster operations for three intervals.

ssages and times need for sending acknowledgment messages if needed. Some traffic types may not require acknowledgement message from cluster head, that it should be declared in REQ message.

The priority of REQs is essentially based on its traffic class, low priority for best effort traffic, medium priority for non-realtime and high priority for streaming video traffics.

Figure 4 shows the pseudo-code of scheduling algorithm that assigns channel-2 and timeslots to active nodes in a cluster. Also announce new_ΔT for next interval. Thereafter, scheduling message broadcasts for sensor nodes in the cluster at network.

If some nodes that were passive get active after this phase, send their REQ messages to cluster head on channel-R. Then, if there is enough unused time on channels, it could be assigned to these nodes. Otherwise, if there is not enough time for some of these nodes, then these nodes only will be announced with new_ΔT. So unsuccessful nodes will go to sleep mode until starting of next request phase.

The network throughput in an interval is given by

$$\tau_j = \sum_{i=1}^{k} P_i / ((N-1)T_j C) \quad (1)$$

where k is the total number of REQ messages in jth interval, P_i is amount of bits of data requested for transmission in ith REQ message. Maximum occupied channel time in jth interval is T_j seconds. In other word T_j is equal to new_ΔT at jth interval that used in **Figure 4**.

N – 1 is the number of contention free channels available at cluster head. And capacity of each these channels are C bps. The total throughput of network in m time intervals is the average of τ_j for $j = 1, 2\ldots m$.

3.3.3. Transmission Phase

After receiving a scheduling message by an active node, it could send its data on assigned channel. If a time slot in a channel is assigned to an active node, it could go to sleep mode until its time slot for sending data approaches.

As mentioned before, some passive nodes may get active at this phase, and send its REQ messages at channel-R. So, if free channels or time slots in channels has been assigned to such nodes, they could send its data at scheduled time.

Scheduling algorithm of Proposed MAC Protocol

1. ***Sort*** *"REQs based on Priority in descending order"*
2. ***Sort*** *"equal priority REQs based on amount of data in descending order"*
3. **Repeat**
4. ***Find*** *"first minimum occupied channel time"*
5. ***Assign*** *"channel or timeslot in channel to REQ(i)"*
6. **Until** *REQ (i) exists, i = 1, 2,…, k*
7. *New_ΔT =* ***Find*** *"maximum occupied channel time"*

Figure 4. Scheduling algorithm.

When cluster head receives packets from its cluster members, it classifies traffics based on its priority then schedules it for sending toward the sink. Such framework illustrated in **Figure 5**.

Nodes receive acknowledgement messages (ACK) for proper sent packets, if specified before at REQ message. Nodes could request unused timeslots, if exist, in an interval for lost packets.

As mentioned earlier the streaming video traffic is assigned the highest priority and the best effort traffic is assigned the lowest priority. We will now analyze the average delay incurred in each of this traffic class. The mean waiting time of a type i customer is denoted by $E(W_i)$ and $E(L_i^q)$ is the number of type i customers waiting in the queue. Further let's assume the processing time of traffic class i is μ_i, with mean $E(\mu_i)$ and residual processing time (R_i), with mean $E(R_i)$. Then the traffic intensity of the system is given by: $\rho_i = \lambda_i E(u_i)$ [12]. Hence, for the highest priority streaming video traffic it holds that

$$E(W_1) = E(L_1^q)E(\mu_1) + \sum_{j=1}^{r} \rho_j E(R_j) = \frac{\sum_{j=1}^{r} \rho_j E(R_j)}{1-\rho_1} \quad (2)$$

where r is the number of different traffic classes whose service is in progress during the arrival of the highest priority traffic class.

And the mean waiting time for lower priority traffics [7] could be estimated as

$$E(W_i) = \frac{\sum_{j=1}^{i} E(L_j^q)E(\mu_j) + \sum_{j=1}^{r} \rho_j E(R_j)}{1-(\rho_1 + \cdots + \rho_{i-1})} \quad (3)$$
$$\forall i : 2 \le i \le n.$$

4. Performance Evaluation

In this section we show different simulation results demonstrating the efficiency of this proposed MAC protocol. We have developed a discrete-event object oriented packet-level simulator in C++. In the simulations presented in this section, the considered packet size is 25 bytes. We assume 3 data channels at sink that each channel capacity is 250 kbps. Experiments are repeated 10 times. The performance of our algorithm is compared with COM-MAC

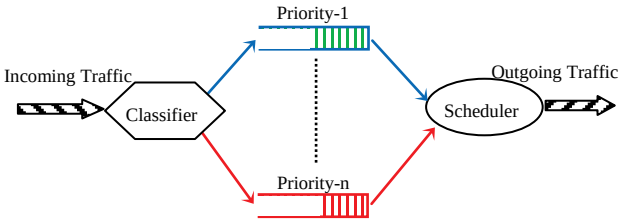


Figure 5. Traffic differentiation and priority queuing in WMSN.

[8] and a baseline protocol, the multichannel TDMA (M-TDMA) protocol. For M-TDMA, the cluster head first evenly distribute the cluster members on the available channels. Then, the cluster head generates a TDMA schedule on each channel and allocates a fixed slot to each cluster members.

Figure 6 compares the network throughput performance of proposed MAC protocol with COM-MAC and M-TDMA for different cluster sizes. Our protocol works well. But, when the number of nodes in cluster increases, the proposed protocol throughput decreases. To find the reason we repeat simulation up to 100 nodes in a cluster, **Figure 7**, I see that our protocol throughput reach to a steady state and it has better throughput than other protocols. This is because the channel tends to be saturated when more nodes are trying to utilize the channel.

As expected, proposed MAC protocol outperforms other two protocols. This is because our protocol is designed to maximize the network throughput by adaptively changing intervals and using unused channels and time slots for passive nodes that get active after request phase.

Figure 8 shows the delay performance compassion of three protocols as cluster size increases. We see that our proposed MAC incurs lower delay when compared to COM-MAC and M-TDMA.

It is because that in our protocol, when a passive node gets active, it could send request from channel-R to use unassigned space on channels for data transmission. So it incurs lower delay than other two protocols. With COM-MAC and M-TDMA such nodes should wait until next intervals thus this increases packet delay. We also notice that the delay performance increases as cluster size increases. This is because that larger cluster size will lead to heavier network load so that a packet has to wait longer to be transmitted.

Figure 9 explains the throughput-dynamics for different traffic classes. The novelty of our MAC protocol is that it first classifies the traffic into different classes depending on the type of service (ToS) then schedules it for data transmission. The streaming video traffic is given the highest priority, the NRT traffic is given the second priority and the BE traffic attains a lowest throughput.

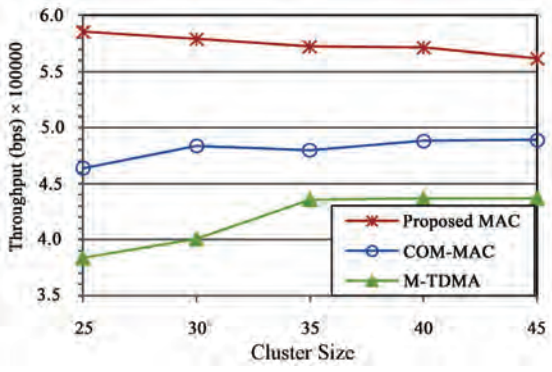


Figure 6. Throughput performance for various cluster sizes up to 45.

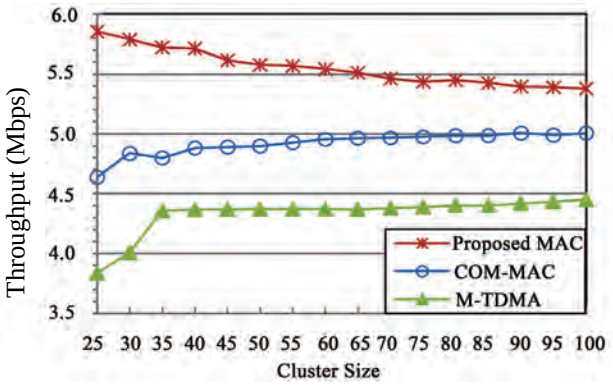


Figure 7. Throughput performance for various cluster sizes up to 100.

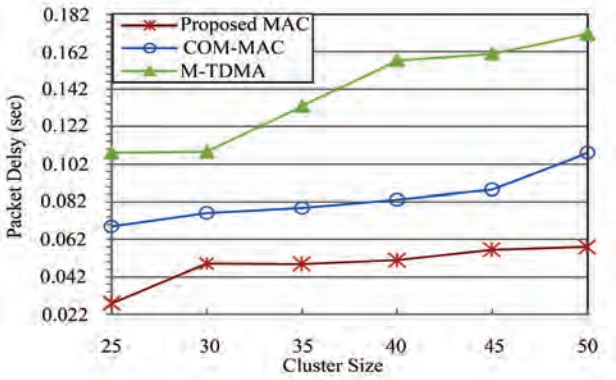


Figure 8. Packet delay performance for various cluster.

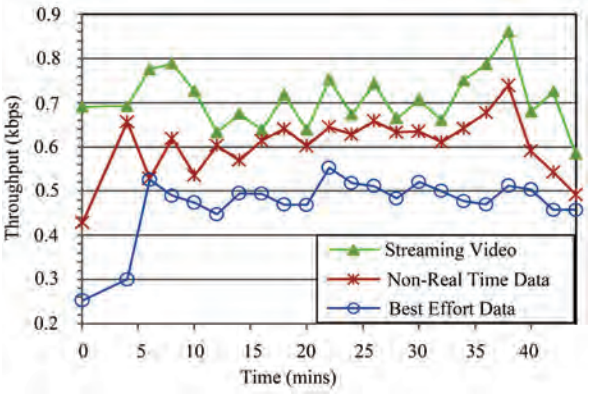


Figure 9. Differentiated MAC-throughput.

5. Conclusions

In this paper we have developed a cross layer multichannel QoS-MAC protocol for wireless multimedia sensor networks which classifies the wireless traffic into different class, and adaptively assigns channel to various traffics. In our proposed protocol nodes get active only when the network layer specifies that there are data for sending. We

verify the advantages of our protocol through network simulation, in terms of network delays, throughput and differentiated throughput of different traffic classes. We see that our proposed MAC protocol provides better energy-efficiency, high-throughput, and data reliability support in WMSNs.

6. Acknowledgements

This work was supported in part by grants from Faculty of Engineering Ferdowsi University of Mashhad under the contracts 18069.

7. References

[1] I. F. Akyildiz, T. Melodia and K. R. Chowdury, "A Survey on Wireless Multimedia Sensor Networks," *Computer Networks* (*Elsevier*), Vol. 51, No. 4, 2007, pp. 921-960.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "A Survey on Sensor Networks," *IEEE Communications Magazine*, Vol. 40, No. 8, 2002, pp. 102-114.

[3] W. Heinzelman, J. Kulik and H. Balakrishnan, "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks," *Proceedings of the* 5*th Annual ACM/ IEEE International Conference on Mobile Computing and Networking*, Seattle, WA, August 1999, pp. 174-185.

[4] Y. Yao and J. Gehrke, "The COUGAR Approach to In-Network Query Processing in Sensor Networks," *ACM SIGMOD Record*, Vol. 31, No. 3, 2002, pp. 9-18.

[5] R. A. F. Mini, M. do V. Machado, A. A. F. Loureiro and B. Nath, "Prediction-Based Energy Map for Wireless Sensor Networks," *Ad Hoc Networks*, Vol. 3, No. 2, 2005, pp. 235-253.

[6] I. F. Akyildiz, T. Melodia and K. R. Chowdhury, "A Survey on Wireless Multimedia Sensor Networks," *The International Journal of Computer and Telecommunications Networking*, Vol. 51, No. 4, 2007, pp. 921-960.

[7] N. Saxena, A. Roy and J. Shin, "Dynamic Duty Cycle and Adaptive Contention Window Based QoS-MAC Protocol for Wireless Multimedia Sensor Networks," *Computer Networks* (*Elsevier*), Vol. 52, No. 13, 2008, pp. 2532-2542.

[8] C. Li, P. Wang, H.-H. Chen and M. Guizani, "A Cluster Based On-demand Multichannel MAC Protocol for Wireless Multimedia Sensor Networks," *IEEE International Conference on Communications*, Beijing, May 19-23, 2008, pp. 2371-2376.

[9] I. Demirkol, C. Ersoy and F. Alagoz, "MAC Protocols for Wireless Sensor Networks: A Survey," *IEEE Communications Magazine*, Vol. 44, No. 4, 2006, pp. 115-121.

[10] T. Kuhn, "A QoS MAC Layer for Ambient Intelligence Systems," *Proceedings of the* 4*th International Conference on Pervasive Computing*, Dublin, 2006, pp. 69-72.

[11] T. V. Dam and K. Langendoen, "An Adaptive Energy-Efficient MAC Protocol for Wireless Sensor Networks," *Proceedings of the* 1*st International Conference on Embedded Networked Sensor Systems*, Los Angeles, 2003, pp. 171-180.

[12] L. Kleinrock, "Queueing Systems, Theory," John Wiley & Sons, New York, 1975.

[13] T. Hea, J. A. Stankovica, C. Lub and T. Abdelzahera, "SPEED: A Stateless Protocol for Real-Time Communication in Sensor Networks," 23*rd IEEE International Conference on Distributed Computing Systems*, Rhode Island, USA, May 2003, pp. 1-10.

[14] S. S. Tang and W. Li, "QoS Supporting and Optimal Energy Allocation for a Cluster Based Wireless Sensor Network," *Computer Communications*, Vol. 29, No. 13-14, 2006, pp. 2569-2577.

[15] Q. Gao, K. J. Blow, D. J. Holding, I. Marshall and X. H. Peng, "Radio Range Adjustment for Energy Efficient Wireless Sensor Networks," *Ad-Hoc Networks*, Vol. 4, No. 1, 2006, pp. 75-82.

[16] Y. Liu, I. Elhanany and H. Qi, "An Energy-Efficient QoS-Aware Media Access Control Protocol for Wireless Sensor Networks," *IEEE International Conference on Mobile Adhoc and Sensor Systems*, Washington, DC, November 7, 2005, pp. 191-193.

[17] K. Paek, J. Kim, U. Song and C. Hwang, "Priority-Based Medium Access Control Protocol for Providing QoS in Wireless Sensor Networks," *IEICE Transaction Letters on Information Systems*, Vol. E90-D, No. 9, 2007, pp. 1448-1451.

[18] Z. Liu and I. Elhanany, "RL-MAC: A QoS-Aware Reinforcement Learning Based Mac Protocol for Wireless Sensor Networks," *IEEE International Conference on Networking, Sensing and Control*, Lauderdale, 2006, pp. 1-6.

[19] N. Sabena, A. Roy and J. Shin, "Dynamic Duty Cycle and Adaptive Contention Window Based QoS-MAC Protocol for Wireless Multimedia Sensor Networks," *Computer Networks*, Vol. 52, No. 13, 2008, pp. 2532-2542.

Int. J. Communications, Network and System Sciences, **2010, 3, 631-637**
doi:10.4236/ijcns.2010.37085 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Scientific Research

Modeling and Analysis of Bandwidth Allocation in IEEE 802.16 MAC: A Stochastic Reward Net Approach

Shanmugam Geetha, Raman Jayaparvathy
Department of EEE, Coimbatore Institute of Technology, Coimbatore, India
E-mail: *geetha_thiagu@yahoo.com, jayaparvathy*14*@gmail.com*
Received April 20, 2010; *revised May* 27, 2010; *accepted July* 2, 2010

Abstract

In this paper, we present a stochastic reward net (SRN) approach to analyse the performance of IEEE 802.16 MAC with multiple traffic classes. The SRN model captures the quality of service requirements of the traffic classes. The model also takes into account pre-emption, priority and timeout characteristics associated with the traffic classes under consideration. The performance of the system is evaluated in terms of mean delay and normalized throughput considering the on-off traffic model. Our analytical model is validated by simulations.

Keywords: Wimax, IEEE 802.16, Stochastic Reward Net, Mean Delay, Throughput

1. Introduction

Over the last few years there has been tremendous increase in the use of broadband access. The deployment has boosted the usage of several multimedia applications such as Voice over IP (VoIP), online gaming and Video on Demand (VoD). However, in the rural and suburban areas, deployment of traditional wired technologies is too expensive. In such cases, broadband wireless access (BWA) based on IEEE 802.16 provides a promising solution [1,2]. One of the key features of IEEE 802.16 is that it supports multiple applications such as HDTV, video conference and conventional internet applications. The challenge for BWA networks is to simultaneously provide quality of service (QoS) to applications with very different characteristics. Hence, a proper resource allocation scheme for packet transmission is imperatively needed.

Performance evaluation of resource allocation mechanisms plays an important role in design of communication systems. Increasing complexity of networks and the way in which they are used, has made it difficult to construct models that are analytically tractable. SRNs are very useful in analytical modeling of complex networks. System operations can be precisely described by means of a graph which translates into a markovian model. Properties such as liveness and deadlock freeness make SRN a reliable analytical modeling tool.

SRN has been used extensively for performance modeling. Performance of opportunistic and non opportunistic schedulers was compared in [3] using analytical model developed with stochastic Petri net (SPN). A protocol of QoS has been developed using Petri net in [4]. The protocol has been verified for service guarantee and effective use of resources. Modeling power, analysis and verification of SPN has been discussed in [5]. Application of Petri net (PN) in performance and availability analysis is discussed in [6]. The authors in [7] presented a SRN approach to model IEEE 802.11 DCF with on-off traffic model. Performance metrics such as mean delay and average system throughput have been evaluated. Reconfigurable PN and their ability to model dynamic systems have been studied in [8].

Several approaches have been used for performance evaluation of IEEE 802.16 networks. Simulation approach has been followed in [9] for evaluating IEEE 802.16 system metrics such as mean delay and throughput. Analytical approach to study bandwidth allocation process has been presented in [10,11]. Packet scheduling scheme for QoS provisioning in WiMax networks is discussed in [12]. The proposed scheme in [12] has been verified using simulations. In [13], authors have proposed a Light WiMAX simulator (LWX) for evaluating performance of IEEE 802.16 bandwidth allocation algorithms. Simulation approach has been adopted in [14] to compare various scheduling schemes such as round-robin, token bucket-based and M-LWDF algorithms. Authors in [15] have proposed an intelligent bandwidth allocation of uplink (IBAU) for WiMax systems. IBAU mechanism is shown to decrease delay and increase throughput of the network. A survey on scheduling

schemes in IEEE 802.16e systems has been presented in [16]. Simulation methodologies to be adopted for MAC and PHY layers of IEEE 802.16 are presented in [17].

In this paper, we propose a SRN approach to model and analyze performance of the IEEE 802.16 MAC with multiple traffic classes. The proposed model incorporates prioritization and pre-emption of traffic classes. Packet drop due to waiting time exceeding threshold is also considered. We compute the average system throughput and mean delay suffered by the first packet (*i.e.*, the packet in the head of line (HOL) of each queue, through the proposed SRN formulation. Mean delay of subsequent packets is determined by modelling each queue as M/G/1 queue [7]. The mean service time for the computation is obtained from the mean delay suffered by the HOL packet. Our analytical model is validated by comparing the results with simulations carried out using event based simulator.

The rest of the paper is organized as follows: Section 2 presents a brief overview of IEEE 802.16 MAC. System model is presented in Section 3. Section 4 discusses the performance evaluation. Results and discussion are presented in Section 5. Conclusions are drawn in Section 6.

2. IEEE 802.16 MAC

IEEE 802.16 system consists of two kinds of fixed stations: subscriber station (SS) and base station (BS). All communication in the network is regulated by BS. Two direction of communication path exists between BS and SS: uplink (from SS to BS) and downlink (from BS to SS). IEEE 802.16 MAC defines QoS signaling mechanisms and functions that control BS and SS data transmissions. Two modes of sharing the wireless medium is possible: Point-to-Multipoint (PMP) and Mesh. In PMP, BS serves a set of SS in a broadcast manner. Coordination of transmissions from SSs is done by BS. In mesh mode, organization of nodes is in ad hoc manner and communication exists between SS. In this paper, we focus on PMP mode.

The IEEE 802.16 MAC defines four different scheduling service flows in order to meet the QoS requirements of multimedia applications [9]. *Unsolicited Grant Service* (UGS) is designed to support real-time applications, with strict delay requirements which generate fixed-size packets at periodic intervals such as T1/E1. *Real-time Polling Service* (rtPS) is designed to support real-time applications with less stringent delay requirements, which generate variable size packets at periodic intervals, such as VoIP with silence suppression. *Non-real-time Polling Service* (nrtPS) support non-real-time variable bit rate services, such as FTP. *Best Effort* (BE) traffic does not have QoS guarantees, such as HTTP. Since rtPS, nrtPS and BE traffic classes have varying bandwidth requirements; bandwidth allocation for these classes is performed dynamically. As UGS is allocated fixed and reserved bandwidth, dynamic reassignment of bandwidth is not required.

SS maintains separate connection for each service flow. The allocation of bandwidth by the BS to SS is based on two modes: grant per subscriber station (GPSS) and grant per connection (GPC). In GPSS, the SS obtains aggregate bandwidth for all its individual flow and in turn reallocates the bandwidth to each flow individually. In GPC, the bandwidth allocation by BS is made on per flow basic. We assume GPSS mode of operation in this paper.

3. System Model

A typical IEEE 802.16 network consists of multiple BSs. Each BS covers several SSs. Every SS is associated with multiple queues corresponding to different traffic classes. We model a single SS with three queues corresponding to rtPS, nrtPS and BE traffic classes as shown in **Figure 1**. The SS is assigned aggregate bandwidth by the BS. The three queues contend for bandwidth from the SS. The objective is to obtain the mean delay and normalized throughput of each traffic class for varying load conditions. The analytical model is required to take into account prioritization, pre-emption and dropping of packets (with waiting time exceeding the threshold) corresponding to various traffic classes.

Packets arrive at each of the queues in random epochs of time. Data packets arriving at a queue gets buffered till they gain access to channel. Newly arriving packets are added to the queue on a first come first serve (FCFS) basis. Delay of a packet is defined as the time spent by a packet till it is successfully transmitted. Normalized throughput of a given traffic class is defined as the ratio of successful packets transmitted to total packets generated. Average system throughput is the sum of throughputs of individual traffic class.

The following assumptions are made in the model.

- There are 3 different traffic classes in the system, namely rtPS, nrtPS and BE denoted as $class_1$, $class_2$ and $class_3$ respectively.
- We consider data-only traffic with on-off traffic model. Data bursts consist of active and idle periods. (Practically, a data burst represents data packet of variable

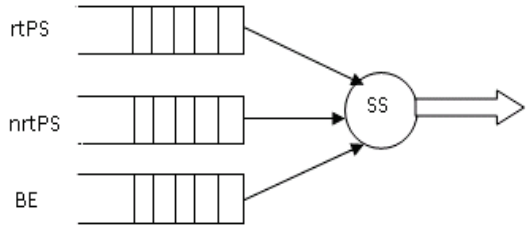


Figure 1. System model.

length, for example an IP packet with zero idle time between finite set of consecutive packets.[7])

- Data bursts arrival at any queue follows a Poisson process with mean arrival rate λ_i.
- Service times of data bursts are exponentially distributed with mean $1/\mu_i$ seconds.
- The SSs are assumed to have negligible mobility.

4. Performance Evaluation

In this section, we present a SRN model to evaluate the performance of the system considered in Section 3. Performance metrics considered are normalized throughput and mean delay suffered by a packets belonging to each traffic class.

4.1. Stochastic Reward Net Model

SRN model for a SS with three queues is shown in **Figure 2**. The model incorporates priority, pre-emption and timeout characteristics of the queues. **Tables 1-3** lists the various places, transitions and the meaning associated with each of them.

Transition usr_i generates packets at a given rate λ_i and deposits them into place q_i. An inhibitor arc with

Table 1. List of places.

Place	Meaning
cap	Total available bandwidth
usg_i	Number of channels currently in use
q_i	Packets in buffer

Table 2. List of timed transitions.

Timed Transition	Meaning
usr_i	Packet arrival at rate λ_i
end_i	Departure of packets after service at rate u_i
$time\_o_i$ *	Removal of time out packets at rate u_{to_i}

* $i = 1, 2$

Table 3. List of immediate transitions.

Immediate Transitions	Meaning
$chchk_i$	Priority transition checking availability of channel
$pre\_empt_{i,j}$	Enable pre-emption

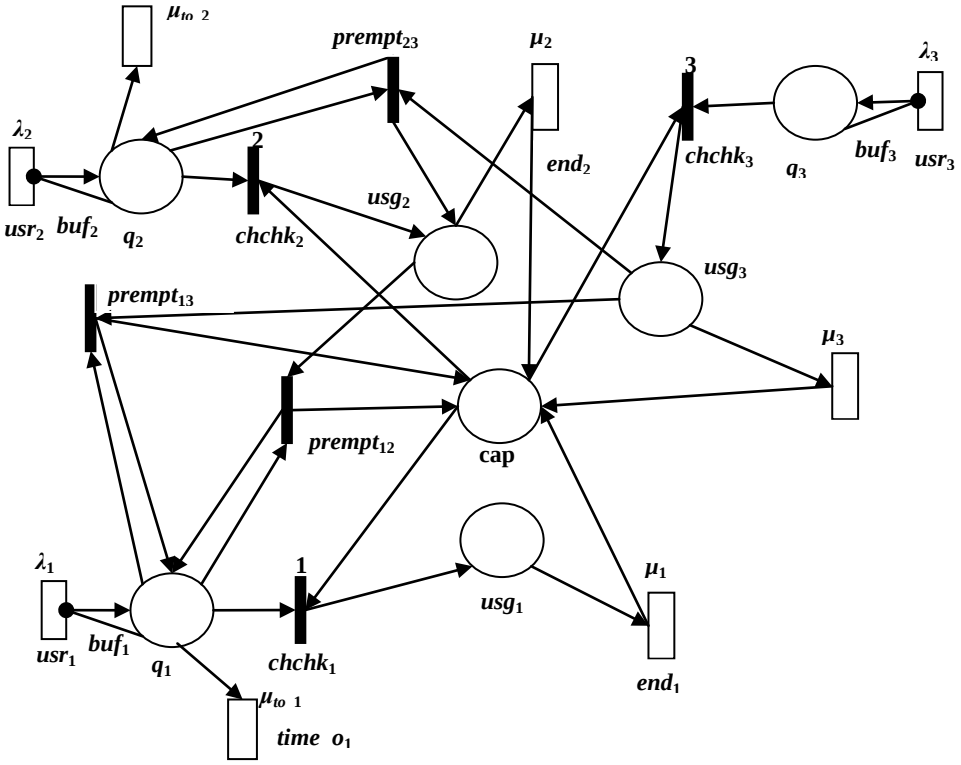


Figure 2. SRN model.

cardinality buf_i is needed to ensure that the number of packets waiting to enter the current queue is finite. If all channels are busy, the data packets are buffered in q_i with buffer size buf_i.

A way to assign priority is to give each transition an integer priority level. Transition $chchk_i$ are modelled as priority transitions. Lower integer value indicates higher priority level. A priority transition is enabled only if no other higher priority transition is enabled. Since, $chchk_1$ is assigned lowest value; $class_1$ has highest priority to gain access to channel, followed by $class_2$ and $class_3$. Firing $chchk_i$ transfers a packet from q_i to usg_i indicating the packet is being served. After completion of service time, transition end_i is fired and the channel is returned to the central pool. Note that $chchk_i$ are modelled as immediate transitions since they represent activity that does not imply time dependency. Although the action of assigning a channel implies time, the time is neglected from the point of view of traffic modelling.

In order to model pre-emption using SRN, it is required to check the simultaneous presence of a packet in place usg_{i+1} and q_i. The meaning of the above condition is that a lower priority packet is being served, when higher priority packet is waiting for resource. Transition $prempt_{i,j}$ are immediate transitions used to model pre-emption. $prempt_{i,j}$ is enabled when packets are available in places q_i and usg_j at the same time, where subscript i and j correspond to higher priority and lower priority traffic class respectively. Arc connecting $prempt_{i,j}$ indicates removal of packet from usg_j, and returning the channel to central pool of channels. Hence, firing $prempt_{i,j}$ pre-empts $class_j$ and enables $class_i$ to access the resource.

The channels available in the central pool of resource are shared by the traffic classes on arrival of data packets and returned to the pool on completion of service. At higher traffic loads, the available channels become insufficient to meet the bandwidth requirement. Under such conditions, packets in buffer wait for availability of resource. Traffic classes, $class_1$ and $class_2$, belong to delay sensitive application with maximum threshold on tolerable delay. Packets exceeding the threshold are dropped. Dropping of packets exceeding the delay limit is incorporated in the model using timed transitions $time\_o_i$. Firing rate of $time\_o_i$ is set to μ_{to_i}, $1/\mu_{to\_i}$ is the maximum tolerable delay for packets belonging to traffic $class_i$. Firing $time\_o_i$ removes a packet from q_i indicating the packet drop. Probability of packet drop depends on the available channels, transmission rates of packets, buffer size etc. Since, $class_3$ traffic is not associated with any such delay limit, we do not include time out feature for $class_3$.

4.2. Mean Delay and Normalized Throughput

The underlying continuous time markov chain (CTMC) of the SRN model discussed can be obtained from extended reachability graph (ERG) [7]. To obtain the desired performance metrics, one has to solve the CTMC. Complexity of CTMC increases with the size of the system. Solution of complex CTMC can be obtained by using standard software packages such as SHARPE [18], SPNica [19] or TimeNET [20]. The average number of packets in each place, and hence the steady state probability of occupancy of each state in the CTMC be determined using the software tools. In this paper, we use SHARPE to construct the SRN model and obtain the performance metrics.

The average throughput of a transition T is defined as the average rate at which packets are deposited by the transition in its output places. If $\hat{O}(t)$ is the average number of packets deposited by transition T in all of its output places up to a time t, then the throughput of a transition T, η_T is defined as

$$\eta_T = \lim_{t\to\infty} \frac{\hat{O}(t)}{t} \tag{1}$$

Since we consider three different traffic classes, the throughput of traffic class i, is given by

$$\eta_i = \frac{\eta_{end_i}}{\eta_{usr_i}} \tag{2}$$

Average system throughput, η is given by,

$$\eta = \sum_{i=1}^{3} \eta_i \tag{3}$$

The mean delay, $\hat{D}_H$, experienced by a HOL packet of traffic class i, is the sum of the mean packet holding time and the sum of mean waiting times in places q_i and usg_i. Let the average number of packets in place P be $\#\hat{P}$. $\hat{D}_H$ can be computed using Little's Theorem [21] as,

$$\hat{D}_{H_i} = \frac{\#(q_i)}{\eta_{usr_i}} + \frac{\#(usg_i)}{\eta_{chchk_i}} + \frac{1}{\mu_i} \tag{4}$$

where μ_i is the mean packet holding time for traffic class i. The buffer in each queue is modelled as M/G/1 queue with mean service time $\hat{D}_{H_i}$. The mean packet delay, $\hat{D}$ can be determined by applying the *Pollackzek-Kinchine* mean value formula [22] as

$$\hat{D}_i = \hat{D}_{H_i}\left[1 + \frac{\rho_{b_i}}{2\left(1-\rho_{b_i}\right)}\left(1+C_{R_i}^2\right)\right] \tag{5}$$

where . If delay of HOL packet is represented by random variable, R_i, then

$$C_{D_i}^2 = \frac{E\left[R_i^2\right]}{\hat{D}_{H_i}^{\,2}} \tag{6}$$

For small loads, $E\left[R_i^2\right]$ can be obtained as

$$E\left[R_i^2\right] = 2\left(\frac{\#usg_i}{\eta_{chchk_i}}\right)^2 \quad (7)$$

5. Results and Discussion

We evaluate the system performance in terms of mean delay and normalized throughput for increasing traffic load, ρ, given by $\sum_{i=1}^{3}\rho_i$, where ρ_i corresponds to traffic load of class$_i$ for $i = 1, 2$ *and* 3. $\rho_i = \lambda_i / \mu_i$, where λ_i is the arrival rate and μ_i is the service rate of each traffic class. Simulation parameters are shown in **Table 4**. Input traffic parameter settings are given in **Table 5.**

We compare the analysis and simulation results for three traffic classes in terms of mean delay and normalized throughput. From the results we find the simulation results match with the analysis, thus validating our analytical approach. We also analyse the performance of the system with varying buffer sizes.

Figure 3 presents a comparison of mean delay for three traffic classes with increasing traffic load. It is observed that the mean delay increases with traffic load. Mean delay suffered by packet of class$_1$ is least followed by class$_2$ and class$_3$. The increase in mean delay is more pronounced for class$_3$ since class$_3$ has the least priority among the competing traffic classes. At higher loads, class$_3$ packets are starved are resources which results in increased mean delay.

We further analyse the system with increased buffer size. **Figure 4** shows the comparison of mean delay for *buf* = 15. From the figure it is observed that with increasing buffer size there is no significant increase the mean delay of class$_1$ traffic because the packets belonging to class$_1$ have to wait for minimum amount of time to gain access to the channel. Further, since class$_1$ and class$_2$ packets are associated with a maximum tolerable delay,

Table 4. Simulation parameters.

Cell Radius	1 km
Duplexing Schemes	TDD
Ratio of Uplink slots to downlink in TDD	50%
Total available bandwidth	50 Mbps
Simulation time	500 s

Table 5. Input traffic parameters.

Traffic Class	Latency (ms)	Packet Size (Bytes)	Packet Interval (ms)	Traffic load (Kbps)	Mean Service Time (ms)
rtPS	8	240	2.6	2.8-20	0.6
nrtPS	10	120	3	2-10	0.5
BE	-	120	5	2-14	0.3

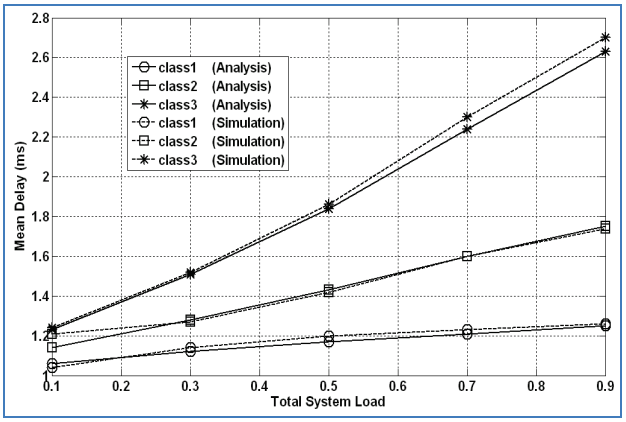


Figure 3. Comparison of mean delay (*buf* = 1).

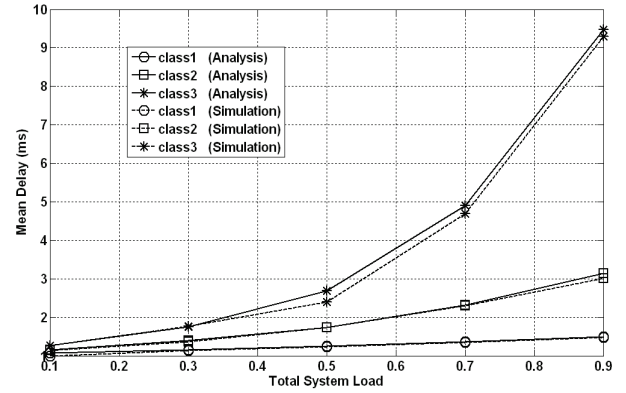


Figure 4. Comparison of mean delay (*buf* = 15).

packets exceeding the tolerable delay are dropped. Dropped packets introduce a decrease in throughput as observed in **Figure 7.**

Mean delay of class$_2$ and class$_3$ for varying buffer sizes is presented in **Figures 5** and **6**. From **Figure 6** we find that for a traffic load of 0.8, mean delay with *buf* = 1 and *buf* = 5 are 2.5 and 5.6 respectively resulting in 55% increase. For the same traffic load mean delay with *buf* = 10 and *buf* =15 are 6.9 and 7.2 respectively producing only a 4% increase. We observe that increase in buffer size does not produce a corresponding increase mean delay, particularly for higher values of buffer sizes. The reason is that the available bandwidth is insufficient to serve all packets in buffer. Hence, the number of packets successfully transmitted, which amounts to mean delay, does not increase significantly with increase in buffer size. Further, existing packets in buffer prevent additional packets entering the system.

Figures 7 and **8** present the normalized throughput of the three traffic classes with buffer size 1 and 15 respectively. From the graphs, it is observed that for a given buffer size, class$_1$ has the highest throughput followed by class$_2$ and class$_3$. Further, throughput of all traffic classes decrease with increase in traffic load. Comparing **Figures 7**

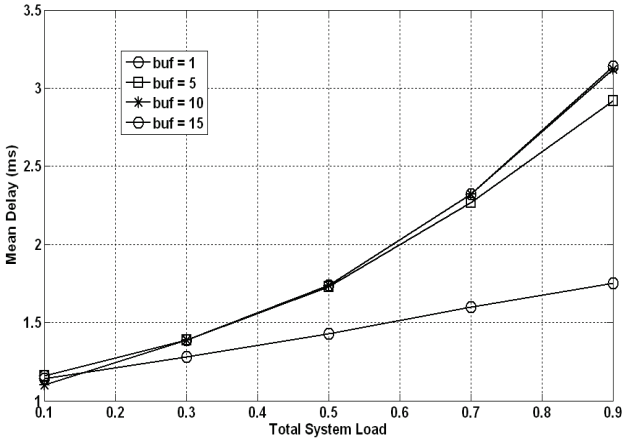


Figure 5. Mean delay of class$_2$ traffic for varying buffer size.

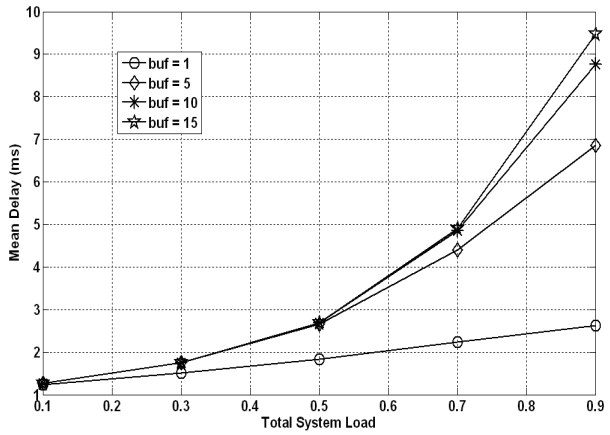


Figure 6. Mean delay of classs$_3$ traffic for varying buffer size.

and **8**, we find that increase in buffer size from 1 to 15 increases the throughput significantly. Decrease in throughput of class$_1$ traffic at higher traffic load is attributed to insufficient bandwidth. Also, class$_2$ and class$_3$ traffic suffer additional decrease in throughput due to pre-emption.

In **Figure 9**, presents the throughput of class$_3$ packets with increasing buffer sizes. From the graph it is observed that increasing buffer size from 1 to 5 increases the throughput significantly. But, further increase in buffer size from 10 to 15 does not produce any considerable increase in throughput. Further increase in buffer size results in saturation of the system with no further increase in throughput.

6. Conclusions

We presented a SRN formulation for performance evaluation of bandwidth allocation in IEEE 802.16 network considering multiple traffic classes. The model includes

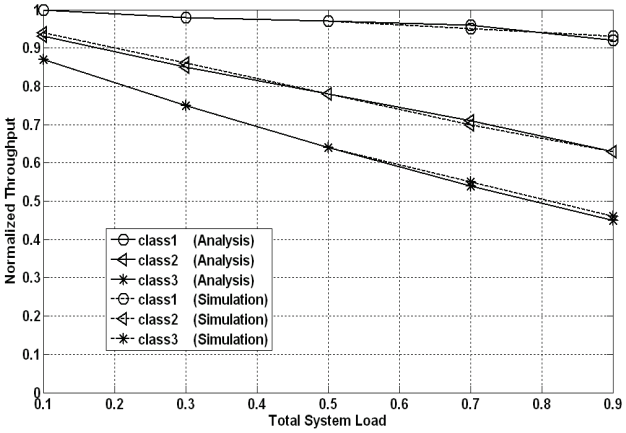


Figure 7. Comparison of normalized throughput (*buf* = 1).

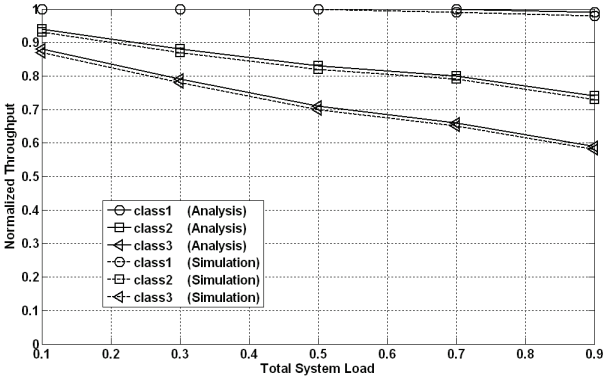


Figure 8. Comparison of normalized throughput (*buf* = 15).

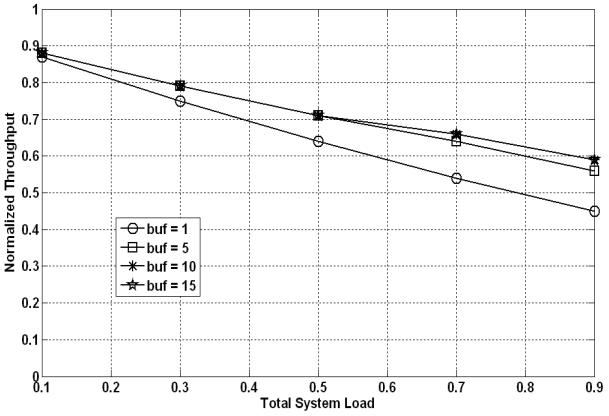


Figure 9. Normalized throughput of BE traffic class for varying buffer size.

priority, pre-emption and time-out characteristics of traffic classes. Performance of the system is evaluated in terms of mean delay and normalized throughput. Our model is validated by using simulations. The model can be extended to include more than three traffic classes. The model can be generalized to incorporate multiple SSs.

7. References

[1] IEEE 802.16-2004, "IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE, October 2004.

[2] IEEE 802.16e-2005, "Amendment and Corrigendum to IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE, February 2006.

[3] L. Lei, C. Lin, J. Cai and X. Shen, "Performance Analysis of Wireless Opportunistic Schedulers Using Stochastic Petri Nets," *IEEE Transactions on Wireless Communications*, Vol. 8, No. 4, 2009, pp. 2076-2087.

[4] D. Lee and J. Baik, "QoS Protocol Verification Using Petri-Net for Seamless Mobility in a Ubiquitous Environment: A Case Study," *International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, Phuket, August 2008, pp. 617-622.

[5] P. J. Haas, "Stochastic Petri Nets for Modelling and Simulation," *Proceedings of the* 2004 *Winter Simulation Conference*, Washington, DC, 2004, pp. 101-112.

[6] Y. Ma, J. J. Han and K. S. Trivedi, "Composite Performance and Availability Analysis of Wireless Communication Networks," *IEEE Transactions on Vehicular Technology*, Vol. 50, No. 5, 2001, pp. 1216-1223.

[7] R. Jayaparvathy, S. Anand, S. Dharmaraja and S. Srikanth, "Performance Analysis of IEEE 802.11 DCF with Stochastic Reward Nets," *International Journal of Communication Systems*, Vol. 20, No. 3, 2007, pp. 273-296.

[8] M. Llorens and J. Oliver, "Structural and Dynamic Changes in Concurrent Systems: Reconfigurable Petri Nets," *IEEE Transactions on Computers*, Vol. 53, No. 9, 2004, pp. 1147-1158.

[9] C. Cicconetti, A. Erta, L. Lenzini and E. Mingozzi, "Performance Evaluation of the IEEE 802.16 MAC for QoS Support," *IEEE Transactions on Mobile Computing*, Vol. 6, No. 1, 2007, pp. 26-38.

[10] Q. Ni, A. Vinel, Y. Xiao, A. Turlikov and T. Jiang, "Investigation of Bandwidth Request Mechanisms under Point-to-Multipoint Mode of WiMAX Networks," *IEEE Communications Magazine*, Vol. 45, No. 5, 2007, pp. 132-138.

[11] Y. P. Fallah, F. Agharebparast, M. Minhas, H. M. Alnuweiri and V. C. M. Leung, "Analytical Modelling of Contention-Based Bandwidth Request Mechanism in IEEE 802.16 Wireless Networks," *IEEE Transactions on Vehicular Technology*, Vol. 57, No. 5, 2008, pp. 3094-3107.

[12] M. Sarkar and H. Sachdeva, "A QoS Aware Packet Scheduling Scheme for WiMAX," *Proceedings of IAENG Conference on World Congress on Engineering and Computer Science, Berkeley*, California, USA, October 2009.

[13] Y.-C. Lai and Y.-H. Chen, "Designing and Implementing an IEEE 802.16 Network Simulator for Performance Evaluation of Bandwidth Allocation Algorithms," *Proceedings of the* 11th *IEEE International Conference on High Performance Computing and Communications*, Seoul, 2009, pp. 432-437.

[14] A. Bestetti, G. Giambene and S. Hadzic, "Fair Traffic Scheduling for WiMAX Systems," 6*th International Symposium on Wireless Communication Systems*, Tuscany, September 7-10, 2009, pp. 254-258.

[15] S. Z. Tao and A. Gani, "Intelligent Uplink Bandwidth Allocation Based on PMP Mode for WiMAX," *Proceedings of the* 2009 *International Conference on Computer Technology and Development*, Malaysia, 2009, pp. 86-90.

[16] C. So-In, R. Jain and A.-K. Tamimi, "Scheduling in IEEE 802.16e Mobile WiMAX Networks: Key Issues and a Survey," *IEEE Journal on Selected Areas in Communications*, Vol. 27, No. 2, 2009, pp. 156-171.

[17] R. Jain, C. So-In and A.-K. Tamimi, "System-Level Modeling of IEEE 802.16E Mobile Wimax Networks: Key Issues," *IEEE Wireless Communications*, Vol. 15, No. 5, 2008, pp. 73-79.

[18] R. A. Sahner, K. S. Trivedi and A. Puliafito, "Performance and Reliability Analysis of Computer Systems: An Example-Based Approach Using the SHARPE Software Package," Kluwer Academic Publishers, Dordrecht, 1996.

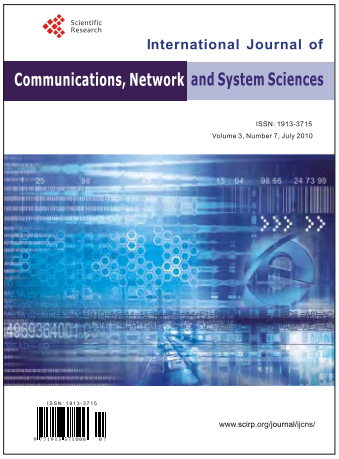
[19] R. German, "Markov Regenerative Stochastic Petri Nets with General Execution Policies: Supplementary Variable Analysis, and a Prototype Tool," *Proceedings of the* 10*th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation*, Palma de Mallorca, Spain, September 1998, pp. 255-266.

[20] R. German, C. Kelling, A. Zimmerman and G. Homel, "TimeNET: A Toolkit for Evaluating Non-Markovian Stochastic Petrinets," *Performance Evaluation*, Vol. 24, No. 1-2, 1995, pp. 69-87.

[21] L. Kleinrock, "Queuing Systems: Volume I, Theory," Kluwer Academic Press, Dordrecht, 1995.

[22] R. Jayaparvathy, S. Dharmaraja and S. Srikanth, "Stochastic Petri Nets in Performance Evaluation of IEEE 802.11 WLANs," *Sixth International Conference of the Association of the Asia Pacific Operational Research Societies*, New Delhi, India, December 2003, pp. 142-150.

Call for Papers



International Journal of

Communications, Network and System Sciences (IJCNS)

ISSN 1913-3715 (Print) ISSN 1913-3723 (Online)

http://www.scirp.org/journal/ijcns/

IJCNS is an international refereed journal dedicated to the latest advancement of communications and network technologies. The goal of this journal is to keep a record of the state-of-the-art research and promote the research work in these fast moving areas.

Editors-in-Chief

Subject Coverage

This journal invites original research and review papers that address the following issues in wireless communications and networks. Topics of interest include, but are not limited to:

MIMO and OFDM technologies	Sensor networks
UWB technologies	Ad Hoc and mesh networks
Wave propagation and antenna design	Network protocol, QoS and congestion control
Signal processing and channel modeling	Efficient MAC and resource management protocols
Coding, detection and modulation	Simulation and optimization tools
3G and 4G technologies	Network security

We are also interested in:

- Short reports—Discussion corner of the journal:
  2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data.
- Book reviews—Comments and critiques.

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. For more details about the submissions, please access the website.

Website and E-Mail

http://www.scirp.org/journal/ijcns ijcns@scirp.org

Int. J. Communications, Network and System Sciences, **2010, 3, 579-637**
Published Online July 2010 in SciRes (http://www.SciRP.org/journal/ijcns/).

TABLE OF CONTENTS

Volume 3 Number 7 **July 2010**

