ISSN: 1949-2421 Volume 2, Number 2, May 2010

Scientific Research

Communications and Network

ISSN: 1949-2421

www.scirp.org/journal/cn

Journal Editorial Board

ISSN Print: 1949-2421 ISSN Online: 1947-3826

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

Communications and Network, **2010, 2, 87-144**
Published Online May 2010 in SciRes (http://www.SciRP.org/journal/cn)

TABLE OF CONTENTS

Volume 2 Number 2 May 2010

Communications and Network (CN)

Journal Information

SUBSCRIPTIONS

The *Communications and Network* (Online at Scientific Research Publishing, www.SciRP.org) is published quarterly 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: cn@scirp.org

Communications and Network, **2010, 2, 87-92**
doi:10.4236/cn.2010.22014 Published Online May 2010 (http://www.SciRP.org/journal/cn)

Performance Analysis of a Threshold-Based Relay Selection Algorithm in Wireless Networks

Hao Niu, Taiyi Zhang, Li Sun
Department of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, China
E-mail: *nhfly*86@gmail.com
Received December 21, 2009; *revised February* 1, 2010; *accepted March* 1, 2010

Abstract

Relay selection is an effective method to realize the cooperative diversity gain in wireless networks. In this paper, we study a threshold-based single relay selection algorithm. A reasonable threshold value is set at each relay node, and the first relay with the instantaneous channel gain larger than the threshold will be selected to cooperate with the source. The exact and closed form expression for its outage probability is derived over independent, non-identically distributed (i.n.i.d) Rayleigh channels. The complexity of the algorithm is also analyzed in detail. Simulation results are presented to verify our theoretical analysis.

Keywords: Relay Selection, Outage Probability, Threshold, Wireless Networks

1. Introduction

User cooperation is a promising technique to improve the performance of wireless networks [1-3]. One possible approach to realize cooperative diversity is to use distributed space-time coding (DSTC) among participating nodes [4]. However, the design of such code is difficult in practice and is still an open area of research.

Aiming at these problems, Bletsas *et al*. introduced a novel scheme called opportunistic relaying in [5], where only the "best" relay among all available candidates is selected to cooperate with the source. Analysis in [5,6] proved that this method can provide the same diversity-and-multiplexing tradeoff (DMT) as DSTC. However, as pointed in [5], a distributed relay selection may lead to packet collision which is a cause to fail the procedure, and the centralized approach requires a large number of channel estimations, which is energy-inefficient and not practical for resource-constrained networks.

In order to overcome the above-mentioned drawbacks, Hwang and Ko proposed a sub-optimal relay selection algorithm in [7], where a pre-determined threshold is set both at the relay and the destination, and the first relay with equivalent channel gain larger than the threshold is selected. This algorithm can significantly reduce the implementation complexity and power consumptions compared with the conventional opportunistic relaying algorithm. However, reference [7] didn't present the exact outage probability formula of the algorithm.

In this paper, we follow the basic ideas of [7] while some detailed analyses are presented. The main contribution of our work is that we derive the exact closed form expression for the outage probability of the algorithm over independent, non-identically distributed (i.n.i.d) Rayleigh channels. We will show that this algorithm can achieve the same diversity order as opportunistic relaying, while its complexity in terms of the amount of channel estimations can be reduced obviously.

2. System Model and Basic Assumptions

We consider a half-duplex dual-hop communication system as shown in **Figure 1**, where there are a source (S), a destination (D) and K relay nodes (R_k , $k \in \{1,...,K\}$).

There are two types of relay selection, reactive and proactive [6], and we only focus on the latter. That is to say, the "best" relay is chosen prior to the source transmission among a collection of K possible candidates. After this has been completed, a two-phase communication starts. During the first phase, the source transmits and the "best" relay listens, while during the second phase, the "best" relay forwards a version of the received

$$\Pr\{outage\} = \begin{cases} \prod_{i=1}^{K}(1-e^{-(\lambda_{si}+\lambda_{id})\gamma}), & th > \gamma \\ \prod_{i=1}^{K}(1-e^{-(\lambda_{si}+\lambda_{id})th}) + \sum_{i=1}^{K}\sum_{n=0}^{i-1}\sum_{\substack{1\le k_1<\ldots<k_n\le i-1 \\ 1\le k_{n+1}<\ldots<k_{i-1}\le i-1 \\ k_1\ne k_2\ne\ldots\ne k_{i-1}}} (-1)^{i-1-n} e^{-\sum_{j=n+1}^{i-1}(\lambda_{sk_j}+\lambda_{k_jd})th} (e^{-(\lambda_{si}+\lambda_{id})th} - e^{-(\lambda_{si}+\lambda_{id})\gamma}), & th \le \gamma \end{cases} \quad (1)$$

signal to the destination using decode-and-forward (DF) protocol [1].

For each link, the channel is assumed to be block flat fading (quasi-static), which remains constant during one frame and varies independently from frame to frame (In our system, one frame is comprised of two phases). The channel gain, *i.e.*, the squared channel strength, between the source and the destination, the source and the *k*th relay, and the *k*th relay and the destination are represented by g_{sd}, g_{sk} and g_{kd}, respectively ($k \in \{1,...,K\}$). The channel coefficients are modeled as zero-mean, independent, circularly-symmetric complex Gaussian random variables, so g_{sd}, g_{sk} and g_{kd} obey exponential distributions and notations λ_{sd}, λ_{sk} and λ_{kd} are introduced to denote their distribution parameters.

In the following analysis, we consider two communication scenarios. For Scenario I, the direct link between the source and the destination doesn't exist, *i.e.*, the source communicates with the destination only via relaying. For Scenario II, the source can communicate with the destination directly, and the destination combines the two received signals coming from the source and the best relay with a maximal ratio combiner (MRC).

3. Review of the Threshold-Based Relay Selection Algorithm

The relay selection procedure is activated before every source transmission, the operation of the algorithm was described in [7] and we summarize it as follows.

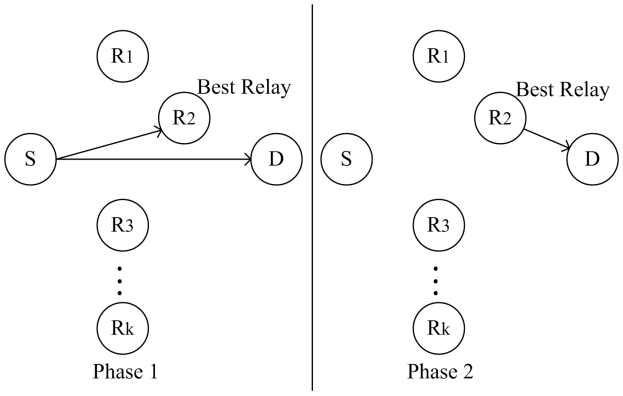


Figure 1. The block diagram of the half-duplex dual-hop system.

Step 1: Initialize $k = 0$.

Step 2: Set $k \leftarrow k + 1$, if $k = K + 1$, go to Step 4.

Step 3: If the instantaneous channel gain of the *k*th relay is larger than the predetermined threshold value, *i.e.* $\min\{g_{sk}, g_{kd}\} > th$, relay k is selected as the best relay and the algorithm terminates: $k^* = k$, otherwise go to Step 2.

Step 4: Evaluate $k^* = \arg\max_k \{\min\{g_{sk}, g_{kd}\}\}$ and select node k^* as the best relay.

In the algorithm above, *th* is the predetermined threshold value and k^* denotes the "best" relay. Note that we use the minimum of the channel gains between the links $S \to k$ and $k \to D$ to describe the channel quality at relay *k*, which is consistent with the statements in [6,7].

4. Research of the Threshold-Based Relay Selection Algorithm for Scenario I

4.1. Outage Probability of the Algorithm

Theorem 1: Denoting the required spectral efficiency by R and the average transmit signal-to-noise by SNR, the outage probability of the threshold-based algorithm can be expressed by (1), where $\gamma = (2^{2R} - 1)/\text{SNR}$.

Proof: Communication through the "best" relay fails due to outage when either of the two hops (from the source to the best relay and from the best relay to destination) fails. Thus the outage probability can be expressed as [8]:

$$\Pr(outage) = \Pr\{\min\{g_{sk^*}, g_{k^*d}\} < \gamma\} \quad (2)$$

For each *i*, letting g_i represent the minimum of g_{si} and g_{id}, *i.e.*, $g_i = \min\{g_{si}, g_{id}\}$, we can see that the probability in (2) depends on the statistical property of g_{k^*}.

Recall that the algorithm selects the first relay with the instantaneous channel gain larger than the threshold as the best relay, and if no such relay exists, the best relay is selected as the opportunistic relaying method does, so we have to consider two cases to calculate the probability.

Case 1: The channel gains of all relays are below the threshold.

By using the fact that the minimum of two independent exponential random variables with parameter λ_{si} and λ_{id} is again an exponential random variable with parameter $\lambda_{si}+\lambda_{id}$, the probability of case 1 can be calculated as

$$\begin{aligned}\Pr\{case\ 1\} &= \Pr\{\max_{i\in\{1,2,\ldots,K\}}\{\min\{g_{si},\ g_{id}\}\} < th\} \\ &= \prod_{i=1}^{K}(1-e^{-(\lambda_{si}+\lambda_{id})th})\end{aligned} \tag{3}$$

For case 1, k^* is the relay node with the largest channel gain and the outage probability conditioned on case 1 is given by

$$\begin{aligned}\Pr\{Outage \mid case\ 1\} &= \Pr\{g_{k^*} < \gamma \mid case1\} \\ &= \begin{cases} 1, & \gamma \ge th \\ \dfrac{\prod_{i=1}^{K}(1-e^{-(\lambda_{si}+\lambda_{id})\gamma})}{\prod_{i=1}^{K}(1-e^{-(\lambda_{si}+\lambda_{id})th})}, & \gamma < th \end{cases}\end{aligned} \tag{4}$$

Case 2: At least one relay has the channel gain lager than *th*.

This case is the complementary event of case 1 and its probability is given in (5).

$$\Pr\{case\ 2\} = 1-\prod_{i=1}^{K}(1-e^{-(\lambda_{si}+\lambda_{id})th}) \tag{5}$$

For case 2, k^* is the first relay with channel gain lager than the threshold and the corresponding outage probability formula is expressed as

$$\begin{aligned}\Pr\{Outage \mid case\ 2\} &= \frac{\Pr\{g_{k^*} < \gamma\}}{\Pr\{case\ 2\}} = \frac{\sum_{i=1}^{K}\Pr\{k^* = i\ \&\ g_i < \gamma\}}{\Pr\{case\ 2\}} \\ &= \frac{\sum_{i=1}^{K}\Pr\{g_1 < th,\ldots,g_{i-1} < th,\ g_i > th, g_i < \gamma\}}{\Pr\{case\ 2\}}\end{aligned} \tag{6}$$

So the outage probability conditioned on case 2 is given in (7).

$$\Pr\{Outage \mid case\ 2\} = \begin{cases} 0, & \gamma < th \\ \dfrac{\sum_{i=1}^{K}(\prod_{j=1}^{i-1}(1-e^{-(\lambda_{sj}+\lambda_{jd})th}))\int_{th}^{\gamma}(\lambda_{si}+\lambda_{id})e^{-(\lambda_{si}+\lambda_{id})y}dy}{1-\prod_{i=1}^{K}(1-e^{-(\lambda_{si}+\lambda_{id})th})}, & \gamma \ge th \end{cases} \tag{7}$$

Finally, the outage probability of the algorithm can be evaluated, using total probability formula, as

$$\Pr\{Outage\} = \sum_{i=1}^{2}\Pr\{case\ i\}\Pr\{Outage \mid case\ i\} \tag{8}$$

Substituting (3), (4), (5), (7) into (8) and with the help of the following multinomial expansion [9], (1) can be obtained. However, we omit the details of this procedure due to the limitation of space.

$$\prod_{i=1}^{K}(x_i+y_i) = \sum_{n=0}^{K}\sum_{\substack{1\le k_1<\ldots<k_n\le K \\ 1\le k_{n+1}<\ldots<k_K\le K \\ k_1\ne k_2\ne\ldots\ne k_K}}\prod_{i=1}^{n}x_{k_i}\times\prod_{i=n+1}^{K}y_{k_i} \tag{9}$$

4.2. Discussion on the Results

The result in (1) may be attractive because it indicates that the threshold-based relay selection algorithm can achieve the same outage probability as opportunistic relaying in [6] when a suitable threshold value is determined, despite its simplicity.

Intuitively, to select the threshold value equal to γ, called *outage threshold*, is a simple but reasonable choice. Note that γ depends only on the required spectral efficiency and average transmit signal-to-noise ratio, so *th* need not to be updated frequently during the communication procedure, which reduces the complexity of the algorithm obviously.

Computer simulations are carried out to validate the analytical expressions. In **Figure 2**, we compare the outage probability of the threshold-based scheme to that of opportunistic relaying [5] over the channel described in Section 2. To see the differences of the two algorithms clearly, we only give the results for low SNR regimes. In this simulation, we assume the nodes of the whole network are distributed in a 1 × 1 rectangular coordinate system. The source node is located at (0,0) and the destination node at (1,1), K = 4 relay nodes are generated randomly and the mean of the fading coefficient between node i and j is determined by the distance d_{ij} between them, *i.e.* $1/\lambda_{ij} = d_{ij}^{-\eta}$, where the path loss exponent is set to be $\eta = 2$. We assume R = 1 and set *th* = 1. As can be seen from the figure, there is an excellent match between the curves of analytical results and simulation ones. The results show that when SNR > 5 dB, *i.e.*, *th* > γ, the outage probability of threshold-based relay selection algorithm is exactly the same as that of opportunistic relaying, whereas

there is a performance loss if SNR < 5 dB, *i.e.*, $th < \gamma$. These observations are consistent with the results in (1).

Figure 2 also shows the outage probability for random selection method. It is noticeable that random selection incurs a substantial penalty loss. This is due to the fact that selecting the "best" relay randomly removes potential selection diversity benefits.

Now we consider the complexity of the threshold-based algorithm in terms of the amount of channel estimations and compare it with that of opportunistic relaying method in [5,6]. The opportunistic relaying needs the 2K number of channel estimations, while the average number of channel estimations of threshold- based method can be evaluated by (10)[1].

$$\bar{N}_e = 2K\left(\prod_{i=1}^{K}(1-e^{-(\lambda_{si}+\lambda_{id})th})\right) + 2\sum_{i=1}^{K}\left(\prod_{j=1}^{i-1}(1-e^{-(\lambda_{sj}+\lambda_{jd})th})\right)e^{-(\lambda_{si}+\lambda_{id})th} i \quad (10)$$

It is not hard to achieve the closed form expression of (10) but this may be not helpful to analyze. Instead, we use numerical calculations to give some intuitive results in **Table 1**, where we set $\lambda_{sk} = 1$, $\lambda_{kd} = 0.7$ ($k \in \{1,2,...,K\}$) for convenience.

Table 1. Complexity comparison.

System Parameter	The Amount of Channel Estimations		
	Opportunistic relaying	*Threshold-based selection* (*average*)	*Reduced by*
$K = 2$, $th = 1$	4	3.6346	9.13%
$K = 4$, $th = 0.7$	8	5.0334	37.08%
$K = 4$, $th = 1$	8	6.0626	24.22%
$K = 4$, $th = 1.5$	8	7.1108	11.11%
$K = 5$, $th = 1$	10	6.9551	30.45%

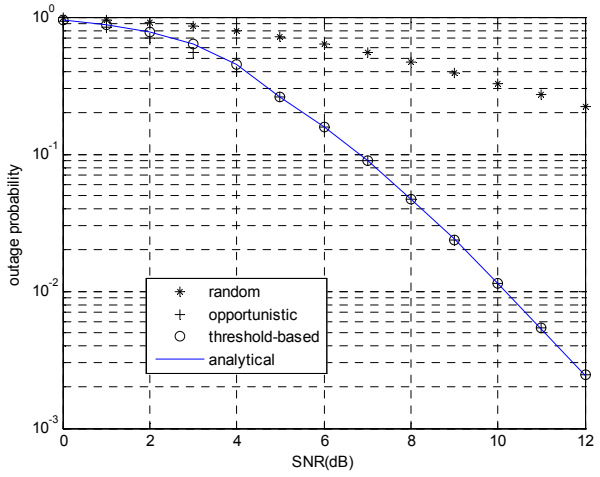


Figure 2. The outage probabilities of random selection, opportunistic relaying, and threshold-based algorithm without the direct link.

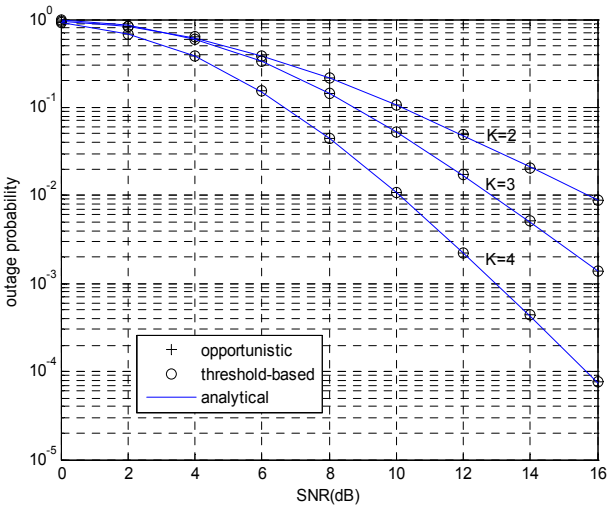


Figure 3. The outage probabilities of opportunistic relaying and threshold-based algorithm without the direct link.

From the results above we can conclude that the threshold-based relay selection algorithm is especially attractive in large networks. Besides that, the lower the threshold value is, the less the number of channel estimations is. However, reducing the threshold will incur a performance loss. Therefore, considering both the performance and complexity, we select the optimal threshold value *th* equal to γ. In **Figure 3**, we compare the outage probability of the threshold-based scheme to that of the opportunistic relaying [5] with $th = \gamma$.

5. Research of the Threshold-Based Relay Selection Algorithm for Scenario II

In this section, we investigate the performance of the algorithm for Scenario II. The channel gain between the source and the destination is denoted by g. For simplicity, we set $th = \gamma$.

Similar to Section 4, we consider two cases and calculate the outage probability as follows.

Case 1: The channel gains of all relays are below the threshold, *i.e.*, $\max_{i\in\{1,2,...,K\}}\{\min\{g_{si}, g_{id}\}\} < th$. The probability of this case can be represented by

$$\begin{aligned}\Pr\{case\ 1\} &= \Pr\{\max_{i\in\{1,2,...,K\}}\{\min\{g_{si}, g_{id}\}\} < th\} \\ &= \prod_{i=1}^{K}(1-e^{-(\lambda_{si}+\lambda_{id})th})\end{aligned} \quad (11)$$

In this case, k^* is the relay node with the largest channel gain and the outage event happens if either of the three situations takes place,

[1]In fact, the actual amount is less than this, however, we use (10) for simplicity.

$$\Pr\{Outage\} = \sum_{i=1}^{2} \Pr\{case\ i\} \Pr\{Outage \mid case\ i\} = \sum_{i=1}^{K} \Pr_{1,i}\{Outage\}$$

$$= \sum_{i=1}^{K} \sum_{n=0}^{K-1} \sum_{\substack{1 \le k_1 < \ldots < k_n \le K \\ 1 \le k_{n+1} < \ldots < k_{K-1} \le K \\ k_1 \ne k_2 \ne \ldots \ne k_{K-1} \ne i}} (-1)^{K-1-n} [-\frac{(\lambda_{si} + \lambda_{id})(1 - e^{-\lambda_{sd} th})}{\sum_{j=n+1}^{K-1} (\lambda_{sk_j} + \lambda_{k_j d}) + (\lambda_{si} + \lambda_{id})} (e^{-(\sum_{j=n+1}^{K-1} (\lambda_{sk_j} + \lambda_{k_j d}) + (\lambda_{si} + \lambda_{id}))th} - 1)$$

$$- \frac{\lambda_{id} e^{-(\lambda_{si} + \lambda_{sd})th}}{\sum_{j=n+1}^{K-1} (\lambda_{sk_j} + \lambda_{k_j d}) + \lambda_{id}} (e^{-(\sum_{j=n+1}^{K-1} (\lambda_{sk_j} + \lambda_{k_j d}) + \lambda_{id})th} - 1) \quad (16)$$

$$+ \frac{\lambda_{id} e^{-(\lambda_{si} + \lambda_{sd})th}}{\sum_{j=n+1}^{K-1} (\lambda_{sk_j} + \lambda_{k_j d}) + (\lambda_{id} - \lambda_{sd})} (e^{-(\sum_{j=n+1}^{K-1} (\lambda_{sk_j} + \lambda_{k_j d}) + (\lambda_{id} - \lambda_{sd}))th} - 1)]$$

$$(a)\ g_{sk^*} < th \,\&\, g_{sk^*} \le g_{k^* d} \,\&\, g < th$$
$$(b)\ th > g_{sk^*} > g_{k^* d} \,\&\, g < th \quad (12)$$
$$(c)\ g_{sk^*} \ge th > g_{k^* d} \,\&\, (g_{k^* d} + g) < th$$

After some calculations, the outage probability is given by (13).

$$\Pr_{1,k^*}\{Outage\}$$
$$= \int_0^{th} (\prod_{\substack{j=1 \\ j \ne k^*}}^{K} (1 - e^{-(\lambda_{sj} + \lambda_{jd}) \cdot x})) [\lambda_{sk^*} e^{-\lambda_{sk^*} x} e^{-\lambda_{k^* d} x} (1 - e^{-\lambda_{sd} th}) \quad (13)$$
$$+ \lambda_{k^* d} e^{-\lambda_{k^* d} x} (e^{-\lambda_{sk^*} x} - e^{-\lambda_{sk^*} th})(1 - e^{-\lambda_{sd} th})$$
$$+ \lambda_{k^* d} e^{-\lambda_{k^* d} x} e^{-\lambda_{sk^*} th} (1 - e^{-\lambda_{sd}(th - x)})] dx$$

Using total probability formula, the outage probability conditioned on case 1 can be expressed as

$$\Pr\{Outage \mid case\ 1\} = \frac{\sum_{i=1}^{K} \Pr_{1,i}\{Outage\}}{\Pr\{case\ 1\}} \quad (14)$$

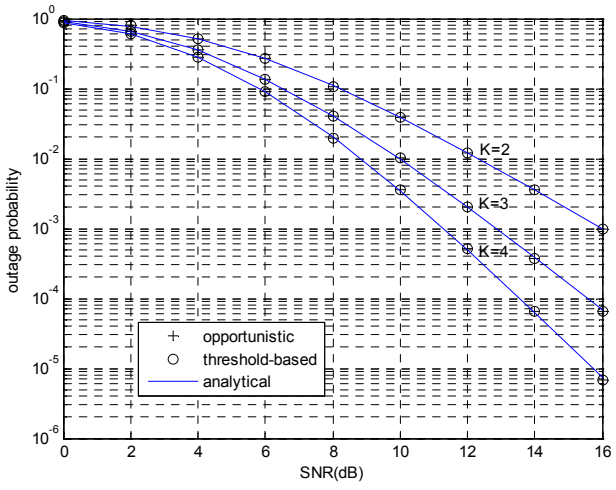


Figure 4. The outage probabilities of opportunistic relaying and threshold-based algorithm with direct link.

Case 2: At least one relay has the channel gain lager than *th*. In this case, the outage probability is obviously zero. That is to say,

$$\Pr\{Outage \mid case\ 2\} = 0 \quad (15)$$

Finally, the outage probability of the algorithm is expressed by (16).

In **Figure 4**, we compare the outage probability of the threshold-based scheme to that of opportunistic relaying [5] with $th = \gamma$ for Scenario II, the simulation conditions are the same as that for **Figure 2**. Simulation results also validate our analytical conclusions.

6. Conclusions

In this paper, we have studied a threshold-based single relay selection algorithm for wireless networks. The analytical closed form expressions of its outage probability are derived. Through computer simulations we analyze the performances of the algorithm. The results show that this method can achieve the same outage probability as that of opportunistic relaying with a suitable threshold, while its complexity is obviously reduced.

7. References

[1] J. N. Laneman, D. N. C. Tse and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Transactions on Information Theory*, Vol. 50, No. 12, December 2004, pp. 3062-3080.

[2] K. J. R. Liu, A. K. Sadek, W. Su and A. Kwasinski, "Cooperative Communications and Networking," Cambridge University Press, Cambridge, 2008.

[3] S. Chen, W. Wang and X. Zhang, "Performance Analysis of Multiuser Diversity in Cooperative Multi-relay Networks under Rayleigh-Fading Channels," *IEEE Transactions on Wireless Communications*, Vol. 8, No. 7, July 2009, pp. 3415-3419.

[4] J. N. Laneman and G. W. Wornell, "Distributed Space-Time Coded Protocols for Exploiting Cooperative Diver-

sity in Wireless Networks," *IEEE Transactions on Information Theory*, Vol. 49, No. 10, October 2003, pp. 2415-2525.

[5] A. Bletsas, A. Khisti, D. P. Reed and A. Lippman, "A Simple Cooperative Diversity Method Based on Network Path Selection," *IEEE Journal on Selected Areas in Communications*, Vol. 24, No. 3, March 2006, pp. 659-672.

[6] A. Bletsas, H. Shin and M. Z. Win, "Cooperative Communications with Outage-Optimal Opportunistic Relaying," *IEEE Transactions on Wireless Communi- cations*, Vol. 6, No. 9, September 2007, pp. 3450-3460.

[7] K.-S. Hwang and Y.-C. Ko, "An Efficient Relay Selection Algorithm for Cooperative Networks," 2007 *IEEE* 66*th Vehicular Technology Conference*, Baltimore, 30 September-3 October 2007, pp. 81-85.

[8] A. Bletsas, "Intelligent Antenna Sharing in Cooperative Diversity Wireless Networks," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, 2005.

[9] F. Xu, C. M. Lau, Q. F. Zhou and D. W. Yue, "Outage Performance of Cooperative Communication Systems Using Opportunistic Relaying and Selection Combining Receiver," *IEEE Signal Processing Letters*, Vol. 16, No. 4, February 2009, pp. 237-240.

Communications and Network, **2010, 2, 93-96**
doi:10.4236/cn.2010.22015 Published Online May 2010 (http://www.SciRP.org/journal/cn)

Method of Carrier Acquisition and Track for HAPS

Mingxiang Guan[1], Fang Yuan[1], Xueyuan Wan[1], Weizhi Zhong[2]
[1]*Department of Electronic Communication Technology*, *Shenzhen Institute of Information Technology*, *Shenzhen*, *China*
[2]*School of Electronic and Information Technology*, *Harbin Institute of Technology*, *Harbin*, *China*
E-mail: *gmx*2020@126.*com*
Received August 30, 2009; *revised April* 14, 2009; *accepted April* 20, 2010

Abstract

This paper emphasizes on the characteristics and schemes of carrier acquisition and track in high dynamic and high information-rate situation. Carrier acquisition model is analyzed theoretically and the design principle of carrier acquisition is deduced and described clearly. An algorithm for carrier acquisition in high dynamic and high information-rate situation is provided. This paper also proves the validity of the algorithm and design scheme in high dynamic and high information-rate situation.

Keywords: HAPS, High Dynamic and High Information-Rate, Acquisition, Track

1. Introduction

Now the HAPS (high altitude platform Station) system has been concentrated on research because of important values in military and application. The HAPS communication system is comprised by the stable HAP as wavelet relay station, controlling devices on the earth, accessing devices and various kinds of wireless users. Recently many research centers including US military are researching the HAPS communication because of many merits such as low expense, fast deployment, little ground devices, convenient retrieve and so on. An information system formed by HAP in stratosphere will be a new generation-system for the wireless communications and the specialized communication system combines the advantages of both terrestrial and satellite communication systems and avoids, to different extents, their disadvantages. Then many countries had spent large manpower and resources to research the HAPS communication system profoundly [1-3].

In the literature [4,5] frequency detector loop was added to the carrier track loop, therefore, the traditional closed-loop technology was substitute with open-loop carrier error estimation technology. A combined method of data phase information and carrier frequency estimation was provided in literature [6] by means of correction of the local numerical control oscillator (NCO) through the frequency and phase error estimation. In the literature [7,8] the author analyzed the frequency and phase error by means of the combined method of self-adaptive filter and flatness technology. In order to implement the carrier acquisition and track in high dynamic and high information-rate situation, the technologies referred to the literatures above can't reach the requirement of the communication system obviously. Therefore this paper provides a new and suitable carrier acquisition and track technology for HAPS communication in high dynamic and high information-rate situation and proved its validity.

2. Principle of Acquisition and Track

To comprehend clearly design method and algorithm this paper in high dynamic and high information-rate situation, the design principle of carrier acquisition and track loop is introduced firstly. The base principle is the non-linear disposal of the suppressed carrier QPSK signal, then the carrier is recovered by phase lock loop and referenced carrier is obtained for coherent demodulation.

$$x(t) = V_i \sin[\omega_0 t + \theta_i + \theta] \tag{1}$$

Phase shift $\theta_i = 2\pi i / N$ may be some discrete value when $lT_s < t < (l+1)T_s$ (i = 1,2,…,N); θ is the phase that need local referenced signal to track. In order to decrease code error resulting from phase offset during the coherent demodulation, therefore phase lock loop is used to track the drift variation when θ is changed with the time. What phase estimated value will result is a key problem. The maximum $\hat{\theta}$ of likelihood function $p(x/\theta)$ is discussed to be regard as the estimated value of θ.

$$\frac{\partial}{\partial\theta}P(x/\theta)|_{\theta=\hat{\theta}}=0 \tag{2}$$

In order to resolve $p(x/\theta)$, suppose the input is Equation (3).

$$x(t)=S(t,\theta)+n(t) \tag{3}$$

$n(t)$ is narrowband gauss white noise. Supposed that T_s is the observed period and θ is one of the N fixed values, therefore θ_i is fixed value. In the period T_s m sampled values corresponding with $x_1, x_2, ..., x_m$ random variable compose m dimensions random vector X. when θ is given, the observed random X is a set which is composed by m independent random variables. The union probability density function (pdf) is:

$$\begin{aligned} p(X/\theta)&=p(x_1,x_2,...,x_m/\theta)\\ &=\frac{1}{(2\pi\sigma^2)^{m/2}}\exp(-\sum_{k=1}^{m}\frac{n_k{}^2}{2\sigma^2})\\ &=\frac{1}{(2\pi\sigma^2)^{m/2}}\exp(-\sum_{k=1}^{m}\frac{(x_k-S_k)^2}{2\sigma^2}) \end{aligned} \tag{4}$$

Make sample interval tend to zero. So $m\to\infty$, Equation (4) is transferred to Equation (5) in the following.

$$\begin{aligned} p(X/\theta)&=A_1\exp\{-\frac{1}{N_0}\int_0^{T_s}[x(t)-S(t,\theta)]^2dt\}\\ &=A_1\exp\{-\frac{1}{N_0}\int_0^{T_s}x^2(t)dt+\\ &\frac{2}{N_0}\int_0^{T_s}x(t)S(t,\theta)dt-\\ &\frac{1}{N_0}\int_0^{T_s}S^2(t,\theta)dt\} \end{aligned} \tag{5}$$

The first integral in Equation (5) has nothing to θ; the third integral in Equation (5) is signal energy which will not vary with the parameter θ. Therefore the limit of the likelihood function depends on the value of the second integral in Equation (5). $p(x/\theta)$ can rewrite to Equation (6) in the following.

$$p(X/\theta)=A_2\exp\{\int_0^{T_s}x(t)S(t,\theta)dt\} \tag{6}$$

A_1, A_2 is constant respectively.

The value $\theta_i=2\pi i/N$ is carried in the period T_s at the same probability and the likelihood function of the N dimension signal can be written as the following:

$$p(X/\theta)=F\sum_{i=1}^{N}\exp\{\frac{V_i^2}{N_0}[\xi_{sl}(\theta)\cos\theta_i+\xi_{cl}(\theta)\sin\theta_i]\} \tag{7}$$

Here, $\xi_{sl}(\theta)=\frac{2}{V_iT_s}\int_{lT_s}^{(l+1)T_s}x(t)\sin(\omega_0t+\theta)dt$, $\xi_{sl}(\theta)=\frac{2}{V_iT_s}\int_{lT_s}^{(l+1)T_s}x(t)\sin(\omega_0t+\theta)dt$, F is a constant.

The solution of the equation $\frac{\partial}{\partial\theta}\ln p(x/\theta)|_{\theta=\hat{\theta}}=0$ is $\theta=\hat{\theta}$ which can be obtained from the following equation:

$$\frac{d}{d\theta}\sum_{i=1}^{N}[\xi_{sl}(\theta)\cos\theta_i+\xi_{cl}(\theta)\sin\theta_i]^N \tag{8}$$

The solution $\hat{\theta}$ obtained from the Equation (8) is a non-offset estimated value of carrier phase θ. Carrier component information is switched to carrier component by means of non-linear N power process. When $|\theta-\hat{\theta}|\le\pi/N$, the Mathematical expectation will be larger than zero if $\theta>\hat{\theta}$; otherwise the Mathematical expectation will be smaller than zero if $\theta<\hat{\theta}$. We can set $N=4$ in the Equation (8) and can get the construction form of QPSK easily. The implementation of carrier link acquisition and track is based on this principle in high dynamic and high information-rate situation for HAPS communication systems.

3. Method of Acquisition and Track

The sampled signals are divided into two path signals which will be processed by Quadrature and frequency down-conversion. The output signal of numerical frequency down-conversion is regarded as input signal of differential demodulation and carrier track loop. DQPSK algorithmic diagram of acquisition and track is shown in **Figure 1.**

The main process of the algorithm is as the following:

The input signal I_Σ and Q_Σ after A/D sample can be written in digital form: $I_\Sigma(k)=A_\Sigma(k)\cos\varphi(k)$, $Q_\Sigma(k)=A_\Sigma(k)\sin\varphi(k)$.

Set $S_{in}(k)=I_\Sigma(k)+jQ_\Sigma(K)=A_\Sigma(k)e^{-j\varphi(k)}$, therefore:

$$\begin{aligned} S_{out}(k)&=S_{in}(k)S_{in}^*(k-1)\\ &=A_\Sigma(k)A_\Sigma(k-1)e^{-j[\phi(k)-\phi(k-1)]} \end{aligned} \tag{9}$$

Definition: $Dok(k)=\mathrm{Re}[S_{out}(k)]$, $Cross(k)=\mathrm{Im}[S_{out}(k)]$, in the condition of QPSK modulation the equation $A_\Sigma(k)=A_\Sigma(k-1)=A_\Sigma$ will exist. Therefore:

$$\begin{aligned} Dok(k)&=A_\Sigma^2\cos[\varphi(k)-\varphi(k-1)]\\ Cross(k)&=-A_\Sigma^2\sin[\varphi(k)-\varphi(k-1)] \end{aligned} \tag{10}$$

set $\Delta\varphi_{\mathrm{mod}}=\varphi(k)-\varphi(k-1)$, whose phase value is $0,\pi/2,\pi,3\pi/2$. But only three logical levels (0, 1 and -1)

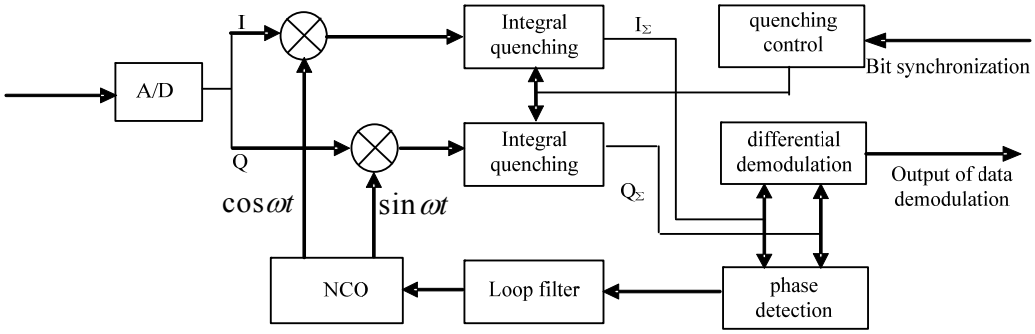


Figure 1. Algorithm of carrier acquisition deriving from DQPSK.

appear when detection is made in Equation (10). Therefore a fixed phase $\pi/4$ is considered here and the detection method is in the following:

$$Dok(k) = A_\Sigma^2 \cos[\varphi(k) - \varphi(k-1) + \pi/4],$$
$$Cross(k) = -A_\Sigma^2 \sin[\varphi(k) - \varphi(k-1) + \pi/4].$$

Combine $\Delta\phi_{\text{mod}} = \phi(k) - \phi(k-1) = 0, \pi/2, \pi, 3\pi/2$ with the two equations above.

Therefore,

$$\begin{aligned} Dok(k) &= \pm\frac{1}{\sqrt{2}} A_\Sigma^2 \\ Cross(k) &= \pm\frac{1}{\sqrt{2}} A_\Sigma^2 \end{aligned} \tag{11}$$

Parallel data output after the differential demodulation is corresponding with the four conditions above. In order to synchronize the output signal of NCO and intermediate frequency signal, the frequency of NCO is adjusted by phase bias of output after differential demodulation. The phase bias control signal is generated by phase detection module. The concrete method is in detail in the following:

Suppose the frequency bias between output of NCO and intermediate frequency signal is $\Delta\omega$ and initial phase bias is $\Delta\varphi$. Therefore the total phase bias is $\Delta\varphi_e = \Delta\omega t + \Delta\varphi$. Then:

$$\begin{aligned} Dok(k) &= A_\Sigma^2 \cos[\Delta\varphi_{\text{mod}} - \Delta\varphi_e + \pi/4] \\ Cross(k) &= -A_\Sigma^2 \cos[\Delta\varphi_{\text{mod}} - \Delta\varphi_e + \pi/4] \end{aligned} \tag{12}$$

$S_{AFC} = Sign\left[Dok(K)\right] Cross(k) - Sign\left[Cross(k)\right] Dok(k)$ is the bias control signal generated by phase detection.

Combine $\Delta\varphi_{\text{mod}} = \varphi(k) - \varphi(k-1) = 0, \pi/2, \pi, 3\pi/2$ with the equation above.

Therefore:

$$S_{AFC} = 2A_\Sigma^2 \sin\Delta\varphi_e \tag{13}$$

NCO frequency is controlled by loop filter and $\Delta\varphi_e$ will come to zero. Therefore the carrier acquisition and track is implemented in high dynamic and high information-rate situation.

4. Experiment and Simulation Analysis

In the design of the carrier acquisition and track loop with high dynamic and high information-rate, the carrier acquisition deriving from DQPSK has a good performance in our experiment. The performance of the algorithm proposed in this paper can't be known directly because of no comparison with other carrier acquisition algorithms. In [5], carrier phase offset, frequency offset and Doppler frequency shift acceleration were simulated in high dynamic situation. So we referred the simulated method in the literature to compare the performance between EKF (Extended Kalman Filter) method and the proposed method in the paper.

With reference to related simulated parameters in [5], the carrier frequency is 10.7 MHz, frequency offset is 200 Hz, and initial phase offset is 0.1°. The simulated results are shown in **Figure 2**. When the frequency is

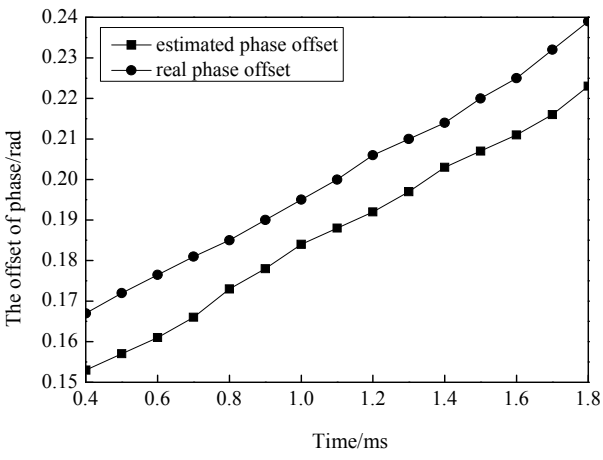


Figure 2. Phase offset estimation of doppler's frequency shift.

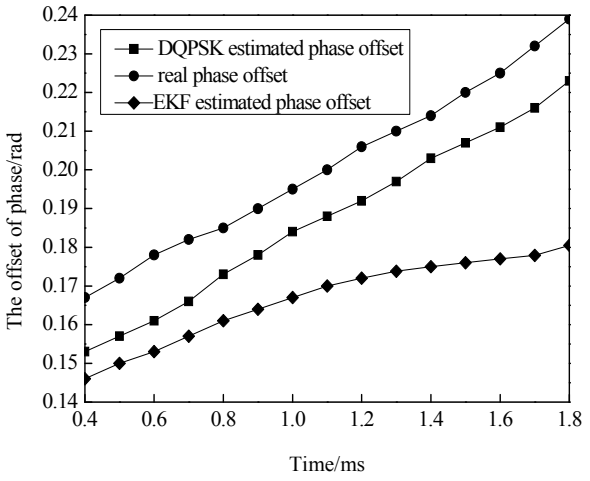


Figure 3. Precision comparison between DQPSK and EFK phase offset estimation.

changed from 200 Hz to 400 Hz, the real time simulated results are shown in **Figure 3**. From the **Figure 2** DQPSK method can estimate the phase valve accurately. From the **Figure 3** phase variation can be tracked less than 20 data time. Therefore the DQPSK method for carrier acquisition is better than EKF method and suitable for high dynamic and high information-rate carrier acquisition and track with a better precision.

5. Conclusions

As a branch of the wireless mobile communications with fast development, HAPS communication attracts more and more attention. This paper emphasizes on the characteristics and schemes of carrier acquisition and track in high dynamic and high information-rate situation. Carrier acquisition model is analyzed theoretically and the design principle of carrier acquisition is deduced and described clearly. An algorithm for carrier acquisition in high dynamic and high information-rate situation is provided. This paper also proves the validity of the algorithm and design scheme in high dynamic and high information-rate situation. By comparison with other methods of carrier acquisition and track, the method proposed this paper is suitable with low complexity of hardware and easy to realization in engineering.

6. Acknowledgements

This paper is supported by the 2nd doctoral innovation foundation of SZIIT (BC2009015). The author would like to thank Prof. Zhao Honglin to help to design high information-rate intermediate frequency de-spread acquisition and track loop, and Dr. Li Lu for his revision of the text, and the editor and the anonymous reviewers for their contributions that enriched the final paper.

7. References

[1] D. Grace, M. H. Capstick, M. Mohorcic and J. Horwath. "Integrating Users into the Wider Broadband Network via High Altitude Platforms," *IEEE Wireless Communications*, Vol. 12, No. 5, 2005, pp. 98-105.

[2] G. P. White and Y. V. Zakharov, "Data Communications to Trains from High-Altitude Platforms," *IEEE Transactions on Vehicular Technology*, Vol. 56, No. 4, July 2007, pp. 2253-2266.

[3] E. Cianca, R. Prasad, M. De Sanctis, A. De Luise, M. A. Antonini, D. M. Teotino and M. Ruggieri, "Integrated Satellite-HAP Systems," *IEEE Transaction on Communications Magazine*, Vol. 43, No. 12, December 2005, pp. 33-39.

[4] B. Wang and H. M. Kwon, 'PN Code Acquisition for DS CDMA Systems Employing Smart Antennas," *IEEE Transactions on Wireless Communication*, Vol. 2, No. 1, 2003, pp. 108-117.

[5] G. Giunta and F. Benedetto, "Spread-Spectrum Code Acquisition in the Presence of Cell Correlation," *IEEE Transactions on Communications*, Vol. 55, No. 2, February 2007, pp. 257-261.

[6] H. M. Yu, T. Kang and C. D. Hong, "A New Adaptive Code-Acquisition Algorithm Using Parallel Subfilter Structure," *IEEE Transactions on Vehicular Technology*, Vol. 55, No. 6, November 2006, pp. 1790-1796.

[7] W. J. Hurd, J. I. Statman and V. A. Vilnrotter, "High Dynamic GPS Receiver Using Maximum Likelihood Estimation and Frequency Tracking," *IEEE Transactions on Aerospace and Electronic System*, Vol. 23, No. 4, July 1987, pp. 425-437.

[8] E. Karami and M. Shiva, "Blind Multi-Input—Multi-Output Channel Tracking Using Decision-Directed Maximum-Likelihood Estimation," *IEEE Transactions on Vehicular Technology*, Vol. 56, No. 3, May 2007, pp. 1447-1454.

Communications and Network, **2010, 2, 97-103**
doi:10.4236/cn.2010.22016 Published Online May 2010 (http://www.SciRP.org/journal/cn)

Maximum Ratio Combining Precoding for Multi-Antenna Relay Systems

Hamid Reza Bahrami[1], Tho Le-Ngoc[2]
[1]*Department of ECE, University of Akron, Akron, USA*
[2]*Department of ECE, McGill University, Montreal, Canada*
E-mail: tho.le-ngoc@mcgill.ca
Received November 16, 2009; *revised January* 30, 2010; *accepted February* 28, 2010

Abstract

This paper addresses the design of practical communication strategies for multi-antenna amplify-and-forward and decode-and-forward relay systems. We show that simple linear techniques at the source and destination in conjunction with maximum ratio combining can provide an optimal transmission strategy in terms of received SNR without imposing a huge computational load over the relay node(s). Besides, the structures of precoding matrices are very similar at the source and relay nodes, which reduce the complexity as all nodes can play the role of source and relay nodes without changing their transmission structure. Numerical results show that the proposed transmission and reception techniques can improve the received SNR, and hence enhance the ergodic capacity.

Keywords: Cooperation, Relay System, Linear Precoding, Maximum Ratio Combining

1. Introduction

Relay networking [1-5] is one of the frameworks in which the concept of cooperation [6] becomes meaningful. In these systems, a source node tries to send its corresponding information to the destination node with the help of one or a number of relay node(s). Cooperative relaying targets additional diversity and coding gain and provides additional level of reliability, particularly when the direct source-destination link has poor quality.

Two natural questions in the context of relay networks are the problems of transmission and reception strategies. Mainly, how the source and relay node(s) should send the information to the destination node and how the destination should optimally combine the source information with replicate version(s) of information from relay node(s). Obviously, one cannot answer to these two questions separately. In other words, transmission and reception strategies should be jointly optimized.

A major practical issue to be addressed in the design of transmission and reception schemes in relay networks is complexity. Unlike point-to-point transmission schemes in which transmitter and receiver are responsible for the recovery of their own information, in the relay systems, other parts of the network are also engaged in the communications. Hence, it is very desirable that the communication strategy imposes minimum level of computational loads on the relay nodes. This point becomes one of the constraints that should be taken into account in the design of strategies for relay systems.

Two popular strategies can be considered for transmission in relay networks: *amplify-and-forward* (AF), in which a relay node does not decode the received signal but forwards it to the destination with a specific weight (e.g., [5,7-10]), and *decode-and-forward* (DF), in which a relay node decodes the received signal from the source and retransmits a decoded version of the signal to the destination (e.g., [11,12]). In this paper, we develop practical transmission and reception for both cases. In addition, throughout this paper, we assume that all terminals (nodes) operate in a *half-duplex* mode.

While there is a vigorous body of work on the relay systems in which each individual terminal is equipped with single antenna, the case of multi-antenna nodes has not been studied extensively. In [13], it was shown that the relay systems with MIMO capability offer a promising capacity and this capacity scales linearly with the number of antennas at source/destination and logarithmically with the number of relay nodes or antennas. For a similar case, a cooperative beamforming approach that can achieve the capacity of the network in the limit of large number of relay nodes was proposed in [14]. Another interesting setup can be found in [15] where the authors elaborate the effect of relay-assisted transmission

on the capacity of rank-deficient MIMO systems.

While all the above results are attractive from a theoretical point of view, the need for practical transmission and reception schemes that can practically realize the ability of multi-antenna relay networks in providing higher capacity and performance compared to that of systems with single-antenna terminals is still pronounced. In [16], three signaling strategies for multi-antenna relay systems are discussed and compared. An optimal hybrid relaying strategy based on a combination of filtering and AF protocol was derived and can outperform AF relaying, especially when Channel State Information (CSI) is available at the relay node. In [17], an AF relaying structure that maximizes the capacity when there is no direct link between source and destination nodes has been introduced. The authors, however, mentioned that their analysis is intractable when there is a direct link between source and destination nodes.

The scarcity of studies on the design of communication schemes for multi-antenna relay systems in the literature is the main motivation of this study. More specifically, throughout this paper, we assume the source, destination and relay nodes are all equipped with multiple antennas. Our goal is to find optimal transmission and reception schemes for this setup while avoiding a huge complexity especially at the relay node. We show that a maximum ratio combining scheme at the receiver in conjunction with suitable linear precoding techniques at transmit and relay node can lead us to this end. Our study shows that the proposed scheme is optimal in terms of received SNR (and capacity) while maintaining an acceptable computational load at all nodes. In addition, the technique can be applied to both AF and DF protocols with small modifications. This feature can facilitate switching between two protocols whenever necessary. On the other hand, structures of source and relay nodes are identical, and, hence, enable a node to play the role of the source or relay in different time instants without the need of additional software or hardware overhead.

We further show that a Generalized Maximum Ratio Combiner (GMRC) at the destination is optimum for both AF and DF protocols in terms of SNR. Furthermore, for DF protocol, the precoders at the source and relay nodes should send the information in the direction of the eigenvectors corresponding to the strongest eigenvalues of the channel matrices. While it is straightforward to derive the precoding structure in the case of DF protocol, the case of AF cannot be elaborated easily. We instead propose a relay selection scheme that can result in the best possible received SNR.

The rest of the paper is organized as follows. In Section 2, we present the system model of the multi-antenna relay network. In Section 3, maximum ratio combining schemes for different scenarios such as point-to-point MIMO, multipoint-to-point system, AF and DF relaying are studied. Section 4 is allocated to the precoder design for transmit and relay nodes and in Section 5, numerical results are presented. Conclusions are given in Section 6.

2. System Model

We consider a relay system composed of one M-antenna source (transmit) node, one L-antenna relay node and one N-antenna destination (receive) node, operating in a half-duplex mode. For simplicity in notations, in the following analysis, we assume $L = M$ and single-symbol transmission, *i.e.*, at a specific time instant, the source tends to transmit a symbol x of a pre-determined code book (or constellation) to the destination[1]. It applies a precoding vector $\mathbf{w}_1$ of size $M \times 1$ to this symbol and sends it to both the destination and relay nodes. In the next time slot, based on the specific protocol (AF or DF), the relay node multiplies the received symbol by another precoding vector $\mathbf{w}_2$ of the same size and resends this precoded version to the destination. Destination then combines the two received signals based on a maximum-ratio-combining strategy.

The received signal in the first time slot can be written as:

$$\mathbf{y}_1 = \sqrt{\gamma_1}\mathbf{H}_1\mathbf{w}_1 x + \mathbf{n}_1 \tag{1}$$

where $\mathbf{H}_1$ is the $N \times M$ forward channel matrix with normalized circularly symmetric Gaussian random entries, γ_1 is its corresponding SNR, $\mathbf{y}_1$ and $\mathbf{n}_1$ are the received and white Gaussian noise vectors of size $N \times 1$, respectively. In the second time slot, the received signal is:

$$\mathbf{y}_2 = \sqrt{\gamma_2}\mathbf{H}_2\mathbf{w}_2\tilde{x} + \mathbf{n}_2 \tag{2}$$

where $\tilde{x}$ is either a detected version of x at relay node for DF or $\tilde{x} = \sqrt{\gamma_G}\mathbf{G}\mathbf{w}_1 x + \mathbf{n}$ for AF scenario. $\mathbf{G}$ is the source-relay $M \times M$ channel matrix, $\mathbf{n}$ is the $M \times 1$ noise vector at relay and γ_G is its corresponding SNR. $\mathbf{H}_2$, γ_2, $\mathbf{y}_2$ and $\mathbf{n}_2$ are defined similar to their counterparts in (1). The destination combines these two signals using two weight vectors $\tilde{\mathbf{w}}_1$ and $\tilde{\mathbf{w}}_2$ to construct the received signal

$$y = \tilde{\mathbf{w}}_1^H\mathbf{y}_1 + \tilde{\mathbf{w}}_2^H\mathbf{y}_2 \tag{3}$$

The decision is made on y to detect the transmit symbol x. Our goal is to find two precoding ($\mathbf{w}_1$ and $\mathbf{w}_2$) and two weight vectors ($\tilde{\mathbf{w}}_1$ and $\tilde{\mathbf{w}}_1$) such that the received SNR is maximized. For limited transmit power, $\mathbf{w}_1$ and $\mathbf{w}_2$ are assumed to be

$$\mathbf{w}_1^H\mathbf{w}_1 = \|\mathbf{w}_1\|^2 \le 1 \qquad \mathbf{w}_2^H\mathbf{w}_2 = \|\mathbf{w}_2\|^2 \le 1 \tag{4}$$

[1]It is not difficult to generalize the discussions to the cases of $L \neq M$ and multiple-symbol transmission. For multiple-symbol transmission, the following analysis depends on the coding used to map the multiple symbols over multiple antennas, e.g., space-time coding, and one should also assume that $N \ge L \ge M$ so that destination (and relay) nodes are able to detect all transmitted symbols.

Maximizing SNR will minimize the probability of wrong decision over x. Also, we show that maximizing SNR in this scenario is equivalent to maximizing the instantaneous mutual information and ultimately the system capacity.

3. Generalized Maximum Ratio Combining

3.1. DF Relay-Assisted MIMO System

We first start consider a point-to-point MIMO system with M transmit antennas and N receive antennas in **Figure 1(a)**. The transmitted symbol x is precoded at transmitter by precoder $\mathbf{w}$ and the received vector is combined using vector $\tilde{\mathbf{w}}$ at the receiver. Therefore, the system model can be written as:

$$\mathbf{y} = \sqrt{\gamma}\tilde{\mathbf{w}}^H\mathbf{H}\mathbf{w}x + \tilde{\mathbf{w}}\mathbf{n} \tag{5}$$

where $\mathbf{H}$, $\mathbf{n}$ and γ are similar to $\mathbf{H}_1$ and $\mathbf{n}_1$ in (1), and the corresponding received SNR is

$$\text{SNR} = \gamma\frac{\mathbf{w}^H\mathbf{H}^H\tilde{\mathbf{w}}\tilde{\mathbf{w}}^H\mathbf{H}\mathbf{w}}{\tilde{\mathbf{w}}^H\tilde{\mathbf{w}}}$$

It is well known that one can select $\tilde{\mathbf{w}} = \mathbf{H}\mathbf{w}$ to achieve the maximum $\text{SNR} = \gamma\mathbf{w}^H\mathbf{H}^H\mathbf{H}\mathbf{w}$ and maximum mutual information between x and $\tilde{x}$,

$$I(x;\tilde{x}) = \log(1+\text{SNR}) = \log(1+\gamma\mathbf{w}^H\mathbf{H}^H\mathbf{H}\mathbf{w}).$$

Now let consider the case when two transmitters send the same information to a receiver but in different time instants as shown in **Figure 1(b)**. Both transmitters are equipped with M antennas while the receiver has N antennas. This case also corresponds to an *ideal* DF scenario when the relay node can always correctly decode the source information. The system model is the same as (1) and (2) except that $\tilde{x}$ is replaced by x in (2). Combining (1), (2) and (3), one can write an *equivalent* compound system equation for the above scenario as

$$y = (\tilde{\mathbf{w}}_1^H\ \tilde{\mathbf{w}}_2^H)\begin{pmatrix}\sqrt{\gamma_1}\mathbf{H}_1 & \mathbf{0}\\ \mathbf{0} & \sqrt{\gamma_2}\mathbf{H}_2\end{pmatrix}\begin{pmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{pmatrix}x + (\tilde{\mathbf{w}}_1^H\ \tilde{\mathbf{w}}_2^H)\begin{pmatrix}\mathbf{n}_1\\ \mathbf{n}_2\end{pmatrix} \tag{6}$$

This is clearly a MIMO system with $2M$ transmit antennas and $2N$ receive antennas, and the generalized maximum ratio combining (GMRC) scheme for such a system is

$$\tilde{\mathbf{w}} = \mathbf{H}'\mathbf{w},\ \tilde{\mathbf{w}}_1 = \mathbf{H}_1\mathbf{w}_1 \text{ and } \tilde{\mathbf{w}}_2 = \mathbf{H}_2\mathbf{w}_2 \tag{7}$$

where

$$\mathbf{H}' = \begin{pmatrix}\sqrt{\gamma_1}\mathbf{H}_1 & \mathbf{0}\\ \mathbf{0} & \sqrt{\gamma_2}\mathbf{H}_2\end{pmatrix}y\ ;\ \tilde{\mathbf{w}} = \begin{pmatrix}\tilde{\mathbf{w}}_1\\ \tilde{\mathbf{w}}_2\end{pmatrix} \text{ and } \mathbf{w} = \begin{pmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{pmatrix}$$

Note that there is a fundamental difference between MRC and GMRC as MRC just considers combining in space domain over elements of receive antennas while GMRC includes combining over both space and time.

In a general DF scenario, the received signal after combining becomes

$$y = (\tilde{\mathbf{w}}_1^H\ \tilde{\mathbf{w}}_2^H)\begin{pmatrix}\sqrt{\gamma_1}\mathbf{H}_1 & \mathbf{0}\\ \mathbf{0} & \sqrt{\gamma_2}\mathbf{H}_2\end{pmatrix}\begin{pmatrix}\mathbf{w}_1x\\ \mathbf{w}_2x'\end{pmatrix} + (\tilde{\mathbf{w}}_1^H\ \tilde{\mathbf{w}}_2^H)\begin{pmatrix}\mathbf{n}_1\\ \mathbf{n}_2\end{pmatrix} \tag{8}$$

where x' denotes the symbol decoded and re-transmitted by the relay, *i.e.*, $x' = x$ for correct detection and $x' \neq x$ for erroneous detection. Let α denote the probability of erroneous detection at the relay. The optimum combining vectors for DF transmission can be written as:

$$\tilde{\mathbf{w}}_1 = \mathbf{H}_1\mathbf{w}_1 \text{ and } \tilde{\mathbf{w}}_2 = \delta\mathbf{H}_2\mathbf{w}_2 \tag{9}$$

where the coefficient δ can be estimated from α. Consider an approximation by assuming that $x' = -x$ for an erroneous detection at the relay. Note that this consideration is exact for binary transmission. For a general M-ary signaling scheme, this assumption represents a pessimistic consideration. Under this assumption, the SNR's corresponding to the cases of correct and erroneous detection at the relay are respectively,

$$\text{SNR}_1 = \frac{(d_1+d_2)(d_1'+d_2')}{d_o},\ \text{SNR}_2 = \frac{(d_1-d_2)(d_1'-d_2')}{d_o}, \tag{10}$$

where

$$d_1 = \sqrt{\gamma_1}\mathbf{w}_1^H\mathbf{H}_1^H\tilde{\mathbf{w}}_1,\ d_2 = \sqrt{\gamma_2}\mathbf{w}_2^H\mathbf{H}_2^H\tilde{\mathbf{w}}_2,\ d_1' = \sqrt{\gamma_1}\tilde{\mathbf{w}}_1^H\mathbf{H}_1\mathbf{w}_1,$$

$$d_2' = \sqrt{\gamma_2}\tilde{\mathbf{w}}_2^H\mathbf{H}_2\mathbf{w}_2, \text{ and } d_o = \tilde{\mathbf{w}}_1^H\tilde{\mathbf{w}}_1 + \tilde{\mathbf{w}}_2^H\tilde{\mathbf{w}}_2.$$

The average received SNR can then be written as (11).

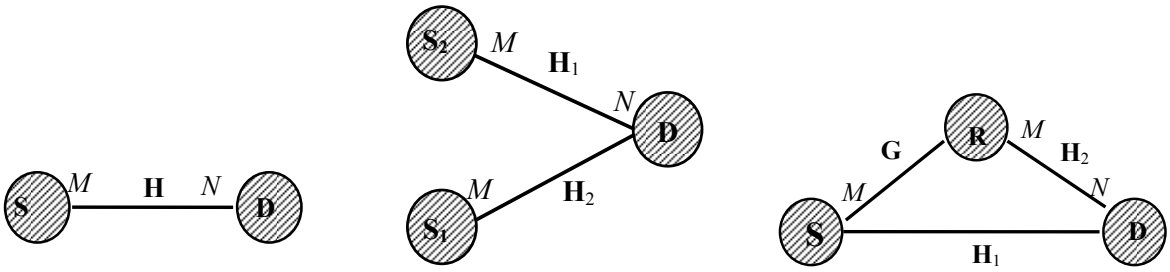
From (11) and considering a system equation similar to (6) and (7), one can find the corresponding equivalent system equation for this general DF scenario as (12).

$$\text{SNR} = (1-\alpha)\text{SNR}_1 + \alpha\,\text{SNR}_2 = \frac{(\gamma_1\mathbf{w}_1^H\mathbf{H}_1^H\tilde{\mathbf{w}}_1\tilde{\mathbf{w}}_1^H\mathbf{H}_1\mathbf{w}_1 + \gamma_2\mathbf{w}_2^H\mathbf{H}_2^H\tilde{\mathbf{w}}_2\tilde{\mathbf{w}}_2^H\mathbf{H}_2\mathbf{w}_2) + (1-2\alpha)\sqrt{\gamma_1\gamma_2}(\mathbf{w}_1^H\mathbf{H}_1^H\tilde{\mathbf{w}}_1\tilde{\mathbf{w}}_2^H\mathbf{H}_2\mathbf{w}_2 + \tilde{\mathbf{w}}_1^H\mathbf{H}_1\mathbf{w}_1\tilde{\mathbf{w}}_2^H\mathbf{H}_2\mathbf{w}_2)}{\tilde{\mathbf{w}}_1^H\tilde{\mathbf{w}}_1 + \tilde{\mathbf{w}}_2^H\tilde{\mathbf{w}}_2} \tag{11}$$

$$y = (\tilde{\mathbf{w}}_1^H\ \tilde{\mathbf{w}}_2^H)\mathbf{H}\begin{pmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{pmatrix}x + (\tilde{\mathbf{w}}_1^H\ \tilde{\mathbf{w}}_2^H)\begin{pmatrix}\mathbf{n}_1\\ \mathbf{n}_2\end{pmatrix} \tag{12}$$

where $\mathbf{H}$ is the channel matrix of the equivalent MIMO system and can be approximated as:

$$\mathbf{H} \approx \begin{pmatrix}\sqrt{\gamma_1}\mathbf{H}_1 & \mathbf{0}\\ \mathbf{0} & \sqrt{\gamma_2}(1-2\alpha)\mathbf{H}_2\end{pmatrix}$$



(a) (b) (c)

Figure 1: Diagrams of different transmission systems: (a) point-to-point MIMO; (b) multipoint-to-point MIMO; (c) relay-assisted MIMO.

In other words, the coefficient δ can be approximated as $\delta \approx 1-2\alpha$ when α is small. Note that one can use other estimations to find δ or the optimum weight ad-hoc. In general, the idea is to reduce the weight of the signal in the second transmit interval as it may contain error and degrade the total received SNR at the destination.

3.2. AF Relay-Assisted MIMO System

The difficulty with this scenario is that the noise in the second time slot is no longer white. In other word, based on (1) and (2), the system equation in this case can be rewritten as:

$$y = (\tilde{\mathbf{w}}_1^H \ \tilde{\mathbf{w}}_2^H)\begin{pmatrix} \sqrt{\gamma_1}\mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \sqrt{\gamma_2\gamma_G}\mathbf{H}_2 \end{pmatrix}\begin{pmatrix} \mathbf{w}_1 \\ \mathbf{W}_2^H\mathbf{G}\mathbf{w}_1 \end{pmatrix} x + (\tilde{\mathbf{w}}_1^H \ \tilde{\mathbf{w}}_2^H)\begin{pmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 + \sqrt{\gamma_2}\mathbf{H}_2\mathbf{W}_2^H\mathbf{n} \end{pmatrix} \quad (13)$$

The difference between the AF and DF protocols is that $\mathbf{W}_2$ is the $M \times M$ precoding matrix (instead of vector) at the relay node, and the noise in the second time interval is no longer white. Assuming $\mathbf{n}$ and $\mathbf{n}_2$ are uncorrelated, the autocorrelation function of the noise in the second time interval can be written as:

$$\begin{aligned}\mathbf{\Lambda} &= E\{(\mathbf{n}_2 + \sqrt{\gamma_2}\mathbf{H}_2\mathbf{W}_2^H\mathbf{n})(\mathbf{n}_2 + \sqrt{\gamma_2}\mathbf{H}_2\mathbf{W}_2^H\mathbf{n})^H\} \\ &= \mathbf{I}_M + \gamma_2\mathbf{H}_2\mathbf{W}_2^H\mathbf{W}_2\mathbf{H}_2^H \end{aligned} \quad (14)$$

where $\mathbf{I}_M$ denotes identity matrix of size M. Applying an eigenvalue decomposition over $\mathbf{\Lambda}$, one can write $\mathbf{\Lambda} = \mathbf{U}\mathbf{D}\mathbf{U}^H$ where $\mathbf{U}$ and $\mathbf{D}$ are $N \times N$ unitary and diagonal matrices, respectively.

Defining the pre-whitening filter as $\mathbf{W}_P = \mathbf{D}^{-1/2}\mathbf{U}^H$ and applying it to the received signal in the second time-slot will make the output noise white. With that, one can apply (8) directly to derive the structure of combining vectors. The system model can be considered as (12) with:

$$\mathbf{H} = \begin{pmatrix} \sqrt{\gamma_1}\mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \sqrt{\gamma_2\gamma_G}\mathbf{D}^{-1/2}\mathbf{U}^H\mathbf{H}_2 \end{pmatrix}$$

Now from (13), the optimum combining vectors for an AF relay-assisted MIMO system can be written as:

$$\tilde{\mathbf{w}}_1 = \mathbf{H}_1\mathbf{w}_1 \ \text{ and } \ \tilde{\mathbf{w}}_2 = \mathbf{D}^{-1/2}\mathbf{U}^H\mathbf{H}_2\mathbf{W}_2^H\mathbf{G}\mathbf{w}_1 \quad (15)$$

In other words, the optimum weight vector in the second time interval is a combination of an MRC vector and a pre-whitening filter. By applying pre-whitening filter, the output noise will be white and therefore one can apply the combining weight vectors in (15) to maximize the SNR, the instantaneous mutual information and ultimately the system capacity.

4. Precoding for Relay-Assisted MIMO Systems

Our goal here is to investigate the design of precoding vectors, $\mathbf{w}_1$ and $\mathbf{w}_2$, in (1) and (2). We start with point-to-point MIMO transmission and then generalize the results to the case of relay-assisted MIMO systems.

Recall that, after applying MRC weight vector, (6) yields $\text{SNR} = \gamma\mathbf{w}^H\mathbf{H}^H\mathbf{H}\mathbf{w}$. To maximize this SNR subject to power constraint over $\mathbf{w}$, similar to (4), one should take $\mathbf{w}$ in the direction of the eigenvector of Hermitian matrix $\mathbf{H}^H\mathbf{H}$ associated with $\lambda_{\max}$ where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{H}^H\mathbf{H}$. In other words, under the total transmit power constraint at source, the optimum precoding vector that maximizes the SNR of an MRC-based MIMO system can be written as

$$\mathbf{w} = \mathbf{u}_{\max}(\mathbf{H}^H\mathbf{H}) \quad (16)$$

where $\mathbf{u}_{\max}$ stands for the eigenvector of $\mathbf{H}$ corresponding the maximum eigenvalue. With this precoding vector and considering the receive combining vector of $\tilde{\mathbf{w}} = \mathbf{H}\mathbf{w}$, the receive SNR of the system can be written as $\text{SNR} = \gamma\lambda_{\max}$. From (8), (9) and (16), it can be shown that

under the total transmit power constraints at source and relay nodes, for a multipoint-to-point MIMO system with GMRC combining vectors at receiver, the optimum precoding vectors are:

$$\mathbf{w}_1 = \mathbf{u}_{\max}(\mathbf{H}_1^H \mathbf{H}_1) \text{ and } \mathbf{w}_2 = \mathbf{u}_{\max}(\mathbf{H}_2^H \mathbf{H}_2) \quad (17)$$

The same conclusion is also valid for the case of DF protocol for relay-assisted MIMO systems.

Now, we are ready to revisit the problem of AF protocol. The difficulty with this case is that the optimization of precoding matrix at the relay node, $\mathbf{W}_2$, is not independent of the optimization of precoding vector at the source node, $\mathbf{w}_1$. This is because the received SNR equation resulted in the second transmit interval is

$$\text{SNR}_2 = \sqrt{\gamma_2 \gamma_G}\, \mathbf{w}_1^H \mathbf{G}^H \mathbf{W}_2 \mathbf{H}_2^H \mathbf{H}_2 \mathbf{W}_2^H \mathbf{G} \mathbf{w}_1 \quad (18)$$

and, therefore, is a function of both $\mathbf{w}_1$ and $\mathbf{W}_2$. Note that (18) comes from applying (15) to calculate the SNR in (6). Although direct maximization of (18) for $\mathbf{W}_2$ would be difficult, one can make a clever guess that if we select $\mathbf{W}_2$ such that it maximizes SNR over the source-relay link, $\mathbf{G}$, this can ultimately result in the maximization of the SNR in (18) over the entire link from source to destination. Therefore, assume that $\mathbf{W}_2$ can be written as

$$\mathbf{W}_2 = \mathbf{G}\mathbf{w}_1 \hat{\mathbf{w}}_2^H \quad (19)$$

where $\mathbf{G}\mathbf{w}_1$ is responsible for maximizing the SNR at relay node while $\hat{\mathbf{w}}_2$ is an independent vector reserved for further optimization of precoding matrix at the relay node. Moreover, for maximizing the SNR over the relay-destination link, using (16), it can be shown that

$$\hat{\mathbf{w}}_2 = \mathbf{u}_{\max}(\mathbf{H}_2) \quad (20)$$

Now, $\mathbf{w}_1$, the precoding vector at the source node is the only remaining parameter to be selected. However, $\mathbf{w}_1$ affects the received SNR from both direct and relayed links. Therefore, to optimize $\mathbf{w}_1$, the SNR equations similar to (10) should be considered, which makes the optimization problem very difficult if not impossible to solve. To resolve this problem, we focus on each of the transmit intervals, separately. Since in a communication system, there are usually a number of available relay terminals (rather than just one), we, ultimately, propose the use of relay selection approach to maximize the overall SNR of the system.

Consider the received SNR in the first transmit interval:

$$\text{SNR}_1 = \gamma_1 \mathbf{w}_1^H \mathbf{H}_1^H \mathbf{H}_1 \mathbf{w}_1$$

Based on (16), SNR_1 is maximized if $\mathbf{w}_1 = \mathbf{u}_{\max}(\mathbf{H}_1)$. On the other hand, in the second transmit interval, by substituting $\mathbf{W}_2$ of (19) into (18), it turns out that the source precoding vector will be responsible for the source-relay portion of the SNR. Therefore, to maximize this portion, one should select $\mathbf{w}_1 = \mathbf{u}_{\max}(\mathbf{G})$. These two equations for $\mathbf{w}_1$ are definitely in contrast with each other. The best scenario is that $\mathbf{u}_{\max}(\mathbf{H}_1) = \mathbf{u}_{\max}(\mathbf{G})$. In this case, based on (16), the overall receive SNR of the system can be written as:

$$\text{SNR} = \lambda_{\max}(\mathbf{H}_1^H \mathbf{H}_1)\,(\gamma_1 + \gamma_G \gamma_2 \lambda_{\max}(\mathbf{H}_2^H \mathbf{H}_2)) \quad (21)$$

Now, let assume that $\mathbf{u}_{\max}(\mathbf{H}_1) \neq \mathbf{u}_{\max}(\mathbf{G})$ but there are K available relay nodes in the system. The best relay should be selected such that the overall receive SNR is maximized. The performance degradation appears as a factor of $\langle \mathbf{u}_{\max}(\mathbf{H}_1), \mathbf{u}_{\max}(\mathbf{G}) \rangle$ in the SNR equation. Two following extreme cases can be considered.

In the first case when the source-destination link is very strong (*i.e.*, asymptotic case of $\lambda_1 \to \infty$), one should choose the precoding vector in the direction of $\mathbf{u}_{\max}(\mathbf{H}_1)$. Therefore, the SNR loss as compared to the optimum case in (21), is due to the *eigen-mismatch* in the *second* transmit interval and the overall SNR can be expressed as:

$$\text{SNR} = \lambda_{\max}(\mathbf{H}_1^H \mathbf{H}_1)\,\gamma_1 + \langle \mathbf{u}_{\max}(\mathbf{H}_1), \mathbf{u}_{\max}(\mathbf{G}) \rangle\, \gamma_G \gamma_2 \lambda_{\max}(\mathbf{G}^H \mathbf{G})\, \lambda_{\max}(\mathbf{H}_2^H \mathbf{H}_2) \quad (22)$$

In the second case when the relay links are much stronger compared to the direct link (*i.e.*, asymptotic case of $\lambda_2, \lambda_G \to \infty$), the natural selection is the source precoding vector in the direction of $\mathbf{u}_{\max}(\mathbf{G})$. In this case, the SNR loss is due to the *eigen-mismatch* in the *first* transmit interval. The overall received SNR can also be written as:

$$\text{SNR} = \langle \mathbf{u}_{\max}(\mathbf{H}_1), \mathbf{u}_{\max}(\mathbf{G}) \rangle \lambda_{\max}(\mathbf{H}_1^H \mathbf{H}_1)\,\gamma_1 + \gamma_G \gamma_2 \lambda_{\max}(\mathbf{G}^H \mathbf{G})\, \lambda_{\max}(\mathbf{H}_2^H \mathbf{H}_2)) \quad (23)$$

Based on (22) and (23), amongst all K candidate relay nodes, the best relay node can be selected as

$$i = \arg \max_{i=1,\ldots,K} \langle \mathbf{u}_{\max}(\mathbf{H}_1), \mathbf{u}_{\max}(\mathbf{G}) \rangle\, \gamma_G \gamma_2 \lambda_{\max}(\mathbf{H}_2^H \mathbf{H}_2)$$

if the source-destination link is stronger than the source-relay and relay-destination links, or

$$i = \arg \max_{i=1,\ldots,K} \langle \mathbf{u}_{\max}(\mathbf{H}_1), \mathbf{u}_{\max}(\mathbf{G}) \rangle \lambda_{\max}(\mathbf{H}_1^H \mathbf{H}_1) + \gamma_G \gamma_2 \lambda_{\max}(\mathbf{H}_2^H \mathbf{H}_2)$$

if the source-relay and relay-destination links are stronger than the source-destination link. $<$, $>$ stands for inner product. This selection maximizes the SNR of the relay-assisted AF system with high probability. In other words, the best relay is the one with strong source-relay and relay-destination links, *i.e.*, large γ_G and γ_2, and when the eigenvector corresponding to the maximum eigenvalue of source-relay link matrix is the closest to $\mathbf{u}_{\max}(\mathbf{H}_1)$.

5. Numerical Results

We study the performance of the proposed MRC-based precoding technique by means of simulation. We consider all source, relay and destination nodes are equipped with two antennas, *i.e.*, $M = N = 2$. We investigate the received SNR and average mutual information for these systems.

Figure 2 shows the received SNR for different setups in **Figure 1**. For the sake of comparison, we also show the performance of the precoder based on equal gain combining in point-to-point MIMO transmission. First, we observe that MRC-based precoding (case *A*) outperforms precoding with equal gain combining in terms of received SNR by about 3 dB. On the other hand, in the case of multipoint-to-point transmission (case *B*), precoding based on MRC can provide an additional gain as compared to the case of point-to-point transmission. This is mainly due to the availability of an additional path between the second transmit and receive nodes, which provides a higher average SNR at the receiver (especially when the link from first transmitter to receive node is poor). Finally, we see that the relay-assisted amplify-and-forward precoding system (case *C*) provides a received SNR comparable to that of an ideal multi-point-to-point system. The slight decrease in the received SNR is due to the noise accumulation in the relay node. The same conclusion is also valid for the case of a decode-and-forward system as there is a decrease in the received SNR due to the probability of wrong decision at the relay node.

Figure 3 also shows that the average mutual informa-

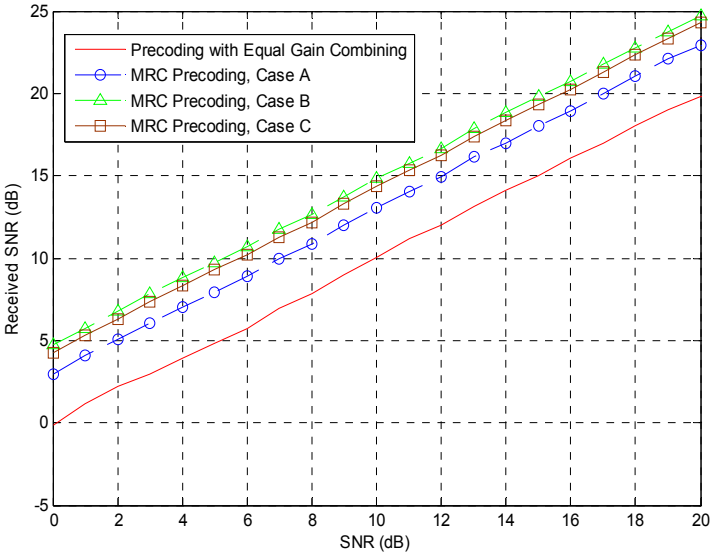


Figure 2. Received SNR in different scenarios.

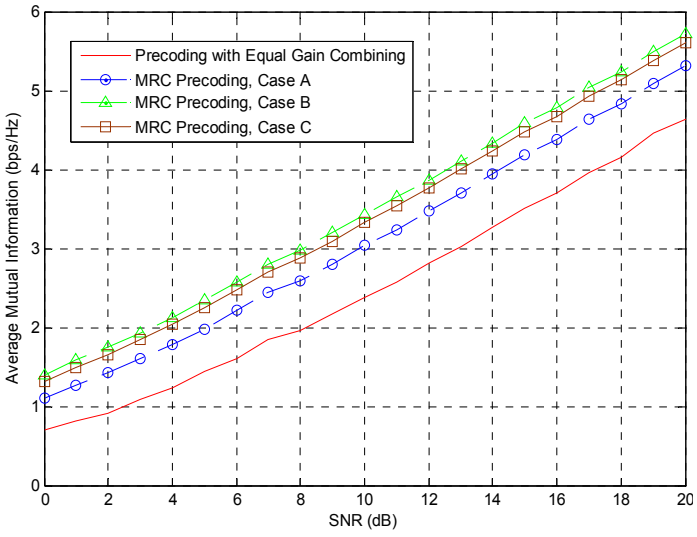


Figure 3. Average mutual information in different scenarios.

tion for all the above scenarios. The same conclusions can be drawn for the average mutual information, *i.e*., the relay-assisted MIMO precoding can provide an average mutual information close to that in the case when two transmitters send identical information to the relay node.

6. Conclusions

We studied optimal linear transmit and receive strategies for a variety of MIMO systems. Our focus was mainly on relay-assisted MIMO systems. These systems are very attractive from both theoretical and practical points of view. However, there are still many open questions concerning transmission and reception schemes in the field. We built a framework based on the well-known MRC scheme on the receiver (destination) side. As we demonstrated, the construction of the optimal combining strategies for both DF and AF relaying protocols can be based on the concept of MRC.

We first derived the optimal linear receiver structure for these systems. Next, based on the structure of optimum receivers, we investigated the optimal linear precoding vectors for the source and relay nodes. Our results show that, for the optimum receiver, the optimal transmit precoding strategy for DF protocol is to send the information in the direction of the eigenvectors of the direct ($\mathbf{H}_1$) and relay ($\mathbf{H}_2$) channel matrices associated with the strongest eigenvalues. This simple result is, however, not valid in the case of AF relaying protocol. Instead, for AF protocol, we propose the use of relay selection scheme to facilitate the design of precoders at the source and relay nodes. Finally, different numerical examples proved that the proposed optimal transmission and reception techniques are indeed effective and provide a meaningful gain in term of received SNR and system capacity while maintaining the complexity very low due to linearity.

7. References

[1] P. Gupta and P. R Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, Vol. 46, No. 2, March 2002, pp. 388-404.

[2] M. Gastpar and M. Vetterli, "On the Capacity of Wireless Networks: The Relay Case," *Proceedings of* 21*th Annual joint Conference of the IEEE Computer and Communications*, New York, Vol. 3, June 2002, pp. 1577-1586.

[3] A. Sendonaris, E. Erkip and B. Aaazhang, "User Cooperation Diversity—Part I: System Description," *IEEE Transactions on Communications*, Vol. 51, No. 11, November 2003, pp. 1927-1938.

[4] A. Sendonaris, E. Erkip and B. Aaazhang, "User Cooperation Diversity—Part II: Implementation Aspects and Performance Analysis," *IEEE Transactions on Communications*, Vol. 51, No. 11, November 2003, pp. 1939-1948.

[5] J. N. Laneman, D. N. C. Tse and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behaviour," *IEEE Transactions on Information Theory*, Vol. 50, No. 12, December 2004, pp. 3062-3080.

[6] F. H. P. Fitzek and M. D. Katz, "Cooperation in Wireless Networks: Principles and Applications," Springer, Netherland, 2006.

[7] R. U. Nabar, H. Bölcskei and F. W. Kneubuhler, "Fading Relay Channels: Performance Limits and Space-Time Signal Design," *IEEE Journal on Selected Areas in Communications*, Vol. 22, No. 6, August 2004, pp. 1099- 1109.

[8] J. N. Laneman and G. W. Wornell, "Distributed Space-Time-Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks," *IEEE Transactions on Information Theory*, Vol. 49, No. 10, October 2003, pp. 2415-2425.

[9] A. Stefanov and E. Erkip, "Cooperative Coding for Wireless Networks," *IEEE Transactions on Communications*, Vol. 52, No. 9, September 2004, pp. 1470-1476.

[10] M. Munoz-Medina, J. Vidal and A. Agustion, "Linear Transceiver Design in Nonregenerative Relays with Channel State Information," *IEEE Transactions on Signal Processing*, Vol. 55, No. 6, June 2007, pp. 2593-2604.

[11] A. Bletsas, A. Khisti, D. P. Reed and A. Lippman, "A Simple Cooperative Diversity Method Based on Network Path Selection," *IEEE Journal on Selected Areas in Communications*, Vol. 24, No. 3, March 2006, pp. 659-672.

[12] A. Bletsas, "Intelligent Antenna Sharing in Cooperative Diversity Wireless Networks," Ph.D. Dissertation, Massachusetts Institute of Technology, 2005.

[13] H. Bölcskei, R. U. Nabar, O. Oyman and A. J. Paulraj, "Capacity Scaling Laws in MIMO Relay Networks," *IEEE Transactions on Wireless Communications*, Vol. 5, No. 6, June 2006, pp. 1433-1444.

[14] S. O. Gharan, A. Bayesteh and A. K. Khandani, "Asymptotic Analysis of Amplify and Forward Relaying in a Parallel MIMO Relay Network," submitted for publication. http://cst.uwaterloo.ca

[15] A. Wittneben and B. Rankov, "Impact of Cooperative Relays on the Capacity of Rank-Deficient MIMO Channels," *Proceedings of the* 12*th IST Summin on Mobile Wireless Communications*, Aveiro, June 2003, pp. 421-425.

[16] Y. Fan and J. S. Thompson, "MIMO Configurations for Relay Channels: Theory and Practive," *IEEE Transactions on Wireless Communications*, Vol. 6, No. 5, May 2007, pp. 1774-1786.

[17] X. Tang and Y. Hua, "Optimal Design of Nonregenerative MIMO Wireless Relays," *IEEE Transactions on Wireless Communications*, Vol. 6, No. 4, April 2007, pp. 1398-1407.

Communications and Network, **2010, 2, 104-112**
doi:10.4236/cn.2010.22017 Published Online May 2010 (http://www.SciRP.org/journal/cn)

Scientific Research

A Survey on Real-Time MAC Protocols in Wireless Sensor Networks

Zheng Teng, Ki-Il Kim*
Department of Informatics, *Research Institute of Computer and Information Communication*,
Gyeongsang National University, *Jinju*, *Korea*
E-mail: *kikim@gnu.ac.kr*
Received January 13, 2010; *revised February* 19, 2010; *accepted March* 10, 2010

Abstract

As wireless sensor network becomes pervasive, new requirements have been continuously emerged. However, the most of research efforts in wireless sensor network are focused on energy problem since the nodes are usually battery-powered. Among these requirements, real-time communication is one of the big research challenges in wireless sensor networks because most of query messages carry time information. To meet this requirement, recently several real-time medium access control protocols have been proposed for wireless sensor networks in the literature because waiting time to share medium on each node is one of main source for end-to-end delay. In this paper, we first introduce the specific requirement of wireless sensor real-time MAC protocol. Then, a collection of recent wireless sensor real-time MAC protocols are surveyed, classified, and described emphasizing their advantages and disadvantages whenever possible. Finally we present a discussion about the challenges of current wireless sensor real-time MAC protocols in the literature, and show the conclusion in the end.

Keywords: Wireless Sensor Networks, Medium Access Control (MAC), Real-Time

1. Introduction

A wireless sensor network [1] consists of a large number of small, inexpensive sensor nodes which are distributed over a geographical area for monitoring physical phenomena like temperature, noise, light intensity and speed etc. Traditionally, the largest challenge of sensor network is the limited lifetime because of the battery-powered node [2]. Specially, applications like military operation, factory automation and so on, need a constraint time of a message transmission from source node to destination for guaranteeing validity of the message. For such kind of cases, the real-time system can play a crucial role.

Real-time system is a computing system that must react within precise time constraints to events in the environment. In the real-time computing system, the primary feature is called the deadline, which is the maximum time which it must complete its execution within. In several critical applications, a massage arrived at the destination after the deadline is both late and wrong. The real-time algorithm not only requires low delay of a packet possess but also to meet the deadline, that is the largest difference between real-time sensor networks and conventional sensor networks. In fact, whereas the objective of the low delay sensor network is to minimize the average response time of a given set of tasks, the objective of real-time sensor network is to meet the individual timing requirement of each task.

Real-time system provides some important features in the critical applications, including:

- Timeliness—Messages have to be transmitted not only by the time they arrive at the destination but also in the time domain.
- Design for peak load—Real-time systems should not collapse when they encounter a peak-load condition, hence they must be designed to manage all anticipated scenarios.
- Predictability—Real-time system should be able to predict the consequences of any scheduling decision for guaranteeing the performance of applications.
- Fault tolerance—Single failure of transmission should not lead the system to crash. Consequently the real-time systems are designed to be fault tolerant.
- Maintainability—The architecture of a real-time system should be designed to ensure that possible system modifications are easy to perform.

In order to adapt the energy constrained applications for prolonging the lifetime of sensor network, the proposed medium access control (MAC) protocols primarily focus on reducing energy consumption related to the wireless medium [3]. Therefore, substantial number of the MAC protocols [4-12] for wireless sensor networks are designed in the literature for the traditional challenge: energy-efficiency. Furthermore, other parameters such as latency or throughput are also important for sensor network transmission. However, during the time critical applications, the largest challenge is that how to let the alarm messages meet their deadline for guaranteeing safety of events in the environment. In addition to those applications, energy consumption is the secondary importance just like WSN is employed in natural disaster monitor system.

Another impact of real-time communications is as follows. The traffic load is not regular for monitoring environment by WSN since the environmental conditions constantly change over time. In general the situation of application environment is calm, however, when the emergencies are detected, plenty of information is sensed and needed to be transmitted to the user. The traffic congestion is easier to be triggered since the suddenly increased traffic load. Then some important messages need to wait for a long time to be transmitted to sink even dropped. Under above case, the real-time system has great potential for reliving or avoiding the phenomena which is mentioned above in many applications.

In this paper, we present an introduction to real-time MAC protocols for wireless sensor networks. As real-time MAC sensor networks differ from traditional wireless MAC networks in many points, the primary parameters of them reflect in timing requirement (deadline), energy, multiple flows, etc. All of these characteristics make the traditional wireless MAC protocols not be suitable for real-time sensor networks.

The remainder of this paper is organized as follows. A survey of the proposed real-time MAC protocol for sensor networks is presented in the Section 2. In Section 3, we discuss the challenges of the current real-time MAC protocol for sensor network and show the future work. In Section 5, we make a conclusion.

2. Real-Time MAC Protocols for Wireless Sensor Network

In this survey we collect recent real-time MAC protocols proposed in the literature. Several typical protocols are included and discussed in this section. According to the difference of applications, real-time MAC protocol for wireless sensor networks can be classified into two classes: hard real-time MAC protocol and soft real-time MAC protocol.

As it is shown in **Figure 1**, there is a classification tree for wireless sensor networks real-time MAC protocols. As shown in this taxonomy, in order to guarantee the constraint time for ensuring the validity of alarm messages in the time critical application, several proposed approaches in the literature are classified as branches from hard and soft real-time MAC protocols. In the following section, we present the description of the collection of wireless sensor real-time MAC protocols and evaluate their advantages and disadvantages whenever possible.

2.1. Hard Real-Time MAC Protocol for Wireless Sensor Networks

2.1.1. TDMA Based Real-Time MAC Protocol

RRMAC [13] is a TDMA based hard real-time MAC protocol for wireless sensor networks. As shown in **Figure 2**, this protocol proposes a tree structure that the packets could flow continuously from leaf-level nodes to the top-level node. The assignment sequence lets the base station of the top of the tree acquire the data from normal sensor nodes in one superframe duration. **Figure 3** shows the superframe structure, which is composed of a beacon only period, contention free period, contention access period, and an inactive period. The protocol uses a beacon frame for synchronizing the sensor network. The nodes adjust their time whenever they receive a beacon frame in the system and then execute the contention free period. The protocol decreases the communication delay by assigning time slots to every node. For multi-hop beacon forwarding, in the beacon only period, each beacon slot is assigned to each coordinator, like the base

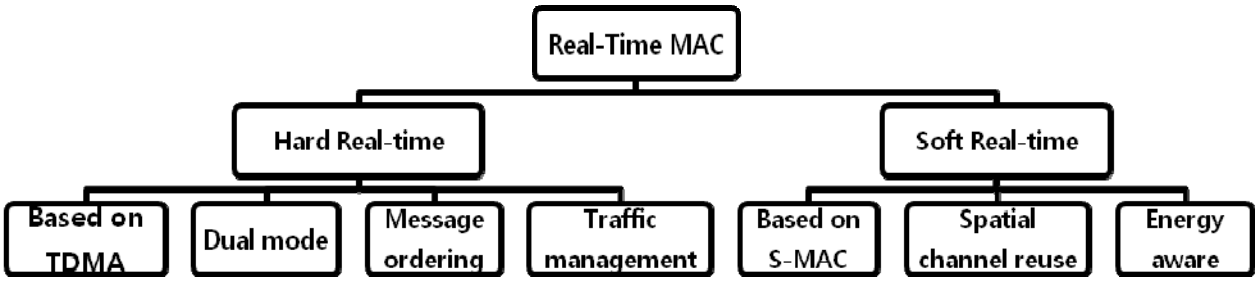


Figure 1. Taxonomy of approaches to real-time MAC protocol in wireless sensor networks

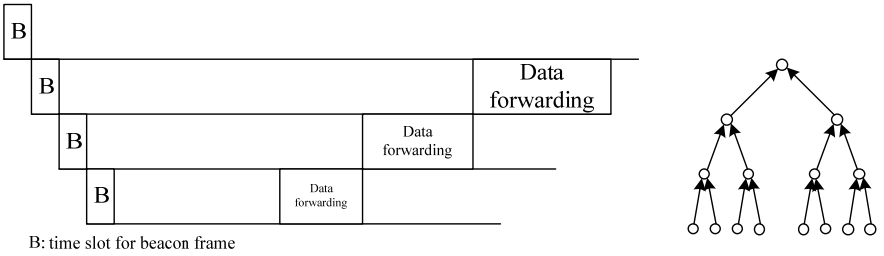


Figure 2. Time slot assignment for a base station and sink node in RRMAC.

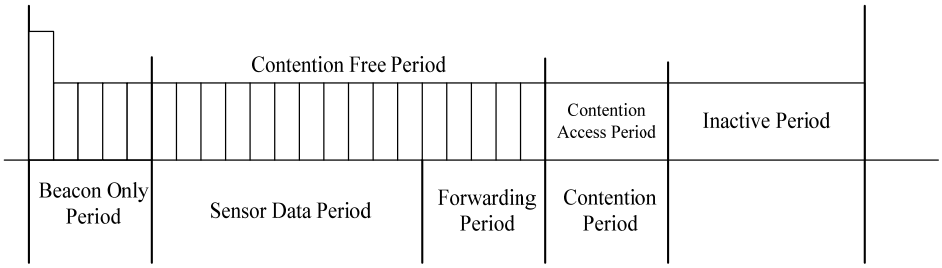


Figure 3. RRMAC superframe structure.

station and sink. This design enables the beacon to be forwarded to the hall network at the beginning of a new superframe.

Several nodes form clusters and each cluster includes a cluster head. RRMAC superframe bases on IEEE 802.15.4 frame structure and only upper level cluster heads can assign time slots in the TDMA superframe. Cluster heads aggregate data collected from lower level sensor nodes and forward the data to upper cluster head in hierarchy. RRMAC nodes are assumed that they have two RF power levels. The sink nodes have high RF power which increases the transmission range of these nodes. Normal nodes have a smaller power and a short communication range. The RRMAC superframe structure is flexible. The superframe can contain only contention period if all of the sensor or sink nodes do not require real time or reliable data transmission. However, the difficulty of RRMAC is maintaining global synchronization in a large randomly distributed multi-hop WSN.

2.1.2 Two Mode-Based Real-Time MAC Protocols

Dual-mode real-time MAC protocol [14,15] is hard real-time MAC protocol for wireless sensor networks. This protocol includes two modes, one is protected mode and another is unprotected mode. The **Figure 4(a)** shows the unprotected mode. In that case, maxrange is the possible max communication range of node. Protocol presents a parameter called backoffunprotected which is the back-off for unprotected mode. When a node sends out an alarm message, nodes that hear its backoffunprotected, which is inversely proportional to their distance to the sending node. A node can be chosen for forwarding the message when no alarm message has been received and backoffunprotected expires. In this way, messages can be transmitted to the sink by a high speed, but not have a high reliability due to collisions. In the protected mode, the protocol provides a high reliability by guaranteeing the collision-free function, however the transmission time is bounded. Every node knows its absolute position since deployment and each message contains the sender's absolute position. In the initial phase, the protocol organizes the network nodes into cells so that all nodes of a cell can communicate with other nodes of two neighboring cells. The unprotected mode does not use cells for transmitting messages. The protected mode uses signaling messages for reserving each cell between source node and sink node (as shown in **Figure 4 (b)**). Once reserved, a cell cannot generate new messages until the transmission is over for avoiding collision. After initialization, the unprotected mode will first be started but when any node detects a collision, it will send a collided alarm message to another node for switching the mode from unprotected to protected mode. The dual-mode real-time MAC protocol supports the randomly deployed wireless sensor network, and can avoid message collision effectively. However, the protocol requires all the nodes need to know their absolute position information, the assumption is hard to achieve for a randomly deployed WSN. Besides, energy efficiency is not designed in the Dual-mode real-time MAC protocol for sensor network.

2.1.3. Message Ordering Based Real-Time MAC Protocol

TOMAC protocol [16] presents hard real-time message ordering at medium access control layer for wireless sensor networks. The hard real-time message ordering mechanism can guarantee the time-order of message delivery in one-hop distance mesh topologies in which the

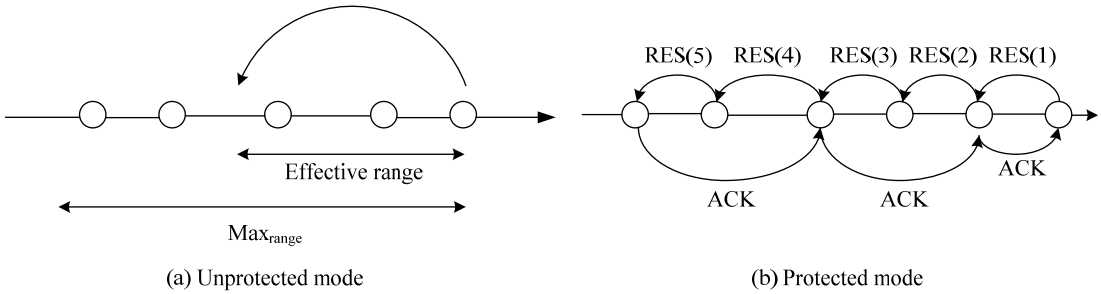


Figure 4. Algorithm of dual-mode real-time MAC protocol.

sensors share the same channel. Ordering messages are based on logical time. For this purpose, each message is assigned a priority, that if a message cannot be transmitted due to the sleep time in a low duty cycle protocol or due to a busy channel, the priority will increase. When the channel is free, the one with the highest priority will gain access. However, TOMAC protocol is difficult to generalize for multi-hop network and other communication topologies.

2.1.4. Traffic Management Based Real-Time MAC Protocol

Supporting components for real time sensors (SUPORTS) [17] base on hard real-time at MAC layer for supporting real-time flows in highly unpredictable sensor network environments. The mechanism is based on a joint traffic regulation and end-to-end scheduling approach. This mechanism attempts to maintain accuracy in a resource-efficient manner even under extremely unstable network conditions where delays are difficult to model and compute.

The goal of SUPORTS is to consider the delay requirement of each arriving packet to maximize the probability of meeting its deadline. SUPORTS implements a least-laxity based scheduler component at each sensor node that determines the order with which each individual packet will be delivered by the MAC service. The protocol computes the laxity value L of a packet as the difference between the deadline and the end-to-end time to transmit the packet from the source to the sink:

$$L = \text{Deadline} - (t_{el} + t_{snk} + D) \quad (1)$$

In the equation mentioned above, t_{el} is the elapsed time since the packet has been initiated at the source and tsnk is the delay that downstream node estimates that will be required until the packet reaches the sink. The D is the local estimation of the projected sojourn time.

In delay sensitive sensor systems the goal of traffic regulation is divided into two parts. First, in cases of congestion, packets need to be dropped to decrease contention and relief overflowing queues in an attempt to reduce delays. Secondly, when a packet is overly delayed, it should not be further forwarded since that would be a waste of transmission energy. Even if the deadline of a packet is large, the packet may still miss its deadline because it might be dropped due to congestion.

2.2. Soft Real-Time MAC Protocol for Wireless Sensor Networks

2.2.1. S-MAC Based Real-Time MAC Protocols

Virtual TDMA for Sensors (VTS) [18] MAC protocol is presented based on soft real-time for WSN applications. VTS protocol is based on sensor-MAC (S-MAC) protocol and provides a Time Division Multiple Access (TDMA) access scheme, in which the number of slots equal to number of nodes in a cell (cluster), the nodes in a cluster will transmit in different time slots. VTS synchronization procedure works as S-MAC, but unlike S-MAC, VTS nodes are only allowed to send data in their captured cycle, a node only sends packets every N_c cycle, N_c is the length of a superframe (as shown in **Figure 5)**. After a number of network setup cycles, the

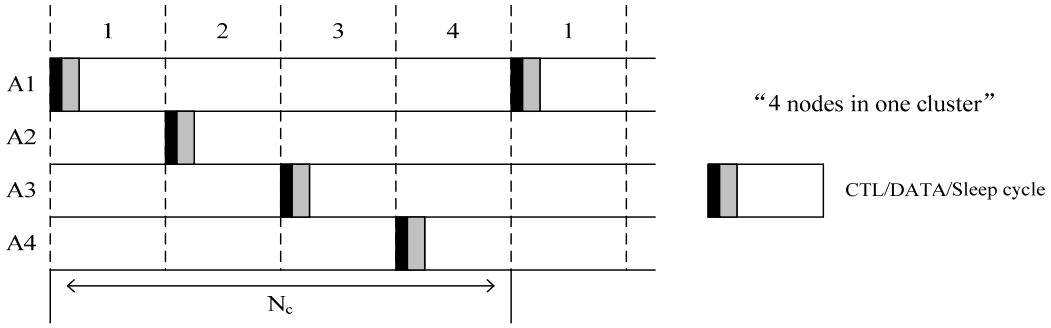


Figure 5. Structure of VTS TDMA frame.

nodes adjust their superframe length counter to their number of known neighbors.

VTS proposes a control packet called CTL as SYNC packet, when all the nodes have sent their first CTL packet, the virtual superframe of N_c timeslots is formed. From then on, the node dynamically adapts to the possibility of nodes joining and leaving the cluster by adjusting the length of superframe. VTS uses the CSMA/CA mechanism for data delivery. At the beginning of each timeslot, all the nodes wake up and listen. The owner of the timeslot performs a carrier sense and broadcasts the CTL. The CTL is used as: synchronization schedule discovery, keep-alive beacon, new node discovery, channel reservation. VTS supports three kinds of transmission: unicast packet transmission, broadcast packet transmission and no data transmission. In the unicast packet transmission, the communication has a sequence as $CTL_{\{RTS\}}$/CTS/DATA/ACK, both nodes go to sleep after the transmission is finished. During the broadcast packet transmission, a $CTL_{\{BCAST\}}$ packet is sent, destination is a broadcast address, without waiting for any CTS reply, sender can send the broadcast packet, after receiving the packet nodes go to sleep and no ACK. When no data transmission, nodes just adjust the clock reference by $CTL_{\{SYNC\}}$ packet and go to sleep.

In addition, VTS proposes to dynamically adjust the duty cycle, VTS uses the control center as synchronizer, the sink node controls the synchronization by CTL packet since it is directly connected to the control center.

Compare with S-MAC, VTS decreases energy consumption and the latency of packet transmission when there are only a few nodes. However, when the number of nodes is higher, the energy consumption is also higher. Additional, since the amount of time-slots, it has a limited packet arrival interval, in some cases VTS is very hard to work for a higher packet generation rate.

A novel real-time MAC layer protocol [19] is designed for soft real-time applications in wireless sensor networks. This protocol bases on S-MAC. The novel protocol uses feedback approach as a medium access mechanism. The novel real-time MAC protocol is for single stream communication.

Working of this protocol is based on use of CC control packet which is used to assign an appropriate value to clear channel flag (CCF) of every sensor node. If CCF equals to 1 the nodes can transmit as well as receive data packets, while it can only receive if its CCF value is 0. Initially all nodes have CCF value as 1. CC control packet has a clear channel counter (CCC), its value ranges from 0 to 3. The value of CCC is 3 at the originating node of CC and decreased by one with on hop transmission of CC. CC is always transmitted from sink to source direction. If value of CCC of CC control packet is 2 or 3 in a node, then CCF of that node will remain 0, if value of CCC of CC is 0 or 1, the CCF of that node will become 1.

Figure 6 explains the novel protocol. As shown in the figure, some data are sent from source node N0 to sink node N9. Duration of one data transfer cycle and one control packet are designated by Tx and Tc respectively. In the first data transfer cycle, RTS/CTS/DATA/ACK will be transmitted by sequence from N0 to N1, after getting ACK, N0 sets its CCF value to 0. From the first duration to fourth duration, the data P0 is forwarded from N0 to N4 and set its own CCF value to 0. Each packet has a Hop Counter (HC) integer variable whose value varies from 0 to 4 for first 4 hops of a communication stream and 0 to 2 for all later 2 hops segments of the communication stream. At N0, the value of HC of P0 is 4 and it is decreased by one each time P0 is transmitted successfully by one hop. Once P0 reaches to N4 node, its HC becomes 0. Then N4 sets HC of P0 to 2. After received ACK from N3, N4 waits for 2Tc duration prior to forwarding P0 to N5. Meantime, in the first and second Tc duration, after receiving ACK, N3 and N2 send CC signal to N2 and N1 respectively. In addition, the CCF of N3 and N2 are set to 0. In the next Tc duration, N1 sends CC signal to N0 and sets its CCF to 1. Thus, after getting CC from N1 node, N0 can transmit new packet P1 to N1 in next one Tx duration. After that, N1 can forward P1 to N2 in next Tx duration and wait there for next CC control packet.

Compare with S-MAC and TMAC by the 99% of duty-cycle, the novel protocol reduces the latency. However, the overhead is higher due to CC control packet. In addition, the novel protocol is difficult to suit to multi-streams communication for WSN.

2.2.2. Spatial Channel Reuse Based Real-Time MAC Protocol

Channel Reuse-based Smallest Latest-start-time First (CR-SLF) [20] algorithm schedules messages at MAC layer for increasing spatial channel reuse in soft real-time

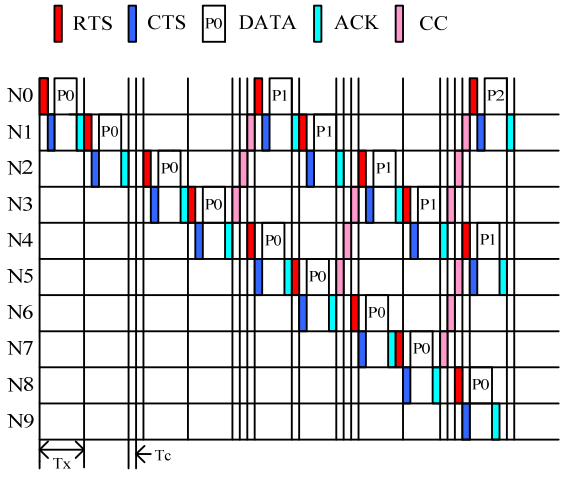


Figure 6. The novel real-time MAC protocol frame format.

multi-hop WSN. This protocol presents an algorithm for mobile wireless sensor network such as a network of mobile robots. The goal of this protocol is to be cognizant of message deadlines at each hop, while avoiding collisions and exploiting spatial reuse. The start time is the time when message is scheduled for transmission and the finish time is the time when the message is completely received by the next hop node. The basic idea is to partition the set of message transmissions into disjoint sets such that transmissions within each set do not interfere with one another and can be executed in parallel.

The algorithm includes three steps. Step 1 selects a transmission to schedule. The scheduler chooses the one with the smallest latest transmission start time (LST), this enables the scheduler to consider the most urgent transmission first. Step 2 assigns this message transmission to a set. The protocol can create n sets: S_1, S_2, …, S_n, the transmissions in different sets are executed in sequence. The scheduler attempts to assign the transmissions to a suitable set in the set list. Step 3 updates the finish time of the feasible set and insert a new transmission for the next hop. If a feasible set S_j is found, then a transmission is inserted into the set, and the new finish time is updated as discussed above.

There is a communication example for describing the algorithm as shown in **Figure 7**. In that communication, m_1 is selected first since the m_1 has the smallest LST. Then m_2 is considered since m_2 has the smallest LST, but m_2 interferes with m1, the scheduler can not be set. By calculating the final schedule is set as: $S_1 = \{m_1, m_3\}$, $S_2 = \{m_2\}$, which means that m_1 and m_3 are transmitted in parallel, followed by the transmission of m_2.

CR-SLF utilizes a centralized scheduling algorithm, in which the centralized scheduler can decide that when and who will transmit or receive messages. Nevertheless, CR-SLF is not scalable as a centralized scheduling algorithm. Moreover, as a wireless sensor protocol, energy consumption is not provided in [20].

2.2.3. Energy Based Real-Time MAC Protocols

Low-power real-time (LPRT) [21] protocol has been proposes at MAC layer for wireless sensing and actuation systems. The LPRT protocol is a hybrid schedule based dynamic TDMA protocol and contention based

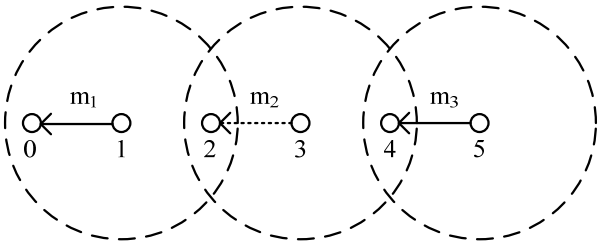


Figure 7. An example for channel reuse in CR-SLF.

CSMA/CA protocol. LPRT considers an infrastructure based star topology, where the stations communicate directly with the base station. If required by the application, the range can be extended with the use of more than one base station, like in a cellular network.

Each superframe of LPRT is divided into a fixed number of mini-slots and starts the transmission by the base station. As shown in **Figure 8**, the superframe includes beacon frame (B), contention period (CP) and contention free period (CFP). The first one is beacon frame, which is followed by the CP. During the CP any station can transmit packets using CSMA/CA protocol. The CFP is allowed to transmit non-real-time asynchronous traffic if it cannot be completed before the beginning of the CFP. The contention free period is placed after the CP. Transmissions during the CFP are determined by the base station using resource grant (RG) information announced previously in the beacon frame of the current superframe. The CFP is composed by an optional retransmission period (RP) and a normal transmission period (NTP), the retransmission procedure helps to increase the reliability of the protocol.

In LPRT, the station decreases the power consumption and coordinates channel. By using the star topology, there is no overhead related with topology discovery and multi-hop communication. However, the application of LPRT is limited, it's very hard to suit the large multi-hop wireless sensor network and other communication topology.

Asynchronous real-time energy-efficient and adaptive MAC (AREA-MAC) protocol [22] is proposed for supporting real-time and energy efficient applications in wireless sensor networks. AREA-MAC based B-MAC, it reduces latency and energy consumption of nodes by

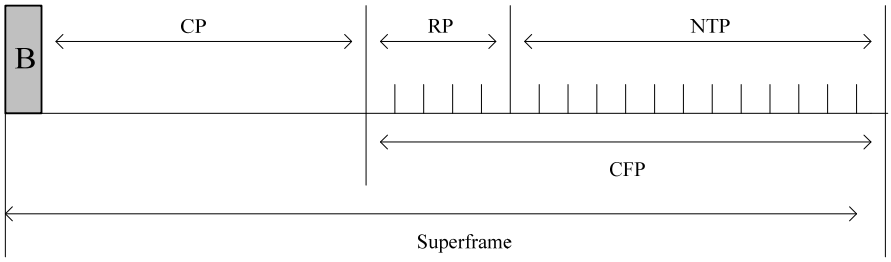


Figure 8. Superframe structure for the LPRT protocol.

using low power listening (LPL) with short preamble messages, where nodes wake up shortly to check the channel activity, if no packet need to receive or forward packets, they will go back to sleep mode immediately.

The main characteristics of AREA-MAC are asynchrony, energy-efficiency, real-time support and adaptability. For real-time data, the source node requests the suitable next-hop neighbor to wake up regardless to its normal schedule for decreasing the delay. Nodes may change their duty cycle according to the real-time request received from their neighbors. The assumption of topology is a grid-based WSN (as shown in **Figure 9**), author assumes that the density of nodes is high enough, so that a node can directly communicate with multiple neighbors, nodes are deployed in an order with the sink node having the highest deployment level. Normal nodes forward data only to up-level direction, *i.e.*, towards sink node.

AREA-MAC considers two types of WSN traffic, one is periodic traffic and the other is non-periodic traffic. For periodic traffic, nodes restrict total energy consumption and send data to 1-level, for non-periodic traffic, nodes restrict delay conditions and sent data to 2-level neighbors. For real-time traffic, the sender directly requests its 2-level neighbor to wake up. It further halves latency and saves more energy at 1-level neighbors. However the nodes need more transmission energy, some nodes will die earlier due to energy exhaustion.

2.3. Real-time MAC Protocol Summary

This is the summary of several real-time MAC protocols which we presented in this section. For showing the performance, **Table 1** is created as a comparison of the real-time MAC protocols.

3. Open Issues in this Research Field

In the last section, we presented several real-time MAC protocols proposed for sensor networks. As a discussion about what we mentioned above, this section consists of open issues of current real-time MAC protocols and the future research directions.

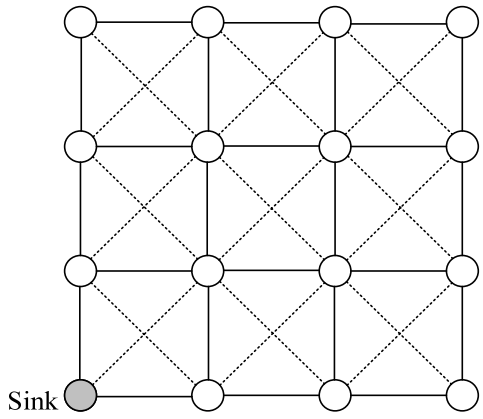


Figure 9. A gird-based topology of WSN in AREA-MAC.

Table 1. Comparing performance of real-time MAC protocols for wireless sensor network.

Protocol	*Energy efficiency*	*Synchrony*	*Timeliness*	*Adaptability*
Dual-mode protocol	No	Yes	Yes	Yes
RRMAC	No	Yes	Yes	No
VTS	Yes	Yes	Yes	No
LPRT	Yes	Yes	Yes	No
AREA-MAC	Yes	No	Yes	Yes
A novel real-time MAC	Yes	Yes	Yes	No
CR-SLF	No	N/A	Yes	No
TOMAC	No	N/A	Yes	No
SUPORTS	No	N/A	Yes	Yes

The keyword of the real-time computing system is deadline. In a real-time control system the packets which are transmitted from source node should arrive at destination before deadline for guaranteeing the timeliness and validity of the alarm messages in the critical and dangerous environments. In order to avoid impact from packets dropping due to miss deadline, packet transmission needs to be scheduled, in other words, the packets need to be allocated priority for transmission sequence since different transmission conditions. For the restricted transmission time, although all of the protocols mention the timeliness which is shown in **Table 1**, most of the real-time MAC protocols for sensor networks design the algorithm without messages transmission sequence by deadline, instead they just design the algorithm for decreasing the packet transmission latency from source to destination node. In those real-time control MAC protocol for sensor network, the great challenge is to often consider that whether it is able to react to external events quickly. According to this interpretation, an algorithm is considered to be real-time if it is low latency. The term low latency, however, has a relative meaning and does not capture the main properties of real-time control systems.

The other important issue is energy. Because the battery-power, wireless sensor networks not have a permanent lifetime for monitoring environment. There is not a higher reliability if the battery of some nodes exhaust power, since the link of communication is interrupted.

Therefore the energy-efficient is an important element for assessing a sensor network real-time MAC protocol whether it adapts monitoring application or not. In wireless sensor networks, there is a more required system than conventional real-time computing system, for prolonging lifetime of wireless sensor networks, the real-time MAC protocols should support an energy-efficient algorithm to increase the reliability of the sensor networks. As a trade-off, it may cause a high latency in return for gaining more energy conservation. However, latency is a very important parameter in real-time transmission system for sensor networks. The trade-off between energy conservation and latency is an obvious challenge for wireless sensor networks.

4. Conclusions and Future Work

This paper has surveyed the real-time medium access control protocols for wireless sensor networks. We have introduced the characteristic of real-time system and discussed the special requirements of wireless sensor network real-time MAC protocols, showed a classification of the research on the current real-time MAC protocols, described architecture of the protocols and discussed the advantage and disadvantage. In addition, we presented the open issues for current sensor networks real-time MAC protocols in the literature.

In summary, most of the existing wireless sensor network real-time MAC protocols focus on decreasing transmission latency, yet still do not adequately consider all of the requirements of sensor networks. In the future, the key challenge which is meeting timing requirement should be guaranteed while establishing a reasonable trade-off and minimizing overhead packets.

5. Acknowledgements

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2010-C1090-1031-0007).

6. References

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

[2] G. Anastasi, M. Conti, M. D. Francesco and A. Passarella, "Energy Conservation in Wireless Sensor Networks: A Survey," *Ad Hoc Networks*, Vol. 7, No. 3, May 2009, pp. 537-568.

[3] K. Kredoand and P. Mohapatra, "Medium Access Control in Wireless Sensor Networks," *Computer Networks*, Vol. 51, No. 4, March 2007, pp. 961-994.

[4] W. Ye, J. Heidemann and D. Estrin, "An Energy-Efficient Mac Protocol for Wireless Sensor Networks," *Proceedings of the* 21*st Annual Joint Conference of the IEEE Computer and Communications Societies*, New York, 23-27 June 2002, pp. 1567-1576.

[5] W. Ye, J. Heidemann and D. Estrin, "Medium Access Control with Coordinated Adaptive Sleeping for Wireless Sensor Networks," *IEEE/ACM Transactions on Networking*, Vol. 12, No. 3, Jun. 2004, pp. 493-506.

[6] T. H. Hsu and J. S. Wu, "An Application-Specific Duty Cycle Adjustment MAC Protocol for Energy Conserving over Wireless Sensor Networks," *Computer Communications*, Vol. 31, No. 17, November 2008, pp. 4081-4088.

[7] P. Lin, C. Qiao and X. Wang, "Medium Access Control with a Dynamic Duty Cycle for Sensor Networks," *Proceedings of IEEE Wireless Communications and Networking Conference*, Atlanta, 21-25 March 2004, pp. 1534-1539.

[8] R. Yadav, S. Varma and N. Malaviya, "Optimized Medium Access Control for Wireless Sensor Network," *International Journal of Computer Science and Network Security*, Vol. 8, No. 2, February 2008, pp. 334-338.

[9] T. van Dam and K. Langendoen, "An Adaptive Energy-Efficient MAC Protocol for Wireless Sensor Networks," *Proceedings of the* 1*st International Conference on Embedded Network Sensor System*, Los Angeles, 5-7 November 2003, pp. 171-180.

[10] S. H. Yang, H. W. Tseng, E. K. Wu and G. H. Chen, "Utilization Based Duty Cycle Tuning MAC Protocol for Wireless Sensor Networks," *Proceedings of Global Telecomminications Conference*, St. Louis, 2 December 2005, pp. 3258-3262.

[11] S. Du, A. K. Saha and D. B. Johnson, "RMAC: A Routing-Enhanced Duty-Cycle MAC Protocol for Wireless Sensor Networks," *Proceedings of* 26*th Annual IEEE Conference on. Computer Communications*, Anchorage, 6-12 May 2007, pp. 1478-1486.

[12] J. Kim and K. H. Park, "An Energy-Efficient, Transport-Controlled MAC Protocol for Wireless Sensor Networks," *Computer Networks*, Vol. 53, No. 11, July 2009, pp. 1879-1902.

[13] J. Kim, J. Lim, C. Pelczar and B. Jang, "RRMAC: A Sensor Network MAC for Real Time and Reliable Packet Transmission," *Proceedings of International Symposium Consumer Electronics*, Vilamoura, 14-16 April 2008, pp. 1-4.

[14] T. Watteyne, I. Augé-Blum and S. Ubéda, "Dual-Mode Real-Time MAC Protocol for Wireless Sensor Networks: A Validation/Simulation Approach," *Proceedings of the* 1*st International Conference on Integrated As Hoc and Sensor Network*, Nice, 30-31 May 2006.

[15] T. Watteyne and I. Augé-Blum, "Proposition of a Hard Real-Time MAC Protocol for Wireless Sensor Networks," *Proceedings of the* 13*th IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication System*, 27-29 September 2005, pp. 533-536.

[16] A. Krohn, M. Beigl, C, Decker and T. Zimmer, "TOMAC-Real-Time Message Ordering in Wireless Sensor Networks Using the MAC Layer," *Proceedings of* 2*nd International Workshop on Networked Sensing Systems*, San Diego, 27-28 June 2005.

[17] K. Karenos and V. Kalogeraki, "Real-Time Traffic Management in Sensor Networks," *Proceedings of IEEE International Real-Time System Symposium*, Rio de Janeiro, 5-8 December 2006, pp. 422-434.

[18] E. Egea-López, J. Vales-Alonso, A. S. Martínez-Sala, J. García-Haro, P. Pavón-Mariño and M. V. B. Delgado, "A Wireless Sensor Networks MAC Protocol for Real- Time Applications," *Personal and Ubiquitous Computing*, Vol. 12, No. 2, February 2008, pp. 111-122.

[19] B. K. Singh and K. E. Tepe, "A Novel Real-Time MAC Layer Protocol for Wireless Sensor Network Applications," *Proceedings of IEEE International Conference on Electro/Information Technogy*, Windsor, 7-9 June 2009, pp. 338-343.

[20] H. Li, P. Shenoy and K. Ramamritham, "Scheduling Messages with Deadlines in Multi-Hop Real-Time Sensor Networks," *Proceedings of IEEE Real Time and Embedded Technology and Applications Symposium*, 7-10 March 2005, pp. 415-425.

[21] J. A. Afonso, L. A. Rocha, H. R. Silva and J. H. Correia, "MAC Protocol for Low-Power Real-Time Wireless Sensing and Actuation," *Proceedings of the* 11*th IEEE International Conference on Electronics*, *Circuits and Systems*, Nice, 10-13 December 2006, pp. 1248-1251.

[22] P. Kumar, M. Gunes, Q. Mushtaq and B. Blywis, "A Real-Time and Energy-Efficient MAC Protocol for Wireless Sensor Networks," *Proceedings of the* 6*th IEEE and IFIP International Conference on Wireless and Optical Communications Networks*, 28-30 April 2009, pp. 1-5.

Communications and Network, **2010, 2, 113-124**
doi:10.4236/cn.2010.22018 Published Online May 2010 (http://www.SciRP.org/journal/cn)

PBB Efficiency Evaluation via Colored Petri Net Models

Peter Vorobiyenko, Kirill Guliaiev, Dmitry Zaitsev, Tatiana Shmeleva
Odessa National Academy of Telecommunications, *Odessa*, *Ukraine*
E-mail: *vorobiyenko@onat.edu.ua*, *k.guliaiev@gmail.com*, *zsoftua@yahoo.com*, *tishtri@rambler.ru*
Received January 6, 2010; *revised February* 5, 2010; *accepted March* 1, 2010

Abstract

Basic components of Provider Backbone Bridge (PBB) network models were constructed: PBB interior switch, PBB edge switch—with the dynamic filling up of address tables. The modeling of PBB networks was implemented. The results of simulation reveal definite imperfections of PBB technology caused by the broadcasting and sensitivity to the ageing time of the address tables' records, which complicates the guaranteeing of a given QoS. The preliminary comparison confirms definite advantages of E6 addressing before PBB.

Keywords: PBB, E6, Colored Petri Net, Simulation, QoS

1. Introduction

Technology of Provider Backbone Bridge (PBB) [1] is aimed to the construction of networks entirely on the base of Ethernet. Corresponding standard IEEE 802.1ah [2], which development was started in 2005 has been prepared in a draft variant, while companies have already started the issuing of PBB switches and providers have started their exploiting. British Telecom has chosen switch-router Nortel Metro 8600 and Metro Ethernet Services Unit of 1850 series as Ethernet components of network in the project 21CN («Network of 21 Century»). The delivery of superproductive PBB switches Black-Diamond 20808 of Extreme Networks Company has been started to Russia.

With the advent of 1 Gbps and 10 Gbps Ethernet standards, new opportunities of Ethernet technology mass employment in provider backbone networks have been opened, but 802.3, 802.1D technology has a series of disadvantages regarding scalability, quality of service, manageability, which a new series of IEEE standards were developed to overcome: 802.1Q—virtual networks, 802.1QinQ—multilevel virtual networks, 802.1ad—provider bridges, 802.1ah—backbone provider bridges, 802.1ag—networks management, 802.1Qay–traffic engineering. Mentioned standards provide the Carrier Ethernet concept to substitute SDH, as well as IP-MPLS solutions in backbone networks, though IETF undertakes active attempts of MPLS and PBB standards integration in virtual private service VPLS [3].

In [4,5], an alternative solution was suggested for Ethernet scalability under encapsulation IP-Ethernet via uniform network hierarchical addresses E6, which are situated into MAC-addresses fields of Ethernet frames. While PBB supposes the enlargement of frame header length adding backbone switches MAC-addresses pairs, E6 has definite advantages because of annulment TCP, UDP, IP protocols, as well as corresponding packet headers and address mapping protocols ARP/RARP. Models of E6 networks presented in [6], creates the basis for comparison of two technologies. However, full-fledged comparative analysis is possible under the construction of rather detailed PBB network models and modeling of IP-Ethernet encapsulation processes.

The purpose of the present work is PBB networks basic components construction in the form of colored Petri nets in the environment of simulating system CPN Tools [7], as well as an evaluation of PBB technology efficiency via PBB networks modeling.

2. The PBB Technology Overview

IEEE 802.1ah frame [1,2] encapsulates IEEE 802.1QinQ and IEEE 802.3 frames. IEEE 802.1ah frame header (**Figure 1**, **Table 1**) contains C-MAC—customer addresses (C-DA, C-SA) and B-MAC—backbone addresses (B-DA, B-SA). Moreover, the recurring encapsulation of PBB frames is stipulated for multilevel backbone networks creation.

Abstracting from header fields of virtual networks, let us consider the interaction of address fields on the example of network shown in **Figure 2**. Let host X with MAC-address AX send a frame to host Y with MAC-address AY. The corresponding 802.3 (802.1ad) frame is created with C-DA = AY, C-SA = AX. The frame is delivered to the nearest PBB edge switch PBBX with MAC-

B-DA	B-SA	B-Tag	I-Tag	C-DA	C-SA	S-Tag	C-Tag	Data	FCS

Figure 1. Format of IEEE 802.1ah frame header.

Table 1. Description of IEEE 802.1ah frame header fields.

Notation	Description
B-DA	Backbone destination address
B-SA	Backbone source address
B-Tag	Backbone VLAN tag
I-Tag	Service instance tag
C-DA	Customer destination address
C-SA	Customer source address
S-Tag	Service provider VLAN tag
C-Tag	Customer VLAN tag
Data	Data
FCS	Frame check sequence

address ABX. By the destination address AY (using address tables), switch PBBX determines address ABY of PBB backbone switch PBBY, which the network containing Y is attached to. PBBX encapsulates 802.3 frame into 802.1ah frame with B-DA = ABY, B-CA = ABX and sends the frame into backbone. PBB backbone switches use only the pair of addresses B-DA, B-SA for the delivery of the frame to PBB edge switch PBBY. At the frame receiving, PBBY extracts the encapsulated 802.3 (802.1ad) frame and implements the frame delivery to host Y using the pair of addresses C-DA, C-SA.

Passive listening is used for the filling in address tables. If the destination address is unknown then broadcasting is implemented. Tree-like network is represented in **Figure 2**; the standard implies modified spanning tree algorithms application for the work on non tree-like topology.

The advantage of PBB technology is the backbone performance increase due to considerable reduction of address table size for PBB interior switches, which contain B-MAC addresses only. But the functioning of PBB edge switches, which implement the mapping of C-MAC addresses into B-MAC addresses and encapsulation of frames, is getting complicated. Usual 802.1D switches work at the network periphery.

3. Scope of the Model

Simulating system CPN Tools [7] was chosen for the construction of models; it was developed in Aarhus University, Denmark and uses the language of colored Petri nets [8] for models description. In Odessa National Academy of Telecommunications named after A. S. Popov, the library of the model components was created for Ethernet, IP, MPLS, Bluetooth, E6 networks as well as the library of measuring fragments for the networks performance and QoS evaluation [6,9-12].

In the present work at the PBB technology modeling, only address part of the frame header without virtual network tags was taken into consideration; moreover, only one level of provider bridges hierarchy and tree-like structure of the network were considered. Modeling of virtual networks tags, multilevel hierarchy and spanning tree algorithms of provider bridges are the directions for future work. Moreover, the investigation of fully connected network structures (without the division into virtual private networks) similar to the Internet relates to the purpose of consequent comparative analysis of PBB and E6 regarding advantages of usage in world-wide networks.

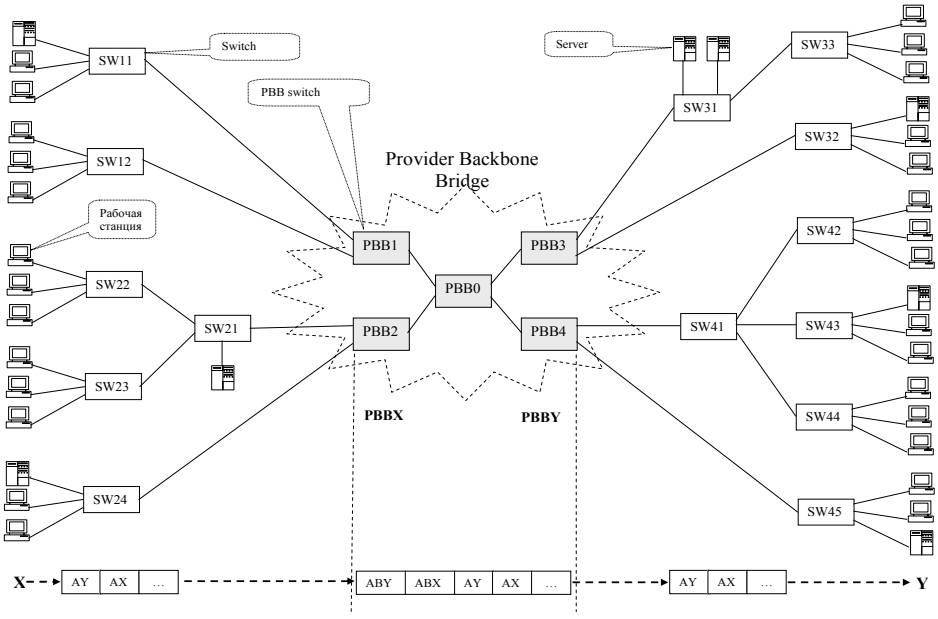


Figure 2. An example of PBB network.

The following components were constructed for the modeling of PBB networks:

—model of interior PBB switch SWB*m*;

—model of edge PBB switch SWB*m*-*n*;

—model of 802.1D (traditional) switch SW*n*;

—models of terminal (subscriber) equipment: WS—workstation, MWS—measuring workstation, S—server.

Variables *m*, *n* denotes the quantities of B-ports and C-ports correspondingly.

At the SW*n* switch model construction (model of port-port), as the base was taken the model [9] with dynamic running of switching tables modified regarding only microsegmented Ethernet usage and running separate queues of frames on ports. Model of interior PBB switch SWB*m* (model of port-PBBport) has distinctions regarding the backbone MAC-addresses processing. The model of edge PBB switch SWB*m*-*n* (models of ports-cport, bport) is the most sophisticated as it provides the mapping of customer C-MAC addresses into backbone B-MAC addresses as well as the broadcasting of two corresponding kinds.

At the traffic modeling, the concept of client-server interaction was used and corresponding components [9,10], which were supplied with counters for the useful and broadcasting traffic estimation as well as various kinds of random functions distribution lows. Note that at the large scale backbones modeling, it is advisable to use the flow traffic models [6] to abstract from the detailed description of the network periphery.

The components were used for the network (**Figure 1**) model construction and analysis; the corresponding main page of the model was named Network. Measuring workstations MWS [9,10] provide the evaluation of network response time; counters represented by fused places shifted to the main page of the model provide the evaluation of useful and broadcasting traffic.

4. Network Model

The network model shown in **Figure 2** is represented by the main page of the model Network in **Figure 3** and by the models of used components in **Figures 5-9**. The main page is constructed on the base of network structure scheme direct mapping principle. One 4-ports interior PBB switch PBB0 of the kind SWB4 and four 3-ports

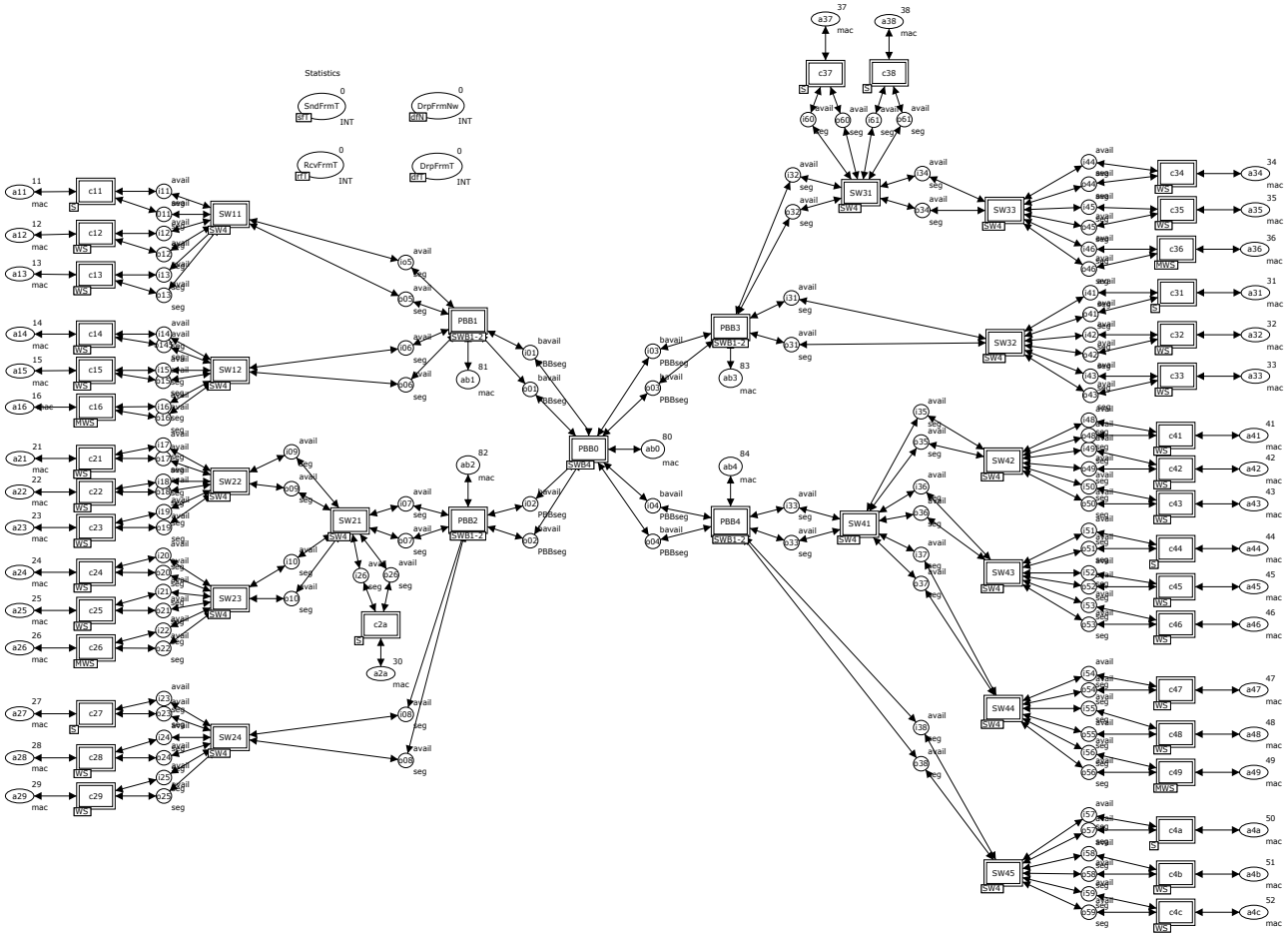


Figure 3. Model of network (Network).

edge switches PBB1-PBB4 of the kind SWB1-2 with one B-port and two C-ports were used for modeling of PBB backbone. 14 4-ports switches of the kind SW4 were used for the periphery networks modeling. The model of terminal (subscriber) equipment is represented by 24 worktations WS, 4 measuring workstations MWS and 8 servers S.

The descriptions of basic data types (color sets) and model functions are represented in **Figure 4**. In the main page of the model the following types of places were used: mac—MAC-address, seg—usual segment, PBBseg—PBB segment; as the indicator of segments availability the constants avail, bavail for usual and PBB segments were used correspondingly.

In the main page of the model, C-MAC addresses of customer equipment are pointed out in places a11-a4c and B-MAC addresses of PBB-switches—in places ab0-ab5; MAC addresses are represented by integer numbers, which does not bound the generality if for instance only the last byte of MAC-addresses coinciding in the first 5 bytes is considered.

Each port of switches is represented by the pair of places modeling the full-duplex mode of work. Place i*k*—input channel of k-th port; place o*k*—output channel.

```
colset mac = INT;
colset mact = mac timed;
colset frm = product mac * mac * nfrm timed;
colset PBBfrm = product mac * mac * mac * mac * nfrm timed;
colset seg = union f:frm + avail timed;
colset PBBseg = union b:PBBfrm + bavail timed;
colset swi = product mac * portnum;
colset swita=list swi;
colset PBBswi = product mac * mac * portnum;
colset PBBswita = list PBBswi;
colset qfrm = list frm;
colset pqfrm = product portnum *qfrm;
colset xfrm = union cf:frm + bf:PBBfrm;
colset qxfrm = list xfrm;
colset pqxfrm = product portnum *qxfrm;
fun eqa a (rr:swi)=((#1 rr)=a);
fun eqaB a (rr:PBBswi)=((#1 rr)=a);
fun eqbaB a (rr:PBBswi)=((#2 rr)=a);
fun grec prd [] = (0,0) | grec prd (q::r) = if prd(q) then q else grec
prd r;
fun xrec prd [] = [] | xrec prd (q::r) = if prd(q) then r else q::(xrec
prd r);
fun grecB prd [] = (0,0,0) | grecB prd (q::r) = if prd(q) then q else
grecB prd r;
fun xrecB prd [] = [] | xrecB prd (q::r) = if prd(q) then r else
q::(xrecB prd r);
fun Delay() = poisson( Delta );
fun Dexec() = poisson( dex );
fun Nsend() = poisson( nse );
fun cT()=IntInf.toInt(!CPN'Time.model_time)
val TCL=100000000;
```

Figure 4. Description of basic data types and functions.

The connection of the equipment according to the network structure scheme is implemented by the fusion (merging) of input and output places of ports. Note that at the two switches connecting, the input channel of one is fused with the output channel of the other one and vice versa; names of ports places are chosen regarding the switch situated more closely to PBB0.

Moreover, the fused places are shifted to the main page of the model for the traffic evaluation: SndFrmT of the kind sfT—counter of sent frames, RcvFrmT of the kind rfT—counter of received frames, DrpFrmNw of the kind dfN—counter of frames dropped in network, DrpFrmT of the kind dfT—counter of frames dropped in terminal equipment.

5. Networking Equipment Models

The networking equipment models are represented by the early mentioned switches models of thee different kinds: SW4, SWB4, SWB1-2. The switch model layout has minor peculiarities and is considered on the example of 802.1D (traditional) switch SW4 shown in **Figure 5**. The model is assembled by cloning required number of ports models port.

Each port is identified by unique number (myport*). The shared data stored in switch memory are used for the interaction of ports. As the most simple, the store-and-forward architecture with obligatory buffering of frames is considered. The frame arrived to the input channel of port A, is stored in the buffer Buf; meanwhile, using the switching table SwT the number of the port B is determined for the frame forwarding. If the frame destination address is not mentioned in the table then the switch implements the broadcasting—the frame is forwarded to all the ports of the switch save the port A. The output channel of the port B extracts the frame out of the buffer and transmits it into corresponding segment. The total quantity of switch ports nport and the number of own port myport* are used in the broadcasting. Place timer contains MAC-addresses, pointed in the table SwT together with time stamps (data type mact); transition ClrSwTa provides the erasing of the corresponding record of the table SwT after the elapsing of ageing time interval (constant TCL); recursive function xrec implements the erasing of the record out of table represented by variable x; function eqa implements the addresses comparison. Port implements the passive listening of traffic (sender addresses) with the aim of filling in the table with new records.

Note that in the buffer Buf of the type qpfrm, separated FIFO queues of frames on switch ports are organized according to the standards. The initial marking creates 4 empty queues (lists); the header of a queue is equal to the number of corresponding port.

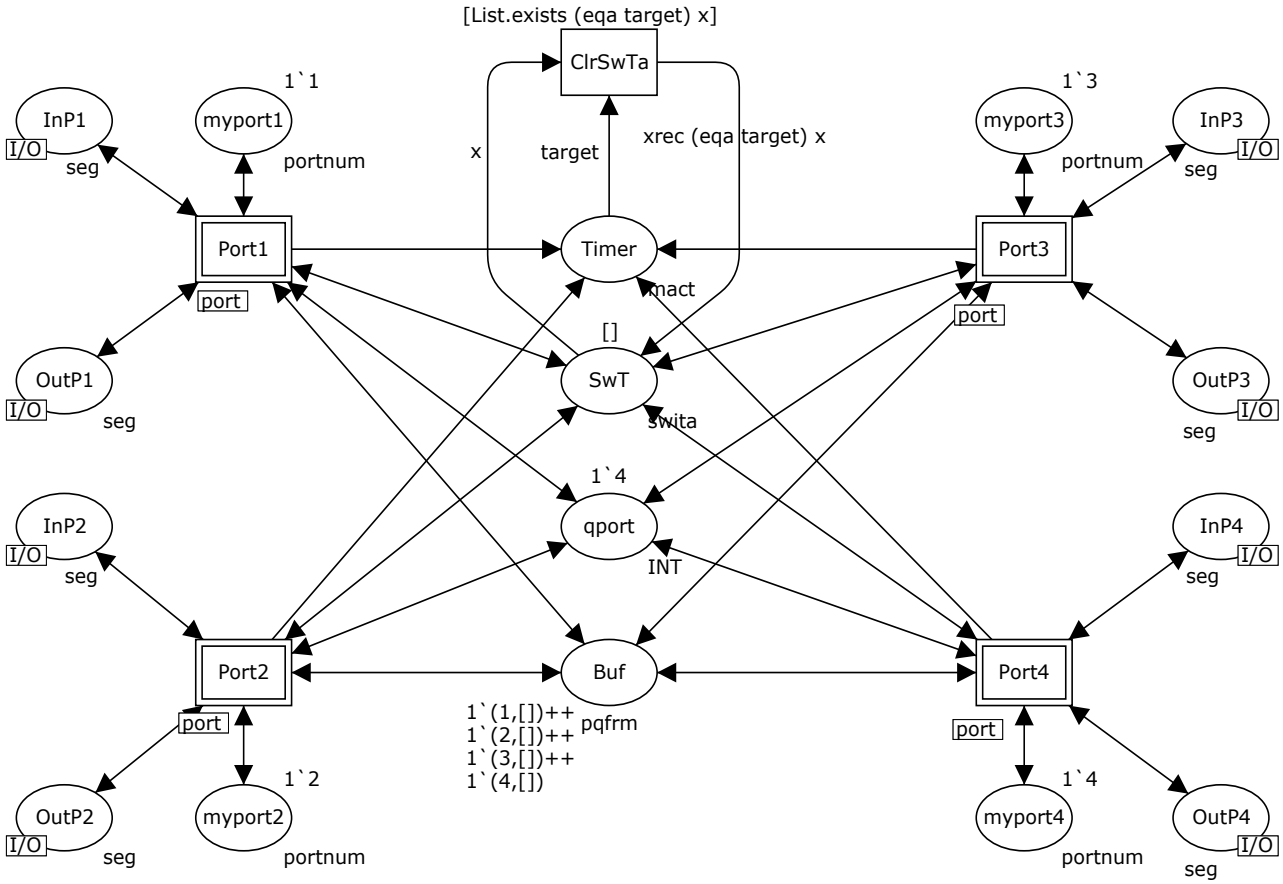


Figure 5. Model of switch 802.1D (SW4).

5.1. Switch Port Model

The model of 802.1D switch port is represented in **Figure 6**. The type seg is used for the description of Ethernet segment channel, which can be either free (the constant avail) or busy with the frame f transmission. The type frm is used for the frame f description consisting of sender address src, destination address dst and frame number nf (the description abstracts from the frame content). The type swita describes the switching table as a list of records swi consisting of destination address dst and port number pnum. Recursive function grec extracts record from switching table. The type pqfrm of the buffer Buf describes queues qfrm enumerated on ports of frames. Input and output arcs inscriptions, which will be described further, realize FIFO discipline of queues.

A frame arrives to input channel of port PortIn; its sender address src can be either new (transition NewSrc) or known (transition OldSrc). Transition NewSrc fills in the switching table with a new record, which contains sender address src and port number m (current port of switch). The frame is put into auxiliary place Aux1, then the frame destination address dst is analyzed, which can be either a new (transition NewDst) or known (transition OldDst). Transition OldDst puts the frame into the buffer Buf, the queue corresponds to the determined output port number. Transition NewDst puts the frame into auxiliary place Aux2 and starts the broadcasting process. Place pnum is used for sequential (i: = i + 1) numeration of broadcasting ports. Transition BroadC implements the broadcasting until the numbers of all the ports are exhausted ($i \le q$); then at ($i > q$) the transition clean is started, which cleans auxiliary places and returns the availability indicator avail into segment. The condition (i<>m) in the inscription of broadcasting arc BroadC→ Buf excludes the broadcasting into own port.

Output channel of a port extracts frames from the buffer Buf redirected into the current port (m) and transmits them into segment via transition Out, which waits and deletes the indicator of segment availability avail.

Let us consider the work with queues of frames in the buffer Buf more closely. The frame insertion is implemented into the tail of port i queue; for that the corresponding queue is extracted from place Buf via arc inscription (i,qu), then the frame f is inserted into the tail of the queue via arc inscription (i,qu^^[f]). The frame extraction is implemented from the head of port m queue;

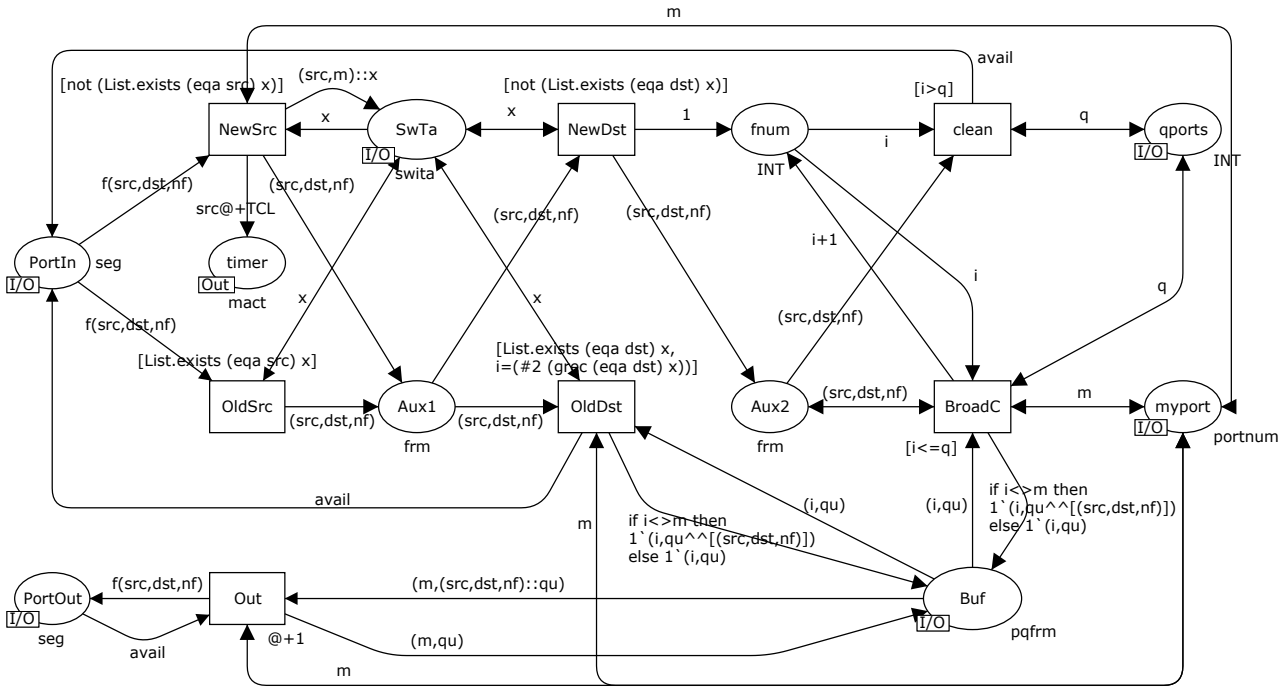


Figure 6. Model of 802.1D switch port (port).

for that the queue is extracted from place Buf with first frame separated via arc inscription (m,f::qu), then the queue without the first frame is returned via arc inscription (m,qu). Operation ^^ implements the queues concatenation; operation :: separates the header element.

5.2. PBB Interior Switch Port Model

Port model PBBport of PBB interior switch is represented in **Figure 7**. The work of PBB interior switch port is in many ways very similar to the work of a usual switch port shown in **Figure 6** with the only difference that B-MAC addresses of 802.1ah frame are used instead of C-MAC addresses.

The type PBBseg is used for the description of PBB segment channel, which can be either free (constant bavail) or busy with transmission of frame b. The type PBBfrm is used for the description of frame b consisting of backbone sender address bsrc, backbone destination address bdst and encapsulated 802.3 frame frm. Switching table contains backbone addresses B-MAC only. The pqxfrm type of the buffer Buf describes queues of frames xfrm for ports; the type xfrm is represented by the union of frames cf of the type frm or frames bf of the type bfrm. PBB interior switch processes bfrm frames only; the possibility of a frame union is used in the models of edge switches.

The main differences in the work of the port are connected with the PBBfrm frame type processing instead of frm and bsrc, bdst addresses usage instead of src, dst addresses correspondingly. Moreover, transition NotMYbdst determines that frame is not addressed to the current switch while transition MYbdst models the processing of (subservient) frame addressed to the current switch via absorption of the frame and increment of counter dfn into place dfPBB.

5.3. PBB Edge Switches Ports Models

The basic difference of PBB edge switch ports consists in the processing both C-MAC addresses and B-MAC addresses and running switching tables, which provide the mapping of C-MAC addresses into B-MAC addresses. Moreover, the broadcasting of two kinds is implemented: on C-ports and on B-ports transmitting frames of different types.

For the description of queues elements of the internal buffer, the union data type xfrm is used, which can store either frame cf of the type frm or frame bf of the type PBBfrm. For flat relative representation of multilevel switching tables and addresses mapping, the data type PBBswi is used, which contains destination address dst, backbone destination address bdst and port number pnum. The field bdst duplication in several records has the advantage of the quick search of complete information by the key dst.

5.3.1. C-port

Model of C-port cport of PBB edge switch is represented in **Figure 8**. The basic difference from ports port, PBBport

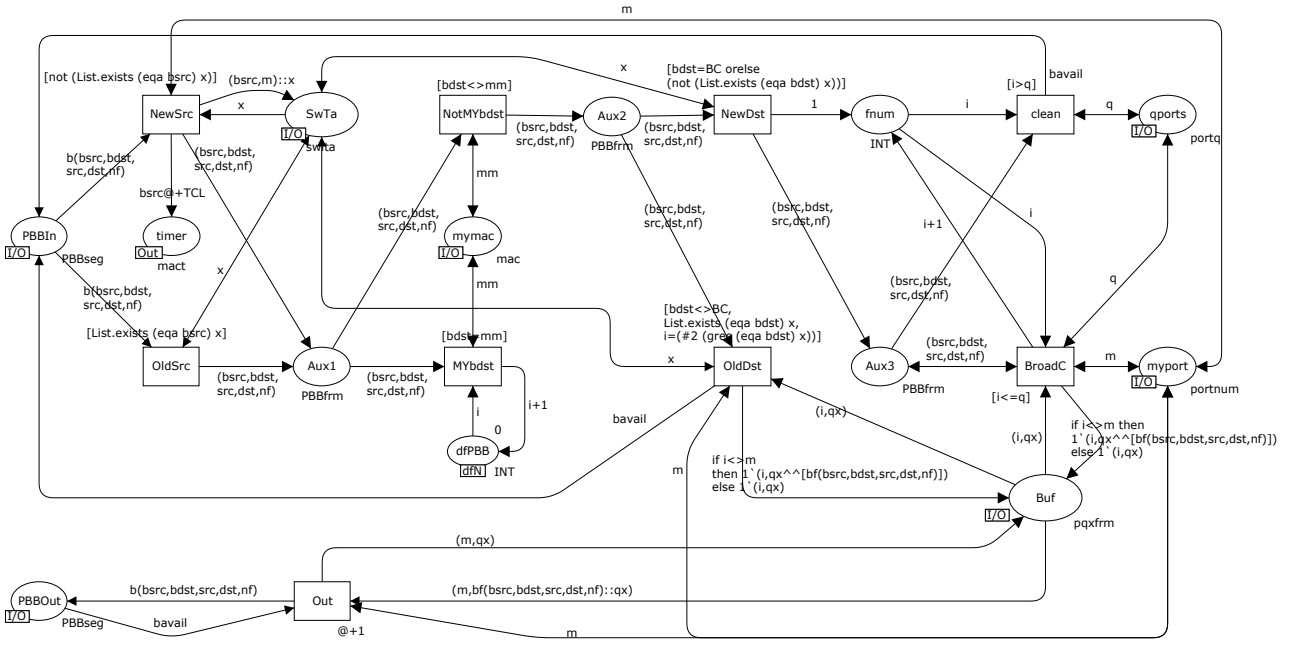


Figure 7. Model of PBB interior switch port (PBBport).

(**Figures 6** and **7**) consists in the refilling of switching table with records containing besides src backbone address bsrc, which coincides with own B-MAC address. Moreover, the function of recursive search grecB is modified as well as the function eqaB of addresses comparison for the records of the type PBBswi processing; the second (#2) and the third (#3) fields of the table PBBswT are employed; an additional checking known destination address on the belonging to own network is implemented via transitions OldDstMy, OldDstNotMy. If a known destination address dst belongs to own network (transition OldDstMy), then the record cf is formed, which is forwarded to C-port. If a known destination address dst does not belong to own network (transition OldDstNotMy), then the record bf is formed, which is forwarded to B-port, while not only the destination port number (#3(grecB(eqaBdst)y)) but also backbone destination address (#2(grecB (eqaBdst)y)) are determined from the table. The broadcasting (transition BroadC) distinguishes ports via place mPBBp, which stores the number of the first B-port; ports are enumerated sequentially: at first all the C-ports, then all the B-ports. That is why the condition I < pbp pointed in the inscription of the arc BroadC- > Buf, separates C-ports only and its alternative (then)—B-ports; depending on this, either cf or bf record is put into the buffer correspondingly. For the broadcasting into backbone a constant BC equaling to 255 is used as the backbone destination address bdst.

5.3.2. B-port

Model of PBB edge switch port bport represented in **Figure 9** is the most complicated that is why it is considered closely. Passive listening (transitions NewSrc, OldSrc) fills in the table SwTa with records containing both customer src and backbone bdst addresses from the current frame. Then the backbone destination address bdst is analyzed; the alternatives are represented by the following transitions: bdstMy—own bdst of the current switch, bdstBC—broadcasting bdst, bdstNotMy—address bdst of some other PBB switch. Then the search in the table is implemented: the key is the destination address dst for own bdst (transitions dstNew, dstOld), the key is bdst address for foreign bdst (transitions bdstNew, bdstOld). In the both cases at the successful search completed, the frame is forwarded (transitions bdstOld, dstOld) into the buffer; in the first case (transition bdstOld)—without changes, in the second case (transition dstOld)—encapsulated 802.3 frame is extracted. The other conditions (transitions dstNew, bdst BC, bdstNew) lead to the start of broadcasting (transition BroadC). Thus, the set of five alternatives is formed: bdstMy& dstNew, bdstMy& dstOld, bdstBC, bdst-NotMy&bdst-New, bdstNotMy& bdstOld. Moreover, the case of possible error is processed separately (transition wrong): the frame is addressed to the current switch (bdstMy), address dst is contained in the table but the corresponding record of the table contains bdst which distinguishes from the address of the current switch.

For the correct forming of the broadcasting, additional Boolean indicators are used in the following places: BCcport—broadcasting on C-ports; BCbport—broadcasting on B-ports. Each of three alternatives of the broadcasting forms its own set of the indicators:

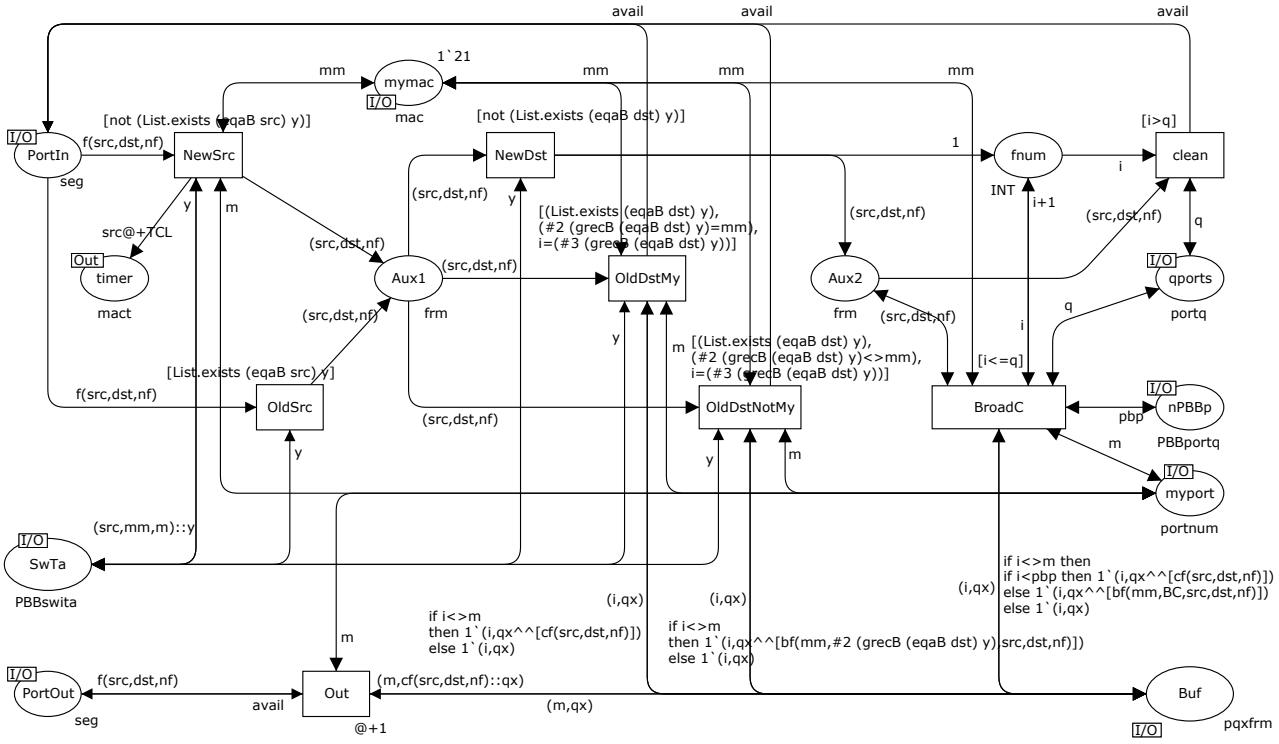


Figure 8. Model of PBB edge switch C-port (cport).

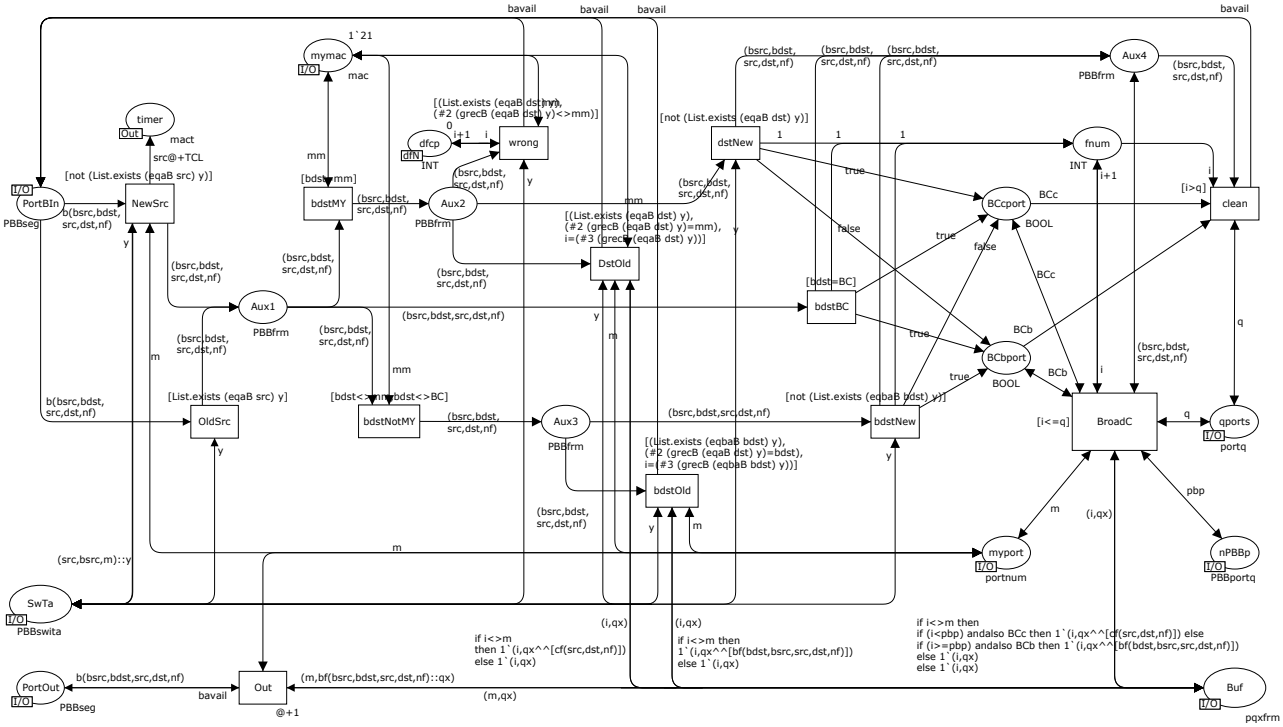


Figure 9. Model of PBB edge switch B-port (bport).

bsdtMy&dstNew—BCcport = true & BCbport = false, bdstBC—BCcport = true & BCbport = true, bdstNotMy &bdstNew–BCcport = false & BCbport = true. Depending on the combination of indicators the inscription of the arc BroadC- > Buf forms broadcasting frames forwarded to C-ports and B-ports.

Note that as the elements of the buffer queues are represented by the union, the output channels of C-ports (transition Out) extracts records cf from the buffer, B-port—records bf and transmit the corresponding frames into the corresponding segments (either of the type seg or PBBseg).

6. Terminal Equipment Models

In the present work, the models of the terminal equipment represented in [9,10] are used. Workstation WS generates queries to servers periodically; the distribution of time between queries is given by the random function Delay(). Server S executes the query of workstation and returns a random number of the reply frames; the distribution of frames number is given by the random function Nsend(); the distribution of query processing time is given by the random function Dexec(). Results of various laws of random values distribution were analyzed: uniform, Poisson, Erlang.

Into the models of the terminal equipment, counters were added represented by the following fused places: SndFrmT of the kind sfT—the counter of sent frames, RcvFrmT of the kind rfT—the counter of received frames, DrpFrmT of the kind dfT—the counter of dropped frames. For the convenience of the model characteristics evaluation, the counters are shifted to the main page (Network). Moreover, measuring workstations MWS implement the evaluation of the network response time directly during the simulation process [9,10].

7. Simulation Results Analysis

Primarily, the separate components debugging was implemented, then the complex debugging of the network model; the tracing of separate frames delivery processes and the filling up of address tables was implemented. Using additional counters it was ascertained that all the sent frames are delivered to their destinations. Dynamically filled up address tables completely correspond to the structural scheme of the network.

The model time unit (MTU) is equal to 1.2 ms, it corresponds to the frame transmission time into 10 Gbps segment. The time of equipment components work was not modeled since it requires nanosecond time scale which complicates the evaluation of the network functioning on prolonged time intervals.

The model measuring fragments are represented by the measuring workstations MWS for the evaluation of the network response time and the frames counters for the evaluation of the performance and the effective performance. The source of overhead in Ethernet technology including PBB is the broadcasting usage as well as the resources spending on the spanning trees construction. The present model allows the estimation of the performance portion spent on the broadcasting only; the evaluation of the spanning tree algorithms work was not implemented.

The counter DrpFrmT of the kind dfT in the main page of the model contains the total number of the broadcasting frames delivered to the terminal equipment; the counter RcvFrmT of the kind rfT—the total number of the delivered useful frames. At the stopping of the model under expired time condition, the content of RcvFrmT is lesser than SndFrmT—the counter of sent frames, which is caused by definite number of frames are in the process of their delivery into the network; but at the generating given number of frames and the model stopping under the absence of events condition, the values of the both counters are equal.

The effective performance of network depends considerably on the records ageing time TCL of the address tables. Moreover, at the network switching on (new subnets connection), a short-term overload is produced because of intensive broadcasting, which leads to temporary decrease of QoS (the network response time). The dynamics of the broadcasting and the response time (as a quality of service characteristic) are shown in **Figure 10**. In the graphs, a splash of broadcasting at the network switching on and its influence on the response time are shown as well as the second wave of the broadcasting after the cleaning of the address tables (after 12 s.), which influence is smoothed on time.

The dependencies of the effective performance and the response time on the records ageing time are shown in **Figure 11**. The increase of the records ageing time leads to the improvement of the network performance as well as QoS but it decreases the capabilities of the adaptation to changes of structure and leads to incorrect frames delivery as the result of irrelevant address table records usage.

The influence of the records ageing time is getting stronger at the traffic intensity increase according to evaluations represented in **Figure 12**. The effective performance of the network is increased at traffic intensity increase (though it leads to the QoS deterioration), which is caused by the increase of frequency of address tables records usage.

Thus, in spite of the considerable advantage, PBB technology possesses the definite imperfections, which complicate the guaranteeing of given QoS and require the performance reservation for the possible overloads smoothing. Note that in the present work the virtual networks, which allow the broadcasting traffic isolation (within the bounds of a virtual network), were not considered.

The preliminary comparative analysis of two technologies is carried out on the base of the models constructed in the present work and the models of E6 networks [6]. The absence of service broadcasting in E6 allows the guaranteed QoS. Hierarchical structure of E6

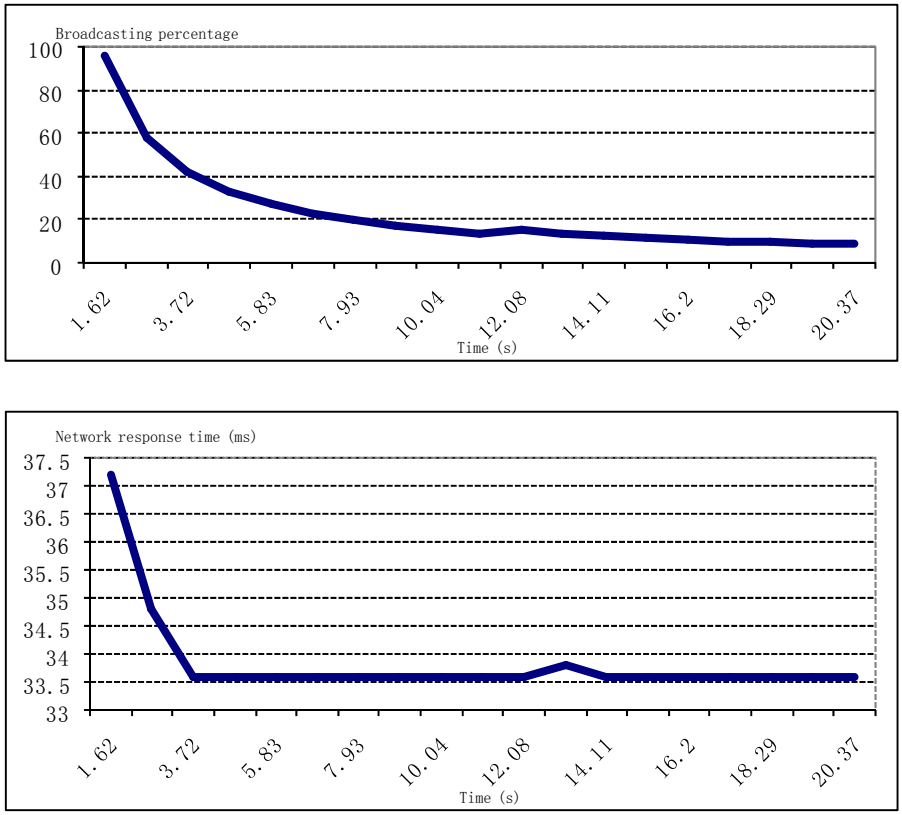


Figure 10. Dynamics of broadcasting and QoS after the switching on.

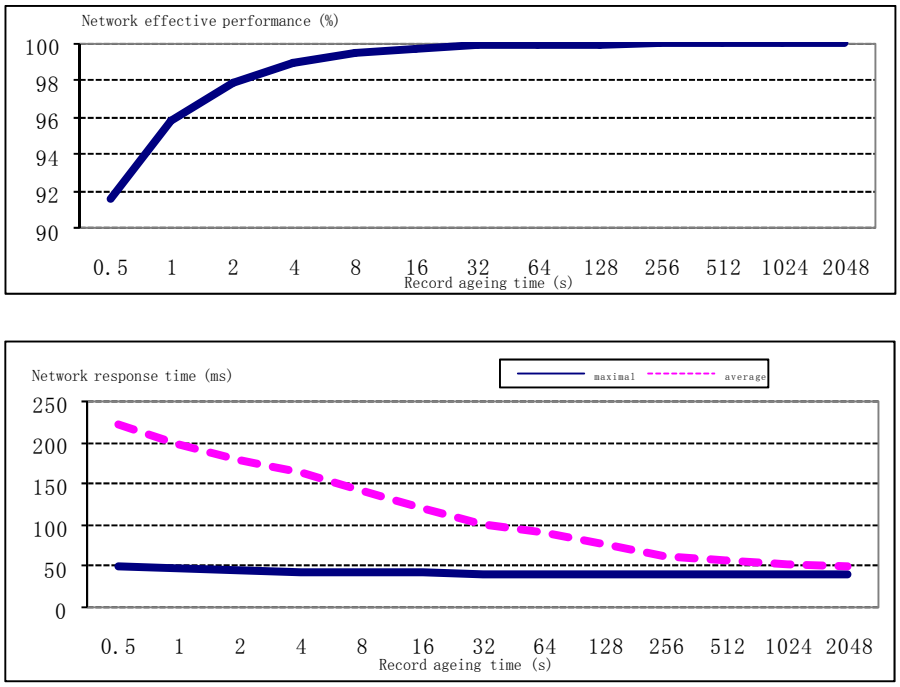


Figure 11. The influence of the record ageing time on the performance and QoS

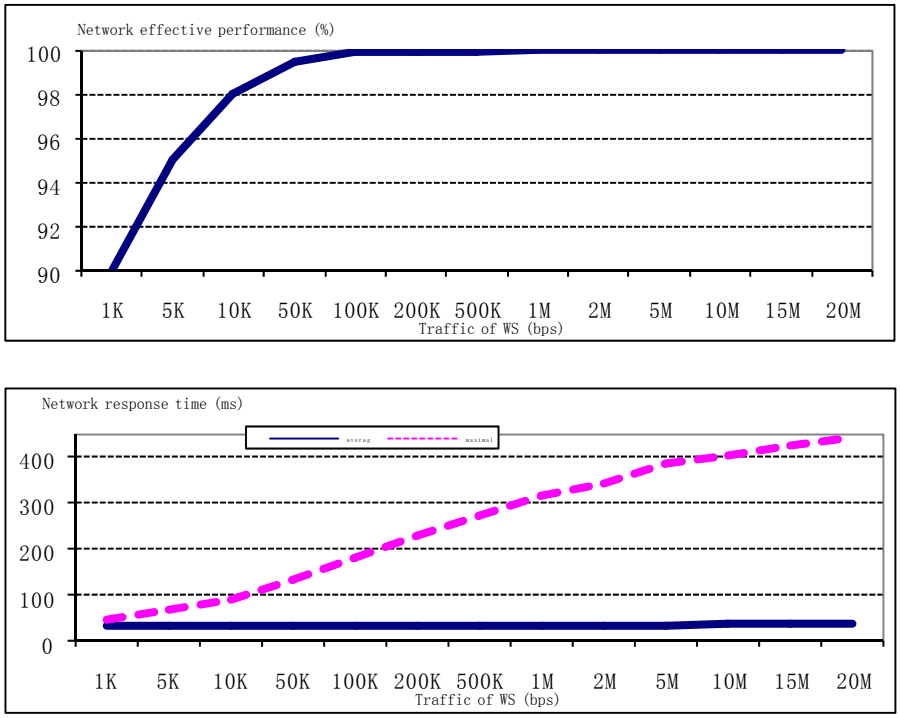


Figure 12. The influence of the traffic intensity on the performance and QoS.

network (including the backbone) is given by the hierarchical E6 addresses and does not lead to the lengthening of the frame header; in PBB each new level of backbone hierarchy requires not less than 12 additional bytes of the frame header (for the pair of B-MAC addresses). Moreover, stack E6 annuls 20-36 bytes of protocols TCP, IP headers for each encapsulated packet. At the encapsulation IP-Ethernet, PBB technology stipulates the duplicate mapping of addresses: IP- > C-MAC, C-MAC- > B-MAC, while E6 completely annuls the address mapping within network due to uniform E6 address usage.

8. Conclusions

In the present work the basic components of PBB network models were constructed: PBB interior switch, PBB edge switch—with dynamic filling up of address tables. The modeling of PBB network functioning was implemented. The analysis of simulation results allows the conclusion regarding definite imperfections of PBB technology caused by the broadcasting and sensitivity to the ageing time of the address tables' records, which complicates the guaranteeing of a given QoS.

The preliminary comparison confirms definite advantages of E6 addressing before PBB. For a thorough comparative evaluation of the two technologies, it is necessary the construction of models displaying the peculiarities of IP-Ethernet encapsulation, the filling up E6 dynamic address tables, the functioning of PBB virtual networks, which are the directions for future research.

9. References

[1] L. Fang, R. Zhang and M. Taylor, "The Evolution of Carrier Ethernet Services—Requirements and Deployment Case Studies," *IEEE Communications Magazine*, Vol. 46, No. 3, March 2008, pp. 69-76.

[2] "IEEE Standard for Local and Metropolitan Area Networks—Virtual Bridged Local Area Networks, Amendment 7: Provider Backbone Bridges," IEEE Std 802.1ah™-2008, 12 June 2008, pp. 1-109.

[3] F. Balus, M. Bocci and M. Aissaoui "VPLS Extensions for Provider Backbone Bridging," Work in Progress, IETF, July 2008, pp. 1-25

[4] P. P. Vorobiyenko, D. A. Zaitsev and O. L. Nechiporuk, "World-Wide Network Ethernet?" *Zviazok* (Communications), No. 5, 2007, pp. 14-19.

[5] P. P. Vorobiyenko, D. A. Zaitsev and K. D. Guliaiev, "Way of Data Transmission within Network with Substitution of Network and Transport Layers by Universal Technology of Data-Link Layer," *Patent of Ukraine on Utility Model*, No. 35773, 2008.

[6] K. D. Guliaiev, D. A. Zaitsev, D. A. Litvin and E. V. Radchenko, "Simulating E6 Protocol Networks Using CPN Tools", *Proceedings of International Conference on IT Promotion in Asia*, Tashkent, 22-26 August 2008, pp. 203-208.

[7] M. Beaudouin-Lafon, W. E. Mackay, M. Jensen, *et al.*, "CPN Tools: A Tool for Editing and Simulating Coloured Petri Nets," *International Conference of Tools and Algo-*

rithms for the Construction and Analysis of Systems, Genova, 2-6 April 2001, pp. 574-580. http://www.daimi.au.dk/CPNTools

[8] K. Jensen, "Colored Petri Nets: Basic Concepts, Analysis Methods and Practical Use," Springer-Verlag, Berlin, 1997.

[9] D. A. Zaitsev and T. R. Shmeleva, "Switched Ethernet Response Time Evaluation via Colored Petri Net Model," *Proceedings of International Middle Eastern Multiconference on Simulation and Modelling*, Alexandria, 28-30 August 2006, pp. 68-77.

[10] D. A. Zaitsev, "An Evaluation of Network Response Time Using a Coloured Petri Net Model of Switched LAN," *Proceedings of 5th Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools*, Aarhus, 8-11 October 2004, pp. 157-167.

[11] D. A. Zaitsev and A. L. Sakun, "An Evaluation of MPLS Efficacy Using Colored Petri Net Models," *Proceedings of International Middle Eastern Multiconference on Simulation and Modelling*, Amman, 26-28 August 2008, pp. 31-36.

[12] M. V. Bereznyuk, K. K. Gupta and D. A. Zaitsev, "Effectiveness of Bluetooth Address Space Usage," *Proceedings of 20th International Conference*, *Software & Systems Engineering and their Applications*, Paris, 4-6 December 2007.

Communications and Network, **2010, 2, 125-133**
doi:10.4236/cn.2010.22019 Published Online May 2010 (http://www.SciRP.org/journal/cn)

An Energy-Efficient Clique-Based Geocast Algorithm for Dense Sensor Networks

Alain Bertrand Bomgni, Jean Frédéric Myoupo
[1]*Department of Computer Science*, *University of Yaounde*, *Yaounde*, *Cameroon*
[2]*Department of Computer Science*, *University of Picardie Jules Verne*, *Amiens*, *France*
E-mail: *jean-frederic.myoupo@u-picardie.fr*
Received February 19, 2010; *revised April* 19, 2010; *accepted April* 23, 2010

Abstract

This paper proposes an energy-efficient geocast algorithm for wireless sensor networks with guaranteed delivery of packets from the sink to all nodes located in several geocast regions. Our approach is different from those existing in the literature. We first propose a hybrid clustering scheme: in the first phase we partition the network in cliques using an existing energy-efficient clustering protocol. Next the set of clusterheads of cliques are in their turn partitioned using an energy-efficient hierarchical clustering. Our approach to consume less energy falls into the category of energy-efficient clustering algorithm in which the clusterhead is located in the central area of the cluster. Since each cluster is a clique, each sensor is at one hop to the cluster head. This contributes to use less energy for transmission to and from the clusterhead, comparatively to multi hop clustering. Moreover we use the strategy of asleep-awake to minimize energy consumption during extra clique broadcasts.

Keywords: Geocast, Wireless Sensor Networks, Clustering, Clique, Energy Consumption

1. Introduction

A wireless sensor network (WSN for short) is a deployment of massive numbers of small, inexpensive, self-powered devices that can sense, compute, and communicate with other devices for the purpose of gathering local information to make global decisions about a physical environment. Commonly monitored parameters are temperature, humidity, pressure, wind direction and speed, illumination intensity, vibration intensity, sound intensity, power-line voltage, and chemical … Routing in a sensor network consists of sending a message from the source to a destination. Routes between two hosts in the network may consist of hops through other hosts in the network. This paper is about multi-geocasting which is a routing protocol based on the position of nodes. The geocast problem consists of sending a message from a sink to all nodes located in a designated region called the geocast region. In the multi-geocast, a message is sent from a sink to all nodes located in multiple geoacst regions. An important objective of geocasting and Multi-geocasting is to achieve guaranteed delivery while maintaining an engergy low cost. Guaranteed delivery ensures that every sensor in a geocast region receives a copy of geoacasting message. Since sensors are generally powered by batteries, the limited energy of sensors requires geocasting and multi-geocasting to consume as less energy as possible.

1.1. Related Work

Flooding is the simplest approach to implement geocasting or multi-geocasting [1,2]. The sink broadcasts the packet to its neighbours that have not received the packet yet, and these neighbors broadcast it to their own neighbours, and so on, until the packet is received by all reachable nodes including the geocast region in the case of the geocasting and the different geocast regions in the case of the multi-geocasting. The earliest work in the geoacasting was proposed by Navas and Imielinski [1] in the context of internet. Their approach integrates geographic coordinates into IP address. It consists of sending the packets to all nodes within a geographic area. They presented a hierarchy of geographically-aware routers that can route packets geographically and use IP tunnels to route through areas not supporting geographic routing. Each router covers a certain geographic area called a service area. When a router receives a packet with a geocast region within its service area, it forwards the packet to its children nodes (routers or hosts) that cover

or are within this geocast region. If the geocast region does not intersect with the router service area, the router forwards the packet to its parent. If the geocast region and the service area intersect, the router forwards to its children that cover the intersected part and also to its parent. Ko and Vaidya [2] proposed geocasting algorithms to reduce the overhead, compared to global flooding, by restricting the forwarding zone for geocast packets. Nodes within the forwarding zone forward the geocast packet by broadcasting it to their neighbors and nodes outside the forwarding zone discard it. Each node has a localization mechanism to detect its location and to decide when it receives a packet, whether it is in the forwarding zone or not (The localized mechanism can be GPS or the techniques of ad hoc positioning systems [3]). When the forwarding zone is empty, the node floods the packet to all its neighbours. To ensure message delivery, face routing was introduced in [4]. In face routing, a planar graph derived from the network topology is used, and the network area is partitioned into a set of faces. To transmit a message from a source *s* to a destination *t*, the message goes through the face intersecting the line segment *st* from *s* to *t*. If an edge *e* on the boundary of the traversed face intersects with *st* and the intersecting point is closer to *t* than to *s*, the face, which is next to *e* and closer to *t* than the currently traversed face, is traversed. This process is repeated until *t* is found. Face routing ensures message delivery, but it might use long forwarding paths [4]. To find a routing path close to the optimal path, the *Geographic-Forwarding-Geocast* (GFG) was proposed in [5]. It has almost optimal minimum overhead and is ideal in dense network. GFG uses the geographic information to forward packets efficiently toward the geoacsat region. It consists of greedy forwarding where perimeter routing is used by nodes outside the region and nodes inside the region broadcast the packet to flood the region. *Geoagraphic-Forwarding-Perimetre-Geoacast* (GFPG) was proposed also in [6]. The algorithm solves the region gap problem in sparse networks. This algorithm combines geocast and perimeter routing to guarantee the delivery of the geocast packets to all nodes in the region. The idea is to use the perimeter routing once the geocast packet reaches a node in geocast region to guarantee delivery of the packet to all the nodes located in the geocast region. An internal node located in the geocast region which has neighbours outside the region, initiates the perimeter routing. The main difference between the algorithm [5] and the one proposed in [6] is that external border nodes in [6] also perform the right-hand based-face traversals with respect to all corresponding neighbors internal border nodes. The authors in [7] proposed Virtual Surrounding Face (VSF). In VSF, the geoacast region is constructed by ignoring the edges intersecting the geocast region in a planar graph. By traversing all the boundary nodes of VSF and performing restricted flooding within the geoacasting region, all nodes are guaranteed to receive the message.

In the case of multi-geocasting problem, several disconnected regions exist. The message will then be delivered from one source to all hosts located in these regions. The flooding algorithm could be executed by sending the message from the sink to all the hosts in the geocast regions. However, this approach generates a huge overhead and then high cost. Multi-geocast protocols that reduce the size of the flooding were proposed in [8]. The scheme proposed by the authors consists of two phases: interest forwarding phase and data forwarding phase. To send interest messages toward multiple regions, a sink first creates a shared path between these regions based on the theorem of Fermat Point. Then, according to this path, interest messages are delivered to each target region. When each node located in a region receives interest messages, it initiates the execution of local flooding algorithm. In [9], the network is supposed to be partitioned geographically. Cellular- Based-Management geographically partitions the network into several disjoints and equally sized cellular regions. A manager is selected in each cell to administrate the cellular which has a unique Cellular-ID. The protocol then creates a shared path for different destinations. Both protocols [7,8] guarantee delivery of the packets only in a dense network and do not guarantee delivery in a sparse network.

Energy-efficient methods for geocast appeared in [10-12]. The protocol in [11] builds a multicast tree connecting geocast node using an energy efficient broadcasting technique without making any restrictions on the shape of the geocast region. The proposed protocol reduces the energy consumption during the phase of sending commands to the sensor nodes in a geocast region and also facilitates in-network data aggregation and, therefore, saves energy during the phase of reporting sensor data. The approach in [12] is based on the construction of a geocast routing tree. As the most existing geometric routing schemes, the protocol in [5] can also discover a non-geometric path to the destination by exploiting the path history of location updates. In addition, their technique employs two location-based optimizations to further reduce the overhead of on-demand route discovery on inevitable routing voids.

Nowadays, some applications in wireless ad hoc or sensor networks are made efficient by partitioning these networks into clusters [13-16]. Consequently complete distributed cluster architectures are proposed mainly to settle a hierarchical routing protocol. In existing solutions for clustering in ad hoc networks, the clustering is performed in two phases: clustering set up and cluster maintenance. The first phase is accomplished by choosing some nodes to act as coordinators of the clustering process (clusterheads). Then a cluster is formed by associating a clusterhead with some of its neighbors (i.e. nodes within the clusterhead's transmission range) that become the ordinary nodes of the cluster. Once the clus-

ter is formed, the clusterhead can continue to be the local coordinator for the operations in its cluster. The existing clustering algorithms differ on the criteria for the selection of the clusterheads. For example, in [14,17], the choice of the clusterhead is based on the unique identifier, say ID, associated to each node: the node with the lowest ID is selected as clusterhead, then the cluster is formed by that node and all its neighbors.

Basagni [18] has been interested in either phases of the clustering process under the common perspective of some desirable clustering properties. The main advantage of its approach is that, by representing with the weights the mobility-related parameters of the nodes, we can choose for the role of clusterhead those nodes that are better suited for that role. For instance, when the weight of a node is inversely proportional to its speed, the less mobile nodes are selected to be clusterheads. In consequence, the clusters are guaranteed to have a longer life, and consequently the overhead associated with the cluster maintenance in the mobile environment is minimized. Although this algorithm can be used in the presence of nodes' mobility (using for instance, the periodical reclustering), the DCA is mainly suitable for ad hoc networks whose nodes do not move or move "slowly" (quasi-static networks). For (possibly highly) mobile networks, Basagni introduced the Distributed Mobility-Adaptive Clustering (DMAC) algorithm. By executing the DMAC routines, each node reacts locally to any variation in the surrounding topology, changing its role (either clusterhead or ordinary node) accordingly. In our former work in [16], we use Basgni clustering technique to derive a geocast algorithm with the guaranteed delivery.

1.2. Our Contribution

This paper proposes an energy-efficient geocast algorithm in wireless sensor networks with guaranteed delivery of packets from the sink to all nodes located in several geocast regions. Our approach is different from those in [10-12]. We first propose a hybrid clustering scheme: in the first step we partition the network in cliques using an existing energy-efficient clustering protocol. Next the set of clusterheads of cliques are in their turn partitioned using energy-efficient hierarchical clustering. Our approach to consume less energy falls into the category of energy-efficient clustering algorithm in which the clusterhead is located near the central area of the cluster. Since each cluster is a clique, each sensor is at one hop to the clusterdhed. This contributes to use less energy for transmission to and from the clusterhead, comparatively to multi hop clustering. Moreover we use the strategy of asleep/awake during extra clique broadcasts.

The rest of this paper is organized as follows. The next section recalls two clustering methods that will be used to derive our geocast algorithm. Section 3 describes in details the geocast algorithm and the analysis of the energy consumption is done in Section 4. It is shown in Section 5 that the same idea holds for multi-geocast. Section 6 gives the curve of the average number of cliques with respect to the number of sensors. A conclusion ends the paper.

2. Preliminaries: Network Clustering

We now describe some tools that are necessary to derive our algorithm. A practical way of tackling the geocast problem would be to build a hierarchical structure above the network in order to simulate a sort of backbone made up of nodes which are more "adapted" than others. This is precisely the goal of clustering. This methodology has already proven its efficiency in the past. In sensor networks the sensor nodes can be partitioned into *clusters* by their physical proximity to achieve better efficiency, and each cluster may elect a *clusterhead* to coordinate the nodes tasks in the cluster. Certain references say that clustering with at most two hops is said to be *node-centric* [18], whereas clustering with over two hops is called *cluster-centric* [15]. In *node-centric* approach, clusterheads are first elected and a procedure indicates how to assign other nodes to different clusters. In *Cluster-centric* approaches, clusters are first formed, and each cluster then elects its clusterhead. Such approaches require that all nodes in one cluster agree on the same membership before electing their clusterhead. We now summarize two clustering schemes that will be helpful to describe our geocast protocol.

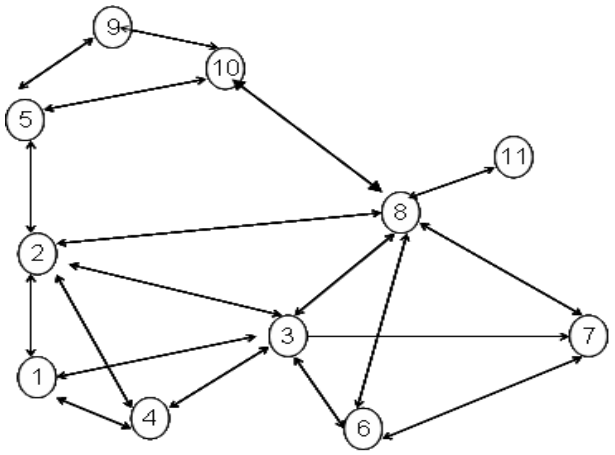
2.1. A Clustering Scheme in Cliques

Our formulation uses one of the protocols from [19,20] to partition network into clusters (cliques). The figure blow illustrates a network in which each clique is a single hop sub network.

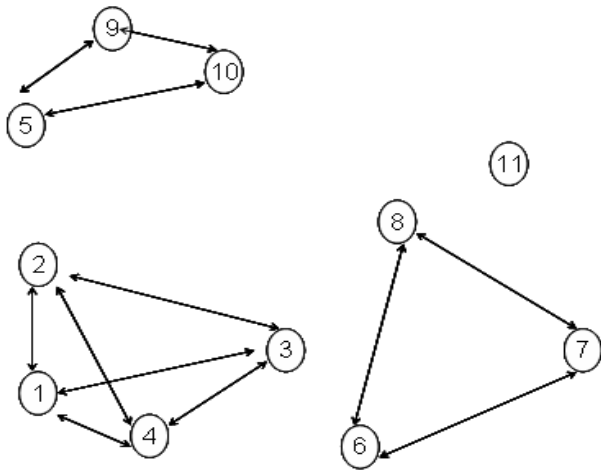
Each clique is a single hop network. Each clusterhead knows the partial IDS of its 1-hop neighbors. Let **G'** be the set of the clusterheads of cliques

2.2. Hierarchical Clustering

Banerjee and Khuller [15] proposed a clustering algorithm for multi-hop sensor networks. Their clustering scheme is motivated by the need to generate an applicable hierarchy for multi-hop wireless environment. Their method yields a multi-stage clustering. To reach their goal they construct a multi-stage depth first search tree such that each level is composed of clusterheads of the immediate low level. These Clusters are disjoint by defi-



(a)



(b)

Figure 1. (a) network with 11 sensors; (b) resulting cluster formation in cliques.

nition and the number of the nodes in a cluster remains between k and $2k$ for some integer k. **Figure 3** shows a hierarchical clustering of a network of 25 sensors with $k = 3$.

3. Geocast with Guaranteed Delivery

We assume that each sensor node is equipped with the GPS or can determine its location using the ad hoc positioning system [3]. Hence each node should know if it is in the geocast region or not. Our Approach to provide geocast in multi-hop sensor network consists of the following four phases:

3.1. Phase 1: Clustering Procedure in Cliques

The sensors run one of the protocols in [19,20] to create cliques like clusters. We assume that this phase yields k cliques (clusters), hence k cluster heads named $CH_{clique\text{-}i}$, $1 \leq i \leq k$, for the clusterhead of clique i.

3.2. Phase 2: Hierarchical clustering

Now we focus only on the set of k clusterheads obtained in phase 1. Consider a network, say G', whose sensors reduce to these k clusterheads. Clearly $|G'| = k$. Partition this network as in subsection 2.2 using the hierarchical clustering such that each resulting cluster contains at least $k/2$ sensors and at most k sensors. Hence the partition will give only one cluster, and thus one clusterhead. This clusterhead knows the IDs of all residents of its cluster, i.e. the IDs of the other k-1 sensors (see **figure 4**).

3.3. Request Phase

When the sink wishes to send a request to all hosts located in the geocast region, it floods a short packet (*REQUEST* (*Message*, *Locaion_Goecast_Region*)) in the backbone (the sensors in G'). This short packet contains the definition of the geocast region. All requests from the network are firstly sent to the ***super clusterhead*** that is the only unit to process or to take a decision on a request. Hence the request packet travels from a clusterhead of the first stage till the ***super clusterhead***.

After receiving the message *REQUEST* (*Message, Locaion_Goecast_Region*)), the ***super clusterhead*** sends a search message containing the definition of the geocast region (*SEARCH* (*Location_Geocast-Region*, β)) to all clique-clusterheads asking them to tell him whether some nodes of their clusters lie on the geocast region. This search message is accompanied by a binary variable β.

Each clique-clusterhead sends the request to each member of its cluster that determines by computation whether it is in the geocast region or not. If it so it sets β to 1 and sends it backwards together with its identifier to its clusterhead. Otherwise no action is taken, which means that it is not in the geocast region. Each clique-clusterhead registers the provenance of the positive answer.

If $\beta = 1$ then the clique-clusterhead sends back to the super clusterhead a small packet (*SEARCH* (*Location_Geocast-Region*, $\beta = 1$)) with β set to 1.

3.4. The Broadcast Phase

On receiving the answers the ***super-clusterhead*** sends the request message *REQUEST* (*Message*, $\beta = 1$) to the clique-clusterheads that send back positive answers. This request travels from clique-clusterheads to clique-clusterheads (which registered β to 1) till the nodes which set β to 1 during the search phase (*i.e.*, those in the geocast region). See the illustrative example of **Figure 5**.

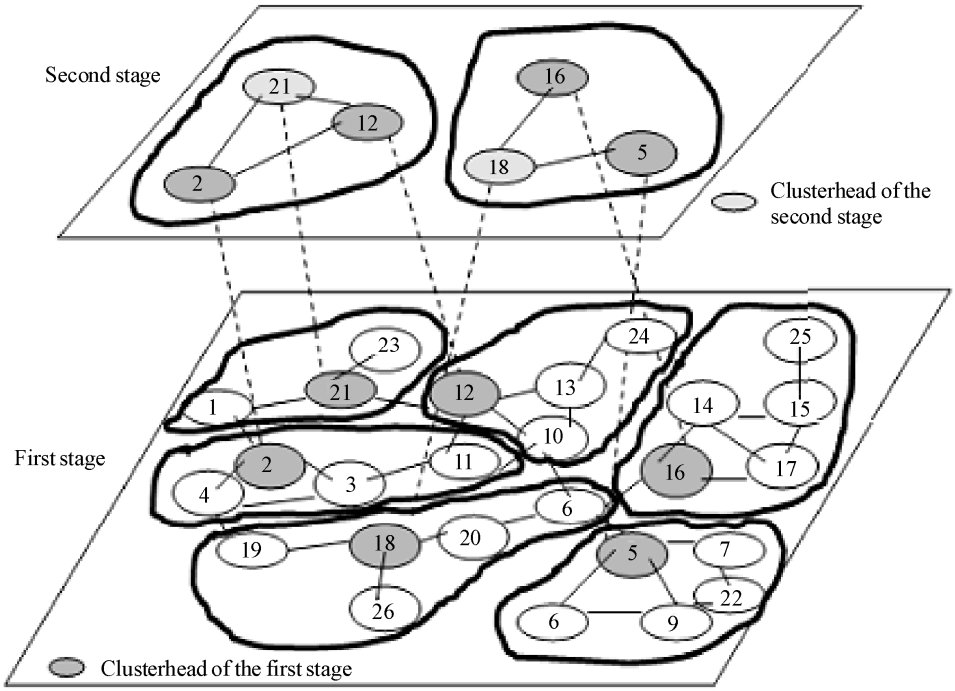


Figure 3. Hierarchical clustering with $k = 3$.

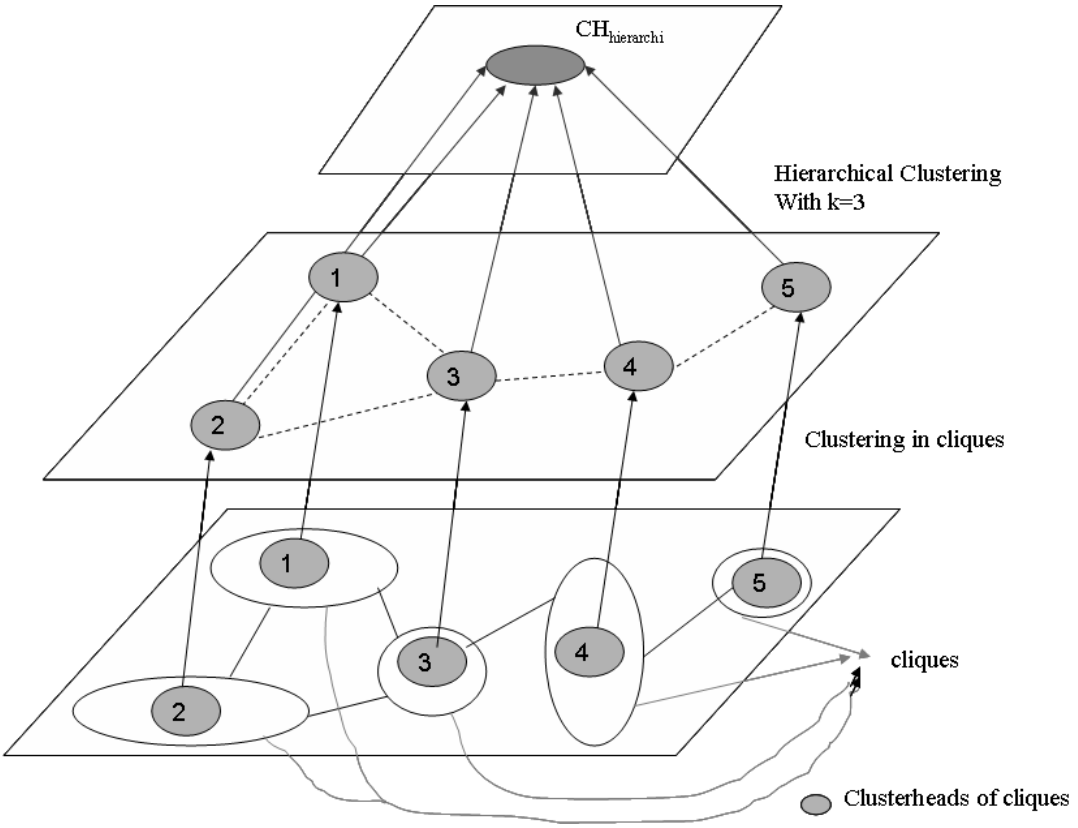


Figure 4. An example hybrid clustering: the first stage shows an example of clustering in cliques. The second stage shows the hierarchical clustering of G'.

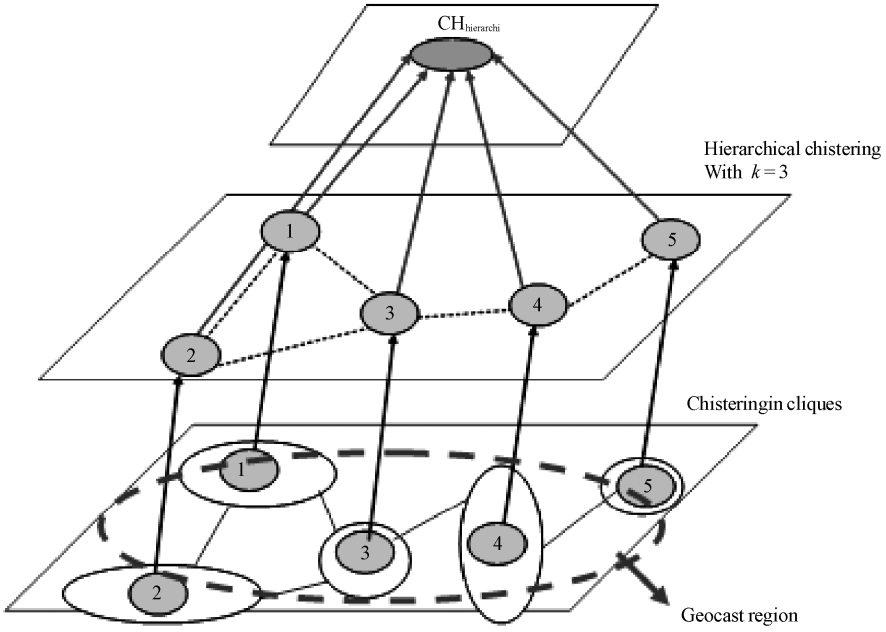


Figure 5. An illustration of geocast.

***Lemma* 1**: *The above multi-stage clustering geocast algorithm guarantees the delivery to all nodes in the geocast region.*

Proof: Assume that there is at least one node in the geocast region that is not reached. Then this node has been disconnected from the network that is no more connected. Therefore it is not a sensor network

4. Analysis of the Energy Consumption

The energy model used here is similar to that used by most existing energy-efficient clusteing model [21-23]

$$E = ET + ER = a \times (e_t + e_{amp} \times d^n) + a \times a_r . \quad (1)$$

where ET and ER are the energy consumptions of transmitting and receiving data items respectively. The energy dissipated in operating the transmitter radio, transmitter amplifier and receiver radio are expressed by e_t, e_{amp} and r_e respectively. And d is the distance between nodes and n is the parameter of the power attenuation with $2 \le n \le 4$.

4.1. Reducing Power Consumption during Clustering in Cliques and Intra Broadcasts

Our approach to consume less energy falls into the category of energy-efficient clustering algorithm in which the clusterhead is located in the central area of the cluster. Here, since each cluster is a clique, each sensor is at one hop to the clusterdhed. This contributes to use less energy for transmission to and from the clusterhead, comparatively to multi hop clustering. Thanks to the energy-efficient clustering algorithm in [24] that matches our requirements.

4.2. Reducing Power Consumption during Hierarchical Clustering

Only clique clusterheads and gateway nodes are involved in this scenario. They are the only sensors that are awake during the hierarchical clustering. The other sensors are asleep and will be waked up using the "Magic packet Tecnology" [13]. It is also known as the "Wake On Lan" (WOL for short). It consists in the hability to switch on remote computer through special network packets. Wol is based on the following principle: when a PC shuts down, the network card still gets power and keeps listening to the network for a magic packet to arrive. This technology was first designed for static wired networks, later a wireless version has been derived [25].

4.3. Reducing Power Consumption: During Intra Clique Broadcasts

Sensors awake only during intra cluster broadcasts, i.e. in Broadcast phase. In this phase the terms ET and ER of the Equation (1) are minimized since all intra cluster broadcasts have sensors at one hop as destinations.

5. Geocast with Multiple Target Regions with Guaranteed Delivery

This section follows similar lines as in Section 3. When the sink wishes to send a request to all hosts located in different geocast regions, it floods a short packet in the

network. This short packet contains the definition of the several geocast regions. It can also send several requests one after another, each for a specific geocast region. It is not difficult to see that the delivery here is also guaranteed.

6. Simulation Results: Average Number of Cliques

In this section, we present simulations results of the clustering algorithm to show the influence of the heuristic used to choose the clusterhead. These benchmarks have been run on a laptop (Pentium-M 1.7 GHz, 1 GO RAM, Windows XP SP2, Cygwin 1.5.19) and programmed in C++. Our main problem has been to establish suitable experiments conditions. As WSNs are supposed to be used in rescue services, we can assume that nodes are static. All nodes are assumed to have the same transmission range. The experiments take place in a geographic square area of side L. Each curve is the average of 100 experiments. We have made the common assumption that two nodes are neighbors if and only if their Euclidean distance is lower than 1 km. The nodes are in a square, which the length is $L = 2$ km. In our implementation, the MAC layer is managed in such a way that a node can only receive one message at a time, yielding delays in the clustering process and so maintaining always a high number of clusters.

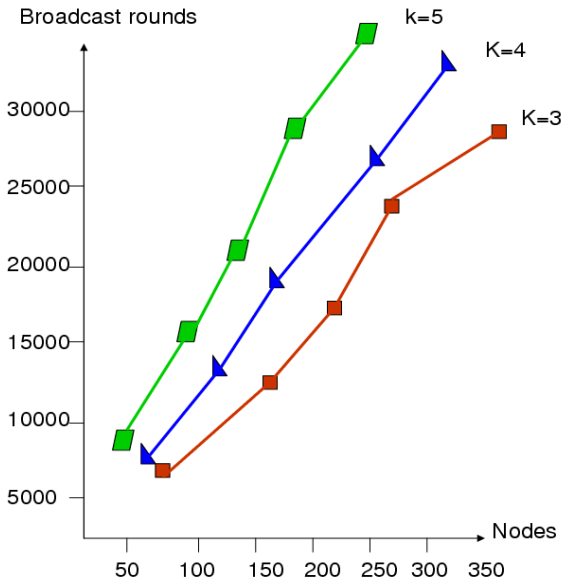


Figure 6. Number of broadcast rounds curves according to *k*.

Figure 6 shows the evolution curve of the number of broadcast rounds with respect to the number of sensors. We have considered 3 values of k, say 3, 4 and 5, yielding 3 curves. We assume that each node has 100 units of energy. A broadcast cost is one unit of energy and a reception cost is one unit of energy. An awake situation cost is 1/10 unit of energy per second. **Figure 7** shows the different

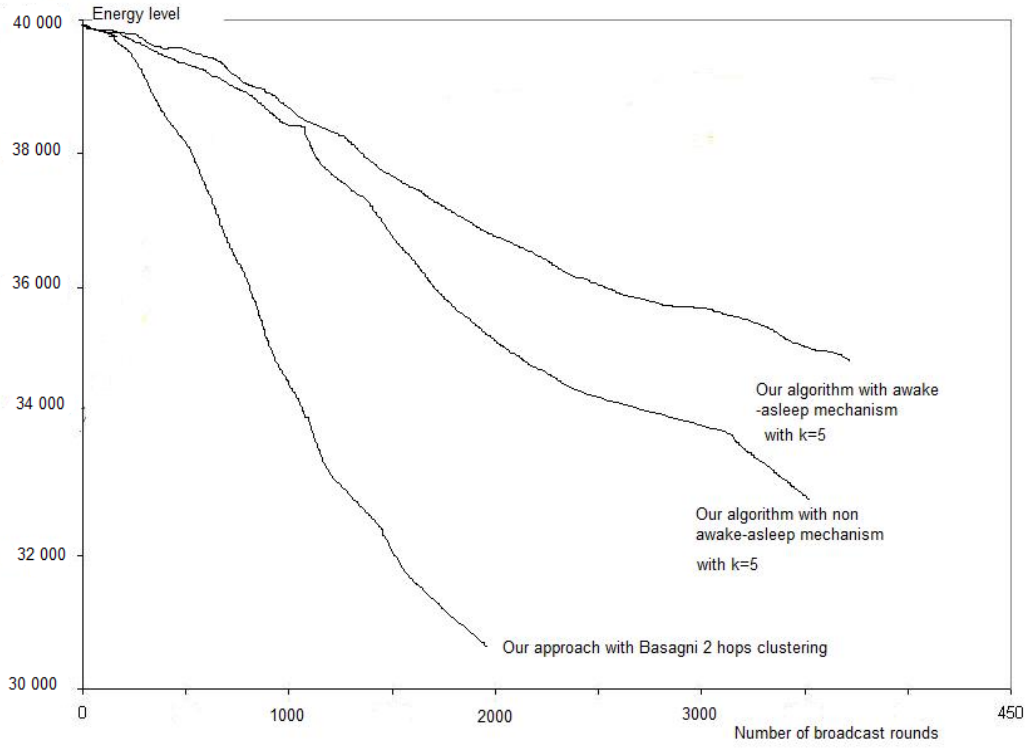


Figure 7. Energy evolutionary curves during the execution of the algorithms wit 400 sensor nodes.

scenarios of the energy evolution. Awake-asleep-based a gorithm performs better than the one not using this mechanism. Furthermore, in general our approach per forms better than the one not using clustering in cliques, like Basagni [18] clustering scheme.

7. Conclusions

This paper presents an energy-efficient geocast algorithm in wireless sensor networks with guaranteed delivery of packets from the sink to all nodes located in several geocast regions. Our approach derives from a hybrid clustering scheme: in the first step we partition the network in cliques using an existing energy-efficient clustering protocol. Next the set of clusterheads of cliques are in their turn partitioned using energy-efficient hierarchical clustering. We show that our protocol falls into the category of energy-efficient clustering algorithm in which the clusterhead is located near the the central area of the cluster. Since each cluster is a clique, each sensor is at one hop to the clusterdhed. This contributes to use less energy for transmission to and from the clusterhead, comparatively to multi hop clustering. Moreover we use the strategy of sleep/awake during extra clique broadcasts to save the energy of non participant sensors. A clique-clusterhead can have a higher burden that that of the local sensors of the clique. Rotating the role of the clique-clusterhead must be operated in order to distribute this higher burden among the local sensors, thereby preventing the clique-clusterhead from dying prematurely [26,27].

However an open problem remains: The derivation from the idea of this paper of a secure protocol for geocast.

When putting last hands on this paper we discovered another work on energy efficiency for geocast [28]. It is just the adaption of the previous work of the authors on a simple goecast algorithm in [29].

8. References

[1] T. Imielinski and J. Navas, "GPS-Based Addressing and Routing," *RFC* 2009 *Computer Science*, Rutgers University Press, Rutgers, March 1996.

[2] Y.-B. Ko and N. H. Vaidya, "Flooding-Based Geocasting Protocols for Mobile Ad Hoc Networks," *MONET*, Vol. 7, No. 6, 2002, pp. 471-480.

[3] D. Niculescu and B. Nath, "Ad Hoc Positioning System (APS)," *Proceedings of IEEE Global Telecommunications Conference*, San Antonio, 25-29 November 2001, pp. 2926-2931.

[4] E. Kranakis, H. Singh and J. Urrutia, "Compass Routing on Geometric Networks," *Proceedings of* 11*th Canadian Conference on Computational Geometry*, Vancouver, 15-18 August 1999, pp. 51-54.

[5] K. Seada and A. Helmy, "Efficient Geocasting with Perfect Delivery in Wireless Networks," *IEEE Wireless Communications and Networking Conference*, Atlanta, 21-25 March 2004, pp. 2551-2556.

[6] I. Stojmenovic, "Geocasting with Guaranteed Delivery in Sensor Networks," *IEEE Wireless Communications*, Vol. 11, No. 6, December 2004, pp. 29-37.

[7] J. Lian, K. Naik, Y. Liu and L. Chen, "Virtual Surrounding Face Geocasting with Guaranteed Message Delivery for Sensor Networks," *Proceedings of the* 14*th IEEE International Conference on Network Protocols*, Santa Barbara, 12-15 November 2006, pp. 198-207.

[8] Y.-M. Song, S.-H. Lee and Y.-B. Ko, "FERMA: An Efficient Geocasting Protocol for Wireless Sensor Networks with Multiple Target Regions," *Lecture Notes on Computer Science*, Vol. 3823, 2005, pp. 1138-1147.

[9] C.-Y. Chang, C.-T. Chang and S.-C. Tu, "Obstacle-Free Geocasting Protocols for Single/Multi- Destination Short Message Services in Ad Hoc Networks," *Wireless Networks*, Vol. 9, No. 2, 2003, pp. 143-155.

[10] L. Choi, J. K. Jung, B.-H. Cho and H. Choi, "M-Geocast: Robust and Energy-Efficient Geometric Routing for Mobile Sensor Networks," *Lecture Notes in Computer Science*, Vol. 5287, 2008, pp. 304-316.

[11] Y.-C. Shim, "Energy Efficient Geocast Protocol for Sensor Networks," *Proceedings of the* 6*th WSEAS International Conference on Electronics*, *Hardware*, *Wireless and Optical Communication*s, Corfu, 16-19 February 2007, pp. 28-34.

[12] W. Zhang, X. Jia and C. Huang, "Distributed Energy-Efficient Geographic Multicast for Wireless Sensor Networks," *International Journal of Wireless and Mobile Computing,* Vol. 1, 2006, pp. 141-147.

[13] Advanced Micro Devices, "White Paper: Magic Packet Technology," November 1995. http://www.amd.com/us-en/assets/content_type/white_papers_and_tech_docs/20213.pdf

[14] D. Baker and A. Ephremides, "The Architectural Organization of a Mobile Radio Network via Distributed Algorithm," *IEEE Transactions on Communications*, Vol. 29, No. 11, November 1981, pp. 1694-1701.

[15] S. Banerjee and S. Khuller. "A Clustering Scheme for Hierarchical Control in Multi-Hop Wireless Networks", *Proceedings of the* 20*th Annual Joint Conference of the IEEE Computer and Communications Societies*, Anchorage, Vol. 2, 22-26 April 2001, pp. 1028-1037.

[16] A. B. Bomgni, J. F. Myoupo and A. O. Cheikhna, "Randomized Multi-Stage Clustering-Based Geocast Algorithms in Anonymous Wireless Sensor Networks," 5*th IEEE*/*ACM International Wireless Communications and Mobile Computing Conference*, Leipzig, 21-24 June 2009, pp. 286-291.

[17] M. Gerla and J. T. C. Tsai, "Multicluster, Mobile, Multimedia Radio Network", *Wireless Networks*, Vol. 1, No. 3, 1995, pp. 255-265.

[18] S. Basagni, "Distributed Clustering for Ad Hoc Networks," *Proceedings of the* 1999 *International Symposium on Parallel Architectures*, *Algorithms and Networks*,

Fremantle, 23-25 June 1999, pp. 310-315.

[19] K. Sun, P. Peng and P. Ning, "Secure Distributed Cluster Formation in Wireless Sensor Networks," 22*nd Annual Computer Security Applications Conference*, Las Vegas, 11-15 December 2006, pp. 131-140.

[20] P. Tosic and G. Agha. "Maximal Clique Based Distributed Coalition Formation for Task Allocation in Large-Scale Multi-Agent Systems," *Lecture Notes in Computer Science*, Vol. 3446, 2005, pp. 104-120.

[21] W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," *Proceedings of the* 33*th IEEE Hawii International Conference on Systems*, Hawii, 4-7 January 2000, pp. 3005-3014.

[22] J. S. Liu and C. H. R. Lin, "Energy-Efficient Clustering Protocol in Wireless Sensor Networks," *Ad Hoc Networks*, Vol. 3, No. 3, May 2005, pp. 371-388.

[23] D. Wei, S. Kaplan and H. A. Chan, "Energy Efficient Clustering Algorithms for Wireless, Sensor Networks," *Proceedings of IEEE Conference on Communications*, Beijing, 19-23 May 2008, pp. 236-240.

[24] Y. Zhou, M. Hart, S. Vadgama and A. Rouz. "A Hierarchical Clustering Method in Wireless Ad Hoc Sensor Networks," *Proceedings of International Conference on Communications*, Glasgow, 24-28 June 2007, pp. 3505-3509.

[25] J. Lewis, "Wake on LAN over Wireless," 2008. http://www.johnlewis.ie/2008/07/10/wake-on-lan-over-wireless

[26] W. Wang and A. Jantsch, "An Algorithm for Electing Cluster Heads Based on Maximum Residual Energy," *Proceedings of International Wireless Communications and Mobile Computing Conference*, Vancouver, 3-6 June 2006, pp. 1465-1470.

[27] O. Younis and S. Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks," *IEEE Transactions on Mobile Computing*, Vol. 3, No. 4, 2004, pp. 366-379.

[28] J. A. Sanchez, P. M. Ruiz and I. Stojmenovic, "Energy-Efficient Geographic Multicast Routing for Sensor and Actuator Networks," *Computer Communications*, Vol. 30, No. 13, September 2007, pp. 2519-2531.

[29] J. A. Sanchez, P. M. Ruiz and I. Stojmenovic, "GMR: Geographic Multicast Routing for Wireless Sensor Networks," *Proceedings of the* 3*rd Sensor and Ad Hoc Communications and Networks*, Reston, 25-28 September 2006, pp. 20-29.

Communications and Network, **2010, 2, 134-137**
doi:10.4236/cn.2010.22020 Published Online May 2010 (http://www.SciRP.org/journal/cn)

An Assessment of WiMax Security

Sanjay P. Ahuja, Nicole Collier
School of Computing, *University of North Florida*, *Jacksonville*, *USA*
E-mail: {*sahuja*, *nicole.collier*}*@unf.edu*
Received February 23, 2010; *revised April* 20, 2010; *accepted April* 29, 2010

Abstract

For a broadband wireless standard such as WiMax, security is important and must be addressed. This is to ensure wide acceptance both from the perspective of the end users and the service providers. In order to compete with existing broadband cable or DSL services, the WiMax network must offer comparable security. We discuss the WiMax security mechanisms for authentication, encryption, and availability. We also discuss potential threats to WiMax security. This paper will also discuss how and why these threats play an important role in the adaptability of WiMax.

Keywords: WiMax, WiMax Security, Service Provider

1. Introduction

With the introduction of Wireless LANs in the 90s network security became a very important subject of discussion among major corporations, service providers, and end users. Security in wireless networks is the maintaining of confidentiality, authentication, non-repudiation, and integrity control [1]. To keep these four areas protected from malicious attacks, certain protocols were put into place. These protocols are constantly being tested and improved as necessary to keep corporations safe from outside, and sometimes inside users. These protocols are also being used by home users as more wireless networks are being brought into the homes of end users.

Worldwide Interoperability for Microwave Access (WiMax) is described as "a standards based technology enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL" [2]. WiMax provides many services using point to point and point to multipoint applications. These applications are cost effective and cover a much larger area than WiFi (IEEE 802.11). WiMax uses a base station to transmit to Customer Premise Equipment (CPE). Using a base station allows WiMax to use applications for fixed, portable, or mobile non-line-of-sight services. With this technology WiMax is able to cover an entire city, not just a coffee shop or office building.

As mentioned, network security is important. In order for WiMax to be an accepted wireless service it must meet or exceed the standards already in place. This is especially important for WiMax as it can cover such large areas. Any flaws in the security and hackers would be able to break in, or there could be interference from one computer to another.

When WiFi was implemented the protocols for network security were developed. The first step for wireless security was Wired Equivalent Privacy (WEP) [3]. As more flaws were found more protocols were developed. To improve key management and initialization WEP 2 was developed. Unfortunately, this was not the answer to all the problems wireless security was faced with. Eventually, wireless security evolved to include many different protocols and encryptions that are used in WiFi today.

The lessons learned with WiFi security paved the way to the security measures used for WiMax. WiMax uses Counter Mode with Cipher Block Chaining Message Authentication Code Protocol (CCMP) to encrypt all traffic on its network. It also uses Advanced Encryption Standard (AES) to transmit data securely. Both of these are used in WiFi today and are strong in encryption and key management. Another protocol developed during the search for network security in WiFi and now used in WiMax, was PKM-EAP (Extensible Authentication Protocol). This protocol is used for end-to-end authentication. WiMax has benefited from the lessons learned with WiFi, but there are still threats and vulnerabilities to be dealt with.

The rest of the paper is organized as follows. WiMax Security standards are discussed in Section 2, Section 3 discusses mobile wireless WiMax Security Architecture, and Section 4 discussed threats to WiMax, Conclusions are listed in Section 5.

2. Standards of WiMax Security

The Protocol Stack used for WiMax is similar to that used for WiFi. The structure is the same, but WiMax uses more sublayers. The standards for WiMax Security are also similar. These standards are discussed in the following sections.

2.1. Data Link Layer Security

The Data Link Layer for WiMax has three sublayers. Privacy and security is handled in the bottom layer. The MAC sublayer is next, which implements secure key exchange and encrypts traffic. The last sublayer is the Service-Specific Convergence sublayer. **Figure 1** shows the Protocol Stack for WiMax. The WiMax MAC layer uses a scheduling algorithm opposed to contention access used in the WiFi MAC layer. For the initial entry into the network, the scheduling algorithm attempts only once from the Subscriber Station (SS) and then it is allocated an access slot by the Base Station (BS). This access slot cannot be used by other subscribers, while assigned to the SS. To ensure confidentiality the MAC layer uses electronic signatures to authenticate the user and the device.

2.2. Authentication

One of the major problems that WiFi faced when first launched was authentication. Since this posed huge security issues, a better standard for authentication was used for WiMax. The WiMax network uses a Privacy Key Management (PKM) protocol for Authentication. This dynamic system makes it harder for hackers to act as a legitimate subscriber [4]. This method of authentication gives three types of protocols, a RSA based authentication, which uses a X.509 certificate with RSA encryption. An EAP based authentication, which also has three types of protocols to choose from. These three types are AKA (Authentication and Key Agreement) for SIM based authentication, TLS for X.509 based authentication, and TTLS for MS-CHAPV2 (Microsoft-Challenge Handshake Authentication Protocol). The third type of PKM protocol is RSA based authentication followed by EAP authentication [5].

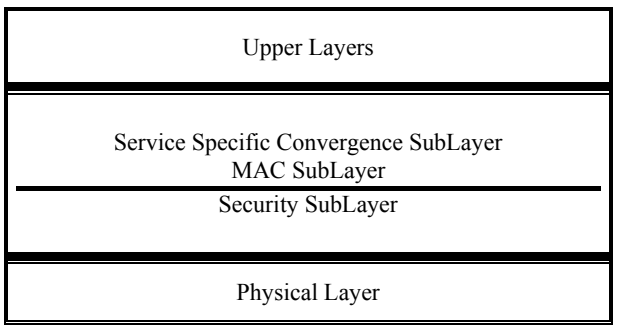


Figure 1. WiMax Protocol Stack.

2.3. Authorization

Authorization works hand in hand with authentication when it comes to security for WiMax. In order for a user to receive authorization, the authentication protocols must be met. Immediately following the authentication process, the SS sends an Authorization Request message to the BS [5]. In return an Authorization Reply message is sent back to the SS with the Authorization Key (AK) encrypted in the SS's public key along with the Security Association ID (SAID) of more Security Associations (SA) the SS is authorized to participate with. A lifetime key is also included with this reply. After the initial Authorization the BS periodically re-authorizes the SS.

2.4. Encryption

WiMax uses 3DES and AES to encrypt data transferred on the network. The Triple Data Encryption Standard (3DES) uses three different keys with a length of 56-bit each. The use of three keys causes for a slower performance in some software. The slow performance and limit on the length of keys is slowly making 3DES obsolete. The main tool for encryption that is used by WiMax is the Advanced Encryption Standard (AES) [6]. AES provides support for 128-bit, 192-bit, or 256-bit encryption keys [4]. AES was built from CCMP and has become a popular algorithm. AES is faster than 3DES, easy to implement and uses very little memory. However, it does require dedicated processors on board the BS, and may not be used by all end-user terminals. Therefore, 3DES remains a vital encryption tool on the WiMax network.

2.5. Availability

WiMax uses Radio Frequency (RF) Spectrum and could function on any frequency below 66 GHz [2]. The highest frequency available in the USA is 2.5 GHz. One of the drawbacks for using RF Spectrum is that the higher the frequency the range of a BS decreases a few hundred meters. Analog TV bands may be available for WiMax use once the rollout of digital TV is complete in February of 2009.

3. Mobile Wireless WiMax Security Architecture

The WiMax Security Architecture is flexible to allow Base Stations of different sizes and Subscriber Stations of different functionality. It also follows the standard end-to-end architecture for the Network Reference

Model (NRM). The network is divided into two main parts, Access Service Network (ASN) and the Connectivity Service Network (CSN). The ASN control and monitors the traffic between the Base Stations and the ASN Gateways. It also maintains the authentication and the key distributions. There are three ASN profiles, A, B, and C. Profiles A and C implements Radio Resource Management (RRM) and Handover functions, using a centralized ASN Gateway. These functions are used in the BS. Profile B embeds the key inside the BS. This eliminates the need for a centralized ASN Gateway. CSN controls the ASNs and end users with services such as AAA, Home Agent Functions, and DHCP Server. CSN also connects to operator's networks and enables inter-operator and inter-technology roaming.

4. Threats to WiMax

Many of the security threats found in WiMax have been addressed. These are issues that were found with the deployment of WiFi. This gives WiMax an advantage. If all, or at least most of the security issues can be addressed before WiMax's mainstream deployment, it will make it a more accepted network than WiFi. This section of the paper will discuss the known security threats in WiMax.

These include:

- Rogue base stations
- DoS attacks
- Man-in-the-middle attacks
- Network manipulation with spoofed management frames
- Threats in the physical layer

4.1. Rogue Base Stations

A Rogue Base Station is defined as an attacker station that imitates a legitimate base station [7]. This kind of attack results in disruptions in service and allows hackers to confuse subscribers. This is more difficult, but not impossible to do in the WiMax network. WiMax uses time division multiple access, therefore the rogue base station must transmit with a stronger strength at the same time the legitimate station transmits. The rogue base station captures the legitimate base station's identity and uses it to authenticate with the subscriber station. The authentication protocols used in WiMax help mitigate this threat. WiMax uses the EAP Protocol as its main protocol for authentication. This protocol forces mutual authentication, therefore the subscriber station would send an authentication message to the rogue base station. This does not completely alleviate the threat of rogue base stations, but it does make it more difficult.

4.2. DoS Attacks

Denial of Service (DoS) attacks is defined as an attempt to make a computer resource unavailable to its intended users [8]. Hackers usually use this type of attack on web servers for banks, credit card payment gateways or DNS root servers. A DoS attack uses the IP address to flood the user's network and obstruct communication between the intended user and the victim. This type of attack is not preventable; however steps can be taken to quickly resolve the attack. Some firewalls have built-in protection from DoS attacks, that monitor the amounts of packets received and the time frame they were received. It has been proposed that a Shared Authentication Information (SAI) protocol could be used to offer a defense mechanism against DoS attacks, without incurring overhead at the ASN gateway and base station. This proposal uses the unused upper 64-bit of the 128-bit Cipher Based Message Authentication Code (CMAC) to calculate a CMAC key [9]. This proposal could be the answer to prevention of DoS threats.

4.3. Man-in-the-Middle Attacks

Man-in-the-Middle attacks are forms of eavesdropping. The hacker establishes separate connections between two victims and relays the messages between them [10]. The hacker intercepts the public key from one of the victims and sends his or her own public key to the intended victim. When that victim responds the hacker then has that public key. The use of the RF Spectrum in WiMax allows for vulnerabilities to the man-in-the-middle attack. However, WiMax uses a three-way handshake scheme that supports re-authentication mechanisms for fast handovers to prevent man-in-the-middle attacks [11]. If the base station is constantly changed the public key changes making it almost impossible for hackers to eavesdrop using public keys.

4.4. Network Manipulation with Spoofed Management Frames

The management frames in WiMax are similar to WiFi's. When WiFi was first deployed vulnerabilities were found in the management frames that allowed DoS attacks by disrupting the wireless session between two nodes. WiMax has cryptographic protections from spoofed identities, but that does not mean it is safe. Replay DoS attacks still remain a threat to WiMax, due to the lack of any mechanisms to specifically detect and discard repeated packets [12].

4.5. Threats in the Physical Layer

Blocking and rushing are the major threats located in the

physical layer. Blocking or jamming activates a strong frequency to lower the capacity of the channel creating a DoS to all stations. This threat is detectable with a radio analyzer device. This device does not prevent this threat, but it alerts the end-user so that steps can be taken to immediately recover. Rushing or scrambling is another type of jamming, but it only activates for short periods of time and only affects certain frames. Jamming can be prevented using an increased signal or using frequency hopping. Control or Management messages are not in danger of rushing or blocking. Scrambling the uplink slots is too difficult for hackers [13].

5. Conclusions

WiMax security has been discussed in this paper. The lessons learned with WiFi's deployment have given WiMax an advantage in providing a safer wireless network. The precautions taken with WiMax were not done for WiFi. These security measures were not taken at WiFi's launch because the threats were unknown. Now that the threats are known and understood, they have been addressed prior to WiMax's deployment. However, this does not mean that WiMax is flawless. There may be new threats that are unknown and will not be addressed until WiMax is launched. WiMax does have much potential. WiMax would allow customers a completely wire free network connection in their homes or businesses. WiMax would also provide better services for mobile devices. There are also some speculations for WiMax to be used in gaming consoles. While security is an important subject and was a main cause of delay in the first stages of development for WiMax, it has now been faced with technological down falls. Sprint Nextel is one of the main companies looking to bring WiMax to the US. The main cause for delay has been the Sprint Nextel backhaul links from the backbone to the towers. It has been unable to support the promised 4Mbps. Until recently there has been little hope that a resolution to this problem would surface. In July 2008, Sprint Nextel announced that a solution has been found. They will be teaming up with DragonWave, a Canadian company, to provide the backhaul links. This will allow all data passing from the backbone to the end user to travel over the air, and the towers will need only power links. With this new development a testing launch of WiMax to Portland, Oregon is ready now, and preparations for a launch in 2009 to LasVegas, Atlanta, and Grand Rapids. If these tests go well we could see a full commercial launch by 2010. However, Clearwire has successfully deployed WiMax to several cities in the US and has been the main WiMax service provider for the last few years. With the formation of the Open Patent Alliance (OPA), which includes Clearwire, Sprint, Alcatel-Lucent, Cisco, Intel Corporation, and Samsung Electronics, WiMax could be global in the next five years.

6. References

[1] A. S. Tanenbaum, "Computer Networks," 4th Edition, Prentice Hall, Inc., New Jersey, 2006.

[2] http://en.wikipedia.org/wiki/WiMax

[3] S. P. Ahuja and P. K. Potti, "Evolution of Wireless LAN Security," *International Conference on Parallel and Distributed Processing Techniques and Applications*, Las Vegas, 24-17 July 2008.

[4] T. Sanders, "Premium Five Essential Elements of WiMax Security," WiMax.com, November 2007.

[5] http://myhsc.pbwiki.com/wimax::aaa

[6] P. Korsenlowski, "Staying Safe in a WiMax World," TechNewsWorld.com, 27 February 2007.

[7] M. Barbeau, J. Hall and E. Kranakis, "Detecting Impersonation Attacks in Future Wireless and Mobile Networks," *Secure Mobile Ad-hoc Networks and Sensors*, Ottawa, 2006, pp. 80-95.

[8] M. McDowell, "Understanding Denial of Service Attacks," *National Cyber Alert System*, 1 August 2007.

[9] K. Youngwook, L. Hyoung-Kyu and B. Saewoong, "Shared Authentication Information for Preventing DDoS Attacks in Mobile WiMax Networks," *Proceedings of the 5th Consumer Communications and Networking Conference*, Las Vegas, 10-12 January 2008.

[10] N. Beacham, "Man in the Middle (MITM) Attacks," *Technology and More*, 28 June 2008.

[11] J. M. Hartley, "WiFi and WiMax Protocols of Security," December 2008. http://softwarecommunity.intel.com/articles/eng/3708.htm

[12] R. Millman, "Security Experts See Vulnerablilites in WiMax," WiMax.com, 17 October 2006.

[13] M. Barbeau, "Threats: Threats to WiMax," http://www.freewimaxinfo.com/physical-layer.html

Communications and Network, **2010, 2, 138-144**
doi:10.4236/cn.2010.22021 Published Online May 2010 (http://www.SciRP.org/journal/cn)

Multiobjective Duality in Variational Problems with Higher Order Derivatives

Iqbal Husain[1], Rumana G. Mattoo[2]
[1]*Department of Mathematics*, *Jaypee Institute of Engineering and Technology*, *Guna*, *India*
[2]*Department of Statistics*, *University of Kashmir*, *Srinagar*, *India*
E-mail: {*ihusain*11, *rumana_research*}*@yahoo.com*
Received December 16, 2009; *revised April* 25, 2010; *accepted April* 30, 2010

Abstract

A multiobjective variational problem involving higher order derivatives is considered and optimality conditions for this problem are derived. A Mond-Weir type dual to this problem is constructed and various duality results are validated under generalized invexity. Some special cases are mentioned and it is also pointed out that our results can be considered as a dynamic generalization of the already existing results in nonlinear programming.

Keywords: Multiobjective Variational Problem; Efficiency, Duality, Pseudoinvexity, Quasinvexity, Nonlinear Programming

1. Introduction

Calculus of variation is a powerful technique for the solution of various problems appearing in dynamics of rigid bodies, optimization of orbits, theory of variations and many other fields. The subjects whose importance is fast growing in science and engineering primarily concern with finding optimal of a definite integral involving a certain function subject to fixed point boundary conditions. In [1] Courant and Hilbert, quoting an earlier work of Friedrichs [2], gave a dual relationship for a simple type of unconstrained variational problem. Subsequently, Hanson [3] pointed out that some of the duality results of mathematical programming have analogous in variational calculus. Exploring this relationship between mathematical programming and the classical calculus of variations, Mond and Hanson [4] formulated a constrained variational problem as a mathematical programming problem and using Valentine's [5] optimality conditions for the same, presented its Wolfe type dual variational problem for validating various duality results under convexity. Later Bector, Chandra and Husain [6] studied Mond-Weir type duality for the problem of [4] for weakening its convexity requirement. In [7] Chandra, Craven and Husain studied optimality and duality for a class of non-differentiable variational problem with non-differentiable term in the integrand of the objective functional while in [8] they derived optimality conditions and duality results for a constrained variational problem having terms with arbitrary norms in the objective as well as constrained functions.

Recently Husain and Jabeen [9] studied a wider class of variational problem in which the arc function is twice differentiable by extending the notion of invexity given in [10]. They obtained Fritz John as well as Karush-Kuhn-Tucker necessary optimality conditions as an application of Karush-Kuhn-Tucker optimality conditions studied various duality results for Wolfe and Mond and Weir type models.

In single objective programming we must settle on a single objective such as minimizing cost or maximizing profit. However, generally any real world problems can be identified with multiple conflicting criteria e.g., the problems of oil refinery scheduling, production planning, portfolio selection and many others can be modelled as multiobjective programming problems.

Duality results are very useful in the development of numerical algorithms for solving certain classes of optimization problems. Duality for multiobjective variational problem has been studied by a number of authors, notably Bector and Husain [11], Chen [12] and many others cited in these references. Applications of duality theory are prominent in physics, economics, management sciences, etc.

Since mathematical programming and classical calculus of variations have undergone independent development, it is felt that mutual adaptation of ideas and tech-

niques may prove useful. Motivated with this idea in this exposition, we propose to study optimality criteria and duality for a wider class of multiobjective variational problems involving higher order derivative. These results not only generalize the results of Husain and Jabeen [9] and Bector and Husain [11] but also present a dynamic generalization of some of the results in multiobjective nonlinear programming already existing.

2. Invexity and Generalized Invexity

Invexity was introduced for functions in variational problems by Mond, Chandra and Husain [10] while Mond and Smart [13] defined invexity for functionals instead of functions. Here we introduce extended forms of definitions of invexity and various generalized invexity for functional in variational problems involving higher order derivatives.

Consider the real interval $I=[a,b]$, and the continuously differentiable function $\phi: I\times R^n\times R^n\times R^n\to R$, where x is twice differentiable with its first and second order derivatives $\dot{x}$ and $\ddot{x}$ respectively. If $x=\left(x^1,x^2,\ldots,x^n\right)^T$, the gradient vectors of f with respect to x, $\dot{x}$ and $\ddot{x}$ respectively denoted by

$$\phi_x=\left[\frac{\partial\phi}{\partial x^1},\ldots,\frac{\partial\phi}{\partial x^n}\right]^T,\quad \phi_{\dot{x}}=\left[\frac{\partial\phi}{\partial \dot{x}^1},\ldots,\frac{\partial\phi}{\partial \dot{x}^n}\right]^T$$

$$\phi_{\ddot{x}}=\left[\frac{\partial\phi}{\partial \ddot{x}^1},\ldots,\frac{\partial\phi}{\partial \ddot{x}^n}\right]^T.$$

DEFINITION 1. (*Invexity*): If there exists vector function $\eta(t,u,\dot{u},\ddot{u},x,\dot{x},\ddot{x})\in R^n$ with $\eta=0$ and $x(t)=u(t),t\in I$ and $D\eta=0$ for $\dot{x}(t)=\dot{u}(t),t\in I$ such that for a scalar function $\phi(t,x,\dot{x},\ddot{x})$, the functional $\Phi(x,\dot{x},\ddot{x})=\int_I\phi(t,x,\dot{x},\ddot{x})dt$ satisfies

$$\Phi(x,u,\dot{u},\ddot{u})-\Phi(x,\dot{x},\ddot{x})\geqq$$
$$\int_I\left\{\eta^T\phi_x(t,x,\dot{x},\ddot{x})+(D\eta)^T\phi_{\dot{x}}(t,x,\dot{x},\ddot{x})+\right.$$
$$\left.\left(D^2\eta\right)^T\phi_{\ddot{x}}(t,x,\dot{x},\ddot{x})\right\}dt$$

Φ is said to be invex in $x,\dot{x}$ *and* $\ddot{x}$ on I with respect to η.

Here D is a differentiation operator defined later.

DEFINITION 2. (*Pseudoinvexity*): Φ is said to be pseudoinvex in $x,\dot{x}$ *and* $\ddot{x}$ with respect to η if

$$\int_I\left\{\eta^T\phi_x(t,x,\dot{x},\ddot{x})+(D\eta)^T\phi_{\dot{x}}(t,x,\dot{x},\ddot{x})\right.$$
$$\left.+\left(D^2\eta\right)^T\phi_{\ddot{x}}(t,x,\dot{x},\ddot{x})\right\}dt\geqq 0$$

implies $\Phi(x,u,\ddot{u})\geqq\Phi(x,\dot{x},\ddot{x})$.

DEFINITION 3. (*Quasi-invex*): The functional Φ is said to quasi-invex in $x,\dot{x}$ *and* $\ddot{x}$ with respect to η if $\Phi(x,u,\ddot{u})\leqq\Phi(x,\dot{x},\ddot{x})$ implies

$$\int_I\left\{\eta^T\phi_x(t,x,\dot{x},\ddot{x})+(D\eta)^T\phi_{\dot{x}}(t,x,\dot{x},\ddot{x})+\right.$$
$$\left.+\left(D^2\eta\right)^T\phi_{\ddot{x}}(t,x,\dot{x},\ddot{x})\right\}dt\leqq 0$$

3. Variational Problem and Optimality Conditions

Before stating our variational problem and deriving its necessary optimality condition, we mention the following conventions for vectors x and y in n-dimensional Euclidian space R^n will be used throughout the analysis of this research.

$$x<y,\quad\Leftrightarrow\quad x_i<y_i,\quad i=1,2,\ldots,n.\quad x$$
$$x\leqq y,\quad\Leftrightarrow\quad x_i\leqq y_i,\quad i=1,2,\ldots,n.$$
$$x\leq y,\quad\Leftrightarrow\quad x_i\leq y_i,\quad i=1,2,\ldots,n,\ \text{but}\ x\neq y$$
$$x\not\leq y,\ \text{is the negation of}\ x\leq y$$

For $x,y\in R$, $x\leq y$ and $x<y$ have the usual meaning.

In this section, we present the following variational problem whose optimality conditions will be derived and duality will be investigated in the subsequent sections:

(**VPE**) Minimize

$$\left(\int_I f^1(t,x,\dot{x},\ddot{x})dt,\ldots,\int_I f^p(t,x,\dot{x},\ddot{x})dt\right)$$

Subject to

$$x(a)=0=x(b) \tag{1}$$
$$\dot{x}(a)=0=\dot{x}(b) \tag{2}$$
$$g(t,x,\dot{x},\ddot{x})\leqq 0\ ,\quad t\in I \tag{3}$$
$$h(t,x,\dot{x},\ddot{x})=0\ ,\quad t\in I \tag{4}$$

where $f^i: I\times R^n\times R^n\times R^n\to R$, $i=1,2,\ldots,p$, $g: I\times R^n\times R^n\times R^n\to R^m$ and $h: I\times R^n\times R^n\times R^n\to R^k$ are continuously differentiable functions, and X designates the space of piecewise functions $x: I\to R^n$ possessing derivatives $\dot{x}$ and $\ddot{x}$ with the norm $\|x\|=\|x\|_\infty+\|Dx\|_\infty+\|D^2x\|_\infty$, where the differentiation operator D is given by

$$u = Dx \Leftrightarrow x(t) = \alpha + \int_a^t u(s)\,ds$$

where α is given boundary value; thus $D \equiv \frac{d}{dt}$ except at discontinuities.

In the results to follow, we use $C(I, R^m)$ to denote the space of continuous functions $\phi : I \to R^k$ with the uniform norm $\|\phi\| = \sup|\phi|_{t\in I}$; the partial derivatives of g and h are $m \times n$ and $k \times n$ matrices respectively; superscript T denotes matrix transpose.

We require the following definition of efficient solution for our further analysis.

DEFINITION 4. (*Efficient Solution*): A feasible solution $\bar{x}$ is efficient for (VPE) if there exist no other feasible x for (VPE) such that for some $i \in P = \{1,2,...,p\}$,

$$\int_I f^i(t,x,\dot{x},\ddot{x})\,dt < \int_I f^i(t,\bar{x},\dot{\bar{x}},\ddot{\bar{x}})\,dt$$

and $\int_I f^i(t,x,\dot{x},\ddot{x})\,dt \leqq \int_I f^i(t,\bar{x},\dot{\bar{x}},\ddot{\bar{x}})\,dt$ for all $j \in P$, $j \neq i$.

In relation to **(**VPE**)**, we introduce the following set of problems $\bar{P}_r$ for each $r = 1,2,\ldots,p$ in the spirit of [14], with a single objective,

$(\bar{P}_r)$ Minimize $\int_I f^r(t,x,\dot{x},\ddot{x})\,dt$

Subject to

$$x(a) = 0 = x(b),$$

$$\dot{x}(a) = 0 = \dot{x}(b),$$

$$g(t,x,\dot{x},\ddot{x}) \leqq 0 \quad , \quad t \in I,$$

$$h(t,x,\dot{x},\ddot{x}) = 0 \quad , \quad t \in I,$$

$$\int_I f^i(t,x,\dot{x},\ddot{x})\,dt \leqq \int_I f^i(t,\bar{x},\dot{\bar{x}},\ddot{\bar{x}})\,dt \ , \ i = 1,2,\ldots,p \ , \ i \neq r$$

The following lemma can be proved on the lines of Chankong and Haimes [14].

LEMMA 1: x^* is an efficient solution of (VPE) if and only if $\bar{x}$ is an optimal solution of $(\bar{P}_r)$ for each $r = 1,2,\ldots,p$.

(P$_0$) Minimize $\int_I \phi(t,x,\dot{x},\ddot{x})dt$

Subject to

$$x(a) = 0 = x(b),$$

$$\dot{x}(a) = 0 = \dot{x}(b),$$

$$g(t,x,\dot{x},\ddot{x}) \leqq 0 \quad , \quad t \in I,$$

$$h(t,x,\dot{x},\ddot{x}) = 0 \quad , \quad t \in I,$$

where $\phi : I \times R^n \times R^n \times R^n \to R$.

PROPOSITION 1. [9]: (*Fritz John Optimality Conditions*) If $\bar{x}$ is an optimal solution of (P$_0$) and $h_x(x(\cdot),\dot{x}(\cdot),\ddot{x}(\cdot))$ maps X into the subspace of $C(I,R^k)$, then there exists Lagrange multiplier $\bar{\tau} \in R$, the piecewise smooth $\bar{y} : I \to R^m$ and $\bar{z} : I \to R^k$, such that

$$\left(\bar{\tau}\phi_x + \bar{y}(t)^T g_x + \bar{z}(t)^T h_x\right) -$$

$$- D\left(\bar{\tau}\phi_{\dot{x}} + \bar{y}(t)^T g_{\dot{x}} + \bar{z}(t)^T h_{\dot{x}}\right)$$

$$+ D^2\left(\bar{\tau}\phi_{\ddot{x}} + \bar{y}(t)^T g_{\ddot{x}} + \bar{z}(t)^T h_{\ddot{x}}\right) = 0 \ , \ t \in I$$

$$\bar{y}(t)^T g(t,\bar{x},\dot{\bar{x}},\ddot{\bar{x}}) = 0 \ , \ t \in I,$$

$$(\bar{\tau}, \bar{y}(t)) \geqq 0 \ , \ t \in I,$$

$$(\bar{\tau}, \bar{y}(t), \bar{z}(t)) \neq 0 \ , \ t \in I.$$

If $\bar{\tau} = 1$, then the above optimality conditions will reduce to the Karush-Kuhn-Tucker type optimality conditions and the solution $\bar{x}$ is referred to as a normal solution.

We now establish the following theorem that gives the necessary optimality conditions for (VPE).

THEOREM 1: (*Fritz-John Conditions*): Let $\bar{x}$ be an efficient solution of **(**VPE**)** and $h_x(x(\cdot),\dot{x}(\cdot),\ddot{x}(\cdot))$ maps X into the subspace of $C(I,R^k)$, then there exist $\bar{\lambda} \in R^k$ and the piecewise smooth $\bar{y} : I \to R^m$ and $\bar{z} : I \to R^k$, such that

$$\left(\bar{\lambda}^T f_x + \bar{y}(t)^T g_x + \bar{z}(t)^T h_x\right) - D\left(\bar{\lambda}^T f_{\dot{x}} + \bar{y}(t)^T g_{\dot{x}} + \bar{z}(t)^T h_{\dot{x}}\right)$$

$$+ D^2\left(\bar{\lambda}^T f_{\ddot{x}} + \bar{y}(t)^T g_{\ddot{x}} + \bar{z}(t)^T h_{\ddot{x}}\right) = 0 \ , \ t \in I, \quad (5)$$

$$\bar{y}(t)^T g(t,\bar{x},\dot{\bar{x}},\ddot{\bar{x}}) = 0 \ , \ t \in I, \quad (6)$$

$$(\bar{\lambda}, \bar{y}(t)) \geqq 0 \ , \ t \in I. \quad (7)$$

$$(\lambda, \bar{y}(t), \bar{z}(t)) \neq 0 \ , \ t \in I. \quad (8)$$

PROOF: Since $\bar{x}$ is an efficient solution of **(**VPE**)** by Lemma 1, $\bar{x}$ is an optimal solution of ($\bar{P}_r$), for each $r = 1,2,\ldots,p$. From Proposition1, it follows that, there exist scalars $\bar{\lambda}^{1r}, \bar{\lambda}^{2r}, \ldots, \bar{\lambda}^{pr}$ and piecewise smooth function $\bar{y} : I \to R^m$ and $\bar{z} : I \to R^k$, such that

$$\bar{\lambda}^{rr} f_x^r + \sum_{\substack{i=1 \\ i\neq r}}^{p} \bar{\lambda}^{ir} f_x^i + \sum_{j=1}^{m} \bar{y}^{jr}(t) g_x^j + \sum_{l=1}^{k} \bar{z}^{lr}(t) h_x^l$$

$$-D\left(\bar{\lambda}^{rr} f_{\dot{x}}^r + \sum_{\substack{i=1 \\ i\neq r}}^{p} \bar{\lambda}^{ir} f_{\dot{x}}^i + \sum_{j=1}^{m} \bar{y}^{jr}(t) g_{\dot{x}}^j + \sum_{l=1}^{k} \bar{z}^{lr}(t) h_{\dot{x}}^l\right)$$

$$+D^2\left(\bar{\lambda}^{rr} f_{\ddot{x}}^r + \sum_{\substack{i=1 \\ i\neq r}}^{p} \bar{\lambda}^{ir} f_{\ddot{x}}^i + \sum_{j=1}^{m} \bar{y}^{jr}(t) g_{\ddot{x}}^j + \sum_{l=1}^{k} \bar{z}^{lr}(t) h_{\ddot{x}}^l\right) = 0 ,$$

$$t \in I,$$

$$\bar{y}^T(t) g\left(t, \bar{x}, \dot{\bar{x}}, \ddot{\bar{x}}\right) = 0 , t \in I ,$$

$$\left(\bar{\lambda}^{1r}, \bar{\lambda}^{2r}, \ldots, \bar{\lambda}^{pr}, \bar{y}^{1r}(t), \bar{y}^{2r}(t), \ldots, \bar{y}^{mr}(t)\right) \geqq 0 , \; t \in I$$

$$\left(\bar{\lambda}^{1r}, \bar{\lambda}^{2r}, \ldots, \bar{\lambda}^{pr}, \bar{y}^{1r}(t), \bar{y}^{2r}(t), \ldots, \bar{y}^{mr}(t), \bar{z}^{1r}(t), \bar{z}^{2r}(t), \ldots, \bar{z}^{lr}(t)\right) \neq 0 , \; t \in I .$$

Summing over *r*, we have

$$\sum_{r=1}^{p}\left(\sum_{i=1}^{p} \bar{\lambda}^{ir}\right) f_x^i + \sum_{r=1}^{p}\left(\sum_{j=1}^{m} \bar{y}^{jr}(t)\right) g_x^j + \sum_{r=1}^{p}\left(\sum_{l=1}^{k} \bar{z}^{lr}(t)\right) h_x^l$$

$$-D\left(\sum_{r=1}^{p}\left(\sum_{i=1}^{p} \bar{\lambda}^{ir}\right) f_{\dot{x}}^i + \sum_{r=1}^{p}\left(\sum_{j=1}^{m} \bar{y}^{jr}(t)\right) g_{\dot{x}}^j + \sum_{r=1}^{p}\left(\sum_{l=1}^{k} \bar{z}^{lr}(t)\right) h_{\dot{x}}^l\right)$$

$$+D^2\left(\sum_{r=1}^{p}\left(\sum_{i=1}^{p} \bar{\lambda}^{ir}\right) f_{\ddot{x}}^i + \sum_{r=1}^{p}\left(\sum_{j=1}^{m} \bar{y}^{jr}\right) g_{\ddot{x}}^j + \sum_{r=1}^{p}\left(\sum_{l=1}^{k} \bar{z}^{lr}(t)\right) h_{\ddot{x}}^l\right)$$

$$= 0 , \; t \in I$$

$$\bar{y}^T(t) g\left(t, \bar{x}, \dot{\bar{x}}, \ddot{\bar{x}}\right) = 0 , \; t \in I ,$$

$$\left(\sum_{r=1}^{p} \bar{\lambda}^{1r}, \ldots, \sum_{i=1}^{p} \bar{\lambda}^{pr}; \sum_{i=1}^{p} \bar{y}^{1r}(t), \ldots, \sum_{i=1}^{p} \bar{y}^{mr}(t)\right) \geqq 0 , \; t \in I$$

$$\left(\sum_{r=1}^{p} \bar{\lambda}^{1r}, \ldots, \sum_{i=1}^{p} \bar{\lambda}^{pr}; \sum_{i=1}^{p} \bar{y}^{1r}(t), \ldots, \sum_{i=1}^{p} \bar{y}^{mr}(t); \sum_{r=1}^{p} \bar{z}^{1r}(t), \ldots, \sum_{r=1}^{p} \bar{z}^{lr}(t)\right) \neq 0 , \; t \in I .$$

Setting $\bar{\lambda}^i = \sum_{r=1}^{p} \bar{\lambda}^{ir}$, $\bar{y}^j(t) = \sum_{r=1}^{p} \bar{y}^{jr}(t)$, $t \in I$ and $\bar{z}^l(t) = \sum_{r=1}^{l} \bar{z}^{lr}(t)$, $t \in I$, we have

$$\left(\bar{\lambda}^T f_x + \bar{y}(t)^T g_x + \bar{z}(t)^T h_x\right) - D\left(\bar{\lambda}^T f_{\dot{x}} + \bar{y}(t)^T g_{\dot{x}} + \bar{z}(t)^T h_{\dot{x}}\right)$$

$$+ D^2\left(\bar{\lambda}^T f_{\ddot{x}} + \bar{y}(t)^T g_{\ddot{x}} + \bar{z}(t)^T h_{\ddot{x}}\right) = 0 , \quad t \in I$$

$$\bar{y}(t)^T g\left(t, \bar{x}, \dot{\bar{x}}, \ddot{\bar{x}}\right) = 0 , \; t \in I ,$$

$$\left(\bar{\lambda}, \bar{y}(t)\right) \geqq 0 , \; t \in I ,$$

$$\left(\bar{\lambda}, \bar{y}(t), \bar{z}(t)\right) \neq 0 , \; t \in I .$$

4. Mond-Weir Type Duality

In this section, we consider the following variational problem involving higher order derivatives, by suppressing the equality constraint in (VPE).

(**VP**) Minimize

$$\left(\int_I f^1(t, x, \dot{x}, \ddot{x}) dt, \ldots, \int_I f^p(t, x, \dot{x}, \ddot{x}) dt\right)$$

Subject to

$$x(a) = 0 = x(b)$$

$$\dot{x}(a) = 0 = \dot{x}(b)$$

$$g(t, x, \dot{x}, \ddot{x}) \leqq 0 \quad , \quad t \in I$$

We formulate the following Mond-Weir type dual to the problem (VP) and establish various duality results under invexity defined in the preceding section.

(**M-WD**) Maximize

$$\int_I f^1(t, u, \dot{u}, \ddot{u}) dt, \ldots, \int_I f^p(t, u, \dot{u}, \ddot{u}) dt$$

Subject to

$$x(a) = 0 = x(b) \tag{9}$$

$$\dot{x}(a) = 0 = \dot{x}(b) \tag{10}$$

$$\left(\lambda^T f_x + y(t)^T g_x\right) - D\left(\lambda^T f_{\dot{x}} + y(t)^T g_{\dot{x}}\right) + D^2\left(\lambda^T f_{\ddot{x}} + y(t)^T g_{\ddot{x}}\right) = 0 \tag{11}$$

$$\int_1 y(t)^T g(t, u, \dot{u}, \ddot{u}) dt \geqq 0 \tag{12}$$

$$y(t) \geqq 0 \quad , \quad t \in I \tag{13}$$

$$\lambda > 0 . \tag{14}$$

THEOREM 2. (*Weak Duality***):** Let $x \in X$ be feasible for (VP) and (u, λ, y) be feasible for (M-WD) if for allfeasible (x, u, λ, y) $\int_I \lambda^T f(t, u, \dot{u}, \ddot{u}) dt$ is pseudoinvex and $\int_I y(t)^T g(t, u, \dot{u}, \ddot{u}) dt$ is quasi-invex with respect to the same η.

Then,

$$\int_I f(t,x,\dot{x},\ddot{x})\,dt \not\leqq \int_I f(t,u,\dot{u},\ddot{u})\,dt .$$

PROOF: The relations $g(t,x,\dot{x},\ddot{x}) \leqq 0$, $y(t) \geqq 0$, $t \in I$ imply

$$\int_I y(t)^T g(t,x,\dot{x},\ddot{x})\,dt \leqq \int_I y(t)^T g(t,u,\dot{u},\ddot{u})\,dt$$

This, because of the quasi-invexity of $\int_I y(t)^T g(t,u,\dot{u},\ddot{u})\,dt$, implies that

$$\begin{aligned} 0 \geqq & \int_I \left\{ \eta^T y^T(t) g_u + (D\eta)^T y^T g_{\dot{u}} + (D^2\eta)^T y^T g_{\ddot{u}} \right\} dt \\ & = \int_I \eta^T y(t)^T g_u\,dt + \eta^T y(t)^T g_{\dot{u}} \Big|_{t=a}^{t=b} \\ & - \int_I \eta^T D y(t)^T g_{\dot{u}}\,dt + (D\eta)^T y(t)^T g_{\ddot{u}} \Big|_{t=a}^{t=b} \\ & - \int_I (D\eta)^T D y(t)^T g_{\ddot{u}}\,dt \end{aligned}$$

(By integration by parts)

Using the boundary conditions which gives $D\eta = 0 = \eta$ at $t = a, t = b$

$$\begin{aligned} & = \int_I \eta^T y(t)^T g_u\,dt - \int_I \eta^T D y(t)^T g_{\dot{u}}\,dt \\ & - \eta^T D y(t)^T g_{\ddot{u}} \Big|_{t=a}^{t=b} + \int_I \eta^T D^2 y(t)^T g_{\ddot{u}}\,dt \end{aligned}$$

(By integration by parts)

Using the boundary conditions which give $D\eta = 0 = \eta$ at $t = a, t = b$

$$\int_I \eta^T y(t)^T g_u\,dt - \int_I \eta^T D y(t)^T g_{\dot{u}}\,dt + \int_I \eta^T D^2 y(t)^T g_{\ddot{u}}\,dt \leqq 0$$

$$\int_I \eta^T \left(y(t)^T g_u - D y(t)^T g_{\dot{u}} + D^2 y(t)^T g_{\ddot{u}} \right) dt \leqq 0$$

From Equation (11) this yields,

$$\int_I \eta^T \left\{ \lambda^T f_x - D\lambda^T f_{\dot{x}} + D^2 \lambda^T f_{\ddot{x}} \right\} dt \geqq 0$$

This by integration by parts and then using boundary conditions gives,

$$\int_I \left\{ \eta^T \left(\lambda^T f_x \right) + (D\eta)^T \left(\lambda^T f_{\dot{x}} \right) + \left(D^2\eta \right)^T \left(\lambda^T f_{\ddot{x}} \right) \right\} dt \geqq 0 ,$$

This, in view of psedoinvexity of $\int_I \lambda^T f(t,x,\dot{x},\ddot{x})\,dt$ implies that

$$\lambda^T \int_I f(t,x,\dot{x},\ddot{x})\,dt \geqq \lambda^T \int_I f(t,u,\dot{u},\ddot{u})\,dt .$$

For this, it follows

$$\int_I f(t,x,\dot{x},\ddot{x})\,dt \not\leqq \int_I f(t,u,\dot{u},\ddot{u})\,dt .$$

THEOREM 3. (*Strong Duality*)**:** If $\bar{x}$ is a feasible solution for (VP) and assume that $\bar{x}$ is an efficient solution and for at least one i , $i \in P$, $\bar{x}$ satisfies a regularity condition for [7] for $\mathrm{P_k}(\bar{x})$**.**

Then there exists one $\bar{\lambda} \in R^p$, $\bar{y} \in R^m$ such that $(\bar{x}, \bar{y}, \bar{\lambda})$ is efficient for (VD). Further if the assumptions of Theorem 2 are satisfied, then $(\bar{x}, \bar{y}, \bar{\lambda})$ is an efficient solution of (VD).

PROOF: Since $\bar{x}$ is efficient solution by Lemma 1, it is an optimal solution of $\mathrm{P_k}(\bar{x})$**.** By Proposition 1, this implies that there exists $\lambda^T = (\lambda^1, \ldots, \lambda^p)$ and piecewise smooth $y : I \to R^m$ such that,

$$\begin{aligned} & \bar{\lambda}_k \left(f_x^k - D f_{\dot{x}}^k + D^2 f_{\ddot{x}}^k \right) + \sum_{i \neq k} \bar{\lambda}_i \left(f_x^i - D f_{\dot{x}}^i + D^2 f_{\ddot{x}}^i \right) \\ & + \left(y(t)^T g_x - D y(t)^T g_{\dot{x}} + D^2 y(t)^T g_{\ddot{x}} \right) = 0 \end{aligned} \tag{15}$$

$$\begin{aligned} & \left(\lambda^T f_x + y(t)^T g_x \right) - D \left(\lambda^T f_{\dot{x}} + y(t)^T g_{\dot{x}} \right) \\ & + D^2 \left(\lambda^T f_{\ddot{x}} + y(t)^T g_{\ddot{x}} \right) = 0 \ , \ t \in I \end{aligned} \tag{16}$$

$$\bar{y}(t)^T g\left(t, \bar{x}, \dot{\bar{x}}, \ddot{\bar{x}}\right) dt = 0 \quad , \ t \in I \tag{17}$$

$$\left(\bar{\lambda}, \bar{y}(t) \right) \geqq 0 \quad , \ t \in I \tag{18}$$

$$\left(\bar{\lambda}, \bar{y}(t) \right) \neq 0 \quad , \ t \in I \tag{19}$$

From (17), we have

$$\int_1 y(t)^T g(t,x,\dot{x},\ddot{x})\,dt = 0 \tag{20}$$

Equations (16), (17) and (18) imply that $(\bar{x}, \bar{\lambda}, \bar{y})$ is feasible for (M-WD). The equality of objective functional of primal and dual problems is obvious from their formulations. Efficiency of $(\bar{x}, \bar{\lambda}, \bar{y})$ is immediate from the application of Theorem 2.

As in [4], by employing chain rule in calculus, it can be easily seen that the expression $\left(\lambda^T f_x + y(t)^T g_x \right) - D\left(\lambda^T f_{\dot{x}} + y(t)^T g_{\dot{x}} \right) + D^2 \left(\lambda^T f_{\ddot{x}} + y(t)^T g_{\ddot{x}} \right)$,may be regarded as a function θ of variables $t, x, \dot{x}, \ddot{x}, \dddot{x}, y, \dot{y}, \ddot{y}$ and λ , where $\dddot{x} = D^3 x$ and $\ddot{y} = D^2 y$. That is, we can write

$$\theta\left(t,x,\dot{x},\ddot{x},\dddot{x},y,\dot{y},\ddot{y},\lambda\right)=\left(\lambda^T f_x+y(t)^T g_x\right)$$
$$-D\left(\lambda^T f_{\dot{x}}+y(t)^T g_{\dot{x}}\right)+D^2\left(\lambda^T f_{\ddot{x}}+y(t)^T g_{\ddot{x}}\right)$$

In order to prove converse duality between (VP) and (M-WD), the space X is now replaced by a smaller space X_2 of piecewise smooth thrice differentiable function $x:I\to R^n$ with the norm $\|x\|_\infty+\|Dx\|_\infty+\|D^2x\|_\infty+\|D^3x\|_\infty$. The problem (M-WD) may now be briefly written as, Minimize

$$\left(-\int_I f^1\left(t,x,\dot{x},\ddot{x}\right)dt,\ldots,-\int_I f^p\left(t,x,\dot{x},\ddot{x}\right)dt\right)$$

Subject to

$$x(a)=0=x(b)$$
$$\dot{x}(a)=0=\dot{x}(b)$$
$$\theta\left(t,x,\dot{x},\ddot{x},\dddot{x},y,\dot{y},\ddot{y},\lambda\right)=0$$
$$\int_1 y(t)^T g\left(t,x,\dot{x},\ddot{x}\right)dt\underset{=}{\geq}0$$
$$y(t)\underset{=}{\geq}0\ ,\quad t\in I$$

Consider $\theta\left(t,x(\cdot),\dot{x}(\cdot),\ddot{x}(\cdot),\dddot{x}(\cdot),y(\cdot),\dot{y}(\cdot),\ddot{y}(\cdot),\lambda\right)=0$ as defining a mapping $\psi:X_2\times Y\times R^p\to B$ where Y is a space of piecewise twice differentiable function and B is the Banach Space. In order to apply Theorem 1 to the problem (M-WD), the infinite dimensional inequality must be restricted. In the following theorem, we use ψ' to represent the Frèchèt derivative

$$\left[\psi_x(x,y,\lambda),\psi_y(x,y,\lambda),\psi_\lambda(x,y,\lambda)\right].$$

THEOREM 4. (*Converse Duality***):** Let D be an efficient solution with $x\in X_2$, $y\in Y_2$ and $\lambda^T\in R^p$ and ψ' have a (weak*) closed range hypothesis. Let f and g be twice continuously differentiable. Assume that

(H$_1$) $\int_I \lambda^T f\,dt$ be pseudoinvex and $\int_I y(t)^T g\,dt$ be quasi-invex with respect to same η.

(H$_2$)
$$\sigma(t)^T\left(\sigma(t)\theta_x-D\sigma(t)\theta_{\dot{x}}+D^2\sigma(t)\theta_{\ddot{x}}-D^3\sigma(t)\theta_{\dddot{x}}\right)=0$$
$$\Rightarrow\sigma(t)=0,\ t\in I.$$

(H$_3$) $f_x^i-Df_{\dot{x}}^i+D^2f_{\ddot{x}}^i$, $i=1,2,\ldots,p$ are linearly independent.

Then $\bar{x}$ is an efficient solution of (VP).

Proof: Since $\left(\bar{x},\bar{\lambda},\bar{y}\right)$ where $\bar{x}\in X$ and ψ' having a closed range, is an efficient solution of (M-WD), by Theorem 2, it implies that there exist $\alpha\in R$, $\gamma\in R$, $\eta\in R^p$ and piecewise smooth $\beta:R\to R^n$ and $\mu:R\to R^m$ satisfying the following conditions .

$$-\alpha\left(f_x-Df_{\dot{x}}+D^2f_{\ddot{x}}\right)$$
$$-\gamma\left(y(t)^T g_x-D\left(y(t)^T g_{\dot{x}}\right)+D^2\left(y(t)^T g_{\ddot{x}}\right)\right)$$
$$+\beta(t)^T\theta_x-D\left(\beta(t)^T\theta_{\dot{x}}\right)+D^2\left(\beta(t)^T\theta_{\ddot{x}}\right)$$
$$-D^3\left(\beta(t)^T\theta_{\dddot{x}}\right)=0 \tag{21}$$

$$\beta(t)^T\theta_y-D\left(\beta(t)^T\theta_{\dot{y}}\right)+D^2\left(\beta(t)^T\theta_{\ddot{y}}\right)-\gamma g-\mu(t)=0 \tag{22}$$

$$\beta(t)^T\left(f_x-Df_{\dot{x}}+D^2f_{\ddot{x}}\right)-\eta=0 \tag{23}$$

$$\gamma\int_1 y(t)^T g\left(t,x,\dot{x},\ddot{x}\right)dt=0 \tag{24}$$

$$\eta^T\lambda=0,\quad \mu(t)^T y(t)=0,\quad t\in I \tag{25}$$

$$\left(\alpha,\gamma,\eta,\mu(t)\right)\underset{=}{\geq}0\ ,\quad t\in I\ \text{ and}$$
$$\left(\alpha,\gamma,\eta,\mu(t),\beta(t)\right)\neq 0,\quad t\in I \tag{26}$$

Since $\lambda>0$, $\eta^T\lambda=0$, which implies $\eta=0$

This yields from (23)

$$\beta(t)^T\left(f_x-Df_{\dot{x}}+D^2f_{\ddot{x}}\right)=0 \tag{27}$$

Using the equality constraint (11) in (21), we have

$$-(\alpha-\gamma\lambda)^T\left(f_x-Df_{\dot{x}}+D^2f_{\ddot{x}}\right)+\beta(t)^T\theta_x$$
$$-D\left(\beta(t)^T\theta_{\dot{x}}\right)+D^2\left(\beta(t)^T\theta_{\ddot{x}}\right)-D^3\left(\beta(t)^T\theta_{\dddot{x}}\right)=0 \tag{28}$$

Postmultiplying Equation (21) by $\beta(t)$ and using (27) in (28) we get,

$$\beta(t)^T\left(\beta(t)^T\theta_{\dot{x}}\right)+D^2\left(\beta(t)^T\theta_{\ddot{x}}\right)-D^3\left(\beta(t)^T\theta_{\dddot{x}}\right)=0\ ,\ t\in I$$

This by hypothesis (H$_2$) implies $\beta(t)=0,\ t\in I$

Also from (28) we have

$$(\alpha-\gamma\lambda)^T\left(f_x-Df_{\dot{x}}+D^2f_{\ddot{x}}\right)=0$$

This, because of linear independence of $f_x^i-Df_{\dot{x}}^i+D^2f_{\ddot{x}}^i$, $i=1,2,\ldots,p$, gives

$$\alpha-\gamma\lambda=0 \tag{29}$$

Now suppose $\gamma=0$, then, from (22) and (29) we have $\mu(t)=0$, $t\in I$ and $\alpha=0$ respectively.

This implies $\left(\alpha,\beta(t),\gamma,\eta,\mu(t)\right)=0$, which is the

contradiction to $\left(\alpha,\beta(t),\gamma,\eta,\mu(t)\right)\neq 0$, $t\in I$.

Hence $\gamma>0$ and by (29) we have, $\alpha>0$.

The relation (22) in conjunction with $\beta(t)=0$, and $\mu(t)\geq 0$, $t\in I$ gives

$$g\left(t,x,\dot{x},\ddot{x}\right)\underset{=}{\leq}0 \quad , t\in I$$

This implies the feasibility of $\bar{x}$ for (VP) and its efficiency is evident from and application of Theorem 2.

5. Natural Boundary Values

The duality results obtained in the preceding sections can easily be extended to the multiobjective variational problems with natural boundary values rather than fixed end points.

Primal (P_1) Minimize

$$\left(\int_I f^1\left(t,x,\dot{x},\ddot{x}\right)dt,\ldots,\int_I f^p\left(t,x,\dot{x},\ddot{x}\right)dt\right)$$

Subject to

$$g\left(t,x,\dot{x},\ddot{x}\right)\underset{=}{\leq}0 \quad , \quad t\in I$$

Dual (D_1) Maximize

$$\left(\int_I f^1\left(t,x,\dot{x},\ddot{x}\right)dt,\ldots,\int_I f^p\left(t,x,\dot{x},\ddot{x}\right)dt\right)$$

Subject to

$$\left(\lambda^T f_x+y(t)^T g_x\right)-D\left(\lambda^T f_{\dot{x}}+y(t)^T g_{\dot{x}}\right)+D^2\left(\lambda^T f_{\ddot{x}}+y(t)^T g_{\ddot{x}}\right)=0 \quad t\in I,$$

$$y(t)^T g_{\dot{x}}=0 \text{ , at } t=a \text{ and } t=b,$$

$$y(t)^T g_{\ddot{x}}=0 \text{ , at } t=a \text{ and } t=b,$$

$$y(t)\underset{=}{\geq}0 \text{ , } t\in I.$$

6. Nonlinear Programming

If the problems (P_1) and (D_1) are independent of t, then they will reduce to the following multiobjective nonlinear programming problems studied in [15]

(NP): Minimize $f(x)$

Subject to

$$g(x)\underset{=}{\leq}0.$$

(ND): Maximize $f(x)$

Subject to

$$\lambda^T f_x+y^T g_x=0$$

$$\lambda>0, \quad y\underset{=}{\geq}0.$$

7. References

[1] R. Courant and D. Hilbert, "Methods of Mathematical Physics," Wiley, New York, Vol. 1, 1943.

[2] K. O. Friedrichs, "Ein Verfrahren der Variations-Rechung das Minimum eines Integrals Maximum eines Anderen Ausdruckes Dazustellan," Göttingen Nachrichten, 1929.

[3] M. A. Hanson, "Bonds for Functionally Convex Optimal Control Problems," *Journal of Mathematical Analysis and Applications*, Vol. 8, No. 1, February 1964, pp. 84-89.

[4] B. Mond and M. A. Hanson, "Duality for Variational Problems," *Journal of Mathematical Analysis and Applications*, Vol. 18, No. 2, May 1967, pp. 355-364.

[5] F. A. Valentine, "The Problem of Lagrange with Differential Inequalities as Added Side Conditions," *Contributions to the Calculus of Variations*, 1933-1937, University of Chicago Press, 1937, pp. 407-448.

[6] C. R. Bector, S. Chandra and I. Husain, "Generalized Concavity and Duality in Continuous Programming," *Utilitas Mathematica*, Vol. 25, 1984, pp. 171-190.

[7] S. Chandra, B. D. Craven and I. Husain, "A Class of Nondifferentiable Continuous Programming Problems," *Journal of Mathematical Analysis Applications*, Vol. 107**,** No. 1, April 1985, pp. 122-131.

[8] S. Chandra, B. D. Craven and I. Husain, "Continuous Programming Containing Arbitrary Norms," *Journal of Australian Mathematical Society* (*Series A*), Vol. 39, No. 1, 1985, pp. 28-38.

[9] I. Husain and Z. Jabeen, "On Variational Problems Involving Higher Order Derivatives," *Journal of Applied Mathematics and Computing*, Vol. 27, No. 1-2, March 2005, pp. 433-455.

[10] B. Mond and S. Chandra and I. Husain, "Duality of Variational Problems with Invexity," *Journal of Mathematical Analysis and Applications*, Vol. 134, No. 2, September 1988, pp. 322-328.

[11] C. R. Bector and I. H. Husain, "Duality for Multiobjective Variational Problems," *Journal of Mathematical Analysis and Applications*, Vol. 166, No. 1, 1 May 1992, pp. 214-224.

[12] X. H. Chen, "Duality for Multiobjective Variational Problems with Invexity" *Journal of Mathematical Analysis and Appliactions*, Vol. 203, No. 1, October 1996, pp. 236-253.

[13] B. Mond and I. Smart, "Duality with Invexity for a Class of Nondifferentiable Static and Continuous Programming Problems," *Journal of Mathematical Analysis and Applications*, Vol. 136, 1988, pp. 325-333.

[14] V. Chankong and Y. Y. Haimes, "Multiobjective Decision Making: Theory and Methodology," North Holland, New York, 1983.

[15] R. R. Egudo and M. A. Hanson, "Multiobjective Duality with Invexity," *Journal of Mathematical Analysis and Appliactions*, Vol. 126, No. 2, September 1987, pp. 469- 477.

Call for Papers

Communications and Network

ISSN 1949-2421 (print) ISSN 1947-3826 (online)
www.scirp.org/journal/cn

CN, an international journal, dedicates 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.

Editor in Chief

Dr. Yi Huang, The University of Liverpool, UK

Executive Editor in Chief

Prof. Renfa Li, Hunan University, China

All manuscripts submitted to CN must be previously unpublished and may not be considered for publication elsewhere at any time during CN's review period. Additionally, accepted ones will immediately appear online followed by printed in hard copy. The topics to be covered by Communications and Network include, but are not limited to:

- ◆ Applications and value-added services
- ◆ Coding, detection and modulation
- ◆ Cognitive radio
- ◆ Communication tools and services
- ◆ Location based services
- ◆ MIMO and OFDM technologies
- ◆ Mobile computing systems
- ◆ Multimedia in wireless networks
- ◆ Network operation, maintenance and management
- ◆ Network protocol, QoS and congestion control
- ◆ Network security
- ◆ QoS and traffic analysis
- ◆ RFID and 802.1x technologies
- ◆ Sensor networks
- ◆ Wave propagation and antenna design
- ◆ WDM, HWDM and OTDM networks

We are also interested in short papers (letters) that clearly address a specific problem, and short survey or position papers that sketch the results or problems on a specific topic. Authors of selected short papers would be invited to write a regular paper on the same topic for future issues of the CN.

Website and E-Mail

http://www.scirp.org/journal/cn **E-Mail:cn@scirp.org**

Communications and Network, **2010, 2, 87-144**
Published Online May 2010 in SciRes (http://www.scirp.org/journal/cn).

TABLE OF CONTENTS

Volume 2 Number 2 **May 2010**

