ISSN: 1945-3078 (Print), 1945-3086 (Online) Volume 2, Number 3, March 2010

Wireless Sensor Network

Chief Editor : Kosai Raoof

www.scirp.org/journal/wsn/

Scientific Research

Journal Editorial Board

ISSN 1945-3078 (Print) ISSN 1945-3086 (Online)

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

Wireless Sensor Network, **2010, 2, 199-266**
Published Online March 2010 in SciRes (http://www.SciRP.org/journal/wsn/)

TABLE OF CONTENTS

Volume 2 Number 3 March 2010

Wireless Sensor Network (WSN)

Journal Information

SUBSCRIPTIONS

The *Wireless Sensor Network* (Online at Scientific Research Publishing, www.SciRP.org) is published monthly by Scientific Research Publishing, Inc., USA.

Subscription rates:

Print: $50 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

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

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: wsn@scirp.org

Wireless Sensor Network, **2010, 2, 199-205**
doi:10.4236/wsn.2010.23026 Published Online March 2010 (http://www.scirp.org/journal/wsn)

A General Algorithm for Biorthogonal Functions and Performance Analysis of Biorthogonal Scramble Modulation System

Yueyun Chen[1,2], Zhenhui Tan[1]
[1]*State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China*
[2]*Beijing Science and Technology University, Beijing, China*
E-mail: *chenyy2@sina.com*
Received August 26, 2009; *revised September* 7, 2009; *accepted November* 6, 2009

Abstract

Applying the theorems of Mobius inverse and Dirichlet inverse, a general algorithm to obtain biorthogonal functions based on generalized Fourier series analysis is introduced. In the algorithm, the orthogonal function can be not only Fourier or Legendre series, but also can be any one of all orthogonal function systems. These kinds of biorthogonal function sets are used as scramble signals to construct biorthogonal scramble modulation (BOSM) wireless transmission systems. In a BOSM system, the transmitted signal has significant security performance. Several different BOSM and orthogonal systems are compared on aspects of BER performance and spectrum efficiency, simulation results show that the BOSM systems based on Chebyshev polynomial and Legendre polynomial are better than BOSM system based on Fourier series, also better than orthogonal MCM and OFDM systems.

Keywords: Biorthogonal Functions, Biorthogonal Scramble Modulation, Generalized, Fourier Series, Mobius Inverse, Wireless Transmission

1. Introduction

In wireless transmission systems, orthogonal signals are often used for transmission information. In order to reduce the complexity of receivers in orthogonal transmission system and improve the bandwidth efficiency, a kind of special biorthogonal modulation was adopted. The biorthogonal signals are consisted of two groups of orthogonal signals [1,2]. In group one, the M/2 functions are directly obtained from orthogonal function system; the M/2 functions in the other group are the inverse functions of group one's. Therefore, only half of the bandwidth and half correlators were required compared with M orthogonal functions modulation system. Biorthogonal CDMA is also helpful to improve spectrum efficiency, reduce multiple-access interference and complexity in receiver [3,4]. However, these systems also belong to orthogonal systems because all the biorthogonal signals are also orthogonal.

In fact, security is difficult to be guaranteed in orthogonal systems, because the signals used in modulation and de-modulation are same in transmitter and receiver (sometimes except the symbol "+" or "–"). Recently, a kind of biorthogonal functions which are no-orthogonal was discussed. In paper [5,6], Wei Yuchuan and Chen Nianxian analyzed several periodic waves based on Fourier series expansions with the theorem of Möbius inverse and the extended Möbius inverse theorem proposed by Chen Nanxian [7], pointed out that a periodic signal is able to be represented as the a superposition of some easily generated periodic functions (such as sawtooth wave and triangular wave) with different frequencies, and gave an algorithm to obtain the biorthogonal functions based on Fourier series analysis. The kind of biorthogonal functions is not orthogonal. However, the condition above is that the Fourier coefficients of the period functions are completely multiplicative. Subsequently, Su Wuxun in paper [8] gave similarly the inverse transform of some often-used symmetrical periodic waveforms based on Fourier series analysis, and suggested a digital communication system architecture using biorthogonal functions as multi-carriers in paper [9]. However, the condition of completely multiplicative Fourier coefficients limited the space of biorthogonal functions.

In paper [10], we introduced an algorithm to obtain biorthogonal functions based on Fourier-Legendre series

analysis, and proposed a wireless transmission system of biorthogonal scramble modulation (BOSM) based on such kind of biorthogonal function sets. It was pointed out that in flat fading channel, the BOSM systems based on Fourier series analysis have the close BER performance, and spectrum efficiency is similar to MCM system; but the system based on Fourier-Legendre series analysis has higher spectrum efficiency better BER performance than other systems.

In this paper, we introduce a universal algorithm to obtain biorthogonal functions based on General Fourier series analysis. The algorithm doesn't need the condition that the series coefficients are completely multiplicative. In the meantime, the orthogonal function can be not only Fourier or Legendre series, but also can be any one of all orthogonal function systems. Then we propose a BOSM (biorthogonal scramble modulation) system using some of these new biorthogonal functions. On the condition of flat fading channel, we analyze and compare the BER performance and spectrum efficiency of different BOSM and orthogonal systems. The simulating results show that the BOSM based on General Fourier series analysis not only has better BER performance, but also has higher spectrum efficiency than the BOSM system based on Fourier series analysis.

2. The Fundamental Theorems of Möbius Inverse Transform and Dirichlet Inverse

Möbius inverse transform and Dirichlet inversion are helpful theorems. The related theorems are briefly introduced in this section.

A basic Möbius inverse formula is described as follows. If $f(n)$ is a number-theoretic function and

$$F(n)=\sum_{d|n} f(d) \tag{1}$$

then, for all positive integers n, there exists

$$f(n)=\sum_{d|n}\mu(d)F(\frac{n}{d}) \tag{2}$$

Inversely, if Equation (2) exists, then Equation (1) can be yielded. In above equation, $\mu(n)$ is a Mobius function defined as

$$\mu(n)=\begin{cases}1, & n=1\\ (-1)^r & \text{if all the primes of } n \text{ are different}\\ 0, & \textit{if } n \text{ has a squared factor}\end{cases}$$

If *f*(n) and $g(n)$ are number-theoretic functions, then their Dirichlet multiplication $h(n)$ is also a number-theoretic functions [11]

$$h(n)=\sum_{d|n} f(n)g\left(\frac{n}{d}\right)\ \ (n=1,2,3,\cdots) \tag{3}$$

and it can be written as

$$h(n)=f(n)*g(n) \tag{4}$$

If $f(1)\neq 0$, there exists a unique number-theoretic function $f^{-1}(n)$ satisfying

$$f(n)*f^{-1}(n)=f^{-1}(n)*f(n)=\delta_{n-1}$$

where δ_{n-1} is Kroneche's delta symbol and $f^{-1}(n)$ is called Dirichlet inverse of $f(n).f^{-1}(n)$ can be calculated by recurrence formula as following

$$f^{-1}(n)=\frac{1}{f(1)}\ \ (\text{n = 1}) \tag{5}$$

$$f^{-1}(n)=\frac{1}{f(1)}\sum_{\substack{d|n\\ d<n}} f\left(\frac{n}{d}\right)f^{-1}(d)\ \ (n>1) \tag{6}$$

Specially, if $f(x)$ is complete multiplicative, then

$$f^{-1}(n)=\mu(n)\,f(n) \tag{7}$$

Following is a useful Mobius inverse formula. It is expressed as follows [8,12]

$$G(x)=\sum_{n=1}^{\infty} r(n)g(nx)\ \ \Leftrightarrow\ \ g(x)=\sum_{n=1}^{\infty} r^{-1}(n)G(nx) \tag{8}$$

where $G(x)$ and $g(x)$ are two functions defined in $(-\infty,\infty)$ and $r^{-1}(d)$ is Dirichlet inverse of $r(d)$.

3. Fourier Series Analysis

Function $f(x)$ belongs to function space $L^2[-\pi,\pi]$ with period 2π. It has a unique Fourier series expressed as

$$f(t)=a(0)+\sum_{n=1}^{\infty}\left[a(n)\cos(nt)+b(n)\sin(nt)\right] \tag{9}$$

where, $a(0)$, $a(n)$ and $b(n)$ are the Fourier series coefficients of $f(t)$.

Supposing that $u(t)$ and $v(t)$ are respectively an even and odd functions with period of 2π, their Fourier series is expressed respectively as

$$u(t)=\sum_{n=1}^{\infty} A(n)\cos(n\omega t)$$

$$v(t)=\sum_{n=1}^{\infty} B(n)\sin(n\omega t)$$

By the theorem of Möbius inverse [12], Equation (9) can be rewritten as the following form

$$\begin{aligned} f(t) &= a(0)+\sum_{n=1}^{\infty}\left[a(n)\sum_{m=1}^{\infty}A^{-1}(m)u(mnt)\right]+\sum_{n=1}^{\infty}\left[b(n)\sum_{m=1}^{\infty}B^{-1}(m)v(mnt)\right] \\ &= a(0)+\sum_{k=1}^{\infty}\sum_{d|k}^{\infty}a(d)A^{-1}\left(\frac{k}{d}\right)u(kt)+\sum_{k=1}^{\infty}\sum_{d|k}^{\infty}b(d)B^{-1}\left(\frac{k}{d}\right)v(kt) \\ &= a(0)+\sum_{k=1}^{\infty}c(k)u(kt)+\sum_{k=1}^{\infty}d(k)v(kt) \end{aligned} \tag{10}$$

where

$$\begin{aligned} c(k) &= \sum_{d|k}^{\infty}a(d)A^{-1}\left(\frac{k}{d}\right) \\ &= \sum_{d|k}^{\infty}\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)\cos(dt)dt\cdot A^{-1}\left(\frac{k}{d}\right) \\ &= \int_{-\pi}^{\pi}f(t)g_k(t)dt \end{aligned} \tag{11}$$

$$\begin{aligned} d(k) &= \sum_{d|k}^{\infty}b(d)B^{-1}\left(\frac{k}{d}\right) \\ &= \sum_{d|k}^{\infty}\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)\sin(dt)dt\cdot B^{-1}\left(\frac{k}{d}\right) \\ &= \int_{-\pi}^{\pi}f(t)h_k(t)dt \end{aligned} \tag{12}$$

Equation (10) shows that a quadratically integrabel function can be decomposed as the superposition of even function $u(t)$ and $v(t)$ and odd function $v(t)$ with different frequencies. In Equation (11) and Equation (12), $g_k(t)$ and $h_k(t)$ is given respectively by

$$g_k(t) = \frac{1}{\pi}\sum_{d|k}^{\infty}A^{-1}\left(\frac{k}{d}\right)\cos dt \tag{13}$$

$$h_k(t) = \frac{1}{\pi}\sum_{d|k}^{\infty}B^{-1}\left(\frac{k}{d}\right)\sin dt \tag{14}$$

It can be proved that

$$\int_{-\pi}^{\pi}u(mt)g_n(t)dt = \delta_{nm} \tag{15}$$

$$\int_{-\pi}^{\pi}v(mt)h_n(t)dt = \delta_{nm} \tag{16}$$

namely, $g_k(t)$ and $u(mt)=\sum_{n=1}^{\infty}A(n)\cos(mnt)$ are biorthogonal functions, the same are $h_k(t)$ and $v(mt)=\sum_{n=1}^{\infty}B(n)\sin(mnt)$. In Equation (10), $u(t)$ and $v(t)$ are called the basic function of biorthogonal functions. In fact, the proof of Equation (15) and Equation (16) do not need the condition that Fourier coefficients $A(n)$ or $B(n)$ of $u(t)$ or $v(t)$ are completely multiplicative. When $A(n)$ or $B(n)$ are not completely multiplicative, *i.e.*, $A^{-1}(n)\neq\mu(n)A(n)$ or $B^{-1}(n)\neq\mu(n)B(n)$, the conclusion of Equation (15) and Equation (16) can also be obtained.

4. General Fourier Series Analysis

Supposing $P_n(x)$ is an infinite orthogonal series with weight function $h(x)$ in the orthogonal interval $[a,b]$, namely

$$\int_a^b h(x)P_m(x)P_n(x)dx = \begin{cases}0, & m\neq n\\ 1 & m\neq n\end{cases} \tag{17}$$

where weight function $h(x)$ is nonnegative and integrable in the interval $[a,b]$ [13]. If $f(x)$ is a function defined in interval $[a,b]$ and satisfies $\int_a^b h(x)f^2(x)dx<\infty$, then it can be decomposed to generalized Fourier series as

$$f(x) = \sum_{n=0}^{\infty}a(n)P_n(x) \tag{18}$$

The coefficient of generalized Fourier series $a(n)$ is defined by

$$a(n) = \int_a^b h(x)f(x)P_n(x)dx \tag{19}$$

Then, a more universal algorithm to obtain biorthogonal functions is introduced by following proposition.

Proposition 1. Supposing $P_n(x)$ is orthogonal series with weighted function $h(x)$ in orthogonal interval $[a,b]$. If $f(x)$ is a function defined in interval $[a,b]$ and satisfies the relation of $\int_a^b h(x)f^2(x)dx<\infty$ (*i.e.* $f(x)$ belongs to $[a,b;h(x)]$). Then, the two group functions

$$F_m(x) = \sum_{n=1}^{\infty}a(n)P_{nm}(x),\quad nm\le K \tag{20}$$

and

$$\tilde{F}_n(x) = \sum_{d|k}a^{-1}\left(\frac{n}{d}\right)h(x)Pd(x) \tag{21}$$

is a biorthogonal function set, namely

$$\int_a^b F_m(x)\tilde{F}_n(x)dx = \delta_{mn} \tag{22}$$

where $a(n)$ is the coefficient of generalized Fourier series of $f(x)$, $a^{-1}(n)$ is its Dirichlet inverse, and K is the order of biorthogonal function set.

Proof.

$$\begin{aligned}\int_a^b F_m(x)\cdot\tilde{F}_n(x)dx &= \int_a^b \sum_{i=1}^{\infty} a(i)P_{im}(x)\cdot\sum_{d|n} a^{-1}\left(\frac{n}{d}\right)h(x)Pd(x)dx \\ &= \sum_{i=1}^{\infty} a(i)\sum_{d|n} a^{-1}\left(\frac{n}{d}\right)\cdot\delta_{im,d} \\ &= \sum_{\substack{d|n\\ m|n}} a\left(\frac{d}{m}\right)a^{-1}\left(\frac{n}{d}\right) \\ &= \sum_{\substack{d|n\\ m|n}} a\left(\frac{d}{m}\right)a^{-1}\left(\frac{n/m}{d/m}\right)=\delta_{mn}\end{aligned}$$

The proof is completed.

In Proposition 1, $P_n(x)$ represents any a kind of orthogonal series expanded in orthogonal function system $\{P_n(x)\}$. With the difference of $h(x)$, $P_n(x)$ can be different orthogonal polynomials series, such as Legendre polynomials, Chabyshev polynomials, Hermite polynomials, Laguarre polynomials, etc. The particular case is Fourier series when $h(x)=1$ and $P_n(x)$ is trigonometric function. Therefore, the algorithm given by Equation (20) and Equation (21) is a general algorithm to obtain biorthogonal functions. In fact, the process of biorthogonalizing corresponds to rotate the orthogonal polynomials coordinate axis, thus one function in bi-orthogonal function set has projection on one or more axis.

There is a fast algorithm for biorthogonal functions. Functions (20) and (21) can be rewritten as matrix forms as

$$\begin{aligned}F' &= \left[F_1'(x)\ \ F_2'(x)\ \ F_3'(x)\ \ F_4'(x)\ \cdots\ F_K'(x)\right]^T \\ &= \begin{bmatrix} a(1) & a(2) & a(3) & a(4) & a(5) & a(6) & \cdots & a(K) \\ 0 & a(1) & 0 & a(2) & 0 & a(3) & \cdots & \\ 0 & 0 & a(1) & 0 & 0 & a(2) & \cdots & \\ 0 & 0 & 0 & a(1) & 0 & 0 & \cdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & 0 & 0 & & \cdots & a(1) \end{bmatrix}\begin{bmatrix} P_1(x) \\ P_2(x) \\ P_3(x) \\ P_4(x) \\ \vdots \\ P_K(x)\end{bmatrix} \\ &= A'P(x)\end{aligned} \tag{23}$$

$$\begin{aligned}\tilde{F}' &= \left[\tilde{F}_1'(x)\ \ \tilde{F}_2'(x)\ \ \tilde{F}_3'(x)\ \ \tilde{F}_4'(x)\ \cdots\ \tilde{F}_K'(x)\right]^T \\ &= h(x)\begin{bmatrix} a^{-1}(1) & 0 & 0 & 0 & \cdots & 0 \\ a^{-1}(2) & a^{-1}(1) & 0 & 0 & \cdots & 0 \\ a^{-1}(3) & 0 & a^{-1}(1) & 0 & \cdots & 0 \\ a^{-1}(4) & a^{-1}(2) & 0 & a^{-1}(1) & \cdots & 0 \\ a^{-1}(5) & 0 & 0 & 0 & \cdots & 0 \\ a^{-1}(6) & a^{-1}(3) & a^{-1}(2) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ a^{-1}(K) & \cdots & \cdots & \cdots & \cdots & a^{-1}(1)\end{bmatrix}\begin{bmatrix} P_1(x) \\ P_2(x) \\ P_3(x) \\ P_4(x) \\ \vdots \\ P_K(x)\end{bmatrix} \\ &= h(x)B'P(x)\end{aligned} \tag{24}$$

where K is called the order of biorthogonal function set. If matrix A and B is normalized individually by $a(1)$ and $a^{-1}(1)$, Equation (23) and Equation (24) are rewritten as

$$F = AP(x) \tag{25}$$

$$\tilde{F} = h(x)BP(x) \tag{26}$$

where $A = A'/a(1)$, $B = B'/a^{-1}(1)$. Then, matrix A and B are sparse triangular matrixes and they are transposes each other (regardless the upper-mark "-1"). Therefore, the calculation speed of the algorithm method is improved rapidly.

Following takes Chabyshev polynomials as an example. Chabyshev polynomials is defined in orthogonal interval [-1,1] with weight function $h(x)=1/\sqrt{1-x^2}$. The series can be expressed as

$$T_n(x)=\cos(n\arccos x)\ \ (n=0,1,2,\cdots) \tag{27}$$

If $f(x)=1\mathrm{n}(1+x)=-1\mathrm{n}2+2\sum_{n=1}^{\infty}\frac{(-1)^n}{n}T_n(x)$, then $a(n)=2(-1)^{n-1}/n,(n\geq 1)$, and $a(n)^{-1}$ can be calculated by Equation (6) and Equation (7). Let $K=6$, then the biorthogonal functions can be directly obtained as follows

$$\begin{bmatrix} F_1(x) \\ F_2(x) \\ F_3(x) \\ F_4(x) \\ F_5(x) \\ F_6(x)\end{bmatrix} = \begin{bmatrix} 1 & -1/2 & 1/3 & -1/4 & 1/5 & -1/6 \\ 0 & 1 & 0 & -1/2 & 0 & -1/4 \\ 0 & 0 & 1 & 0 & 0 & -1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix} = \begin{bmatrix} P_1(x) \\ P_2(x) \\ P_3(x) \\ P_4(x) \\ P_5(x) \\ P_6(x)\end{bmatrix} \tag{28}$$

$$\begin{bmatrix} \tilde{F}_1(x) \\ \tilde{F}_2(x) \\ \tilde{F}_3(x) \\ \tilde{F}_4(x) \\ \tilde{F}_5(x) \\ \tilde{F}_6(x)\end{bmatrix} = \frac{1}{\sqrt{1-x^2}}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 0 & 0 & 0 \\ -1/3 & 0 & 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 1 & 0 & 0 \\ -1/5 & 0 & 0 & 0 & 1 & 0 \\ -1/6 & -1/3 & 1/2 & 0 & 0 & 1\end{bmatrix} = \begin{bmatrix} P_1(x) \\ P_2(x) \\ P_3(x) \\ P_4(x) \\ P_5(x) \\ P_6(x)\end{bmatrix} \tag{29}$$

It can be proved that functions $\{F_m(x)\}$ and $\{\tilde{F}_m(x)\}$ are biorthogonal, and the functions in one group are not orthogonal. So, this is a kind of more general biorthogonal functions.

5. Transmission System of BOSM

The biorthogonal function sets discussed above have a magnetic feature that the functions in same one group are

not orthogonal in time and frequency region. If the functions of one group are mixed up, it is very difficult to apart each one of them from the mixture if without the other group of biorthogonal functions. Profiting from this feature, a wireless security transmission system based this kind of comprehensive biorthogonal function can be constructed. Using the biorthogonal function sets as scramble modulation signal, the systems have outstanding security performance. We call the system as Bi-Orthogonal Scrambling Modulation (BOSM) transmission system. The principle diagram is shown in **Figure 2**.

The input sequence d_i is converted to K parallel sub-sequences $d_{pi}\,(1 \le i \le K)$ which is individually scrambled by the biorthogonal signal $\{F_1(x), F_2(x), \cdots, F_K(x)\}$ called BOSS (biorthogonal scramble signal) and the variable x is treated as time t. The transmitted symbol $s(t)$ is performed by mixing all the outputs of K parallel branches. It is expressed as

$$s(t) = \sum_{i=1}^{K} d_{pi}(t) F_i(t) \tag{30}$$

Supposing that the mixed signal $s(t)$ is transmitted over flat fading channel and the transmitted signal is compensated effectively in receiver. Then, the received signal $r(t)$ is expressed as

$$r(t) = s(t) + n(t) = \sum_{i=1}^{K} d_{pi}(t) F_i(t) + n(t) \tag{31}$$

where $n(t)$ is AWGN with $mean = 0$, variance σ_0^2. Further, supposing that the scramble signal in receiver has been already synchronized with transmitter's. The received signal $r(t)$ is correlated individually by $\{\tilde{F}_1(t), \tilde{F}_2(t), \cdots, \tilde{F}_L(t)\}$, and the output of the j-th correlator is

$$d'_{pi} = \int_0^T r(t)\tilde{F}_j(t)\,dt = \sum_{i=1}^{K} d_{pi}(t) \int_0^T F_i(t)\tilde{F}_j(t)\,dt + \int_0^T n(t)\tilde{F}_j(t)\,dt = \begin{cases} \alpha_i d_{pi}(t) & i = j \\ n'(t) & i \ne j \end{cases} \tag{32}$$

where α_i is a constant. When the correlating process is finished, the de-scramble process is completed also.

6. Simulation Analysis for Transmission Performance of BOSM System

In this section, we discuss BER performance, spectrum efficiency and the impact of synchronization precision in BOSM system. The simulation system is established according to **Figure 1**. Considering the scene of flat fading channel in narrow band condition, let $K = 6$ and symbol rate $R_s = 4800$ baud/s. In receiver, the decision criterion is ML (maximum likelihood). Because parallel transmission depresses ISI (inter-symbol interference) remarkably, the BER performance is mainly affected by channel and ICI (inter-channel interference).

We compare different BOSM and orthogonal systems on the performance of BER and spectrum efficiency. They are based on different BOSS which are obtained respectively by Chebyshev polynomial, Legendre polynomial (general Fourier series) and Fourier series analysis.

The basic function for Chebyshev polynomial analysis is the example given in Section 4. For Legendre polynomial analysis, the basic function is as following

$$f(x) = \begin{cases} 1, & 0 < x \le 1 \\ -1, & -1 \le x < 0 \end{cases} \tag{33}$$

An even symmetrical trapezoid is used as the basic function for Fourier series analysis, expressed as

$$Tra(t) = \begin{cases} t + \pi & -\pi \le t \le -\pi/2 \\ \pi/2 & -\pi/2 \le t \le \pi/2 \\ -t + \pi & \pi/2 \le t \le \pi \end{cases} \tag{34}$$

Correspondingly, the obtained biorthogonal functions sets are marked as BOSS-C, BOSS-L, and BOSS-F, individually. **Figure 2** shows the transmitted signal waveform of the example of BOSS-C. It seems like noise.

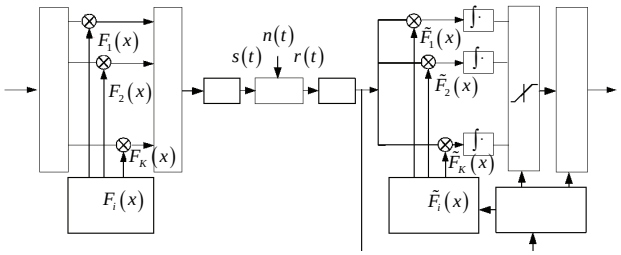


Figure 1. Block diagram of baseband BOSM system.

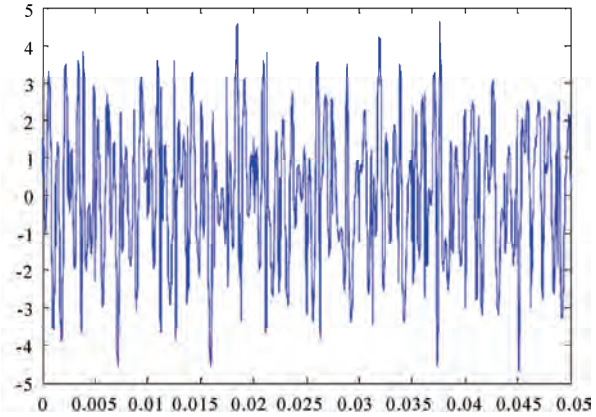


Figure 2. Waveform of mixed signal.

6.1. Spectrum Efficiency

In BOSS-F, the bandwidth of mixed signal is $B \approx 14kHz$ (see **Figure 3**), the spectrum efficiency is 0.34 baud/Hz approximately, and equal to orthogonal MCM (Multiple Carriers Modulation) system regardless the guard band.

In BOSS-C and BOSS-L, the PSD (power spectrum density) of mixed signal is shown respectively in **Figures 4(a)** and **4(b)**. The energy of mixed signal is mainly concentrated in interval of 0~3.2kHz and 0~4.8kHz, and the spectrum efficiency approximately is 1 baud/Hz and 1.5 baud/Hz respectively. In fact, the PSD of mixed signals are changed a bit with the change of basic function $f(t)$.

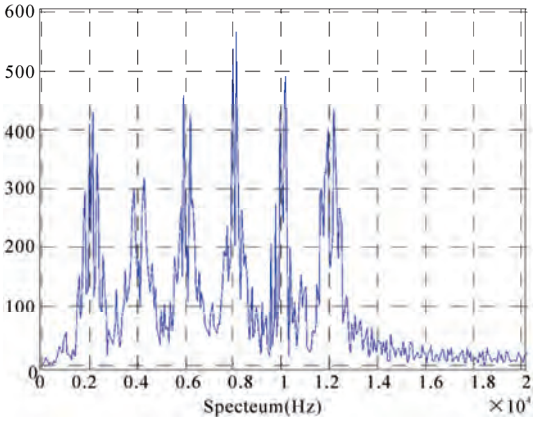


Figure 3. Spectrum of mixed signal for BOSS-F.

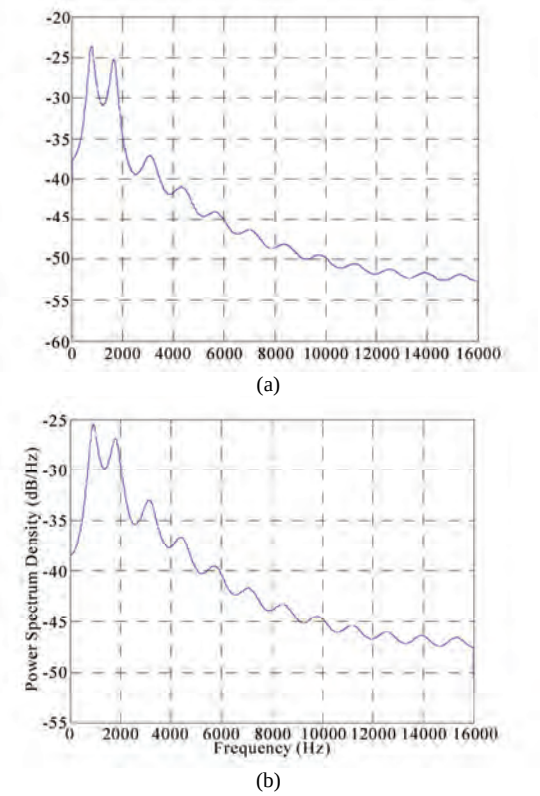


Figure 4. Single-side PSD of mixed signal. (a) BOSS-C; (b) BOSS-L.

In OFDM system, the spectrum efficiency is improved approaching 1 times than orthogonal MCM system. Compared with OFDM system, the BOSM system based on 10 Legendre polynomial analyses has similar performance of spectrum efficiency. Therefore, the BOSM systems based on general Fourier series analysis, including Chebyshev series and Legendre series analysis, have higher spectrum efficiency.

6.2. BER Performance

In paper [10], we had given a conclusion that the BOSM systems based on Fourier series with different basic functions have close BER performance. Supposing synchronization is exactly established. The BER performance of different systems is shown in **Figure 5**. It reveals that, in flat fading channel, the BER performance of BOSS-L and BOSS-C systems are the best, and BOSS-F and orthogonal-M systems (or OFDM system) have the close BER performance. Therefore, the BER performance of the BOSM systems based on general Fourier series analysis is satisfying.

6.3. Synchronization Property

Synchronization error of biorthogonal signals between transmitter and receiver leads to ICI. Supposing the error is $\triangle\tau$, then, the output of j-th correlator is ($n(t)$ is omitted)

$$\begin{aligned} d'_{pi} &= \sum_{i=1}^{K} d_{pi}(t) \int_0^T F_i(t)\tilde{F}_j(t+\triangle\tau)dt \\ &= \sum_{i=1}^{K} d_{pi}(t) \int_0^T \sum_{k=1}^{K} A_{ik}P_i \sum_{l=1}^{K} B_{jl}P_j(t+\triangle\tau)dt \\ &= \sum_{i=1}^{K} d_{pi}(t) \sum_{k=1}^{K}\sum_{l=1}^{K} A_{ik}B_{jl} \int_0^T P_i(t)P_j(t+\triangle\tau)dt \end{aligned} \quad (35)$$

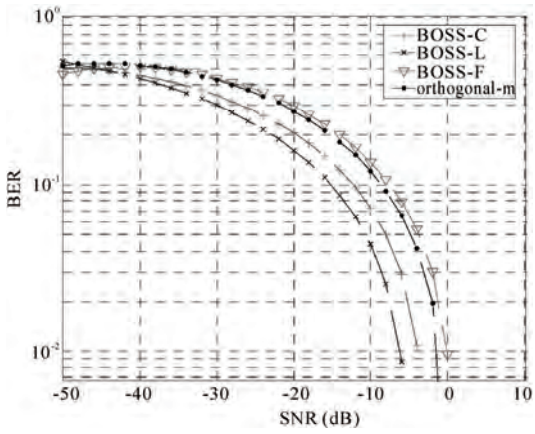


Figure 5. Spectrum of mixed signal.

In above expression, the output is impacted by all other K-1 channels, because the biorthogonal feature is damaged in different degree with different $\Delta\tau$. When using the example of BOSS-C, **Figure 6** shows the impact of synchronization error to BER performance. The BER curve is changed periodically with the synchronization error, and the period is the length of interval [a,b] (sample number 40 represents the interval length). Therefore, synchronization precision needed to be guarantied in BOSM system, just same as OFDM systems.

7. Conclusions

The biorthogonal functions were introduced into ordinary orthogonal function systems, and pointed out that the condition of complete multiplication is not needed. In the meantime, a more general algorithm for biorthogonal functions called general Fourier series analysis was proposed. Because the obtained biorthogonal functions are not orthogonal each other, the BOSS that the obtained biorthogonal functions were used as scramble signal to be modulated by transmitted symbols has outstanding security performance, especially when the order is enough big. This is very different from orthogonal system and traditional biorthogonal modulation system. By simulation analysis, in flat fading channel, the BOSM systems based on general Fourier series analysis, such as Chebyshev polynomial and Legendre polynomial analysis, have better BER performance and spectrum efficiency than the BOSM system based on Fourier series analysis, orthogonal MCM system or OFDM system.

However, the demand of high synchronization precision and having high PAPR (peak average power rate) in BOSM systems are two primary problems, just as OFDM or multi-carriers systems, and the degree of PAPR rises with increasing of the order of BOSS. Because smaller amplitude of polynomial coefficients in biorthogonal functions appears with the increasing of the order of biorthogonal function set, the performance of demodulation error probability will be deteriorate when the energy in a symbol interval is significantly weaker than the considered noise level. Therefore, the order K should be not too big.

To properly choose the basic function can change the PDF and PAPR of transmitted signal. In order to make BOSM systems with better performance, how to construct a proper basic function in an orthogonal function system to obtain the biorthogonal functions, it is still a challenge.

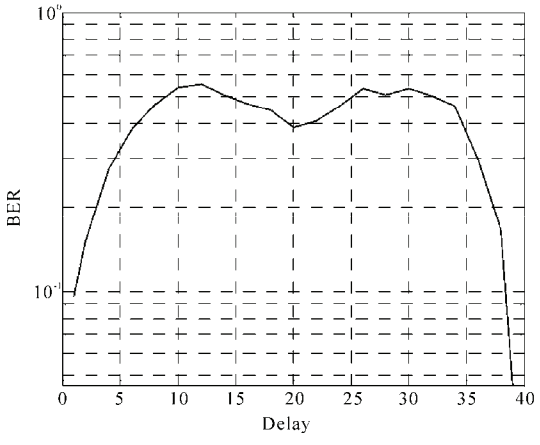


Figure 6. Impact on BER by synchronization error.

8. References

[1] R. B. Blizard, "Comparison of biorthogonal and bisimplex codes," IEEE Transactions on Communications Technology, Vol. 15, No. 4, pp. 657–658, August 1967.

[2] M. Kim and S. Tretter, "Performance of sequential decoding with biorthogonal modulation and Q-level quantization," IEEE Transactions on Communications, Vol. 19, pp. 88–92, February 1971.

[3] S.-H. Hong and J.-S. No, "Performance analysis of CDMA systems by using biorthogonal codes," IEEE 45th Conference on Vehicular Technology, Vol. 2, pp. 694–698, July 1995.

[4] S.-J. Kang, D.-K. Hong, *et al.*, "Constant-amplitude multicode-biorthogonal modulation," IEEE Transactions on Communications, Vol. 55, No. 1, pp. 69–75, January 2007.

[5] Y. Wei and N. Chen, "Square wave analysis," Journal of Mathematical Physics, Vol. 39, No. 8, pp. 4226–4245, August 1998.

[6] Y. Wei, "Dirichlet multiplication and easily-generated function analysis," Computers & Mathematical with Applications, Vol. 39, pp. 173–199, 2000.

[7] N. Chen, "Modified Möbius inverse formula and its applications in physics," Physics Review Letters, Vol. 64, No. 11, pp. 1193–l195, 1990.

[8] W. Su, W. Zhang, and J. Wang, "The evaluations of the inverse transform of eight often-used waveforms by Möbius transform—The inverse transform of their Fourier series," Chinese of Journal Electronics, Vol. 14, No. 3, pp. 513–518, July 2005.

[9] C. Ling, F. Chen, and W. Su, "The Chen-mobius multi-carriers digital communication system and its simulation," IEEE International Workshop on Anti-Counterfeiting, Security and Identification, pp. 16–18, April 2007.

[10] Y. Chen, J. Zhang, and Z. Tan, "A kind of biorthogonal functions based on Fourier-Legendre series and its application in wireless transmission system," IEEE 9th International Conference on Signal Processing, 2008.

[11] K. H. Rosen, "Elementary number theory and its application," 5th edition, Pearson Education Inc. Press, 2005.

[12] M. R. Schroeder, "Number theory in science and communication," 4th edition, Springer-Verlag Press, New York, 2006.

[13] Z. Liu, "Orthogonal function and its application," National Defense Industry Publisher, Beijing, 1982.

Wireless Sensor Network, **2010, 2, 206-207**
doi:10.4236/wsn.2010.23027 Published Online March 2010 (http://www.scirp.org/journal/wsn)

Cardiac Pacemaker and Wireless Capsule Endoscopy Interference: Case Report in a Patient with Gastric Vascular Ectasias

A. G. Gravina[1], R. Bozzi[2], I. J. Romano[3], E. Pezzullo[3], A. Miranda[1], M. G. Merola[1], M. Romano[1], A. Pezzullo[1]
[1]*Dipartimento Medico Chirurgico di Internistica Clinica e Sperimentale, Gastroenterologia e C.I.R.A.N.A.D., Seconda Università degli Studi di Napoli, Napoli, Italy*
[2]*U.O.C. Chirurgia ed Endoscopia Digestiva e, Napoli, Italy*
[3]*Cardiologia A.O.R.N. V. Monaldi, Napoli, Italy*
E-mail: *marco.romano@unina2.it*
Received December 29, 2009; *revised January* 13, 2010; *accepted January* 17, 2010

Abstract

Wireless capsule endoscopy is a new endoscopic tool for the diagnosis and management of small bowel diseases. The main indication at present is the evaluation of GI bleeding of obscure origin, Crohn's disease, coeliac disease and small bowel tumors. Studies suggest that capsule endoscopy is associated with few adverse events. Whether cardiac pacemaker may interfere with capsule endoscopy is still a controversial issue. We here report a case showing that there is a possibility of interference between the two procedures, cardiac pacemaker affecting the proper functioning of capsule endoscopy and that this is related to the distance between the pacemaker and the recorder.

Keywords: Capsule Endoscopy, Cardiac Pacemaker, Interference, Vascular Ectasias

1. Summary

Wireless capsule endoscopy is a new endoscopic tool for the diagnosis and management of small bowel diseases. The main indication at present is the evaluation of GI bleeding of obscure origin, Crohn's disease, coeliac disease and small bowel tumors. Studies suggest that capsule endoscopy is associated with few adverse events. Whether cardiac pacemaker may interfere with capsule endoscopy is still a controversial issue. We here report a case showing that there is a possibility of interference between the two procedures, cardiac pacemaker affecting the proper functioning of capsule endoscopy and that this is related to the distance between the pacemaker and the recorder.

2. Case Report

A 75 year old man with ischemic cardiopathy with dilatative evolution treated with aorto-coronaric bypass, complicated by cardiac cirrhosis and mild ascites and renal failure came to our observation because of melena. He had been implanted a cardiac pacemaker because of a complete atrioventricular block (pacemaker model Biotronik-Kalos 05 VVI n/s 36012460 with electrode Biotronik HL 150, n/s 56756B). Conventional diagnostic interventions, including upper and lower endoscopy, failed to identify the site of bleeding. An enteroclysis was performed and resulted normal. The patient underwent a small bowel study by wireless capsule endoscopy (Endocapsule Olympus ®). The recorder was positioned next to external abdominal sites near the pulse generator while the electrocardiogram was continuously recorded and reviewed by a cardiologist, to confirm the absence of pacing inhibition by the video transmission. During videocapsule examination, we revealed an interference lasting 20 minutes (**Figure 1**), with complete absence of any image on the real time viewer. The interference with videocapsule proper recording was subsequently confirmed after downloading on the workstation. In particular, we found this interference only when the recorder was positioned near the cardiac pacemaker while no interferences were found when the recorder was positioned at a distance from the pacemaker. The patient had continuous cardiac monitoring for the duration of the procedure and no arrhythmias or other adverse cardiac events

were noted during videocapsule endoscopy. Wireless capsule endoscopy revealed the presence of two vascular lesions in the gastric antrum (**Figure 2**), suggestive for gastric vascular ectasias. This was confirmed by an EGDS, performed on the same day (**Figure 3**).

3. Discussions

There is a theorical risk of electromagnetic interference between the UHF digital radiofrequency of the wireless videocapsule (433.8 MHz), and mobile phones, implanted cardiac pacemaker or defibrillator devices. However, interference between wireless capsule endoscopy and cardiac pacemaker has been observed only in one case [1], but has not been confirmed by others [2–5].

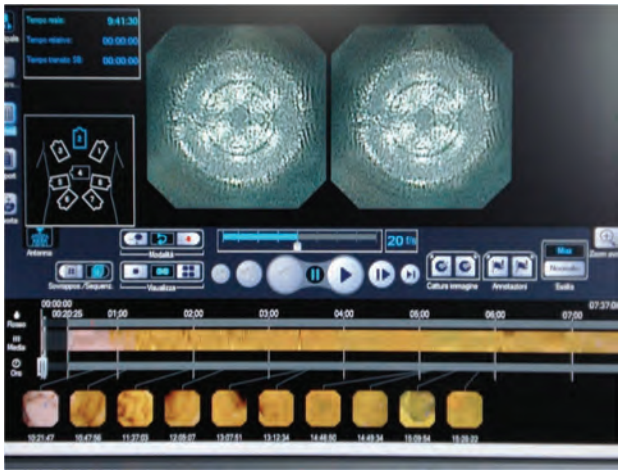


Figure 1. Workstation screen showing significant interference in the capsule recording.

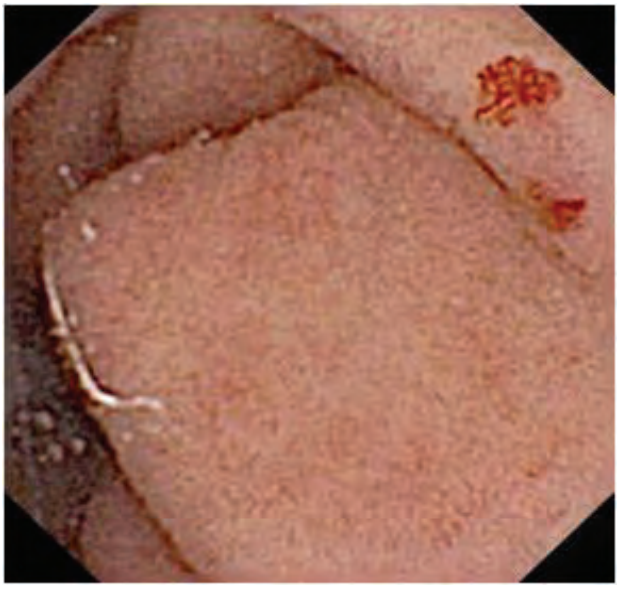


Figure 2. Wireless videocapsule imaging of the stomach showing two gastric lesions suggestive of vascular ectasias.

Figure 3. Esophagogastroduodenoscopy showing two vascular ectasias in the gastric antrum.

In our case, which is the first described using Endocapsule Olympus ®, we show that cardiac pacemaker may interfere with videocapsule proper functioning leading in our case to a "blank" period of approximately 20 minutes and this depends upon the distance between the recorder and the cardiac pacemaker. In particular no interferences were observed when a distance of about 20 cm between the two devices was maintained. Also, we confirm that wireless capsule endoscopy does not interfere with the correct function of cardiac pacemaker as assessed by continuous cardiac monitoring of the patient. Interestingly, wireless capsule endoscopy helped identify the gastric source of the bleeding then confirmed by a subsequent EGDS.

4. References

[1] Y. Guyomar, L. Vandeville, S. Heuls, *et al*., "Interference between pacemaker and video capsule endoscopy," Pacing and Clinical Electrophysiology, Vol. 27, pp. 1329–1330, 2004.

[2] J. A. Leighton, V. K. Sharma, K. Srivathsan, *et al*., "Safety of capsule endoscopy in patients with pacemakers," Gastrointestinal Endoscopy, Vol. 59, pp. 567, 2004.

[3] G. Payeras, J. Piquersas, V. J. Morena, *et al*., "Effects of capsule endoscopy on cardiac pacemakers," Endoscopy, Vol. 37, pp. 1181–1185, 2005.

[4] M. H. Dirks, F. Costea, and E. G. Seidman, "Successful videocapsule endoscopy in patients with an abdominal cardiac pacemaker," Endoscopy, Vol. 40, pp. 73–75, 2008.

[5] D. Bandorski, W. Irnich, M. Bruck, *et al*., "Capsule endoscopy and cardiac pacemaker: Investigation for possible interference," Endoscopy, Vol. 40, pp. 36–39, 2008.

Wireless Sensor Network, **2010, 2, 208-217**
doi:10.4236/wsn.2010.23028 Published Online March 2010 (http://www.scirp.org/journal/wsn)

Modeling of Circuits within Networks by fMRI

G. de Marco, A. le Pellec
Université Paris X, Laboratoire du contrôle moteur et d'analyse du mouvement, Paris, France
E-mail: *demarco.giovanni@gmail.com*
Received November 24, 2009; *revised December* 19, 2009; *accepted January* 8, 2010

Abstract

In this review, the authors describe the most recent functional imaging approaches used to explore and identify circuits within networks and model spatially and anatomically interconnected regions. After defining the concept of functional and effective connectivity, the authors describe various methods of identification and modeling of circuits within networks. The description of specific circuits in networks should allow a more realistic definition of dynamic functioning of the central nervous system which underlies various brain functions.

Keywords: fMRI, CNS, Modeling, Network, Effective Connectivity

1. Introduction

Imaging can be used to locate the brain areas involved in various forms of motor behavior, attention, vision or emotion, self-awareness and awareness of others, but brain network modeling probably remains the greatest challenge in the field of imaging data analysis [1]. Neuroimaging first allowed researchers to describe the cortical and subcortical activity of regionally segregated functional regions during a variety of experimental or cognitive tasks. More recently, functional integration studies have described how these functionally specialized areas, *i.e.*, areas whose activity is temporally modified, interact within a highly distributed neural network. By using functional magnetic resonance imaging (fMRI), which has become the method most commonly used to investigate human brain functions and define neural populations as distributed local networks transiently, linked by large-scale reciprocal dynamic connections [2].

After defining the concept of functional and effective connectivity, various approaches to the identification and modeling of circuits into networks will be presented in order to more realistically define the dynamics of the central nervous system which underlies various cerebral functions. A distinction should be made between methods that only consider correlations and ignore issues of causality and influence and methods that attempt to describe or draw inferences concerning the direction of influence between regions. Methodological approaches to the study of connectivity using fMRI data may be broadly divided into those that are more data-driven and attempt to map connectivity in the whole brain and those that use prior knowledge or hypotheses-driven, limited to a restricted set of regions [3]. These two categories of analysis are described, as indicated below, as functional connectivity and effective connectivity, respectively [4-6]. Techniques in the first group that consider only correlations between regions include mapping using seed-voxel correlations. Techniques in the second group use more elaborate models and additional assumptions applied to calculate correlations or covariances to address questions about directional influences and include mapping based on structural equation modeling (SEM), multivariate autoregressive (MAR) modeling, dynamic causal modeling (DCM).

2. Functional and Effective Connectivity

The dichotomy between local and large-scale networks serves as a neural basis for the key assumption that brain functional architecture abides by two principles: functional segregation and functional integration [2,3,7]. A large-scale brain network can be defined as a set of segregated and integrated regions that share strong anatomical connections and functional interactions. Whether top-down or bottom-up, connections and interactions are quintessential aspects of networks [8,9]. Cognitive and sensorimotor processes depend on complex dynamics of temporally and spatially segregated brain activities. While the segregation principle states that some functional processes specifically engage well-localized and specialized brain regions, it is now thought that brain functions are most likely to emerge through integration of information flows across widely distributed regions [2,10,11]. According to this approach, it is not only isolated brain areas that are presumed to process informa-

tion but rather a large-scale network, *i.e.* a set of brain regions interacting in a coherent and dynamic way. Hence, according to the functional integration concept, cortical areas and therefore functions are integrated within specific dynamic networks.

This concept supposes the existence of a dynamic interaction between interconnected, active areas and that the brain areas are expressed as networks within integrated systems. In such a system, localized areas are included in networks which become dynamic according to the cognitive task. Brain areas underlie several functions and can belong successively to several different functional networks. In other words, a given brain area does not have a single function; its resources can be exploited in several different cognitive strategies. The principle of functional integration which is also known in the field of electrophysiology was used to analyze the event potentials obtained from multielectrode recordings [12]. Thus, based on the functional integration principle, the relationships between several brain areas may be examined.

Effective connectivity, closer to the intuitive notion of a connection, can be defined as the influence that one neural system exerts over another, either at a synaptic level (synaptic efficacy) or a cortical level [13,14]. This approach emphasizes that determining effective connectivity requires a causal model of the interactions between the elements of the neural system of interest. In electrophysiology, there is a close relationship between effective connectivity and synaptic efficacy [15]. Effective connectivity can be estimated from linear models to test whether a theoretical model seeking to explain a network of relationships can actually fit the relationships estimated from the observed data. In the case of fMRI, the theoretical model is an anatomically constrained model and the data are interregional covariances of activity [16].

Consequently, effective connectivity represents the dynamic influence that cortical and subcortical regions exert on each other via a putative network of interdependent areas [5,12]. This approach might be based on linear time-invariant models that relate the time-course of experimentally controlled manipulations to BOLD signals in a voxel-specific fashion. Although various statistical models have been proposed [17], these standard models treat the voxels throughout the brain as isolated black boxes, whose input-output functions are characterized by BOLD responses evoked by various experimental conditions [18]. fMRI provides simultaneous recordings of activity throughout the brain evoked by cognitive and sensorimotor challenges, but at the expense of ignoring temporal information, *i.e.*, the history of the experimental task (input) or physiologic variable (signal). This is important, as interactions within the brain, whether over short or long distances, take time and are not instantaneous which is implicit within regression models. Furthermore, the instantaneous state of any brain system that conforms to a dynamic system will depend on the history of its input.

3. Data-Driven Approaches

The first category of methods includes seed-voxel correlations, Granger causality derived autoregressive models [19], fuzzy clustering which assumes that brain voxels can be grouped into clusters sharing similar activity patterns [20–22], hierarchical clustering [23,24], psychophysiologic interactions which test for changes in the regression slope of activity at every voxel on a seed voxel that are induced by an experimental manipulation [25], and spectral analysis [26–28]. Other techniques, such as principal component analysis [29–31] and independent component analysis (ICA) [32–35], suppose that fMRI data are a linear mixing of a given number of temporal factors with an associated factor-specific spatial distribution. Among all of these methods, we propose to briefly describe the ICA method (time analysis of the BOLD response) and the spectral method (frequency analysis of the time response) that are two interesting methods to spatially identify circuits within networks in the brain.

3.1. Independent Component Analysis

Independent component analysis (ICA) is a data-based multivariate statistical technique that uses higher order statistics to perform decomposition of linearly combined statistically independent sources [36]. Each statistically independent component represents a hemodynamic map of the whole brain. Each independent component is supposed to describe a particular functional activity of the brain with its deployment over time [37–39]. Each independent component extracted by applying a spatial ICA is spatially independent of all other independent components [35]. Therefore, the contribution of a spatial independent component to each voxel is given by the independent component magnitude at that point modulated over time by the associated time-course. The main advantage of ICA is that it requires little knowledge about the nature of the data. The only necessary hypothesis concerns the presence of a sufficient amount of independent sources (temporal or spatial), which are linearly mixed. Conversely, one of the main drawbacks of ICA is the large amount of brain activations resulting from this kind of decomposition [40]. At some point, hypotheses are necessary to select relevant from spurious activations.

For this reason, ICA can be used in conjunction with other well-established techniques [41] or further information may be associated with the reference time-course, such as the spatial localization of activities [42] and the covariate relation of independent component time-course

[43]. ICA could be combined with SEM to extend the explanatory power of each technique. SEM is a well developed, computationally minimally intensive connectivity analysis technique suitable for neuroimaging data, especially when it is combined with other data-driven methods such as ICA. In this case, SEM coupled with ICA is capable to handle data from a large number of subjects [32]. The biological relevance and cortical connections of the SEM models have also been evaluated with reference to available knowledge based on animal and human circuitries. The main advantage of spatial group ICA is its ability to identify the distinct functional elements involved in the circuitry [33]. Functionally connected brain regions encompassed in each independent component are active at the same time, suggesting that one or more anatomical connections are in use during performance of the task. Although this reasoning is more in line with the "connectionist" approach to brain functions based on parallel processing mechanisms performed by a group of connected functional elements, the ICA approach lacks a statistical method to model the functional connections assumed to exist between regions. The addition of ICA to SEM can address this issue. Each ICA map or part of the map corresponds to one component in an SEM.

3.2. Spectral Analysis

The description of a correlation structure in the frequency domain can be a promising approach to investigate interregional strengths of interactions of a functional network. As time-dependent correlations may vary between fMRI signals and across the space independently of the underlying neural dynamics, a method of analysis of frequency-dependent correlations would be one way to overcome this interregional variability of the BOLD response and would also be crucial for extracting the fine detail of information hidden within the fMRI signal. Functional connectivity analysis in the presence of major physiologic noise sources is a pitfall especially when the correlation (or covariance) between BOLD signals is performed in the time domain. In this case, these noise sources may artificially increase the magnitude of cross-correlation. Estimation of coherence between pairs of voxels at a specific frequency or at a limited range of frequencies can therefore be one way to deal with noxious physiologic noise.

The frequency domain approach can be used to analyze a limited range of linear relationships within a restricted frequency band [44]. Consequently, measurement of the correlation between fMRI data can be enhanced and can help to resolve the problem of false connectivity derived from cardiac and respiratory cycles and/or vascular differences. This approach can be performed by using spectral analysis, which allows examination of the structure of covariance and provides certain voxel-based parameters such as coherence which assesses the dependence between voxel signals [26].

The spectral theory for multivariate time series has already been used in several fMRI studies [27,28,45]. By using fMRI signals, these authors demonstrated that time domain approaches may be sufficiently susceptible to substantially high frequency artefacts, whereas the spectral domain is essentially resistant to these artefacts. They also demonstrated that the frequency-dependent correlation is higher than that measured in the temporal domain. In other fields of neuroscience, for instance in electroencephalography (EEG), coherence analysis is widely used to investigate correlated oscillatory activities between various areas in the brain [46–49]. In magnetoencephalography (MEG), coherence analysis has also been demonstrated to be a useful technique in clinical studies for discriminating different rhythmic behaviors in various brain regions [50–52]. However, although the relationship between neuronal currents and hemodynamic response is poorly understood, simultaneous intracortical neural recordings and fMRI signals acquired in animals recently revealed a significant correlation between local field potential and vascular response [53]. The feasibility of a correlation between the synchrony of low frequency BOLD fluctuations in functionally related brain regions and neuronal connections that facilitate coordinated activities has been demonstrated in various applications [54,55].

4. Hypothesis-Driven Approaches

The alternative to data-based approaches is to use a model that attempts to describe the relationships between a set of selected regions, in which region-specific measurements such as BOLD time series are extracted from whole-brain data prior to the connectivity modeling stage. This category includes structural equation modeling (SEM) [56–62], multivariate autoregressive (MAR) modeling [63,64], dynamic causal modeling (DCM) [65–67], generative models including neural mass models [68,69] and large-scale neural models [70–72].

4.1. Structural Equation Modelling

Path analysis, also referred to as structural equation modeling (SEM), was originally developed in the early 1970s by Jöreskog, Keesling, and Wiley, when they combined factor analysis with econometric simultaneous equation models [73–76]. In the early 1990s, McIntosh introduced SEM to neuroimaging [56,59,77–79] for modeling, testing, and comparison of directional effective connectivity of the brain. SEM rapidly became popular in this field [31,57,80–86]. Structural models can be used to analyze linear relationships between variables

from analysis of the covariance among the variables. Structural models were developed from two principal methods of analyses: factorial analysis (for a review: [75]) and multiple regression or causal path analysis (a method developed in the 1930s by Wright e.g., (for a review: [87]). Structural models examine multiple sources of influence on the dependent variable in an experiment [88,89].

Structural Equation Modeling (SEM) is a hypothesis-based multivariate statistical technique of data analysis that can be used with neuroimaging data. An increasing number of PET, fMRI and transcranial magnetic stimulation (TMS) studies have used SEM to investigate large-scale functional brain networks [90–93] and show specific networks involved in either working memory [94–100], attentional processes [64,101–103], face perception [104–106], motor movement processing [61, 107–112], language [32,113,114] or processing of painful stimuli [62].

SEM methods, in comparison with classical approaches such as linear regression, allow simultaneous analysis of several types of interrelationships between variables in an experiment [13,115–117]. The nature of the relationship between variables is given by the regression coefficient; it describes how much the dependent variable changes when an independent variable changes by one unit. SEM directly integrates measurement errors into a statistical model, so that estimates of regression coefficients are consequently more precise than with classical methods such as multiple regression, factorial analysis, or analysis of variance. The older methods examine only one linear relationship at the same time between independent and dependant variables and only within a range of values set by the investigator [14]. In contrast with classical methods, SEM analyzes a structure of variances and covariances in a dataset of observed variables and can be used to predict dependences between variables. In other words, SEM seeks to explain as much of the variance in dependant variables as it can from simultaneous measurement of the variances of the independent variables included in the model. Similarly, SEM incorporates measurement errors of the independent variables into calculation of the estimate, which reinforces the statistical power of the method and provides more precise estimates of regression coefficients. A model of measurement can therefore be validated from a theoretical model or empirical data [99]. The objective of effective connectivity analysis is to estimate parameters that represent influences between regions that may change over time and with respect to experimental tasks.

In order to describe a functional network, network nodes and anatomical connections must therefore be proposed in conjunction with a SEM model to explain interregional covariances and determine the intensity of the connections. When applied to PET or fMRI data, SEM allows modeling of connection pathways between cortical or subcortical areas and reveals relationships, interdependencies and covariance between the various areas. In a given anatomical model, SEM shows the effects of an experimental task on a specific network of connections [14,118–120]. In this type of statistical analysis, normalized variables are considered in terms of the structure of their covariances. SEM therefore allows inference of interregional dependencies between various cerebral cortical areas.

SEM is a simple and pragmatic approach to effective connectivity when dynamic aspects can be disregarded. A linear model is sufficient and the observed variables can be measured precisely, the input is unknown but stochastic and stationary. SEM comprises a set of regions and a set of directed connections. Importantly, a causal relationship is ascribed to these connections. Causal relationships are therefore not inferred from the data, but are assumed a priori. The strengths of connections can therefore be set so as to minimize the discrepancy between observed and implied correlations and thereby fit a model to the data. Changes in connectivity can be attributed to experimental manipulation by partitioning the data set. If, for example, a given fMRI data set is partitioned into those scans obtained for different levels of an experimental factor, differences in connectivity can then be attributed to that factor leading to the conclusion that a pathway has been activated. An SEM with particular connection strengths implies a particular set of instantaneous correlations between regions. Structural equation models posit a set of theoretical causal relationships between variables and model instantaneous correlations *i.e.*, correlations between regions at the same time-point. Instantaneous activity is assumed to be the result of local dynamics and connections between regions.

4.2. Multivariate Autoregressive (MAR) Models

To overcome the difficulties of SEM, Harrison *et al.* proposed the use of multivariate autoregressive (MAR) models for the analysis of fMRI data [63]. They were the first to introduce multivariate autoregressive (MAR) models into brain pathway analyses to characterize interregional dependence. MAR models are time-series models and consequently model temporal order within measured brain activity. Goebel *et al.* [19] and Roebroeck *et al.* [121] subsequently generalized the MAR approach by incorporating Granger causality between two time series. MAR models posit a set of causal relationships between variables; they incorporate cross-covariances between regions (covariances at multiple lags) and exploit temporal relationships between different scans to allow conclusions about predominant directions of influence between regions as well as their strength [18, 122,123].

An autoregressive approach is used to characterize a structure in a time series, whereby the current value of a time series is modeled as a weighted linear sum of previous values. Consecutive measurements within a given time series contain information about the process that generated this series. This is an autoregressive process and is a very simple, yet effective, approach to time series characterization. This is distinct from regression techniques that quantify instantaneous correlations, but is similar to the SEM model in that it estimates the relative influences over time. Autoregressive models of fMRI data address the temporal aspect of causality in a BOLD time series, focusing on the causal dependence of the present on the past. Each data point of a time series is explained as a linear combination of past data points. This approach contrasts with SEM regression-based models in which the time series can be permuted without changing the results. MAR models contain directed influences among a set of regions whose causal interactions, expressed at the BOLD level, are inferred via their mutual predictability from past time points.

4.3. Dynamic Causal Modeling

A major criticism of SEM or MAR with regard to neuroimaging data is that they model effective connectivity changes at the "hemodynamic level" rather than the "neuronal level". This is a serious problem because the causal architecture of the system that we want to identify is expressed in terms of neuronal dynamics, which are not directly observed using noninvasive techniques. In the case of fMRI data, previous models of effective connectivity have been fitted to the measured time series which result from a hemodynamic convolution of the underlying neural activity. Since classical statistical models do not include the forward model linking neuronal activity to the measured hemodynamic data, analyses of interregional connectivity performed on hemodynamic responses are problematic. For example, different brain regions can exhibit marked differences in neurovascular coupling, and these differences, expressed in different latencies (see above) may lead to false inferences about connectivity [124].

Dynamical Causal Modeling (DCM) has recently been developed as a generalization of both convolution models and SEM [66,67]. As described in Penny *et al*. [66], SEM can be shown to be a simplified version of DCM which also depends on the definition of a structural model. DCM model assumes a dynamic neuronal model of interacting brain regions, whereby neuronal activity in a given brain region causes changes in neuronal activity in other regions according to the structural model. This neuronal model is then supplemented with a forward model of how neuronal activity generates a measured BOLD response through the balloon model which was initially formulated by Buxton *et al.* [125] and later extended by Friston *et al.* [126]. A Bayesian inference scheme is devised to infer the model parameters from the data. The mathematical framework of DCM takes into account nonlinearities and temporal correlations. It also quantifies the interaction strength that one brain region exerts on another brain region at the neuronal level, whereas SEM only concerns the observed BOLD signal. DCM is suspected to be less sensitive than SEM to the number of degrees of freedom. Unlike SEM, DCM also models the effect of experimental, external, and modulatory inputs on network dynamics. Since DCM models neurobiologically plausible neural activities and takes into account dynamics and modulations, this mathematical framework would appear to be more advantageous than SEM.

4.4. Diffusion Tensor Imaging

While fMRI provides detailed information about the spatial location of functionally active cortical areas, the question of anatomical interdependency between cortical areas remains elusive. A key tool to assess the validity of large-scale distributed networks in fMRI is knowledge of the underlying anatomical connections. The original idea behind SEM and functional neuroimaging was to combine two data sets: a functional set with an anatomical set (connections between regions), based on the assumption that anatomy was the source of spatial causal relationships. Our understanding of the connections between regions is limited, but since the advent of newer tractography methods, the main white matter tracts can be described. Diffusion Tensor Imaging (DTI) is a powerful MRI technique [127,128] that can be used to translate self-diffusion, or microscopic motion of water molecules in tissue into a MRI measure of tissue integrity and structure (white matter fibers). Data from diffusion tensor imaging (DTI) and fMRI have been combined in a few previous studies [129–131]. These studies showed that a combination of techniques can give additional information about brain organization which may give more specific information about organization of brain functions and brain injuries. In this latter case, a DTI-driven SEM would integrate information about white matter changes (e.g. maturation, aging) [100,132]. The prospect of using information derived from tractography could be used to constrain structural models. DTI and fMRI combinations will be essential to discover to what extent the brain functional organization as investigated with fMRI reflects structural features of the brain and, hence, to more accurately assess the relevance of fMRI to examine the relationship between functional and large-scale anatomical networks. However, more studies are still needed to investigate anatomical correlates which would be related to effective connectivity.

5. Conclusions

This article describes the most recent imaging approaches used to explore and identify circuits within networks and to spatially and anatomically model interconnected regions. Structural equation modeling is the most widely used method to model effective connectivity [56,82,133]. The relevance of applying SEM to fMRI neuroimaging data has been discussed in detail elsewhere [58,66,82,134]. SEM allows one to start with simpler models and then progress to more complex models by repeatedly testing the model fit to real data. SEM is useful when some information is available, such as a small set of potential structural models or partial information concerning connectivity. Newer, more sophisticated effective connectivity analysis methods such as Dynamic Causal Modeling might circumvent the drawbacks of SEM and may shed more insight into how brain regions interact in information processing. Nevertheless, SEM is a well developed, computationally less intensive connectivity analysis technique suitable for neuroimaging data especially for block designs and combined with other methods such as independent component analysis, partial correlation or DTI. The use of SEM may be justified by the fact that, unlike DCM, the statistical model underlying SEM is quite simple and not computationally demanding.

6. References

[1] K. E. Stephan, *et al*., "Models of functional neiroimaging data," Center for Molecular Imaging Research, Vol. 2, pp. 15–34, 2006.

[2] F. Varela, *et al*., "The brainweb: Phase synchronization and large-scale integration," Nature Reviews Neuroscience, Vol. 2, No. 4, pp. 229–239, 2001.

[3] G. Marrelec, P. Bellec, and H. Benali, "Exploring large-scale brain networks in functional MRI," Journal of Physiology Paris, Vol. 100, No. 4, pp. 171–181, 2006.

[4] B. Horwitz, *et al*., "Investigating the neural basis for functional and effective connectivity. Application to fMRI," Philosophical Transactions of the Royal Society B: Biological Sciences, Vol. 360, No. 1457, pp. 1093–1108, 2005.

[5] K. J. Friston, C. D. Frith, and R. S. J. Frackowiak, "Time-dependent changes in effective connectivity measured with PET," Human Brain Mapping, Vol. 1, pp. 69–79, 1993.

[6] A. M. Kelly and H. Garavan, "Human functional neuroimaging of brain changes associated with practice," Cerebral Cortex, Vol. 15, No. 8, pp. 1089–1102, 2005.

[7] B. Horwitz, M. A. Tagamets, and A. R. McIntosh, "Neural modeling, functional brain imaging, and cognition," Trends in Cognitive Sciences, Vol. 3, No. 3, pp. 91–98, 1999.

[8] M. M. Mesulam, "From sensation to cognition," Brain, Vol. 121, No. Pt 6, pp. 1013–1052, 1998.

[9] S. L. Bressler, "Large-scale cortical networks and cognition," Brain Research Reviews, Vol. 20, No. 3, pp. 288–304, 1995.

[10] G. Tononi, G. M. Edelman, and O. Sporns, "Complexity and coherence: Integrating information in the brain," Trends in Cognitive Sciences, Vol. 2, pp. 474–484, 1998.

[11] O. Sporns, *et al*., "Organization, development and function of complex brain networks," Trends in Cognitive Science, Vol. 8, No. 9, pp. 418–425, 2004.

[12] G. L. Gerstein and D. H. Perkel, "Simultaneously recorded trains of action potentials: Analysis and functional interpretation," Science, Vol. 164, No. 881, pp. 828–830, 1969.

[13] K. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," Human Brain Mapping, Vol. 2, pp. 56–78, 1994.

[14] A. McIntosh and F. Gonzalez-Lima, "Structural equation modeling and its application to network analysis in functional brain imaging," Human Brain Mapping, Vol. 2, pp. 2–22, 1994.

[15] A. Aersten and H. Preissl, "Dynamics of activity and connectivity in physiological neuronal networks," In: H. G. Schuster, Ed., NonLinear Dynamics and Neuronal Networks, VCH Publishers Inc., New York, pp. 281–302, 1991.

[16] C. Buchel and K. Friston, "Assessing interactions among neuronal systems using functional neuroimaging," Neural Networks, Vol. 13, No. 8–9, pp. 871–82, 2000.

[17] R. Henson, "Analysis of fMRI time series: Linear time-invariant models, event-related fMRI, and optimal experimental design," In: R. Frackowiack *et al*. Ed., Human Brain Function, 2nd edition, Elsevier, San Diego, pp. 793–823, 2004.

[18] K. E. Stephan, *et al*., "Biophysical models of fMRI responses," Current Opinion Neurobiology, Vol. 14, No. 5, pp. 629–635, 2004.

[19] R. Goebel, *et al*., "Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping," Magnetic Resonance Imaging, Vol. 21, No. 10, pp. 1251–1261, 2003.

[20] R. Baumgartner, *et al*., "Comparison of two exploratory data analysis methods for fMRI: Fuzzy clustering vs. principal component analysis," Magnetic Resonance Imaging, Vol. 18, No. 1, pp. 89–94, 2000.

[21] R. Baumgartner, *et al*., "Resampling as a cluster validation technique in fMRI," Journal of Magnetic Resonance Imaging, Vol. 11, No. 2, pp. 228–231, 2000.

[22] R. Baumgartner, C. Windischberger, and E. Moser, "Quantification in functional magnetic resonance imaging: Fuzzy clustering vs. correlation analysis," Magnetic Resonance Imaging, Vol. 16, No. 2, pp. 115–125, 1998.

[23] D. Cordes, *et al*., "Hierarchical clustering to measure connectivity in fMRI resting-state data," Magnetic Resonance Imaging, Vol. 20, No. 4, pp. 305–317, 2002.

[24] C. Goutte, *et al*., "On clustering fMRI time series," Neuroimage, Vol. 9, No. 3, pp. 298–310, 1999.

[25] K. J. Friston, *et al*., "Psychophysiological and modulatory interactions in neuroimaging," Neuroimage, Vol. 6, No. 3, pp. 218–229, 1997.

[26] S. Fall and G. de Marco, "Assessment of brain interactivity in the motor cortex from the concept of functional connectivity and spectral analysis of fMRI data," Biological Cybernetics, Vol. 98, No. 2, pp. 101–114, 2008.

[27] F. T. Sun, L. M. Miller, and M. D'Esposito, "Measuring temporal dynamics of functional networks using phase spectrum of fMRI data," Neuroimage, Vol. 28, No. 1, pp. 227–237, 2005.

[28] K. Muller, *et al*., "On multivariate spectral analysis of fMRI time series," Neuroimage, Vol. 14, No. 2, pp. 347–356, 2001.

[29] K. J. Friston, *et al*., "Functional connectivity: The principal-component analysis of large (PET) data sets," Journal of Cerebral Blood Flow & Metabolism, Vol. 13, No. 1, pp. 5–14, 1993.

[30] A. H. Andersen, D. M. Gash, and M. J. Avison, "Principal component analysis of the dynamic response measured by fMRI: A generalized linear systems framework," Magnetic Resonance Imaging, Vol. 17, No. 6, pp. 795–815, 1999.

[31] E. T. Bullmore, *et al*., "Functional magnetic resonance image analysis of a large-scale neurocognitive network," Neuroimage, Vol. 4, No. 1, pp. 16–33, 1996.

[32] P. R. Karunanayaka, *et al*., "Age-related connectivity changes in fMRI data from children listening to stories," Neuroimage, Vol. 34, No. 1, pp. 349–360, 2007.

[33] N. Correa, T. Adali, and V. D. Calhoun, "Performance of blind source separation algorithms for fMRI analysis using a group ICA method," Magnetic Resonance Imaging, Vol. 25, No. 5, pp. 684–694, 2007.

[34] V. D. Calhoun, *et al*., "Latency (in) sensitive ICA. Group independent component analysis of fMRI data in the temporal frequency domain," Neuroimage, Vol. 20, No. 3, pp. 1661–1669, 2003.

[35] M. J. Jafri, *et al*., "A method for functional network connectivity among spatially independent resting-state components in schizophrenia," Neuroimage, Vol. 39, No. 4, pp. 1666–1681, 2008.

[36] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," IEEE Transactions on Neural Networks, Vol. 10, pp. 626–634, 1999.

[37] F. Esposito, *et al*., "Spatial independent component analysis of functional MRI time-series: To what extent do results depend on the algorithm used?" Human Brain Mapping, Vol. 16, No. 3, pp. 146–157, 2002.

[38] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," IEEE Transactions on Medical Imaging, Vol. 23, No. 2, pp. 137–152, 2004.

[39] M. J. McKeown and T. J. Sejnowski, "Independent component analysis of fMRI data: Examining the assumptions," Human Brain Mapping, Vol. 6, No. 5–6, pp. 368–372, 1998.

[40] M. J. McKeown, *et al*., "Analysis of fMRI data by blind separation into independent spatial components," Human Brain Mapping, Vol. 6, No. 3, pp. 160–188, 1998.

[41] D. Hu, *et al*., "Unified SPM-ICA for fMRI analysis," Neuroimage, Vol. 25, No. 3, pp. 746–755, 2005.

[42] B. Hong, G. D. Pearlson, and V. D. Calhoun, "Source density-driven independent component analysis approach for fMRI data," Human Brain Mapping, Vol. 25, No. 3, pp. 297–307, 2005.

[43] M. J. McKeown, "Detection of consistently task-related activations in fMRI data with hybrid independent component analysis," Neuroimage, Vol. 11, No. 1, pp. 24–35, 2000.

[44] F. T. Sun, L. M. Miller, and M. D'Esposito, "Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data," Neuroimage, Vol. 21, No. 2, pp. 647–458, 2004.

[45] L. J. Marchini and B. D. Ripley, "A new statistical approach to detecting significant activation in funtional MRI," Neuroimage, Vol. 12, pp. 366–380, 2000.

[46] C. Andrew and G. Pfurtscheller, "Event-related coherence as a tool for studying dynamic interaction of brain regions," Electroencephalography and Clinical Neurophysiology, Vol. 98, No. 2, pp. 144–148, 1996.

[47] J. Classen, *et al*., "Integrative visuomotor behavior is associated with interregionally coherent oscillations in the human brain," Journal of Neurophysiology, Vol. 79, No. 3, pp. 1567–1573, 1998.

[48] F. G. Andres, *et al*., "Functional coupling of human cortical sensorimotor areas during bimanual skill acquisition," Brain, Vol. 122, No. Pt 5, pp. 855–870, 1999.

[49] P. Rappelsberger and H. Petsche, "Probability mapping: Power and coherence analyses of cognitive processes," Brain Topography, Vol. 1, No. 1, pp. 46–54, 1988.

[50] T. Locatelli, *et al*., "EEG coherence in Alzheimer's disease," Electroencephalography and Clinical Neurophysiology, Vol. 106, No. 3, pp. 229–237, 1998.

[51] K. Le Roc'h, "EEG coherence in Alzheimer disease, by Besthorn *et al*.," Electroencephalography and Clinical Neurophysiology, Vol. 91, No. 3, pp. 232–233, 1994.

[52] C. Besthorn, *et al*., "EEG coherence in Alzheimer disease," Electroencephalography and Clinical Neurophysiology, Vol. 90, No. 3, pp. 242–245, 1994.

[53] N. K. Logothetis, *et al*., "Neurophysiological investigation of the basis of the fMRI signal," Nature, Vol. 412, No. 6843, pp. 150–157, 2001.

[54] V. Haughton and B. Biswal, "Clinical application of basal regional cerebral blood flow fluctuation measurements by fMRI," Advances in Experimental Medicine and Biology, Vol. 454, pp. 583–590, 1998.

[55] D. Cordes, *et al*., "Mapping functionally related regions of brain with functional connectivity MR imaging," American Journal of Neuroradiology, Vol. 21, No. 9, pp.

1636–1644, 2000.

[56] A. R. McIntosh and F. Gonzalez-Lima, "Structural equation modeling and its application to network analysis in functional brain imaging," Human Brain Mapping, Vol. 2, pp. 2–22, 1994.

[57] M. Glabus, *et al*., "Interindividual differences in functional interactions among prefrontal, parietal and parahippocampal regions during working memory," Cerebral Cortex, Vol. 13, pp. 1352–1361, 2003.

[58] M. S. Goncalves and D. A. Hall, "Connectivity analysis with structural equation modelling: An example of the effects of voxel selection," Neuroimage, Vol. 20, No. 3, pp. 1455–1467, 2003.

[59] A. McIntosh, *et al*., "Network analysis of cortical visual pathways mapped with PET," Journal of Neuroscience, Vol. 14, pp. 655–666, 1994.

[60] C. Büchel and K. Friston, "Assessing interactions among neuronal systems using functional neuroimaging," Neural Networks, Vol. 13, No. 8–9, pp. 871–882, 2000.

[61] T. Taniwaki, *et al*., "Age-related alterations of the functional interactions within the basal ganglia and cerebellar motor loops in vivo," Neuroimage, Vol. 36, No. 4, pp. 1263–1276, 2007.

[62] J. G. Craggs, *et al*., "Functional brain interactions that serve cognitive-affective processing during pain and placebo analgesia," Neuroimage, Vol. 38, No. 4, pp. 720–729, 2007.

[63] L. Harrison, D. Penny, and K. Friston, "Multivariate autoregressive modeling of fMRI time series," Neuroimage, Vol. 19, No. 4, pp. 1477–1491, 2003.

[64] J. Kim, *et al*., "Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data," Human Brain Mapping, Vol. 28, No. 2, pp. 85–93, 2007.

[65] W. D. Penny, *et al*., "Comparing dynamic causal models," Neuroimage, Vol. 22, No. 3, pp. 1157–1172, 2004.

[66] W. D. Penny, *et al*., "Modelling functional integration: A comparison of structural equation and dynamic causal models," Neuroimage, Vol. 23, Supplement 1, pp. S264–S274, 2004.

[67] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modeling," Neuroimage, Vol. 19, No. 4, pp. 1273–1302, 2003.

[68] O. David and K. J. Friston, "A neural mass model for MEG/EEG: Coupling and neuronal dynamics," Neuroimage, Vol. 20, No. 3, pp. 1743–1755, 2003.

[69] O. David, D. Cosmelli, and K. J. Friston, "Evaluation of different measures of functional connectivity using a neural mass model," Neuroimage, Vol. 21, No. 2, pp. 659–673, 2004.

[70] B. Horwitz, "Relating fMRI and PET signals to neural activity by means of large-scale neural models," Neuroinformatics, Vol. 2, No. 2, pp. 251–266, 2004.

[71] F. T. Husain, *et al*., "Relating neuronal dynamics for auditory object processing to neuroimaging activity: A computational modeling and an fMRI study," Neuroimage, Vol. 21, No. 4, pp. 1701–1720, 2004.

[72] M. A. Tagamets and B. Horwitz, "Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study," Cerebral Cortex, Vol. 8, No. 4, pp. 310–320, 1998.

[73] J. Loehlin, "Latent variable models: An introduction to factor, path, and structural analysis," 4th edition, Lawrence Erlbaum, Mahwah, New Jersey, 1998.

[74] K. G. Jöreskog and D. Sörbom, "LISREL 8.5 user's reference guide," Scientific Software International, Chicago, 2000.

[75] K. Bollen, "With new incremental structural index for general equation models made," Sociological Methods and Research, Vol. 17, pp. 303–316, 1989.

[76] K. Bollen and J. Long, "Testing structural equation models," Sage, Thousand Oaks, California, 1993.

[77] A. McIntosh, "Understanding neural interactions in learning and memory using function neuroimaging," Annals of the New York Academy of Sciences, Vol. 855, pp. 556–571, 1998.

[78] A. McIntosh and F. Gonzales-Lima, "Structural modeling of functional neural pathways mapped with 2-deoxyglucose; effects of acoustic startle habituation on the auditory system 7," Brain Research, Vol. 547, pp. 295–302, 1991.

[79] A. McIntosh and F. Gonzales-Lima, "Network interactions among limbic cortices, basal forebrain, and cerebellum differentiate a tone conditioned as a Pavlovian excitor or inhibitor: Fluorodeoxyglucose mapping and covariance structural modeling," Journal of Neurophysiology, Vol. 72, pp. 1717–1733, 1994.

[80] C. Büchel, J. Coull, and K. Friston, "The predictive value of changes in effective connectivity for human learning," Science, Vol. 283, pp. 1538–1541, 1999.

[81] C. Büchel and K. J. Friston, "Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI," Cerebral Cortex, Vol. 7, No. 8, pp. 768–778, 1997.

[82] E. Bullmore, *et al*., "How good is good enough in path analysis of fMRI data?" Neuroimage, Vol. 11, No. 4, pp. 289–301, 2000.

[83] P. Fletcher, *et al*., "Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging," Cerebral Cortex, Vol. 9, pp. 168–178, 1999.

[84] S. Grafton, *et al*., "Network analysis of motor system connectivity in Parkinson's disease: Modulation of thalamocortical interactions after pallidotomy," Human Brain Mapping, Vol. 2, pp. 45–55, 1994.

[85] G. Honey, *et al*., "Effects of verbal working memory load on corticocortical connectivity modeled by path analysis of functional magnetic resonance imaging data," Neuroimage, Vol. 17, pp. 573–582, 2002.

[86] J. Jennings, A. McIntosh, and S. Kapur, "Mapping neural interactivity onto regional activity: An analysis of semantic processing and response mode interactions," Neuroimage, Vol. 7, pp. 244–254, 1998.

[87] M. Hollander, "Nonparametric statistical methods," Wiley Series in Probability and Statistics: Applied Probability and Statistic, Wiley, New York, 2nd edition, 1999.

[88] R. Kline, "Principles and practice of structural equation modeling (methodology in the social sciences)," The Guilford Press, 2nd edition, 2004.

[89] B. Byrne, "Structural equation modeling with Amos: BASIC concepts, applications, and programming," Deliver, Lawrence Erlbaum Associates, Reprint edition, 2001.

[90] G. Marrelec, *et al*., "Using partial correlation to enhance structural equation modeling of functional MRI data," Magnetic Resonance Imaging, Vol. 25, No. 8, pp. 1181–1189, 2007.

[91] G. Marrelec, *et al*., "Large-scale neural model validation of partial correlation analysis for effective connectivity investigation in functional MRI," Human Brain Mapping, 2008.

[92] B. P. Rogers, *et al*., "Assessing functional connectivity in the human brain by fMRI," Magnetic Resonance Imaging, Vol. 25, No. 10, pp. 1347–1357, 2007.

[93] B. P. Rogers, *et al*., "Comment on 'Assessing functional connectivity in the human brain by fMRI'," Magnetic Resonance Imaging, Vol. 26, No. 1, pp. 146, 2008.

[94] M. V. Au Duong, *et al*., "Modulation of effective connectivity inside the working memory network in patients at the earliest stage of multiple sclerosis," Neuroimage, Vol. 24, No. 2, pp. 533–538, 2005.

[95] M. F. Glabus, *et al*., "Interindividual differences in functional interactions among prefrontal, parietal and parahippocampal regions during working memory," Cerebral Cortex, Vol. 13, No. 12, pp. 1352–1361, 2003.

[96] R. G. Schlosser, G. Wagner, and H. Sauer, "Assessing the working memory network: Studies with functional magnetic resonance imaging and structural equation modeling," Neuroscience, Vol. 139, No. 1, pp. 91–103, 2006.

[97] R. Schlosser, *et al*., "Altered effective connectivity during working memory performance in schizophrenia: A study with fMRI and structural equation modeling," Neuroimage, Vol. 19, No. 3, pp. 751–763, 2003.

[98] H. Kondo, *et al*., "Functional roles of the cingulo-frontal network in performance on working memory," Neuroimage, Vol. 21, No. 1, pp. 2–14, 2004.

[99] J. B. Krause, *et al*., "Imaging and neural modeling in episodic and working memory processes," Neural Networks, Vol. 13, No. 8–9, pp. 847–859, 2000.

[100] R. A. Charlton, *et al*., "A structural equation modeling investigation of age-related variance in executive function and DTI measured white matter damage," Neurobiology of Aging, 2007.

[101] J. Rowe, *et al*., "Attention to action: Specific modulation of corticocortical interactions in humans," Neuroimage, Vol. 17, No. 2, pp. 988–998, 2002.

[102] F. M. Mottaghy, *et al*., "Systems level modeling of a neuronal network subserving intrinsic alertness," Neuroimage, Vol. 29, No. 1, pp. 225–233, 2006.

[103] K. I. Erickson, *et al*., "A structural equation modeling analysis of attentional control: An event-related fMRI study," Cognitive Brain Research, Vol. 22, No. 3, pp. 349–357, March 2005.

[104] M. N. Rajah, A. R. McIntosh, and C. L. Grady, "Frontotemporal interactions in face encoding and recognition," Cognitive Brain Research, Vol. 8, No. 3, pp. 259–269, 1999.

[105] G. de Marco, *et al*., "Changes in effective connectivity during incidental and intentional perception of fearful faces," Neuroimage, Vol. 30, No. 3, pp. 1030–1037, 2006.

[106] J. L. Stein, *et al*., "A validated network of effective amygdala connectivity," Neuroimage, Vol. 36, No. 3, pp. 736–745, 2007.

[107] B. P. Rogers, J. D. Carew, and M. E. Meyerand, "Hemispheric asymmetry in supplementary motor area connectivity during unilateral finger movements," Neuroimage, Vol. 22, No. 2, pp. 855–859, 2004.

[108] S. T. Grafton, *et al*., "Network analysis of motor system connectivity in Parkinson's disease: Modulation of thalamocortical interactions after pallidotomy," Human Brain Mapping, Vol. 2, pp. 45–55, 1994.

[109] I. Toni, *et al*., "Changes of cortico-striatal effective connectivity during visuomotor learning," Cerebral Cortex, Vol. 12, No.10, pp. 1040–1047, 2002.

[110] T. Taniwaki, *et al*., "Functional network of the basal ganglia and cerebellar motor loops in vivo: Different activation patterns between self-initiated and externally triggered movements," Neuroimage, Vol. 31, No. 2, pp. 745–753, 2006.

[111] J. Zhuang, *et al*., "Connectivity exploration with structural equation modeling: An fMRI study of bimanual motor coordination," Neuroimage, Vol. 25, No. 2, pp. 462–470, 2004.

[112] A. R. Laird, *et al*., "Modeling motor connectivity using TMS/PET and structural equation modeling," Neuroimage, 2008.

[113] V. Quaglino, *et al*., "Differences in effective connectivity between dyslexic children and normal readers during a pseudoword reading task: An fMRI study," Clinical Neurophysiology, Vol. 38, No. 2, pp. 73–82, 2008.

[114] C. H. Fu, *et al*., "Modulation of effective connectivity by cognitive demand in phonological verbal fluency," Neuroimage, Vol. 30, No. 1, pp. 266–271, 2006.

[115] B. Horwitz, K. J. Friston, and J. G. Taylor, "Neural modeling and functional brain imaging: An overview," Neural Networks, Vol. 13, No. 8–9, pp. 829–846, 2000.

[116] L. Lee, L. Harrison, and A. Mechelli, "A report of the functional connectivity," Workshop, Dusseldorf, pp. 457–465, 2003.

[117] A. McIntosh, "Towards a network theory of cognition," Neural Networks, Vol. 13, pp. 861–876, 2001.

[118] C. Buchel and K. J. Friston, "Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI," Cerebral Cortex, Vol. 7, No. 8, pp. 768–778, 1997.

[119] B. Horwitz, "The elusive concept of brain connectivity," Neuroimage, Vol. 19, No. 2 Pt 1, pp. 466–470, 2003.

[120] F. Gonzalez-Lima and A. McIntosh, "Analysis of neural network interactions related to associative learning using structural equation modeling," Mathematics and Computers in Simulation, Vol. 40, No. 1–2, pp. 115–140, 1995.

[121] A. Roebroeck, E. Formisano, and R. Goebel, "Mapping directed influence over the brain using Granger causality and fMRI," Neuroimage, Vol. 25, No. 1, pp. 230–242, 2005.

[122] I. Korhonen, *et al*., "Linear multivariate models for physiological signal analysis: Theory," Computer Methods Programs in Biomedicine, Vol. 51, No. 1–2, pp. 85–94, 1996.

[123] I. Korhonen, R. Takalo, and V. Turjanmaa, "Multivariate autoregressive model with immediate transfer paths for assessment of interactions between cardiopulmonary variability signals," Medical & Biology Engineering & Computer, Vol. 34, No. 3, pp. 199–206, 1996.

[124] K. E. Stephan, *et al*., "Dynamic causal models of neural system dynamics: Current state and future extensions," Journal of Bioscience, Vol. 32, No. 1, pp. 129–144, 2007.

[125] R. B. Buxton, E. C. Wong, and L. R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: The balloon model," Magnetic Resonance in Medicine, Vol. 39, No. 6, pp. 855–864, 1998.

[126] K. J. Friston, *et al*., "Nonlinear responses in fMRI: The Balloon model, Volterra kernels, and other hemodynamics," Neuroimage, Vol. 12, No. 4, pp. 466–477, 2000.

[127] P. J. Basser and D. K. Jones, "Diffusion-tensor MRI: Theory, experimental design and data analysis—a technical review," NMR Biomedicine, Vol. 15, No. 7–8, pp. 456–467, 2002.

[128] P. J. Basser, J. Mattiello, and D. LeBihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo," Journal of Magnetic Resonance B, Vol. 103, No. 3, pp. 247–254, 1994.

[129] D. J. Werring, *et al*., "The structural and functional mechanisms of motor recovery: Complementary use of diffusion tensor and functional magnetic resonance imaging in a traumatic injury of the internal capsule," Journal of Neurology, Neurosurgery & Psychiatry, Vol. 65, pp. 863–869, 1998.

[130] D. J. Werring, *et al*., "A direct demonstration of both structure and function in the visual system: Combining diffusion tensor imaging with functional magnetic resonance imaging," Neuroimage, Vol. 9, pp. 352–361, 1999.

[131] U. C. Wieshmann, *et al*., "Combined functional magnetic resonance imaging and diffusion tensor imaging demonstrate widespread modified organization in malformation of cortical development," Journal of Neurology, Neurosurgery & Psychiatry, Vol. 70, pp. 521–523, 2001.

[132] P. J. Olesen, *et al*., "Combined analysis of DTI and fMRI data reveals a joint maturation of white and grey matter in a fronto-parietal network," Cognitive Brain Research, Vol. 18, No. 1, pp. 48–57, 2003.

[133] F. Gonzalez-Lima and A. R. McIntosh, "Analysis of neural interactions related to associative learning using structural equation modeling," Mathematics and Computers in Simulation, Vol. 40, pp. 115–140, 1995.

[134] A. Mechelli, *et al*., "Effective connectivity and intersubject variability: Using a multisubject network to test differences and commonalities," Neuroimage, Vol. 17, No. 3, pp. 1459–1469, 2002.

Wireless Sensor Network, **2010, 2, 218-226**
doi:10.4236/wsn.2010.23029 Published Online March 2010 (http://www.scirp.org/journal/wsn)

Performance Analysis of Multi-Parametric Call Admission Control Strategies in Un-Buffered Multi-Service Cellular Wireless Networks

Jang Hyun Baek[1], Che Soong Kim[2], Agassi Melikov[3], Mehriban Fattakhova[3]
[1]*Department of Industrial and Information Systems Engineering, Chonbuk National University, Jeonju, Republic of Korea*
[2]*Department of Industrial Engineering, Sangji University, Wonju, Republic of Korea*
[3]*Department of Aerospace Information Technologies & Control Systems, National Aviation Academy, Azerbaijan*
E-mail: *jbaek@chonbuk.ac.kr, dowoo@sangji.ac.kr, {agassi.melikov, meri-fattax}@rambler.ru*
Received December 28, 2009; *revised January* 12, 2010; *accepted January* 15, 2010

Abstract

In this paper model of integrated voice/data cellular wireless networks (CWN) are investigated. The unified approximate approach to calculate the desired Quality of Service (QoS) metrics in an isolated cell of such networks under two multi-parametric call admission control (CAC) strategies is developed. One of them is based on the guard channels scheme while the second is based on a threshold scheme. Results of the numerical experiments are given and a comparison of QoS metrics under different CAC strategies is carried out.

Keywords: Cellular Networks, Call Admission Control, Quality of Service, Calculation Algorithm

1. Introduction

A cellular wireless network (CWN) consists of radio access points, called base stations (BS), each covering a certain geographic area. With distance, the power of radio signals fade away (fading or attenuation of signal occurs) which makes it possible to use the same frequencies over several cells, but in order to avoid interference, this process must be carefully planned. For better use of frequency resources, existing carrier frequencies are grouped, and the number of cells, in which this group of frequencies is used, defines the so called, frequency reuse factor. Therefore, in densely populated areas with a large number of mobile subscribers (MS), small dimensioned cells (micro-cells and pico-cells) are to be used.

In connection with the limitation of transmission spectrum in a CWN, problems of allocation of common spectrum among cells are very important. A unit of the wireless spectrum, necessary for serving a single user is called a channel (for instance, time slots in TDMA are considered as channels). There are three solutions for the channels allocation problem: Fixed Channel Allocation (FCA), Dynamic Channel Allocation (DCA) and Hybrid Channel Allocation (HCA). Advantages and disadvantage of each of these are well known [1–3]. At the same time, owing to realization simplicity, the FCA scheme is widely used in existing cellular networks. In this paper models with FCA schemes are considered.

Quality of service (QoS) in the certain cell with the FCA scheme could be improved if appropriate call ad-mission control (CAC) strategies for the heterogeneous traffic are provided. The use of such an access strategy doesn't require many resources, therefore this method could be considered operative and more effective in view of problems relating to resource shortages.

Apart from original (or new) calls (o-calls) flows, additional classes of calls that require a special approach also exist in wireless cellular networks. These are so-called handover calls (h-calls). This is specific only for wireless cellular networks. The essence of this phenomenon is that a moving MS, that already established connection with a network, passes boundaries between cells and gets served by a new cell. From the new cell's point of view this is an h-call, and since the connection with the MS has already been established, the handling of the transfer to a new cell must be transparent for the user. In other words, in wireless networks the call may occupy channels from different cells, several times during the duration of the call. This means that the channel occupation period is not the same as the call duration.

Mathematical models of call handling processes in multi-service CWN can be developed adequately enough

based on a queuing theory of networks with different types of calls and random topology. Such models are researched poorly in literature [4–6]. This is explained by the fact, that despite the elegance of those models, in practice they are useful only for small dimensional networks and with some limiting simplifying assumptions that are contrary to fact in real functioning wireless networks. In connection with that, in the majority of the research work models of an isolated cell are analyzed.

In the overwhelming majority of available work, one-dimensional queuing models of call handling processes in an isolated cell of mono-service CWN are proposed. However these models cannot describe the studying processes in multi-service CWN since in such net-works calls of heterogeneous traffic differ significantly with respect to their bandwidth requirement and arrival rate and channel occupancy time. In connection with that in the given paper new two-dimensional (2-D) queuing models of multi-service networks are developed. In order to be specific we consider integrated voice/data CWN. In such networks voice calls (v-calls) are more susceptible to possible losses and delays than data (original or handover) call (d-calls). That is why a number of different CAC strategies for prioritization of v-calls are suggested in various works, mostly implying the use of guard channels (cutoff strategy) for high priority calls [7–8] and/or threshold strategies [9] which restrict the number of low priority calls in channels.

In this paper we introduce a unified approach to approximate performance analysis of two multi-parametric CAC in a single cell of un-buffered integrated voice/data CWN which differs from known works in this area. Our approach is based on the principles of theory of phase merging of stochastic systems [10].

The proposed approach allows overcoming an assumption made in almost all of the known papers about equality of handling intensities of heterogeneous calls. Due to this assumption the functioning of the CWN is described with one-dimensional Markov chain (1-D MC) and authors managed simple formulas for calculating the QoS metrics of the system. However as it was mentioned in [11] the assumption of the same mean channel occupancy time even for both original and handover calls of the same class is unrealistic. The presented models are more general in terms of handling intensities and the equality is no longer required.

This paper is organized as follows. In Section 2, we provide a simple algorithm to calculate approximate values of desired QoS metrics of the model of integrated voice/data networks under CAC based on the guard channels strategy. A similar algorithm is suggested in Section 3 for the same model under CAC based on a threshold strategy. In Section 4, we give results of numerical experiments which indicate high accuracy of the proposed approximate algorithms as well as a comparison of QoS metrics in different CAC strategies. In Section 5 we provide some conclusion remarks in conclusion.

2. The CAC Based on Guard Channels Strategy

It is well-known that in an integrating voice/data CWN voice calls of any type (original or handover) have a high priority over data calls and within each, flow handover calls have high priority over original calls.

As a means of assigning priorities to handover v-calls (hv-call) in such networks a back-up scheme that involves reserving a particular number of guard channels of a cell, expressly for calls of this type, is often utilized. According to this scheme any hv-call is accepted if there exists at least one free channel, while calls of the remaining kind are accepted only when the number of busy channels does not exceed some class-dependent threshold value.

We consider a model of an isolated cell in an integrated voice/data CWN without queues. This cell contains N channels, $1 < N < \infty$. These channels are used by Poisson flows of hv-calls, original v-calls (ov-calls), handover d-calls (hd-calls) and original d-calls (od-calls). Intensity of x-calls is λ_x, $x \in \{\text{ov,hv,od,hd}\}$. As in almost all cited works the values of handover intensities are considered known hereinafter, although it is apparent that definition of their values depending on the intensity of original calls, shape of a cell, mobility of an MS and etc. is rather challenging and complex. However, if we consider the case of a uniform traffic distribution and at most one handover per call, the average handover intensity can be given by the ratio of the average call holding time to the average cell sojourn time [12].

To handle any narrow-band v-call (either original or handover) only one free channel is required, while one wide-band d-call (either original or handover) requires simultaneously $b \geq 1$ channels. Here it is assumed that wide-band d-calls are inelastic, *i.e.* all b channels are occupied and released simultaneously (though can be investigated as can models with elastic d-calls).

Note that the channels' occupancy time considers both components of occupancy time: the measure of calls duration, and their mobility. Distribution functions of channel occupancy time of heterogeneous calls are assumed to be independent and exponential, but their parameters are different, namely the intensity of handling of voice (data) calls equals μ_v (μ_d), and generally speaking $\mu_v \neq \mu_d$. If during call handling the handover procedure is initiated, the remaining handling time of this call in a new cell (yet as an h-call) is also exponentially distributed with the same mean due to the memory-free property of exponential distribution.

In a given CAC the procedure by which the channels are engaged by calls of different types is realized in the

following way. As was mentioned before, if upon arrival of an hv-call, there is at least one free channel, this call seizes one of any free channels; otherwise this call is blocked. With the purpose of defining the proposed CAC for calls of other types, three parameters N_1, N_2 and N_3 (where $1 \le N_1 \le N_2 \le N_3 \le N$) are introduced. It is assumed that N_1 and N_2 are multiples of b.

Arrived ov-call is accepted if the number of busy channels is less than N_3, otherwise it is blocked. Arrived od-call (respectively, hd-call) is accepted only in the case at most N_1-b (respectively, N_2-b) busy channels, otherwise it is blocked.

Consider the problem of finding the major QoS metrics of the given multi-parametric CAC strategy – blocking (loss) probabilities of calls of each type and overall channels utilization. For simplicity of intermediate mathematical transformations first we shall assume that b=1. The case $b > 1$ is straightforward (see below).

By adopting an assumption for the type of distribution laws governing the incoming traffics and their holding times it becomes possible to describe the operation of an isolated cell by means of a two-dimensional Markov chain (2-D MC), *i.e.* in a stationary regime the state of the cell at an arbitrary moment of time is described by a 2-D vector $\boldsymbol{n}=(n_d, n_v)$, where n_d (respectively, n_v) is the number of data (respectively, voice) calls in the channels. Then the state space of the corresponding Markov chain describing this call handling scheme is defined as follows:

$$S := \{\boldsymbol{n} : n_d = 0,1,\cdots,N_2,\ n_v = 0,1,\cdots,N,\ n_d + n_v \le N\}. \quad (1)$$

Elements of generating matrix of this MC $q(\mathbf{n},\mathbf{n}'), \mathbf{n},\mathbf{n}' \in S$ are determined from the following relations:

$$q(\boldsymbol{n},\boldsymbol{n}') = \begin{cases} \lambda_d & \text{if } n_d + n_v \le N_1 - 1,\ \boldsymbol{n}' = \boldsymbol{n} + \boldsymbol{e}_1, \\ \lambda_{hd} & \text{if } N_1 \le n_d + n_v \le N_2 - 1,\ \boldsymbol{n}' = \boldsymbol{n} + \boldsymbol{e}_1, \\ \lambda_v & \text{if } n_d + n_v \le N_3 - 1,\ \boldsymbol{n}' = \boldsymbol{n} + \boldsymbol{e}_2, \\ \lambda_{hv} & \text{if } N_3 \le n_d + n_v \le N - 1,\ \boldsymbol{n}' = \boldsymbol{n} + \boldsymbol{e}_2, \\ n_d\mu_d & \text{if } \boldsymbol{n}' = \boldsymbol{n} - \boldsymbol{e}_1, \\ n_v\mu_v & \text{if } \boldsymbol{n}' = \boldsymbol{n} - \boldsymbol{e}_2, \\ 0 & \text{in other cases,} \end{cases} \quad (2)$$

where $\lambda_d := \lambda_{od} + \lambda_{hd}$, $\lambda_v := \lambda_{ov} + \lambda_{hv}$, $\boldsymbol{e}_1 = (1,0)$, $\boldsymbol{e}_2 = (0,1)$.

State diagram of the model and the system of global balance equations (SGBE) for the steady state probabilities $p(\boldsymbol{n})$, $\boldsymbol{n} \in S$ are shown in [13]. Existence of stationary regime is proved by the fact that all states of finite-dimensional state space S are communicating.

Desired QoS metrics are determined via stationary distribution of the initial model. Let P_x denote the blocking probability of the x-calls, $x \in \{\text{hv,ov,hd,od}\}$. Then by using the PASTA theorem [14] we obtain:

$$P_{hv} := \sum_{\boldsymbol{n}\in S} p(\boldsymbol{n})\delta(n_d + n_v, N)\ , \quad (3)$$

$$P_{ov} := \sum_{\boldsymbol{n}\in S} p(\boldsymbol{n})I(n_d + n_v \ge N_3)\ , \quad (4)$$

$$P_{hd} := \sum_{\boldsymbol{n}\in S} p(\boldsymbol{n})I(n_d + n_v \ge N_2)\ , \quad (5)$$

$$P_{od} := \sum_{\boldsymbol{n}\in S} p(\boldsymbol{n})I(n_d + n_v \ge N_1)\ , \quad (6)$$

where $I(A)$ denoted the indicator function of event A and $\delta(i, j)$ represents Kronecker's symbols.

The mean number of busy channels $\tilde{N}$ is also calculated via stationary distribution as follows:

$$\tilde{N} := \sum_{k=1}^{N} kp(k)\ , \quad (7)$$

where $p(k) = \sum_{\boldsymbol{n}\in S} p(\boldsymbol{n})\delta(n_d + n_v, k), k = \overline{1, N}$, are marginal probability mass functions.

Stationary distribution is determined as a result of the solution of an appropriate SGBE of the given 2-D MC. However, to solve the last problem one requires laborious computation efforts for large values of N since the corresponding SGBE has no explicit solution. Very often the solution of such problems is evident if the corresponding 2-D MC has reversibility property [15] and hence there exists stationary distribution in a multiplicative form. Given the SGBE has a multiplicative solution only in a special case when $N_1 = N_2 = N_3 = N$ (even in this case there are known computational difficulties). However, by applying Kolmogorov criteria [15] it is easily verified that the given 2-D MC is not reversible. Indeed, according to the mentioned criteria the necessary reversibility condition of 2-D MC consists in the fact that if the transition from state (i,j) into state (i',j') exists, then there must also be the reverse transition from state (i',j') to state (i,j). However, for MC considered this condition is not fulfilled. So by the relations (2) in the given MC the transition $(n_d,n_v) \to (n_d\text{-}1,n_v)$ exists with intensity $n_d\mu_d$ where $n_d + n_v \ge N_2$, but the inverse transition not existing.

In [13], a recursive technique has been proposed as the solution to the above-mentioned SGBE. It requires multiple inversion calculations of certain matrices of sufficiently large dimensions that in itself is a complex calculating procedure. To overcome the mentioned difficulties, a new, efficient and refined approximate method for the calculation of the stationary distribution of the given model is suggested below. The proposed method, due to right selection of state space splitting of corresponding 2-D MC allows one to reduce the solution of the problem considered to calculation by explicit formulae which contain the known (even tabulated) stationary distributions of classical queuing models.

For correct application of phase merging algorithms (PMA) it is assumed below that $\lambda_v \gg \lambda_d$ and $\mu_v \gg \mu_d$. This assumption is not extraordinary for an integrating voice/data CWN, since this is a regime that commonly occurs in multimedia networks, in which wideband d-calls have both longer holding times and significantly smaller arrival rates than narrowband v-calls, e.g. see [16, 17]. Moreover, it is more important to note, that the final results as shown below, are independent of traffic parameters, and are determined from their ratio, *i.e.* the developed approach can provide a refined approximation even when parameters of heterogeneous traffic are only moderately distinctive.

The following splitting of state space (1) is examined:

$$S = \bigcup_{k=0}^{N_2} S_k, \quad S_k \bigcap S_{k'} = \varnothing, \quad k \neq k', \tag{8}$$

where $S_k := \{\boldsymbol{n} \in S : n_d = k\}$.

Further state classes S_k combine into separate merged states <*k*> and the following merging function in state space *S* is introduced:

$$U(\boldsymbol{n}) =< k > \quad \text{if } \boldsymbol{n} \in S_k,\ k = \overline{0, N_2}. \tag{9}$$

Function (9) determines the merged model which is a one-dimensional Markov chain (1-D MC) with the state space $\tilde{S} := \{< k >: \quad k = \overline{0, N_2}\}$. Then, according to PMA, the stationary distribution of the initial model approximately equals:

$$p(k,i) \approx \rho_k(i)\pi(< k >), \quad (k,i) \in S_k, \quad k = \overline{0, N_2}, \tag{10}$$

where $\{\rho_k(i) : (k,i) \in S_k\}$ is stationary distribution of a split model with state space S_k and $\{\pi(< k >) :< k > \in \tilde{S}\}$ is stationary distribution of a merged model, respectively.

By using (2) we conclude that the elements of generating matrix of this 1-D birth-death processes (BDP) $q_k(i,j)$ are obtained as follows:

$$q_k(i,j) = \begin{cases} \lambda_v & \text{if } i \le N_3 - k - 1,\ j = i+1, \\ \lambda_{hv} & \text{if } N_3 - k \le i < N,\ j = i+1, \\ i\mu_v & \text{if } j = i-1, \\ 0 & \text{in other cases.} \end{cases}$$

So, stationary distribution within class S_k is the same as that $M/M/N-k/N-k$ queuing system where the service rate of each channel is constant, μ_v and arrival rates are variable quantities. Hence

$$\rho_k(i) = \begin{cases} \dfrac{\nu_v^i}{i!}\rho_k(0) & \text{if } 1 \le i \le N_3 - k, \\ \left(\dfrac{\nu_v}{\nu_{hv}}\right)^{N_3-k} \dfrac{\nu_{hv}^i}{i!}\rho_k(0) & \text{if } N_3 - k + 1 \le i \le N - k, \end{cases} \tag{11}$$

where $\rho_k(0) = \left(\sum_{i=0}^{N_3-k} \dfrac{\nu_v^i}{i!} + \left(\dfrac{\nu_v}{\nu_{hv}}\right)^{N_3-k} \sum_{i=N_3-k+1}^{N-k} \dfrac{\nu_{hv}^i}{i!}\right)^{-1}$,

$\nu_v := \lambda_v / \mu_v, \nu_{hv} := \lambda_{hv} / \mu_v$.

Then, from (2) and (11) by means of PMA elements of generating matrix of a merged model $q(< k >,< k' >)$, $< k >,< k' >\in \tilde{S}$ are found:

$$q(< k >, < k' >) = \begin{cases} \lambda_d \sum_{i=0}^{N_1-k-1} \rho_k(i) + \lambda_{hd} \sum_{i=N_1-1}^{N_2-k-1} \rho_k(i) \\ \text{if } 0 \le k \le N_1 - 1, k' = k+1, \\ \lambda_{hd} \sum_{i=0}^{N_2-k-1} \rho_k(i) \\ \text{if } N_1 \le k \le N_2 - 1, k' = k+1, \\ k\mu_d \\ \text{if } k' = k-1, \\ 0 \\ \text{in other cases.} \end{cases} \tag{12}$$

The latter formula allows determining the stationary distribution of a merged model. It coincides with an appropriate distribution of state probabilities of a 1-D BDP, for which transition intensities are determined in accordance with (12). Consequently, stationary distribution of a merged model is determined as

$$\pi(< k >) = \frac{\pi(< 0 >)}{k!\mu_d^k} \prod_{i=1}^{k} q(< k-1 >, < k >), \ k = \overline{1, N_2}, \tag{13}$$

where, $\pi(< 0 >) = \left(1 + \sum_{k=1}^{N_2} \dfrac{1}{k!\mu_d^k} \prod_{i=1}^{k} q(< k-1 >, < k >)\right)^{-1}$

Then by using (11) and (13) from (10) stationary distribution of the initial 2-D MC can be found. So, summarizing the above given and omitting the complex algebraic transformations the following approximate formulae for the calculation of QoS metrics (3)–(7) can be suggested:

$$P_{hv} \approx \sum_{k=0}^{N_2} \pi(< k >) \rho_k(N-k); \tag{14}$$

$$P_{ov} \approx \sum_{k=0}^{N_2} \pi(< k >) \sum_{i=N_3-k}^{N-k} \rho_k(i); \tag{15}$$

$$P_{hd} \approx \sum_{k=0}^{N_2} \pi(< k >) \sum_{i=N_2-k}^{N-k} \rho_k(i); \tag{16}$$

$$P_{od} \approx \sum_{k=0}^{N_1-1} \pi(< k >) \sum_{i=N_1-k}^{N-k} \rho_k(i) + \sum_{k=N_1}^{N_2} \pi(< k >); \tag{17}$$

$$\tilde{N} \approx \sum_{i=1}^{N} i \sum_{k=0}^{f_{N_2}(i)} \pi(< k >) \rho_k(i-k). \tag{18}$$

Hereinafter $f_k(x) = \begin{cases} x & \text{if } 1 \le x \le k, \\ k & \text{if } k \le i \le N. \end{cases}$

Now we can develop the algorithm to calculate the QoS metrics of the investigated multi-parametric CAC for the similar model with wide-band d-calls (due to the limited volume of work this algorithm does not present here).

3. The CAC Based on Threshold Strategy

Now we consider an alternative CAC in the integrated voice/data networks which is based on a threshold strategy. A more detailed description of the given CAC, follows. As in the CAC based on guard channels, we assume that an arrived hv-call is accepted as long as at least one free channel is available; otherwise it is blocked. For the purpose of definition of CAC based on a threshold strategy for calls of other types, three parameters R_1, R_2 and R_3, where $1 \le R_1 \le R_2 \le R_3 \le N$ are introduced. Then the proposed CAC defines the following rules for admission of heterogeneous calls: an od-call (respectively, hd-call and ov-call) is accepted only if the number of calls of the given type in progress is less than R_1 (respectively, R_2 and R_3) and a free channel is available; otherwise it is blocked.

For the sake of simplicity we shall assume that $b=1$. The case $b > 1$ is straightforward (see Section 2). The state of the system under the given CAC at any time is also described by 2-D vector $\boldsymbol{n} = (n_d, n_v)$, where n_d (respectively, n_v) is the number of data (respectively, voice) calls in the channels. Then state space of appropriate 2-D MC is given by:

$$S := \left\{ \boldsymbol{n} : n_d = \overline{0, R_2},\ n_v = \overline{0, N};\ n_d + n_v \le N \right\}. \qquad (19)$$

The elements of generating matrix of the appropriate 2-D MC in this case is determined as follows:

$$q(\boldsymbol{n}, \boldsymbol{n}') = \begin{cases} \lambda_d & \text{if } n_d \le R_1 - 1,\ \boldsymbol{n}' = \boldsymbol{n} + \boldsymbol{e}_1, \\ \lambda_{hd} & \text{if } R_1 \le n_d \le R_2 - 1,\ \boldsymbol{n}' = \boldsymbol{n} + \boldsymbol{e}_1, \\ \lambda_v & \text{if } n_v \le R_3 - 1,\ \boldsymbol{n}' = \boldsymbol{n} + \boldsymbol{e}_2, \\ \lambda_{hv} & \text{if } R_3 \le n_v \le N - 1,\ \boldsymbol{n}' = \boldsymbol{n} + \boldsymbol{e}_2, \\ n_d \mu_d & \text{if } \boldsymbol{n}' = \boldsymbol{n} - \boldsymbol{e}_1, \\ n_v \mu_v & \text{if } \boldsymbol{n}' = \boldsymbol{n} - \boldsymbol{e}_2, \\ 0 & \text{in other cases.} \end{cases} \qquad (20)$$

Blocking probability of hv-calls and the mean number of busy channels are defined similarly to (3) and (7), respectively. The other QoS metrics are defined as the following marginal distributions of initial chain:

$$P_{ov} := \sum_{\boldsymbol{n} \in S} p(\boldsymbol{n}) I(n_v \ge R_3)\ , \qquad (21)$$

$$P_{hd} := \sum_{\boldsymbol{n} \in S} p(\boldsymbol{n}) \delta(n_d, R_2) + \sum_{n \in S} p(\boldsymbol{n}) \delta(n_d + n_v, N) I(n_d < R_2)\ , \qquad (22)$$

$$P_{od} := \sum_{\boldsymbol{n} \in S} p(\boldsymbol{n}) I(n_d \ge R_1) + \sum_{n \in S} p(\boldsymbol{n}) \delta(n_d + n_v, N) I(n_d < R_1)\ . \qquad (23)$$

Unlike the CAC based on the guard channel strategy, it is easy to see that under this one there is no circulation flow in the state diagram of the underlying 2-D MC, *i.e.*, reversible [15]. In other words, there is a general solution to the system of local balance equations (SLBE) in this chain. Therefore, by choosing any path between these states in the state diagram, we can express any state probability $p(n_d,n_v)$ using the state probability $p(0,0)$. So, in case $R_2+R_3 \le N$ we get the following multiplicative solution for stationary distribution of the underlying 2-D MC:

$$p(n_d, n_v) = \begin{cases} \dfrac{\nu_d^{n_d}}{n_d!} \cdot \dfrac{\nu_v^{n_v}}{n_v!} \cdot p(0,0), \\ \text{if } n_d \le R_1,\ n_v \le R_3, \\ \dfrac{\nu_d^{n_d}}{n_d!} \cdot \dfrac{\nu_{hv}^{n_v}}{n_v!} \left(\dfrac{\nu_v}{\nu_{hv}} \right)^{R_3} \cdot p(0,0), \\ \text{if } n_d \le R_1,\ R_3 < n_v \le N, \\ \dfrac{\nu_{hd}^{n_d}}{n_d!} \cdot \dfrac{\nu_v^{n_v}}{n_v!} \cdot \left(\dfrac{\nu_d}{\nu_{hd}} \right)^{R_1} \cdot p(0,0), \\ \text{if } R_1 < n_d \le R_2,\ n_v \le R_3, \\ \dfrac{\nu_{hd}^{n_d}}{n_d!} \cdot \dfrac{\nu_{hv}^{n_v}}{n_v!} \cdot \left(\dfrac{\nu_d}{\nu_{hd}} \right)^{R_1} \cdot \left(\dfrac{\nu_v}{\nu_{hv}} \right)^{R_3} \cdot p(0,0), \\ \text{if } R_1 < n_d \le R_2,\ R_3 < n_v \le N, \end{cases} \qquad (24)$$

where $p(0,0)$ is determined from normalizing condition,

$$p(0,0) = \left(\sum_{n \in S_1} \frac{\nu_d^{n_d}}{n_d!} \cdot \frac{\nu_v^{n_v}}{n_v!} + \left(\frac{\nu_v}{\nu_{hv}} \right)^{R_3} \sum_{n \in S_2} \frac{\nu_d^{n_d}}{n_d!} \cdot \frac{\nu_{hv}^{n_v}}{n_v!} + \left(\frac{\nu_d}{\nu_{hd}} \right)^{R_1} \sum_{n \in T_3} \frac{\nu_{hd}^{n_d}}{n_d!} \cdot \frac{\nu_v^{n_v}}{n_v} + \left(\frac{\nu_d}{\nu_{hd}} \right)^{R_1} \left(\frac{\nu_v}{\nu_{hv}} \right)^{R3} \sum_{n \in S_2} \frac{\nu_{hd}^{n_d}}{n_d!} \cdot \frac{\nu_{hv}^{n_v}}{n_v!} \right)^{-1}$$

Here we use the following notations: $\nu_d := \lambda_d/\mu_d$, $\nu_{hd} := \lambda_{hd}/\mu_d$;

$$S_1 := \left\{ \boldsymbol{n} \in S : n_d \le R_1, n_v \le R_3 \right\},$$
$$S_2 := \left\{ \boldsymbol{n} \in S : n_d \le R_1, R_3 + 1 \le n_v \le N \right\},$$
$$S_3 := \left\{ \boldsymbol{n} \in S : R_1 + 1 \le n_d \le R_2, n_v \le R_3 \right\},$$
$$S_4 := \left\{ \boldsymbol{n} \in S : R_1 + 1 \le n_d \le R_2, R_3 + 1 \le n_v \le N \right\}.$$

In the case $R_2 + R_3 > N$ stationary distribution has the following form:

$$p(n_d,n_v)=\begin{cases}\dfrac{\nu_d^{n_d}}{n_d!}\cdot\dfrac{\nu_v^{n_v}}{n_v!}\cdot p(0,0),\\ \text{if } 0\le n_d\le R_1, 0\le n_v\le R_3,\\ \dfrac{\nu_{hd}^{n_d}}{n_d!}\cdot\dfrac{\nu_v^{n_v}}{n_v!}\cdot\left(\dfrac{\nu_d}{\nu_{hd}}\right)^{R_2}\cdot p(0,0),\\ \text{if } R_1+1\le n_d\le R_2, 0\le n_v\le N-n_d,\\ \dfrac{\nu_d^{n_d}}{n_d!}\cdot\dfrac{\nu_{hv}^{n_v}}{n_v!}\cdot\left(\dfrac{\nu_v}{\nu_{hv}}\right)^{R_3}\cdot p(0,0),\\ \text{if } 0\le n_d\le N-R_3-1, R_3+1\le n_v\le N,\end{cases} \tag{25}$$

where

$$p(0,0)=\left(\sum_{n\in T_1}\frac{\nu_d^{n_d}}{n_d!}\cdot\frac{\nu_v^{n_v}}{n_v!}+\left(\frac{\nu_d}{\nu_{hd}}\right)^{R_1}\sum_{n\in T_2}\frac{\nu_{hd}^{n_d}}{n_d!}\cdot\frac{\nu_v^{n_v}}{n_v!}+\left(\frac{\nu_v}{\nu_{hv}}\right)^{R_3}\sum_{n\in T_3}\frac{\nu_d^{n_d}}{n_d!}\cdot\frac{\nu_{hv}^{n_v}}{n_v}\right)^{-1};$$

$$T_1:=\{\boldsymbol{n}\in S: 0\le n_d\le R_1, 0\le n_v\le R_3\}$$

$$T_2:=\{\boldsymbol{n}\in S: R_1+1\le n_d\le R_2, 0\le n_v\le N-n_d\},$$

$$T_3:=\{\boldsymbol{n}\in S: 0\le n_d\le N-R_3-1, R_3+1\le n_v\le N\}$$

The exact method to determine the steady state probabilities, in terms of a multiplicative representation (25) (or (26)) for large values of *N*, encounters numerical problems such as imprecision and overflow. These are related to the fact that with such a method the entire state space has to be generated, and large factorials and powers, close to zero, of the quantities (for low loads) or large values (for high loads) have to be calculated, *i.e.* there arises the problem of exponent overflow or underflow. Hence we can use a developed approximate method to determine the QoS metrics of the model, under the use of the proposed CAC based on threshold strategy, even when state space (19) is large.

As in Section 2, we assume that $\lambda_v \gg \lambda_d$ and $\mu_v \gg \mu_d$ and examine the following splitting of the state space

$$S=\bigcup_{k=0}^{R_2}S_k,\quad S_k\bigcap S_{k'}=\varnothing,\quad k\ne k',$$

where $S_k:=\{\boldsymbol{n}\in S: n_d=k\}$.

Next classes of states S_k are combined into individual merged states <*k*> and in (19) the merged function with range $\tilde{S}:=\{<k>:\ k=0,1,\cdots,R_2\}$ which is similar to (27) is introduced. As in the exact algorithm, in order to find the stationary distribution within splitting classes S_k we will distinguish two cases: 1) $R_2+R_3\le N$ and 2) $R_2+R_3 > N$. In the first case, the elements of the generating matrix, of appropriate 1-D BDP, are the same for all splitting models, *i.e.*

$$q_k(i,j)=\begin{cases}\lambda_v & \text{if } i\le R_3-1,\ j=i+1,\\ \lambda_{hv} & \text{if } R_3\le i\le N-1,\ j=i+1,\\ i\mu_v & \text{if } j=i-1,\\ 0 & \text{in other cases.}\end{cases}$$

From the last formula we conclude that the stationary distribution within class S_k is the same as that of the *M/M/N – k/N – k* queuing system with state-dependent arrival rates and constant service rate of each channel, *i.e.*

$$\rho_k(i)=\begin{cases}\dfrac{\nu_v^i}{i!}\rho_k(0) & \text{if } 1\le i\le R_3,\\ \left(\dfrac{\nu_v}{\nu_{hv}}\right)^{R_3}\dfrac{\nu_{hv}^i}{i!}\rho_k(0) & \text{if } R_3+1\le i\le N-k,\end{cases} \tag{26}$$

where $\rho_k(0)=\left(\sum_{i=0}^{R_3}\frac{\nu_v^i}{i!}+\left(\frac{\nu_v}{\nu_{hv}}\right)^{R_3}\sum_{i=R_3+1}^{N-k}\frac{\nu_{hv}^i}{i!}\right)^{-1}$.

So, from (20) and (26) we conclude that elements of the generating matrix, of the merged model, are

$$q(<k>,<k'>)=\begin{cases}\lambda_d(1-\rho_k(N-k))\\ \text{if } 0\le k\le R_1-1, k'=k+1,\\ \lambda_{hd}(1-\rho_k(N-k))\\ \text{if } R_1\le k\le R_2-1, k'=k+1,\\ k\mu_d\\ \text{if } k'=k-1,\\ 0\\ \text{in other cases.}\end{cases} \tag{27}$$

Distribution of the merged model is calculated by using (27) and has the following form:

$$\pi(<k>)=\frac{\pi(<0>)}{k!\mu_d^k}\prod_{i=1}^{k}q(<k-1>,<k>),\ k=\overline{1,R_2}, \tag{28}$$

where, $\pi(<0>)=\left(1+\sum_{k=1}^{R_2}\frac{1}{k!\mu_d^k}\prod_{i=1}^{k}q(<k-1>,<k>)\right)^{-1}$.

Finally the following approximate formulae to calculate the desired QoS metrics, under the use of the proposed CAC based on the threshold strategy, are obtained:

$$P_{hv}\approx\sum_{k=0}^{R_2}\pi(<k>)\rho_k(N-k); \tag{29}$$

$$P_{ov}\approx\sum_{k=0}^{R_2}\pi(<k>)\sum_{i=R_3}^{N-k}\rho_k(i); \tag{30}$$

$$P_{hd}\approx\pi(<R_2>)+\sum_{k=0}^{R_2-1}\pi(<k>)\rho_k(N-k); \tag{31}$$

$$P_{od} \approx \sum_{k=R_1}^{R_2} \pi(<k>) + \sum_{k=0}^{R_1-1} \pi(<k>)\rho_k(N-k); \quad (32)$$

$$N_{av} \approx \sum_{k=1}^{N} k \sum_{i=0}^{f_{R_2}(k)} \pi(<i>)\rho_i(k-i). \quad (33)$$

In the second case (*i.e.* when $R_2 + R_3 > N$) distributions for splitting models with state space S_k for $k = 0,1, \cdots, N-R_3-1$ are calculated by using relations (26) while distributions for splitting models with state space S_k for $k = N-R_3, \cdots, R_2$ coincides with distributions of model $M/M/N-k/N-k$ with load ν_v erl, see (22). And all stages of the developed procedure, to calculate the QoS metrics, are the same as in the first case except the calculating of P_{ov}. The last QoS metric in this case is calculated as follows:

$$P_{ov} \approx \sum_{k=0}^{N-R_3} \pi(<k>) \sum_{i=R_3}^{N-k} \rho_k(i) + \sum_{k=N-R_3+1}^{R_2} \pi(<k>)\rho_k(N-k). \quad (34)$$

4. Numerical Results

First briefly consider some results for the CAC based on the guard channels strategy in the integrated voice/data model with four classes of calls. The developed approximate formulas allow, without any computing difficulties, to carry out the authentic analysis of the QoS metrics in any change of the values of the loading parameters in the heterogeneous traffic, satisfying the assumption concerning their ratio (*i.e.* when $\lambda_v \gg \lambda_d$ and $\mu_v \gg \mu_d$) and also at any number of channels in a cell. Some results are shown in **Figures 1-3** where $N = 16$, $N_3 = 14$, $N_2 = 10$, $\lambda_{ov} = 10$, $\lambda_{hv} = 6$, $\lambda_{od} = 4$, $\lambda_{hd} = 3$, $\mu_v = 10$, $\mu_d = 2$. Behavior of the studied curves fully confirms all theoretical expectations.

In the given model, at the fixed value of the total number of channels (N), it is possible to change values of three threshold parameters (N_1, N_2 and N_3). In other words, there is three degrees of freedom. Note, that the increase in value of one of the parameters (in admissible area) favorably influences the blocking of the probability of calls of the corresponding type only (see **Figures 1** and **2**). So, in these experiments, the increase in value of the parameter N_1 leads to a reduction of the blocking probability of od-calls, but other blocking probabilities (*i.e.* P_{hv}, P_{ov} and P_{hd}) increase. At the same time, the increase in value of any parameter leads to an increase in the overall channels utilization (see **Figure 3**).

Research in other directions consists of an estimation of the accuracy of the developed approximate formulas to calculate the QoS metrics. Exact values (EV) of QoS metrics are determined from SGBE. It is important to note, that under fulfilling the mentioned assumptions related to the ratio of the loading parameters of hetero

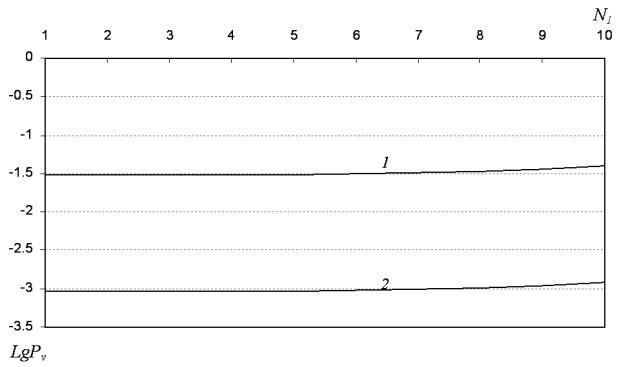


Figure 1. Blocking probability of v-calls versus N_1: 1-P_{ov}; 2-P_{hv}.

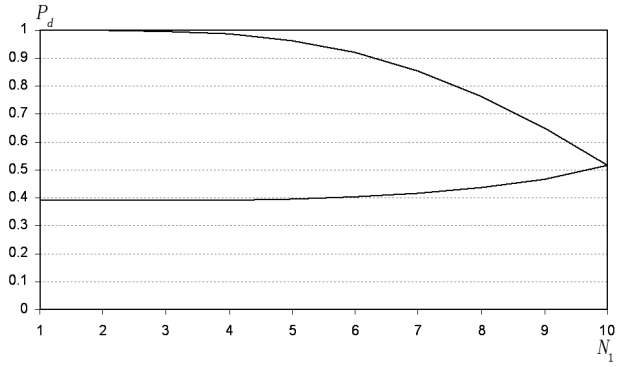


Figure 2. Blocking probability of d-calls versus N_1: 1-P_{od}; 2-P_{hd}.

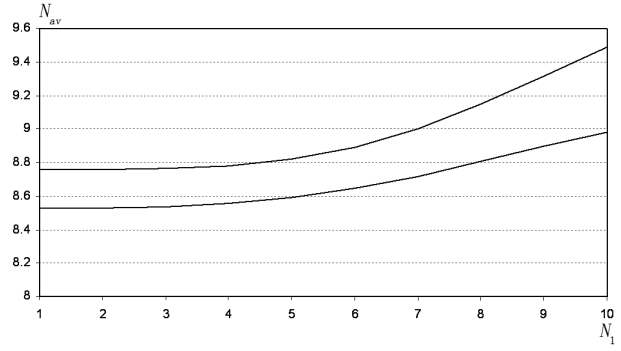


Figure 3. Average number of busy channels versus N_1: 1-N_3 = 15; 2-N_3 = 11.

geneous traffic, the exact and approximate values (AV) almost completely coincide in all QoS metrics. Therefore these comparisons are not shown here. At the same time, it is obvious that finding the exact values of QoS metrics on the basis of the solution of SGBE appears effective only for models with moderate dimension.

It is important to note the sufficiently high accuracy of the suggested formulae even in the case where the accepted assumption about the ratio of the traffic loads is not fulfilled. To facilitate the computation efforts, as exact values of QoS metrics, we use their values calcu-

lated from explicit formulas, see [21]. In mentioned work, appropriate results are obtained in the special case $b = 1$ and $\mu_v = \mu_d$. Let's note, that condition $\mu_v = \mu_d$ contradicts our assumption $\mu_v \gg \mu_d$. Highest accuracy of the developed approximate formulas is observed at the calculation of the QoS metric for v-calls, since the maximal difference between exact and approximate values does not exceed 0.001. Small deviations take place in the calculation of the QoS metrics for d-calls, but also thus in the worst case scenario the absolute error of the proposed formulas does not exceed 0.09, that are quite comprehensible in an engineering practice. Similar results are observed for an average number of occupied channels of a cell. It is important to note, that numerous numerical experiments have shown, that at all admissible loads, the accuracy of the proposed approximate formulas grows with the increase in the value of the total number of channels. It is clear that in terms of simplicity and efficiency, the proposed approach is emphatically superior to the approach based on the use of balance equations to calculate the QoS metrics of the given CAC in the model with a non-identical channel occupancy time.

Also note, that high accuracy in the calculation of the QoS metrics for v-calls is observed even at those loadings which do not satisfy any of the accepted assumptions above concerning the ratio of intensities of heterogeneous traffic. So, for example, at the same values of number of channels and parameters of strategy, at $\lambda_{ov} = 4$, $\lambda_{hv} = 3$, $\lambda_{od} = 10$, $\lambda_{hd} = 6$, $\mu_v = \mu_d = 2$ (*i.e.* when assumptions $\lambda_v \gg \lambda_d$, $\mu_v \gg \mu_d$ are not fulfilled) the absolute error for the mentioned QoS metric does not exceed 0.002. Similar results are observed for an average number of occupied channels of a cell. However, the proposed approximate formulas show low accuracy for d-calls since for them the maximal absolute error exceeds 0.2.

Numerical experiments with the CAC based on the threshold strategy are carried out also. Due to the limited volume of work these results are not presented here. As in CAC based on guard channels, the increase in value of one of the parameters (in admissible area), favorably affect the blocking probability of calls of the corresponding type only. So, the increase in value of parameter R_1 leads to a reduction of the blocking probability of od-calls but other blocking probabilities (*i.e.* P_{hv}, P_{ov} and P_{hd}) increase. At the same time, the increase in the value of any parameter leads to an increase in the overall channels utilization.

At the end of this section we conducted research on comparative analysis of QoS metrics of two schemes: CAC based on the guard channels scheme and the CAC based on the threshold strategy. Comparison was done in the broad range of the number of channels and load parameters. In each access strategy the total number of channels is fixed and controllable parameters are N_1, N_2, N_3 (for CAC based on the guard channels scheme) and R_1, R_2, R_3 (for CAC based on the threshold strategy). As mentioned above, the behavior of the QoS metrics, with respect to the indicated controllable parameters, in different CAC, are the same.

It is important to note that with the given number of channels, loads and QoS requirements in either of the CAC strategies may or may not meet the requirements. For instance, in the model of mono-service CWN for the given values of $N = 100$, $\nu_o = 50$ erl, $\nu_h = 35$ erl following requirements $P_o \le 0.1$, $P_h \le 0.007$ and $\tilde{N} \ge 80$ are not met with CAC based on guard channels irrespective of the value of the parameter g (number of guard channels), whereas CAC based on individual pool only for h-calls (*i.e.* $r_o = 0$) meets the requirements at $r_h = 40$. However, for the same given initial data, requirements $P_o \le 0.3$, $P_h \le 0.0001$ and $\tilde{N} \ge 60$ are only met by CAC based on the guard channel scheme at $g = 20$, and never met by CAC based on the individual pool strategy irrespective of the value of its parameter r_h. Thus it is possible to find the optimal strategy (in a given context) at the given loads without changing the number of channels.

Apparently, both strategies have the same implementation complexity. That is why the selection of either of them, at each particular case, must be based on the answer to the following question: does it meet the given QoS requirements? These issues are a subject to a separate investigation.

5. Conclusions

In this paper, an effective and refined approximate approach to the performance analysis of the un-buffered integrated voice/data CWN, under different multi-parametric CAC, has been proposed. Note that many well-known results related to the mono-service CWN are special cases of such proposed ones. In almost all of the available work on devoted mono-service CWN, the queuing model is investigated with assumption that both handover and original calls are identical in terms of channel occupancy time. This assumption is rather limiting and unrealistic. Here, models of the un-buffered integrated voice/data CWN are explored with more general parameter requirements. Performed numerical results demonstrate high accuracy of the developed approximate method.

It is important to note that the proposed approach may facilitate the solution of problems related to the selection of the optimal (in given sense) values of parameters in the investigated multi-parametric CAC. These problems are subjects to separate investigation.

6. Acknowledgements

This research has been supported by Sangji University Research Fund 2009.

7. References

[1] S. Tekinay and B. Jabbari, "Handover policies and channel assignment strategies in mobile cellular networks," IEEE Communication Magazine, Vol. 29, No. 11, pp. 42–46, 1991.

[2] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems," IEEE Personal Communications, pp. 10–31, June 1996.

[3] S. DasBit and S. Mitra, "Challenges of computing in mobile cellular environment—a survey," Computer Communications, Vol. 26, pp. 2090–2105, 2003.

[4] R. J. Boucherie and M. Mandjes, "Estimation of performance measures for product form cellular mobile communications networks," Telecommunication Systems, Vol. 10, pp. 321–354, 1998.

[5] R. J. Boucherie and N. M. Van Dijk, "On a queuing network model for cellular mobile telecommunications networks," Operation Research, Vol. 48, No. 1, pp. 38–49, 2000.

[6] W. Li and X. Chao, "Modeling and performance evaluation of cellular mobile networks," IEEE/ACM Transactions on Networking, Vol. 12, No. 1, pp. 131–145, 2004.

[7] D. Hong and S. S. Rapoport, "Traffic model and performance analysis of cellular mobile radio telephones systems with prioritized and non-prioritized handoff procedures," IEEE Transactions on Vehicular Technology, Vol. 35, No. 3, pp. 77–92, 1986.

[8] G. Haring, R. Marie, R. Puigjaner, and K. Trivedi, "Loss formulas and their application to optimization for cellular networks," IEEE Transactions on Vehicular Technology, Vol. 50, No. 3, pp. 664–673, 2001.

[9] B. Gavish and S. Sridhar, "Threshold priority policy for channel assignment in cellular networks," IEEE Transactions on Computers, Vol. 46, No. 3, pp. 367–370, 1997.

[10] V. S. Korolyuk and V. V. Korolyuk, "Stochastic model of systems," Kluwer Academic Publishers, Boston, 1999.

[11] W. Yue and Y. Matsumoto, "Performance analysis of multi-channel and multi-traffic on wireless communication networks," Kluwer Academic Publishers, Boston, 2002.

[12] S. Nanda, "Teletraffic models for urban and suburban microcells: Cell sizes and hand-off rates," IEEE Transactions on Vehicular Technology, Vol. 42, No. 4, pp. 673–682, 1993.

[13] S. E. Ogbonmwan and L. Wei, "Multi-threshold bandwidth reservation scheme of an integrated voice/data wireless network," Computer Communications, Vol. 29, No. 9, pp. 1504–1515, 2006.

[14] R. W. Wolff, "Poisson arrivals see time averages," Operations Research, Vol. 30, No. 2, pp. 223–231, 1992.

[15] F. P. Kelly, "Reversibility and stochastic networks," John Wiley & Sons, New York, 1979.

[16] V. Casares-Giner, "Integration of dispatch and interconnect traffic in a land mobile trunking system. Waiting time distributions," Telecommunication Systems, Vol. 16, No. 3–4, pp. 539–554, 2001.

[17] A. G. Greenberg, R. Srikant, and W. Whitt, "Resource sharing for book-ahead and instantaneous-request calls," IEEE/ACM Transactions on Networking, Vol. 7, No. 1, pp. 10–22, 1999.

[18] A. Z. Melikov and A. T. Babayev, "Refined approximations for performance analysis and optimization of queuing model with guard channels for handovers in cellular networks," Computer Communications, Vol. 29, No. 9, pp. 1386–1392, 2006.

[19] R. L. Freeman, "Reference manual for telecommunications engineering," John Wiley & Sons, New York, 1994.

[20] A. Z. Melikov, M. I. Fattakhova, and A. T. Babayev, "Investigation of cellular communication networks with private channels for service of handover calls," Automatic Control and Computer Sciences, Vol. 39, No. 3, pp. 61– 69, 2005.

[21] H. Chen, L. Huang, S. Kumar, an d C. C. Kuo, "Radio resource management for multimedia QoS supports in wireless networks," Kluwer Academic Publishers, Boston, 2004.

Wireless Sensor Network, **2010, 2, 227-232**
doi:10.4236/wsn.2010.23030 Published Online March 2010 (http://www.scirp.org/journal/wsn)

Scientific Research

Combined Nodal Method and Finite Volume Method for Flow in Porous Media

Abdeslam Elakkad[1], Ahmed Elkhalfi[1], Najib Guessous[2]
[1]*Laboratoire Génie Mécanique, Faculté des Sciences et Techniques, Route d'Imouzzer, Fès, Morocco*
[2]*Département de mathématiques et informatique, Ecole normale Supérieure de Fès, Bensouda, Fès, Morocco*
E-mail: *elakkadabdeslam@yahoo.fr*
Received December 4, 2009; *revised December* 14, 2009; *accepted December* 16, 2009

Abstract

This paper describes a numerical solution for two dimensional partial differential equations modeling (or arising from) a fluid flow and transport phenomena. The diffusion equation is discretized by the Nodal methods. The saturation equation is solved by a finite volume method. We start with incompressible single-phase flow and move step-by-step to the black-oil model and compressible two phase flow. Numerical results are presented to see the performance of the method, and seem to be interesting by comparing them with other recent results.

Keywords: Saturation Equation, Nodal Methods, Finite Volume Method, Two-Phase Simulation

1. Introduction

Nodal methods have long been one of the most popular discretization techniques employed within the reactor physics community to solve multigroup diffusion problem [1,2]. A survey of these methods can be founds in [3].

The Finite volume method (FV) has been proposed initially by Durlofsky *et al.* in 1990 for the advection equations and Burgers. Other works have been introduced by J. P. Cioni, in 1995, R. Eymard *et al.* in 1997 [4] and A. Shamsai and H. R. Vosoughifar [5].

In this paper we consider the model for incompressible two-phase flow in a porous medium. We consider Nodal methods for solving the diffusion equation and a general class of explicit finite volume upwind schemes for solving purely advective transport in the absence of gravity and capillary forces. We shall employ a Newton-Raphson method to solve the implicit system.

Section 2 presents the model problem used in this paper. The discretization by Nodal methods described is in Section 3. The discretization by finite volume method for the diffusion equation described is in Section 4. Section 5 shows the discretization by finite volume method for the saturation equation. Numerical experiments carried out within the framework of this publication and their comparisons with other results are shown in Section 6.

2. Governing Equations

Here we consider incompressible two-phase flow in domain $\Omega \subset IR^2$.

The pressure equation is given by

$$\begin{cases} \operatorname{div}(U) = f & on \quad \Omega, \\ U = -K\lambda(S)\nabla P & on \quad \Omega, \\ P = \overline{P} & on \quad \Gamma^D, \\ U \cdot n = 0 & on \quad \Gamma^N. \end{cases} \tag{1}$$

We consider the saturation equation in its simplest form (neglecting gravity and capillary forces)

$$\varphi \frac{\partial S}{\partial t} + \nabla \cdot (g(S)U) = \frac{f_w}{\rho_w} \quad on \quad \Omega \times]0,T[\tag{2}$$

We shall assume that the phases are oil (o) and water (w) and that the two phases together fill the void space completely so that

$$S_w + S_O = 1 \tag{3}$$

The source term for the saturation equation becomes:

$$\frac{f_w}{\rho_w} = \max(f,0) + g(S)\min(f,0).$$

$$\frac{f_w}{\rho_w} = \max(f,0) + g(S)\min(f,0). \tag{4}$$

where $g(S)=\dfrac{S^2}{S^2+(1-S^2)}$

To close the model, we must also supply expressions for the saturation-dependent quantities. Here we use simple analytical expressions:

$$\lambda_w(S)=\frac{(S^*)^2}{\mu_w},\quad \lambda_0(S)=\frac{(1-S^*)^2}{\mu_0},\quad S^*=\frac{S-S_{wc}}{1-S_{or}-S_{wc}}$$

K is the permeability, ϕ is the Rock porosity, μ_0 and μ_w Viscosities, S_{or} is the irreducible oil saturation, S_{wc} is the connate water saturation, and the total mobility is given by $\lambda=\lambda_w+\lambda_0$.

3. Discretization with Nodal Methods

The discretizations which follow assume that Ω is a rectangular domain and employ an $L\times M$ tensor product mesh having cells $\Omega_{i,j}=(x_{i-\frac{1}{2}},x_{i+\frac{1}{2}})\times(y_{j-\frac{1}{2}},y_{j+\frac{1}{2}})$. It is convenient to denote the mesh spacing by $\Delta x_i=x_{i+\frac{1}{2}}-x_{i-\frac{1}{2}}$ and $\Delta y_j=y_{j+\frac{1}{2}}-y_{j-\frac{1}{2}}$.

Common to all nodal discretizations is the choice of cell-and edge-based unknowns. Generally, these are taken to be moments up to some specified order; hence in the lowest-order case simple averages are employed. Specifically, the cell-based unknowns are averages defined by

$$P_{i,j}=\frac{1}{\Delta x_i\Delta y_j}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}P(x,y)dxdy \qquad (5)$$

while the edge-based scalar unknowns, namely, edge averages of the scalar flux, are given by

$$P_{i+\frac{1}{2},j}=P_j(x_{i+\frac{1}{2}}),\ P_j(x)=\frac{1}{\Delta y_j}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}P(x,y)dy \qquad (6)$$

with the analogous definitions of $P_{i,j+\frac{1}{2}}$ and $P_i(y)$. Similarly, the edge-averaged currents are written as

$$U^{\pm}_{i+\frac{1}{2},j}=\lim_{x\to x^{\pm}_{i+\frac{1}{2}}}U_j(x),\ U_j(x)=\frac{1}{\Delta y_j}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}\left[-(K\lambda)\frac{\partial}{\partial x}\{P(x,y)\}\right]dy$$

while $U^{\pm}_{i,j+\frac{1}{2}}$ is defined analogously.

All lowest-order members of the nodal method family, such as the nodal expansion methods (NEM) [6], the nodal integration method (NIM) [7], the nodal Green's function method (NGFM) [8] and the coarse mesh expansion methods [9] yield equivalent discretization of (1). We have chosen to present a brief discussion of the NIM. The first step in the NIM discretization consists of posing the cell-based trans-verse integrated ODEs which govern $P_j(x)$ and $P_i(y)$. To this end we assume that a homogenized diffusion coefficient $(K\lambda)_{ij}$ is defined on each cell, and transverse integrate (1) to obtain

$$-\frac{\partial}{\partial x}\left\{(K\lambda)_{ij}\frac{\partial}{\partial x}P_j(x)\right\}=-\frac{1}{\Delta y_j}\left\{U^-_{j+\frac{1}{2}}(x)-U^+_{j-\frac{1}{2}}(x)\right\}+f_j(x),\quad x_{i-\frac{1}{2}}\le x\le x_{i+\frac{1}{2}},$$

$$-\frac{\partial}{\partial y}\left\{(K\lambda)_{ij}\frac{\partial}{\partial y}P_i(y)\right\}=-\frac{1}{\Delta x_i}\left\{U^-_{i+\frac{1}{2}}(y)-U^+_{i-\frac{1}{2}}(y)\right\}+f_i(y),\quad y_{j-\frac{1}{2}}\le y\le x_{j+\frac{1}{2}},$$

where $U^{\mp}_{j\pm\frac{1}{2}}(x)$ and $U^{\mp}_{i\pm\frac{1}{2}}$ are one-sided limits of the normal currents along the edges of the cell, and f_j is defined in analogy with the transverse averaged unknowns.

Thus, we have reduced the discretization of the PDE given in (1) to that of two ODEs that are coupled through pseudo source terms. The definition of the edge averages (6) naturally yields Dirichlet boundary conditions for each cell. Moreover, for lowest-order or constant-constant NIM the pseudo source term (*i.e.* $U^{\pm}_{j\mp\frac{1}{2}}(x)$ and $U^{\pm}_{i\mp\frac{1}{2}}(y)$) are assumed to be constant along their respective cell edges. We assume that the source $f(x,y)$ is constant over each cell.

With expressions for $P_j(x)$ and $P_i(y)$ in hand we construct the discretization. First, note that two independent definitions of the cell average are possible:

$$P_{i,j}=\frac{1}{\Delta x_i}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}P_j(x)dx=\frac{1}{\Delta y_j}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}P_i(y)dy$$

Yielding 2LM equations, e.g.

$$P_{i,j}=\frac{1}{2}(P_{i+\frac{1}{2},j}+P_{i-\frac{1}{2},j})-\frac{\Delta x_i^2}{12}(K\lambda)^{-1}_{i,j}\left\{\frac{1}{\Delta y_j}\left(U^-_{i,j+\frac{1}{2}}-U^+_{i,j-\frac{1}{2}}\right)-f_{i,j}\right\}.$$

Furthermore, under the assumption of an homogenized diffusion coefficient and utilizing $U=-K\lambda(S)\nabla P$ we obtain expressions for $U^{\pm}_{i\mp\frac{1}{2},j}$, $U^{\pm}_{i,j\mp\frac{1}{2}}$, on each cell, e.g.

$$U^-_{i+\frac{1}{2},j}=-\frac{(K\lambda)_{i,j}}{\Delta x_i}(P_{i+\frac{1}{2},j}+P_{i-\frac{1}{2},j})+\frac{1}{2\Delta x_i}\left\{\frac{1}{\Delta y_j}(U^-_{i,j+\frac{1}{2}}-U^+_{i,j-\frac{1}{2}})-f_{i,j}\right\}$$

Although these comprise for equations per cell, only three of them are linearly independent, as the same balance equation arises from both $U^-_{i+\frac{1}{2},j}-U^+_{i-\frac{1}{2},j}$ and $U^-_{i,j+\frac{1}{2}}-U^+_{i,j-\frac{1}{2}}$. Imposing continuity of $J\cdot n$ yields an equation for each interior edge (*i.e.* $(L\text{-}1)M+L(M\text{-}1)$ equations) while the boundary conditions give rise to $2L+2M$ discrete boundary equations. Thus we have $7LM+L+M$ equations as many unknowns.

Although the constant-constant NIM discretization is complete at this point, it is seldom used in this form. Typically in the literature one proceeds by eliminating

the edge currents $U^{\pm}_{i\mp\frac{1}{2},j}$, $U^{\pm}_{i,j\mp\frac{1}{2}}$, followed by the trivial elimination of the averages $P_{i,j}$ to obtain the 7-point nearest neighbor hexagonally coupled stencils that govern the edge-based fluxes $P_{i+\frac{1}{2}j}$:

$$
\begin{aligned}
&(K\lambda)_{i,j}\left(\frac{\Delta y_j}{\Delta x_i}\right)\left(P_{i+\frac{1}{2},j}-P_{i-\frac{1}{2},j}\right)-(K\lambda)_{i+1,j}\left(\frac{\Delta y_j}{\Delta x_{i+1}}\right)\left(P_{i+\frac{3}{2},j}-P_{i+\frac{1}{2},j}\right)\\
&+\frac{3(K\lambda)_{i,j}\,\Delta x_i\Delta y_j}{\Delta x_i^2+\Delta y_j^2}\left\{P_{i+\frac{1}{2},j}+P_{i-\frac{1}{2},j}-P_{i,j+\frac{1}{2}}-P_{i,j-\frac{1}{2}}\right\}\\
&+\frac{3(K\lambda)_{i+1,j}\,\Delta x_{i+1}\Delta y_j}{\Delta x_{i+1}^2+\Delta y_j^2}\left\{P_{i+\frac{3}{2},j}+P_{i+\frac{1}{2},j}-P_{i+1,j+\frac{1}{2}}-P_{i+1,j-\frac{1}{2}}\right\}\\
&=\frac{1}{2}\frac{\Delta x_i\Delta y_j^3}{\Delta x_i^2+\Delta y_j^2}f_{i,j}+\frac{1}{2}\frac{\Delta x_{i+1}\Delta y_j^3}{\Delta x_{i+1}^2+\Delta y_j^2}f_{i+1,j}
\end{aligned}
$$

With the y-oriented rotated analogue at $P_{i,j+\frac{1}{2}}$.

4. Finite Volume Method for the Diffusion Equation

In a finite volume method for (1) one introduces a family of control volumes (typically the given mesh) and imposes mass conservation locally for each control volume E: $\int_{\partial E}-K\lambda\nabla P.nds=\int_E f dx$, where n is the outward unit normal on ∂E. The cell-centered finite volume method, which may be expressed as a pressure stencil for each cell E_i,

$$\sum_j T_{ij}\left(P_i-P_j\right)=\int_{E_i} f dx, \tag{7}$$

where j loops over all cells neighboring cell E_i. The transmissibilities T_{ij} are associated with cell interfaces, and the expression $T_{ij}\left(P_i-P_j\right)$ is a discrete form of $-\int_{\partial E_i\cap\partial E_j}K\lambda\,\nabla P.n_{ij}ds$, where n_{ij} is the unit normal on $\Gamma_{ij}=\partial E_i\cap\partial E_j$ pointing into E_j. Thus, $T_{ij}\left(P_i-P_j\right)$ estimates the total flux across $\partial E_i\cap\partial E_j$. The system (7) is clearly symmetric, and a solution is, as for the continuous problem, defined up to an arbitrary constant. The system is made positive definite and symmetry is preserved, by forcing $P_1=0$, for instance. That is, by adding a positive constant to the first diagonal of the matrix $A=[a_{ik}]$,
where

$$a_{ik}=\begin{cases}\sum_j T_{ij} & if\ k=i,\\ -T_{ik} & if\ k\neq i.\end{cases} \tag{8}$$

The matrix A is sparse, consisting of a tridiagonal part corresponding to the *x*-derivative, and two off-diagonal bands corresponding to the *y*-derivatives.

5. Finite Volume Method for the Saturation Equation

For the saturation Equation (2), we use a finite volume scheme on the form

$$S_i^{n+1}=S_i^n+\left(\delta_x^t\right)_i\left(\max(f_i,0)-\sum_j g(S^m)_{ij}\,U_{ij}+g\left(S^m_i\right)\min(f_i,0)\right) \tag{9}$$

where $\left(\delta_x^t\right)_i=\frac{\Delta t}{\left(\varphi_i|\Omega_i|\right)}$, ϕ_i is the porosity in Ω_i, f_i denotes the source term, Δt is the time step, and S_i^k is the cell-average of the water saturation at time $t=t_k$,

$$S_i^k=\frac{1}{|\Omega_i|}\int_{\Omega_i}S(x,t_k)dx.$$

By specifying for time level m in the evaluation in the fractional flow functions, we obtain either an explicit ($m = n$) or implicit ($m = n + 1$) scheme. Here U_{ij} is the total flux (for oil and water) over the edge Γ_{ij} between two adjacent cells Ω_i and Ω_j, and g_{ij} is the fractional flow function at Γ_{ij}

$$g_w(S)_{ij}=\begin{cases}g_w(S_i) & if\ U.n_{ij}\geq 0,\\ g_w(S_j) & if\ U.n_{ij}<0.\end{cases} \tag{10}$$

The explicit solver is obtained by using $m = n$ in (9). Explicit schemes are only stable provided that the time step Δt satisfies a CFL condition. For the homogeneous transport equation (with $f\equiv 0$), the CFL condition for the first-order upwind scheme reads

$$\max_{S\in(0,1)}\left|g'(S)\right|\max_i\left(\delta_x^t\right)_i\sum_j\left|U_{ij}\right|\leq 2\left(1-S_{wc}-S_{or}\right) \tag{11}$$

For the inhomogeneous equation, we can derive a stability condition on Δt using a more heuristic argument.

Physically, we require that $S_{wc}\leq S_i^{n+1}\leq 1-S_{or}$.

This implies that the discretization parameters $|\Omega_i|$ and Δt must satisfy the following dynamic conditions:

$$S_{wc}\leq S_i^n+\left(\delta_x^t\right)_i\left(\max(f_i,0)-\sum_j g(S^n)_{ij}\,U_{ij}+g\left(S_i^n\right)\min(f_i,0)\right)\leq 1-S_{or} \tag{12}$$

The interface fluxes U_{ij} satisfy the following mass balance property:

$$U_i^{in} = \max(f_i,0) - \sum_j \min(U_{ij},0) = -\min(f_i,0) + \sum_j \max(U_{ij},0)$$
$$= U_i^{out}.$$

Thus, since $0 \le g(S) \le 1$ it follows that

$$-g(S_i^n)U_i^{in} = \max(f_i,0) - \sum_j g(S^n)_{ij}\, U_{ij} + g(S_i^n)\min(f_i,0)$$
$$\le (1-g(S_i^n))U_i^{in}$$

Then the general saturation dependent stability condition holds in Ω_i if the following inequality is satisfied:

$$\max\left(\frac{g(S_i^n)-0}{S_i^n - S_{wc}}, \frac{1-g(S_i^n)}{1-S_{or}-S_i^n}\right)\left(\delta_x^t\right)_i U_i^{in} \le 1 \quad (13)$$

Finally to remove the saturation dependence from (13), we invoke the mean value theorem and deduce that (13) holds whenever

$$\Delta t \le \frac{\varphi|\Omega_i|}{U_i^{in} \max\left\{g'(S)\right\}_{0\le S\le 1}} \quad (14)$$

6. Numerical Simulations

In this section we illustrate the performance of the Nodal methods. We restrict to quarter-five spot problems in two dimensions with uniform rectangular grids [10,11]. We compare the solutions obtained using Nodal Methods with the reference fine scale solution obtained using a two point finite volume method.

We compute a mean pressure error in the following manner:

$$\varepsilon(P) = \frac{\left\|P - P^{ref}\right\|^2}{\left\|P^{ref}\right\|^2} \quad (15)$$

where P and P^{ref} are array vectors that contain the average pressure in each fine element (Nodal Methods solution and reference finite volume solution, respectively).

We measure velocity solution errors using the following measure:

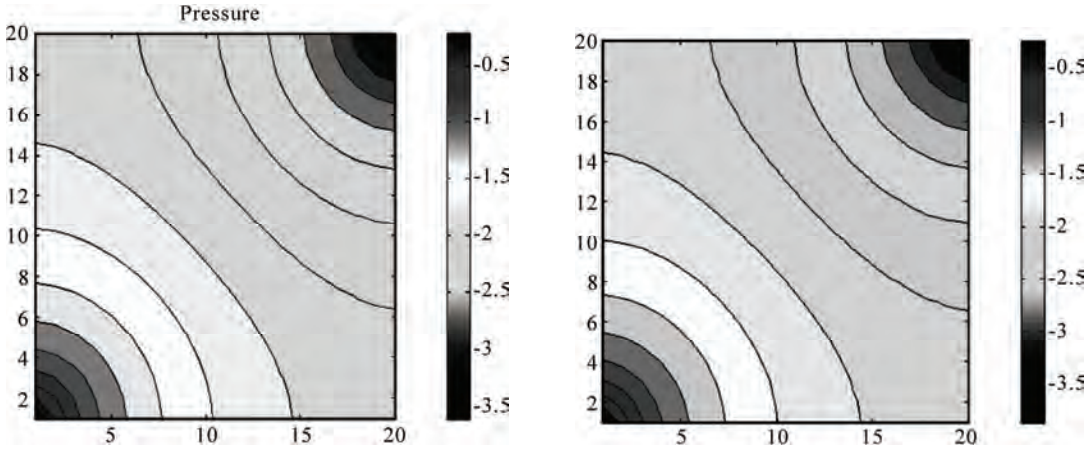
$$\varepsilon(U) = \frac{\left\|U_x - U_x^{ref}\right\|^2}{\left\|U_x^{ref}\right\|^2} + \frac{\left\|U_y - U_y^{ref}\right\|^2}{\left\|U_x^{ref}\right\|^2} \quad (16)$$

U_x and U_y are vectors containing the average velocities in the x-and y-directions, respectively, $\|.\|$ and is the Euclidean norm.

The pressure equation is discretized by the Nodal volume method (FV). We use an Explicit and Implicit upwind finite volume discretization for the saturation Equation (2).

Example: Here, we present a simulator for incompressible and immiscible Two-Phase flow system. We first want to look at the so called the quarter-five spot problems. The reservoir is $\Omega = [0,1]\times[0,1]$ with a unit injection well placed at (0; 0) and a production well with unit production rate placed at (1; 1). Let us revisit the quarter-five spot problems, but now assume that the reservoir is initially filled with pure oil. To produce the oil in the upper-right corner, we inject water in the lower left. We assume unit porosity, unit viscosities for both phases, and set $S_{or} = S_{wc} = 0$. We impose no-flow boundary conditions on $\partial\Omega$.

We obtain a similar result with J. E. Aarnes *et al.* [10]. The water saturation is increasing monotonically toward the injector, meaning that more oil is gradually displaced as more water is injected, and the pressure fields are symmetric.



Figuret 1. The left plot shows pressure contours for a homogeneous quarter-five spot obtained using Nodal Methods and the right plot shows pressure contours for a homogeneous quarter-five spot with the reference solution obtained by J. E. Aarnes *et al.* [10], with a 20 × 20 square grid.

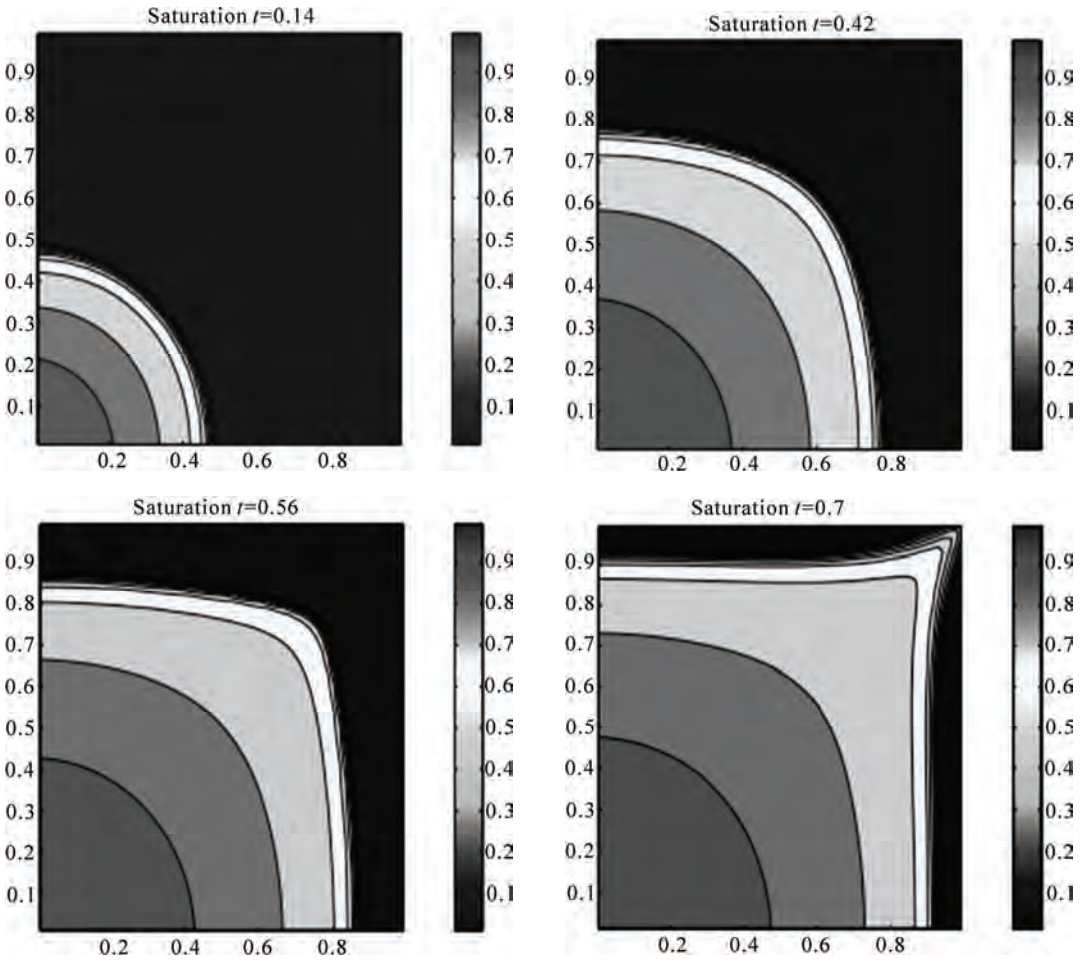


Figure 2. Saturation profiles for the homogeneous quarter-five spot at several different times: t = 0.14, t = 0.42, t = 0.56 and t = 0.7.

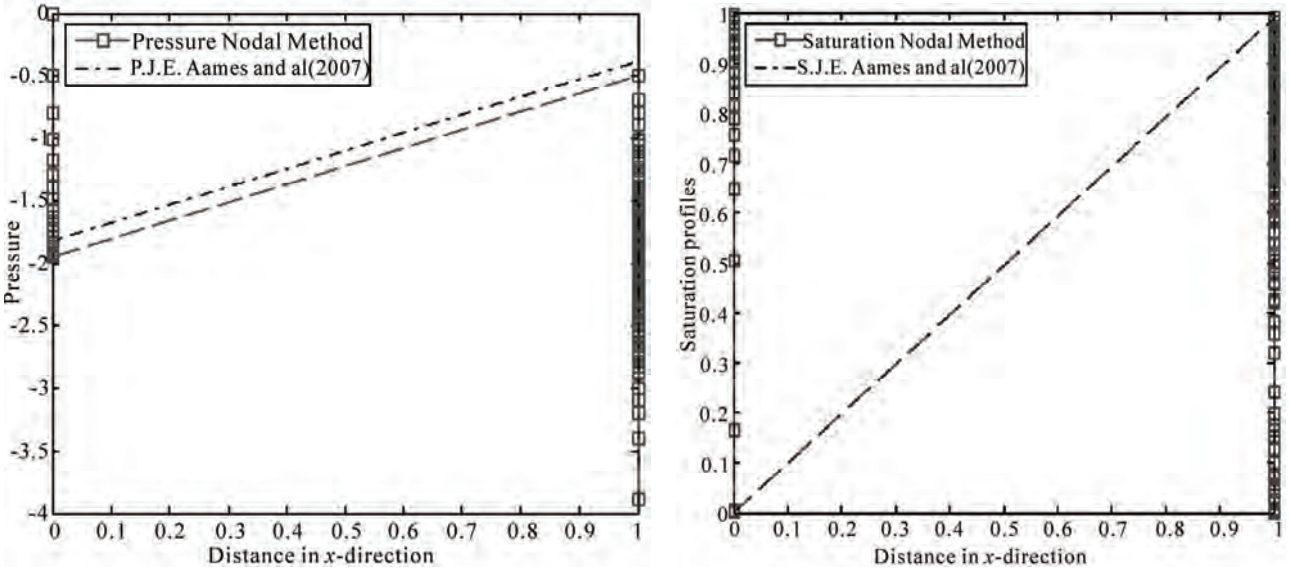


Figure 3. The left shows pressure obtained with Nodal Method and pressure obtained by J. E. Aarnes *et al.*, and the right shows saturation profiles obtained with Nodal Method and saturation profiles obtained by J. E. Aarnes *et al.*, with a 20 × 20 square grid.

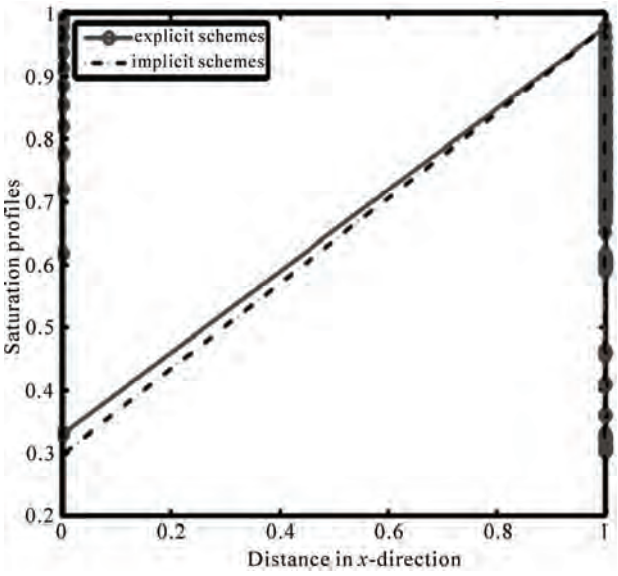


Figure 4. Saturation profiles for explicit and implicit schemes for the homogeneous quarter-five spot.

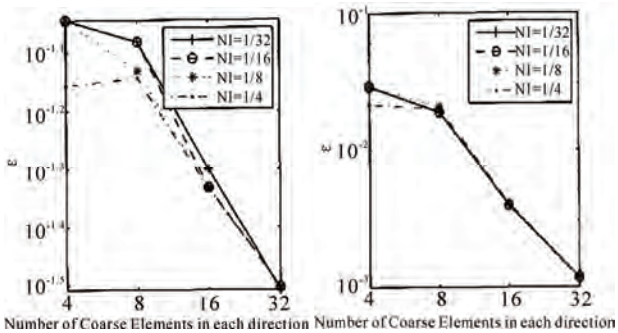


Figure 5. Pressure (left) and Velocity (Right) errors for different coarse discretizations and well length scales used to compute the fine well contributions. The error decreases with increasing numbers of coarse cells.

7. Conclusions

We were interested in this work in the numeric solution for two dimensional partial differential equations modeling a fluid flow and transport phenomena. In this article, numerical approximation of two-phase incompressible problems is studied. The diffusion equation is discretized by the Nodal Methods. The saturation equation is solved by a finite volume method. The pressure and velocity field are symmetric about both diagonals for the homogeneous field. The water saturation is increasing monotonically toward the injector for the homogeneous field, meaning that more oil is gradually displaced as more water is injected.

Numerical results are presented to see the performance of the method, and seem to be interesting by comparing them with other recent results.

8. Acknowledgments

The authors would like to express their sincere thanks for the referee for his/her helpful suggestions.

9. References

[1] J. W. Song and J. K. Kim, "An efficient nodal method for transient calculations in light water reactors," Nuclear Technology, Vol. 103, pp. 157–167, 1993.

[2] N. K. Gupta, "Nodal methods for three-dimensional simulators," Progress in Nuclear Energy, Vol. 7, pp. 127–149, 1981.

[3] R. D. Lawrence, "Progress in nodal methods for the solution of the neutron diffusion and transport equations," Progress in Nuclear Energy, Vol. 17, pp. 271–301, 1986.

[4] R. Eymard, T. Gallouet, and R. Herbin, "Finite volume methods," In: P. G. Ciarlet and J. L. Lions, Ed., Handbook of Numerical Analysis, Elsevier Science B.V., Amsterdam, Vol. 7, pp. 713–1020, 1997.

[5] A. Shamsai and H. R. Vosoughifar, "Finite volume discretization of flow in porous media by the MATLAB system," Scientia Iranica, Vol. 11, No. 1–2, pp. 146–153, 2004.

[6] H. Finnemann, F. Bennewitz, and M. R. Wagner, "Interface current techniques for multidimensional reactor calculations," Atomkernenergie, Vol. 30, pp. 123–128, 1977.

[7] H. D. F. Fisher and H. Finnemann, "The nodal integration method—A diverse solver for neutron diffusion problems," Atomkernenergie-Kerntechnik, Vol. 39, pp. 229–236, 1981.

[8] R. D. Lawrence and J. J. Dorning, "A nodal Green's function method for mutidimensional neutron diffusion calculations," Nuclear Science and Engineering, Vol. 76, pp. 218–231, 1980.

[9] B. Montagnini, P. Soraperra, C. Trantavizi, M. Sumini, and D. M. Zardini, "A well-balanced coarse-mesh flux expansion method," Annals of Nuclear Energy, Vol. 21, pp. 45–53, 1994.

[10] G. J. E. Aarnes, T. Gimse, and K.-A. Lie, "An introduction to the numerics of flow in porous media using Matlab," In: G. Hasle, K.-A. Lie, and E. Quak, Ed., Geometrical Modeling, Numerical Simulation and optimization: Industrial Mathematics at SINTEF, Springer Verlag, pp. 265–306, 2007.

[11] D. Burkle and M. Ohlberger, "Adaptive finite volume methods for displacement problems in porous media," Computing and Visualization in Science, Vol. 5, No. 2, 2002.

Wireless Sensor Network, **2010, 2, 233-238**
doi:10.4236/wsn.2010.23031 Published Online March 2010 (http://www.scirp.org/journal/wsn)

Scientific Research

A Novel Energy Aware Clustering Technique for Routing in Wireless Sensor Networks

Ouadoudi Zytoune[1], Youssef Fakhri[2], Driss Aboutajdine[1]
[1]*LRIT unite associée au CNRST, Faculty of Sciences Agdal, Rabat, Morocco*
[2]*University Ibn Tofail, Kenitra, Morocco*
E-mail: *zytoune@gmail.com, yousseffakhri@yahoo.fr, aboutaj@fsr.ac.ma*
Received December 20, 2009; *revised December* 28, 2009; *accepted December* 30, 2009

Abstract

Cluster-based architectures are one of the most practical solutions in order to cope with the requirements of large-scale wireless sensor networks (WSN). Cluster-head election problem is one of the basic QoS requirements of WSNs, yet this problem has not been sufficiently explored in the context of cluster-based sensor networks. Specifically, it is not known how to select the best candidates for the cluster head roles. In this paper, we investigate the cluster head election problem, specifically concentrating on applications where the energy of full network is the main requirement, and we propose a new approach to exploit efficiently the network energy, by reducing the energy consumed for cluster forming.

Keywords: Wireless Sensor Networks, Clustering Protocol, Energy Efficiency, Cluster-Head Selection, Information Routing

1. Introduction

Wireless Sensor Networks can offer unique benefits and versatility with respect to low-power and low-cost rapid deployment for many applications, which do not need human supervision. The nodes in WSNs are usually battery operated sensing devices with limited energy resources and replacing or replenishing the batteries is usually not an option. Thus energy efficiency is one of the most important issues and designing power-efficient protocols is critical for prolonging the lifetime. The latest developments in time critical, low cost, long battery life, and low data rate wireless applications have led to work on WSNs. These WSNs have been considered for work in certain applications with limited power, reliable data transfer, short communication range, and reasonably low cost such as industrial monitoring and control, home automation and security, and automotive sensing applications [1]. The WSNs consist of a set of sensors that communicate with each other to form a sensor field. These large numbers of nodes, which have the ability to communicate wirelessly, to perform limited computation, and to sense their surroundings, form the WSNs [2]. Specific functions can be obtained through cooperation between these nodes: functions such as sensing, tracking, and alerting [3]. These functions make these wireless sensors very useful for monitoring natural phenomena, environmental changes, controlling security, estimating traffic flows, monitoring military application, and tracking friendly forces in the battlefields.

For this work, we make some assumptions:

1) The network area is $M \times M$.

2) The number of the network nodes is N.

3) The base station is located in the center of the network.

4) The data packets length is L bits.

5) All network nodes can reach the Base Station.

6) The clustered nodes transmit to their cluster-head, and the not clustered nodes transmit directly to the sink.

7) And, the traffic generation model is supposed a uniform event generation which mean that every sensor has a data packet to be reported in a fixed time or round.

The remainder of this paper is organized as follows. In Section 2, we briefly review related work. Section 3 describes the energy consumption model. Section 4 presents the detail of our algorithm. Section 5 shows the performance of the proposed algorithm by simulations and compares it with literature technique. Finally, Section 6 gives concluding remarks and provides some future work.

2. Related Work

In order to enhance the network lifetime by the period of a particular mission, many routing protocols have been devised. One of these is network clustering, in which network is partitioned into small clusters and each cluster is monitored and controlled by a node, called Cluster Head (CH). These cluster heads can communicate directly with the base station (BS). Other nodes send the data, sensed from the environment to these CHs. CHs first aggregate the data from the multiple sensor nodes, and then finally send it directly to the BS. Hence the CH should be powerful, closer to the cluster-centroid a less vulnerable [4]. Heinzelman *et al*. proposed LEACH [5] a protocol based on network clustering. Each cluster has a cluster-head that aggregates all the data received from the near nodes and send them to the base station. The cluster-head are selected following a distributed algorithm for each round. The [6] proposed an algorithm called TB-LEACH which is an improvement of the LEACH one. This algorithm permits to dominate the number of clusters heads to have at any transmission round, the optimal cluster-heads amount that modifies the cluster-head selection algorithm to improve the partition of cluster. This algorithm assumes that all nodes receive the messages broadcasted by the nodes selected as cluster-heads. On one hand, if a node is not reachable by a cluster head it assumes that the number of clusters heads is insufficient, and elects them to be cluster head, therefore the number of cluster-heads may be not dominated, on the other hand, this is not real with the large networks because the those messages can not reach all the network. PEGASIS [7] is an improvement of the LEACH protocol. Rather than forming multiple clusters, PEGASIS forms chains from sensor nodes so that each node transmits and receives from a neighbor and only one node is selected from that chain to transmit to the base station (sink). Gathered data moves from node to node, aggregated and eventually sent to the base station. The chain construction is performed in a greedy way. However, PEGASIS introduces excessive delay for distant node on the chain. In addition the single leader can become a bottleneck. All the previous techniques are used for homogenous networks where all the nodes have the same initial battery energy. SEP [8] is a proposed scheme for heterogeneous wireless sensor networks, which is composed of two types of nodes according to the initial energy. The advance nodes are equipped with more energy than the normal nodes at the beginning. This technique prolongs the stability period, which is defined as the time until the first node failure.

DEEC [9] is a distributed clustering scheme for heterogeneous wireless sensor networks. In DEEC the cluster-heads are elected by a probability based on the ratio between residual energy of each node and the average energy of the network. The epochs of being cluster-heads for nodes are different according to their initial and residual energy. The nodes with high initial and residual energy will have more chances to be the cluster-heads than the nodes with low energy. In the last cited works, for each round, a new cluster-heads are chosen, so, many control messages are exchanged between these CHs and their closest nodes to form the clusters. These control messages makes some energy lost.

3. Energy Consumption Model

Recently, there is a significant amount of work in the area of building low-energy radios. For the purpose of this study we use similar energy model and analysis as proposed in [5] as shown in **Figure 1**. For the experiments described here, both the free space (d^2 power loss) and the multi path fading (d^4 power loss) channel models were used, depending on the distance between the transmitter and the receiver. If the distance is less than a threshold, the free space (fs) model is used; otherwise, the multi path (mp) model is used.

Thus, to transmit an $L-$ bits message over a distance d, the radio expends (1):

$$E_{TX}(l,d) = E_{TX-elec}(l) + E_{TX-amp}(l,d) \quad (1)$$

And then:

$$E_{TX}(l,d) = \begin{cases} lE_{elec} + l\varepsilon_{fs}d^2 & \text{if } d<d_0 \\ lE_{elec} + l\varepsilon_{mp}d^4 & \text{if } d \geq d_0 \end{cases} \quad (2)$$

where the threshold d_0 is done by (3):

$$d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \quad (3)$$

The electronics energy E_{elec} depends on many factors such as the digital coding, the modulation, the filtering, and the spreading of the signal, whereas the amplifier energy, $\varepsilon_{fs}d^2$ or $\varepsilon_{mp}d^4$, depends on the distance to the receiver and the acceptable bit-error rate.

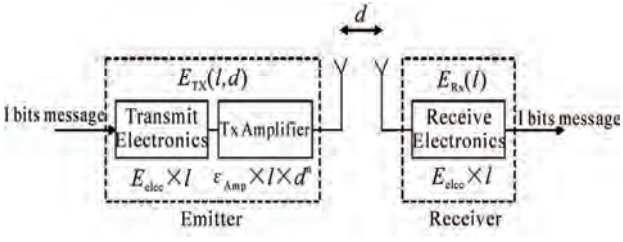


Figure 1. Radio energy dissipation model.

To receive an l-bit message, the radio expends (4):

$$E_{RX}(l) = E_{RX-elec}(l) = lE_{elec}. \quad (4)$$

It is also assumed that the radio channel is symmetric, which means the cost of transmitting a message from A to B is the same as the cost of transmitting a message from B to A. The used parameter values in our work are given in the following **Table 1**.

4. Clustering Technique for Routing in Wireless Sensor Networks

All proposed clustering techniques in literature, use a cluster head rotation in order to balance the transmission energy cost over the network nodes, because the cluster head role is energy expansive. That permits to grant approximately, the same lifetime until the battery energy depletion. So, in every transmission round, some new nodes play concurrence to be elected as cluster head. Each node selected, has to advertise its status to its neighbor nodes, to know the nodes which will belong to its cluster and to schedule the TDMA intervals [5]. Then, some energy is consumed in this state. This energy for clustering control is considerable, and it is important to reduce this energy to use it to exploit the total network energy to extend the network lifetime.

Our contribution consists in reducing the control energy for cluster formation by keeping each selected cluster head for more than one transmission round. So, each node selected as cluster head, play this role for m consecutive transmission rounds before conceding it for upcoming selection nodes. The proposed algorithm, called Clustering Technique for Wireless Sensor Networks (CTRWSN) is a self-organizing, dynamic clustering method that divides dynamically, the network on a number of a priori fixed clusters. Each cluster has one cluster-head. In this work, we use two-level heterogeneous networks, in which there are two types of sensor nodes: the advanced nodes and normal nodes. Let E_0 the initial energy of the normal nodes and, f the fraction of the advanced nodes, which own a times more energy than the normal ones. Thus there are $f.N$ advanced nodes equipped with initial energy of $(1+a)E_0$ and $(1-f)N$ normal nodes equipped with initial energy of E_0.

Table 1. Radio parameter values.

Description	Symbol	Value
Energy consumed by the amplifier to transmit at a shorter distance	ε_{fs}	10pJ/bit/m^2
Energy consumed by the amplifier to transmit at a longer distance	ε_{mp}	0.0013pJ/bit/m^4
Energy consumed in the electronics circuit to transmit or receive the signal	E_{elec}	50nJ/bit
Energy consumed for beam forming	E_{DA}	5nJ/bit/signal

We can compute the total initial energy of the networks which is given by:

$$E_{total} = N(1-f)E_0 + Nf(1+a)E_0.$$

The node n becomes cluster-head for t_n rounds. In homogenous networks, to guarantee that there are average $P_{opt}N$ cluster-heads every round, LEACH let each node n becomes a cluster-head once every $t_n = 1/P_{opt}$ rounds. The network nodes will have different residual energy when the network evolves.

If the rotating epoch t_n is the same for all the nodes as proposed in LEACH, the energy will be not well distributed and the low-energy nodes will die more quickly than the high-energy nodes. DEEC protocol; choose different t_n based on the residual energy $E_n(r)$ of node n at round r.

The probability threshold that each node n use to determine whether itself to become a cluster-head in each round, is given as follow Equation (5):

$$T(n) = \begin{cases} \dfrac{p_n}{\left(1 - p_n.\left(r \bmod \left(\dfrac{1}{p_n}\right)\right)\right)} & \text{if } n \in G \\ 0 & \text{Otherwise} \end{cases} \quad (5)$$

where G is the set of nodes that are eligible to be cluster-heads at round r.

If node n has not been a cluster-head during the most recent $\dfrac{1}{p_n}$ rounds, we have $n \in G$. The p_n parameter is given by Equation (6) from the [9].

$$p_n = \begin{cases} \dfrac{p_{opt}E_n(r)}{(1+af)\bar{E}(r)} & \text{if } n \text{ is a normal node} \\ \dfrac{(1+a)p_{opt}E_n(r)}{(1+af)\bar{E}(r)} & \text{if } n \text{ is an advanced node} \end{cases} \quad (6)$$

where $E_n(r)$ is the residual energy of the node n at the round r, $\bar{E}(r)$ denotes the average energy of the network at round r, which can obtained by (7)

$$\bar{E}(r) = \frac{E_{total}}{N}\left(1 - \frac{r}{R}\right), \quad (7)$$

$$R = \frac{E_{total}}{E_{round}}, \quad (8)$$

$$E_{round} = L\left(2NE_{elec} + NE_{DA} + k\varepsilon_{mp}d_{toBS}^4 + N\varepsilon_{mp}d_{toCH}^2\right), \quad (9)$$

k is the number of clusters, E_{DA} is the data aggregation cost which is expended in the cluster-heads, d_{toBS} is the average distance between the cluster-head and the base station, and d_{toCH} is the average distance between the cluster members and the cluster-head. If the nodes are uniformly distributed, from [5,10] we can give:

$$d_{toCH} = \frac{M}{\sqrt{2k\pi}} \text{ and } d_{toBS} = 0.765\frac{M}{2}.$$

By the Equation (9), we can find the optimal value of k that minimizes E_{round}, which is (10):

$$k_{opt} = \frac{\sqrt{N}}{\sqrt{2\pi}}\frac{\sqrt{\varepsilon_{fs}}}{\sqrt{\varepsilon_{mp}}}\frac{M}{d_{toBS}^2}. \quad (10)$$

This value of k is used to determine E_{round}, and therefore, by (7) and (8) each node n can find the value of the parameter p_n used in $T(n)$ calculation.

And, for each round r, when node n finds it is eligible to be a cluster-head, it will choose a random number between 0 and 1. If the number is less than threshold $T(n)$, the node n becomes a cluster-head during the current round. The operation of CTRWSN is broken up into rounds where each round consists of a set-up phase and steady-state phase. In the following sub-sections we will detail each of these phases.

4.1. A Set-Up Phase

Every transmission round, each node n uses the Formula (5) to calculate the $T(n)$ value and choose a random number between 0 and 1. If this chosen number is less than the calculated threshold $T(n)$, the node n becomes a cluster-head. The selected cluster heads broadcast an advertisement message to the network to declare themselves as cluster heads. Receiving this message, the not cluster head nodes belong to the cluster which the energy to join is minimum among all selected cluster-heads. The node can determine the needed energy to transmit to the cluster head based on the received signal strong.

Once the nodes decide to which cluster belong, they inform the cluster-head transmitting a join-request message to it, using CSMA/CA MAC protocol.

A header, the node ID and the cluster-head ID, forms this message, which is a short one. This message size grants to reduce the time channel access and the transmission energy cost. Receiving all nodes join-messages, the cluster-head schedule a TDMA allocating a slot time to each cluster's nodes. After that, the cluster nodes are informed by a broadcasted message containing the TDMA schedule. The choice of TDMA technique in the cluster allows a energy saving, because no collisions caused and the node can pass to sleep mode when it is not transmitting; in this way, the clusters are formed.

4.2. Steady-State Phase

Once the clusters are established, the nodes transmit their data messages towards the cluster-head. Within the cluster, the communication uses TDMA, as described in the set up phase. When the cluster-head receives all the nodes data, it performs its compression, to form a new message that sent to the base station.

Figure 2 gives the flowchart that explains the work of the proposed algorithm.

The network function is divided into cycles, each cycle lasts for m transmission rounds. Then, selected nodes for cluster-heads play this role for *m* consecutive transmission rounds.

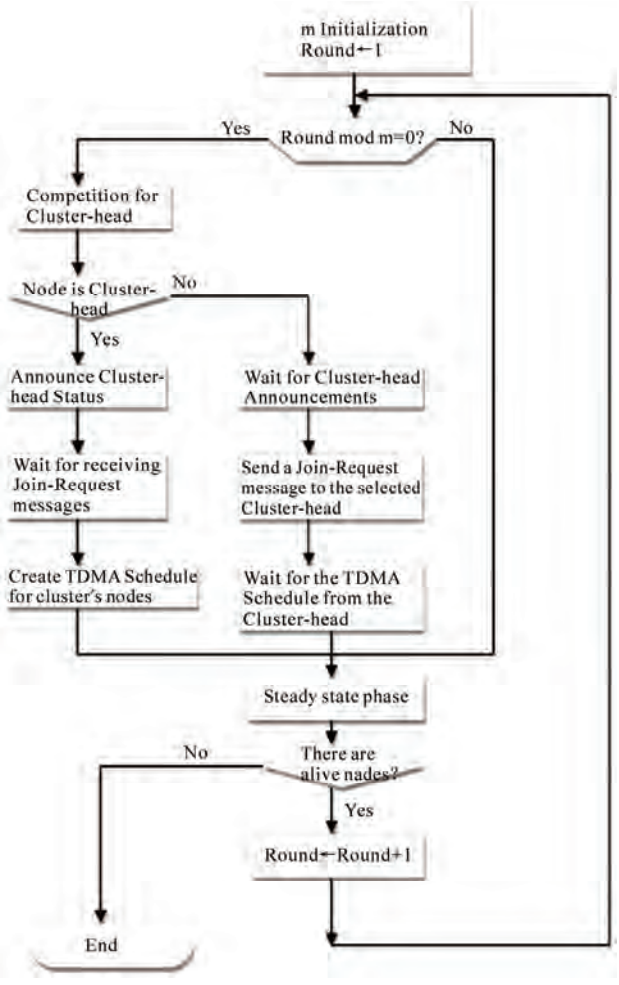


Figure 2. Protocol flowchart.

It is so difficult to determine analytically the parameter *m*, because the nodes deployment is random and the cluster-heads position is also stochastic, then we determine this optimal value based on simulation.

5. Simulation and Results

The simulation parameters values used in our work are given in the **Table 2**.

For the simulation bellow, we fix f = 0.2 and a = 3. **Figure 3** represents the DEEC energy consumed in the network for the clusters forming for every transmission round. This energy is expended by the messages exchanged between the cluster-heads and its belonging nodes to form the clusters. As we can see, a lot of the network energy is lost to control the clusters formations. The total of this energy is evaluated as 33.3824J which is 41.73% of the total network energy, and in average 0.0057J per round.

To find the optimal value of the parameter *m*, we do simulation by varying it. **Figure 4** gives the relative network lifetime. As depicted in this figure, the relative network lifetime becomes approximately constant when the parameter *m* is greater than 10. This lifetime is maximal when the cluster-head duration m is equal to 32; each node selected as cluster-head, plays this role for m consecutive transmission rounds that permits to economize the energy consumed by the cluster-heads to form their clusters, because the number of control messages is reduced. As we can see in the figure, if the cluster-head period duration becomes longer than a threshold, even if the network control energy is reduced, oppositely, the network lifetime decreases slightly. The reason of that is the unbalancing of the cluster-heads energy load over the network nodes. So, some nodes are highly solicited when they belong to a cluster. Continuing to belong to the same cluster, some unlucky nodes have to transmit for larger distance, and then there energy drains quickly. We can see in **Figure 5** the energy for forming the clusters in the CTRWSN. **Figure 6** gives the network lifetime of the proposed protocol compared to DEEC one. The first node dies in the 718th round but in CTRWSN it occurs in the 1271st round, which means that the stable region is extended by up to 77%. The half network nodes die in DEEC at the 832nd round and in CTRWSN it happens in the 1898th round and the last node die in the 5865th in DEEC, and 9494th round in CTRWSN, which is approximately 62% longer than DEEC.

Table 2. Simulation parameter values.

Description	Symbol	Value
Network dimension	M	100m
Number of network nodes	N	100
Data packet length	L	4000bits
Control packet length	L_{crt}	200bits
Optimal probability	p_{opt}	0.1
Advanced Nodes percentage	f	Variable
Fraction of advanced nodes energy to normal nodes	a	Variable

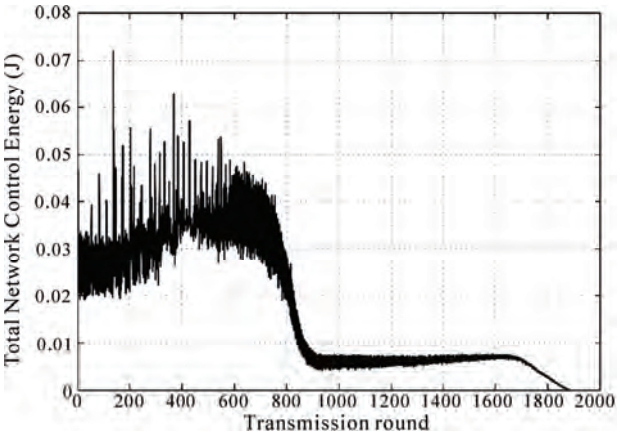


Figure 3. Total network energy for Clusters forming in DEEC.

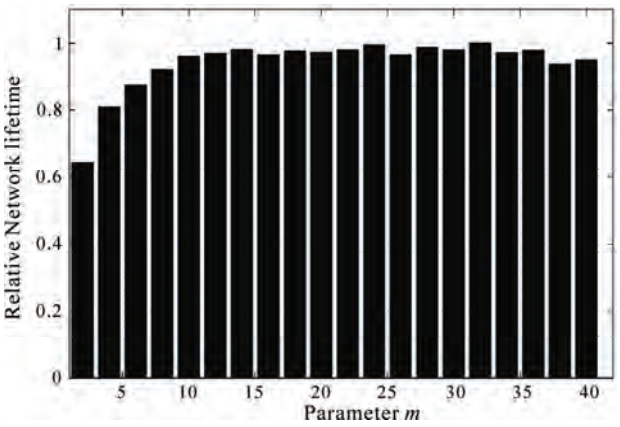


Figure 4. Relative network lifetime.

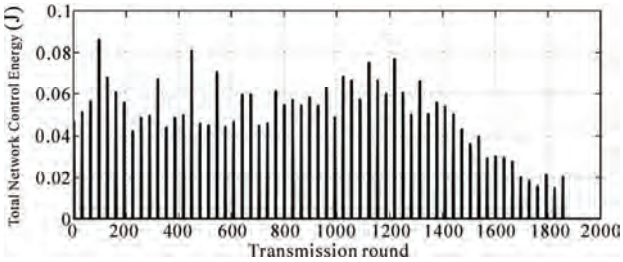


Figure 5. Total network energy for clusters forming in CTRWSN.

Figure 7 gives the network remaining energy. The average network remaining energy in DEEC is 10.2149J/round and in CTRWSN is 13.9502/round that means that the CTRWSN consumes little energy compared to DEEC, which helps to extend the network lifetime for many extra-rounds.

Figures 8(a) and **8(b)** gives the network lifetime defined until the first node dies when *a* and *f* are varying.

This figure shows that the proposed technique provides best results when the network parameters change.

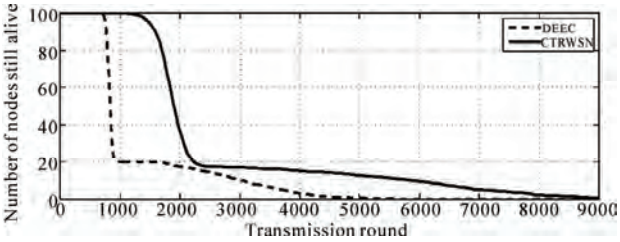


Figure 6. Network lifetime comparison.

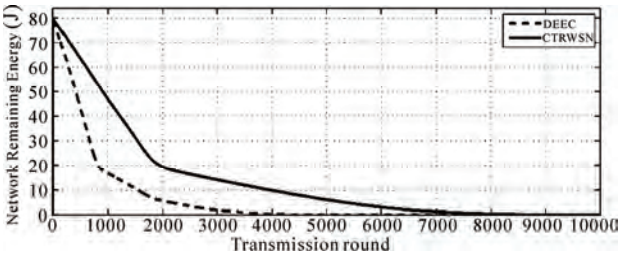


Figure 7. Network remaining energy comparison.

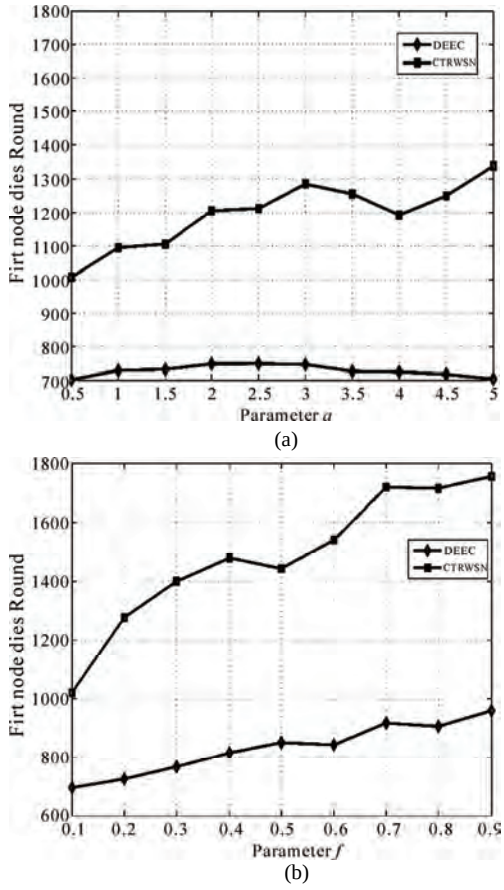


Figure 8. Network lifetime until the first node dies. (a) *f* = 0.2 and *a* varying from 0.5 to 5; (b) *a* = 3 and *f* varying from 0.1 to 0.9.

6. Conclusions and Future Works

In this paper, we have proposed a clustering based routing protocol for heterogeneous WSNs, which is entirely distributed. We have interested in reducing the number of control message, and so, the protocol overhead.

Through the simulation, we demonstrate that the proposed algorithm allows a large stable network lifetime compared to the most known clustering algorithms in this area. As future work, we will reconsider the probability of becoming cluster-head, to extend yet the network lifetime.

7. Acknowledgements

This work was supported by the Hassan II Académie des Sciences et Techniques.

8. References

[1] F. Akyidiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor network: A survey," Computer Networks, Vol. 38, No. 4, pp. 393–422, 2002.

[2] K. Romer, O. Kastin, and F. Mattern, "Middleware challenges for wireless sensor networks," ACM SIGMOBILE Mobile Computing and Communications Review, Vol. 6, No. 4, pp. 59–61, 2002.

[3] R. Shorey, A. Ananda, and W. T. Ooi, "Mobile, wireless, and sensor networks," 1st edition, IEEE Press, John Wiley & Sons, 2006.

[4] Z. Khalid, G. Ahmed, N. M. Khan, and P. Vigneras, "A real-time energy-aware routing strategy for wireless sensor networks," Asia-Pacific Conference on Communications, Bangkok, Thailand, pp. 381–384, 2007.

[5] W. R. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Transactions on Wireless Communications, Vol. 1, No. 4, pp. 660–670, 2002.

[6] J. Hu, Y. Jin, and L. Dou, "A time-based cluster-head selection algorithm for LEACH", Proceeding of IEEE Symposium on Computers and Communications, Marrakech, Morocco, 6–9 July 2008.

[7] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power efficient gathering in sensor information systems," Proceedings of the IEEE Aerospace Conference, Big Sky, Montana, March 2002.

[8] G. Smaragdakis, I. Matta, and A. Bestavros, "SEP: A stable election protocol for clustered heterogeneous wireless sensor networks," Second International Workshop on Sensor and Actor Network Protocols and Applications, 2004.

[9] L. Qing, Q. X. Zhu, and M. W. Wang, "Design of a distributed energy—efficient clustering algorithm for heterogeneous wireless sensor networks," Computer Communication, Elsevier, Vol. 29, No. 12, pp. 2230–2237, 2006.

[10] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," Proceeding of International Conference on Computer Communication, Vol. 3, pp. 1713–1723, April 2003.

Wireless Sensor Network, **2010, 2, 239-242**
doi:10.4236/wsn.2010.23032 Published Online March 2010 (http://www.scirp.org/journal/wsn)

WEP and WPA Improvement

Mustafa ElGili, Samani A. Talab, Awad H. Ali
Department of Information Technology, University of Neelain, Khartoum, Sudan
E-mail: *mustgili@hotmail.com*
Received September 24, 2009; *revised October* 21, 2009; *accepted October* 25, 2009

Abstract

This paper aims to describe a solution to improve wireless network security protocols WEP and WPA based on a modified RC4 algorithm for encryption, and based on initialization vector (IV) with secret key for a session key exchange, and new mutual authentication mechanism.

Keywords: Wireless, Security, Authentication, WEP, WPA, RC4

1. Introduction

Wireless communications offer organizations and users many benefits such as portability and flexibility, increased productivity, and lower installation costs. However, risks are inherent in any wireless technology. The loss of confidentiality and integrity and the threat of denial of service attacks are risks typically associated with wireless communications. WEP and WPA are designed to protect, but they still have weaknesses discussed in [1–8].

2. Related Work

Many solutions have been proposed for the remedy of the WEP and WPA encryption, key exchange and mutual authentication problems. Reference [9] proposes a scheme similar to WEP. The difference is that in EWEP it encrypts the concatenation of the message and IV with RC4. Encrypting IV aims to hide it from eavesdropping. This proposal has many weaknesses represented below:

1) Because IV [*i*+1] depends on IV [*i*], if some frames are lost, all frames coming from the same sender would not be decryptable.

2) It uses Diffe-Hellman [10] to agree IV; this is difficult because there is no clear share value.

3) Rekey is done using special messages. No protection mechanism for the messages contains the new key.

4) No new authentication mechanism described which means the authentication process is still weak.

5) It suffers from following attacks:

Disassociation and Deauthentication Attacks.
Shared Key Authentication Attacks.
FMS Attack.
Session Hijacking.
Reply Attack.

Reference [11] proposes new scheme which modifies the process of WEP Key generation on TKIP. Comparing with TKIP, the proposed protocol neither changes nor increases hardware cost. Moreover, it narrows down hardware computing quantities. Further, SEWTP reduces authentication frequency. It also provides security function as well as TKIP and does not affect the performance of throughput. But it has limitations represented as:

1) Over flooding, any client will send random number. New secret key has been generated before authentication process that leads to AP overload, because every client joining the AP domain will take a time to generate his secret key.

2) The clear IV still there.

3) Every $2^{\wedge 24}$ packet there will be random number from client to AP, this will lead to:

Over load traffic.
Overload in process time.

4) It suffers from the following attacks:

Disassociation and Deauthentication Attacks.
Shared Key Authentication Attacks.
FMS Attack.
Session Hijacking.
SYCH attack.

2.1. Proposed Solution: WEP and WPA Improvement

We will propose a novel schema which essentially consists of three mechanisms:

1) Mutual Authentication between AP and Station.

2) Session Key Exchange mechanism.

3) Strong Encryption Protocol using modified RC4, and IV shadow.

2.1.1. Mutual Authentication Process and Session Key Exchange

This is a solution for weak authentication in Wireless LAN which uses pre-shared key authentication, and also can be used in enterprise solution without any third party. The mutual authentication mechanism and Session Key Exchange mechanism consisting in the following steps:

1) The first step stores the challenge value in the Station like secret key similar to Access point.

2) The second step: When the Station sends associate request to the Access point, the Access Point sends encrypted challenge with the secret key to Station.

3) The Station receives the Encrypted challenge and decrypts it using the secret key and Access point IV, and compares it with the challenge it has. If they are equal that means this Access point is trusted.

4) If this Access Point is trusted Calculate K', IV' using Equations 1, 2 respectively, and uses them with K (secret key), to encrypt the challenge again and sends it to the access point.

5) Update the Station challenge by adding IV' to old challenge.

6) The Access Point receives the Encrypted challenge and calculates the K', IV' and uses them with K (secret key), to decrypt challenge if it is equal to the challenge sent. This means the Station is trusted.

7) Update Access Point challenge by adding IV' to old challenge.

2.1.2. Session Key Exchange Mechanism

During Authentication process, we can calculate the session key, by using the equation below:

$$k' = (IV_{cli} + IV_{acc}) + K^0 + k[i+3] \quad (1)$$

The result of using Equation(1) to calculate K' shown in **Figures 1, 2** and **3**, they explain that there is no linear relation between old IV[*i*] and K'[*i*].

2.1.3. Strong Encryption Using Modified RC4

We modified RC4 to be suitable for use with our approach, by using K' instead of feedback j as it is shown in algorithm below. To prevent first byte attack [9], and inverse attack:

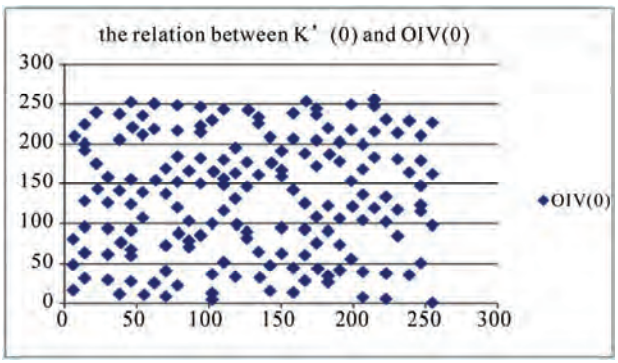


Figure 1. Shows the relation between k'[0] and old IV[0].

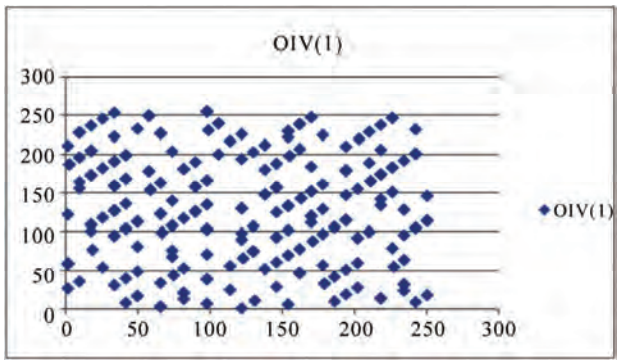


Figure 2. Shows the relation between k'[1] and old IV[1].

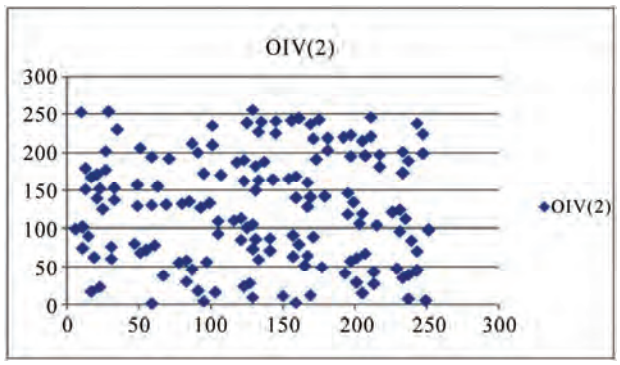


Figure 3. Shows the relation between k'[2] and old IV[2].

RC4 Modified key scheduling algorithm:

1) {initialization}
2) for *i* from 0 to *n* − 1 do
3) S[*i*]: = *i*
4) end for
5) *j*: = 0
6) {generate a random permutation}
7) for *i* from 0 to *n* − 1 do
8) *j*: = (K'[I mod e]+S[*i*]+ K[I mod l])mod *n*
9) swap S[*i*] and S[*j*]
10) end for

As a result of simulation test there is no linear relation between KSA[i] and old IV[i] as shown in **Figures 4, 5** and **6**. Which means the first byte attack cannot succeed.

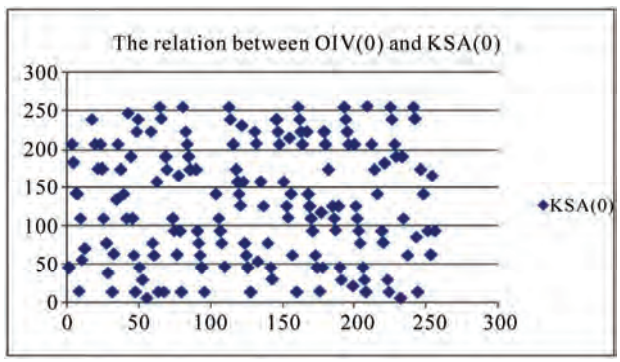


Figure 4. shows the relation between old IV[0] and KSA[0].

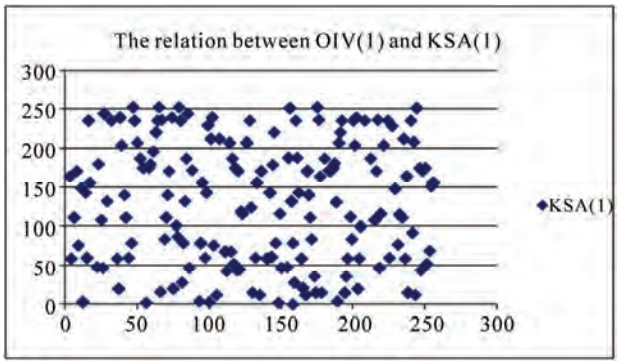


Figure 5. Shows the relation between old IV[1] and KSA[1].

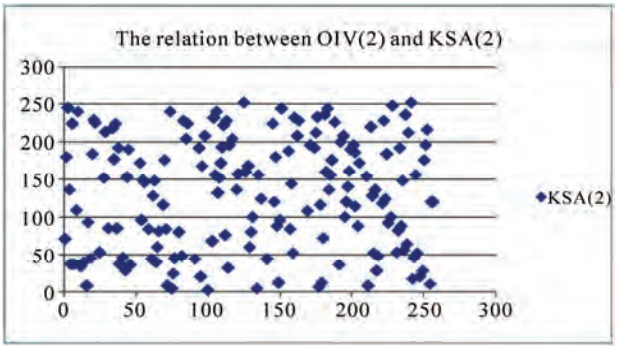


Figure 6. Shows the relation between old IV[2] and KSA[2].

1) An encryptor and Decryptor that share a RC4 secret key (K) agree session key or day key depending on the security level that they need. Using Equation (1).
Where is K' represent session key or day key.

2) The secret key K uses to confuse IV to generate IV' which will combine with K to generate RC4 key seed as follows:

For i=o; IV length;

$$\{ IV' = ((IV * k + IV^{k[i \bmod 11]i}) \bmod 256 \} \qquad (2)$$

From the results of Equation (2) there is no linear relation between new IV[i] and old IV[i] as shown in **Figures 7, 8** and **9**. This means that we can send the old IV as plaintext. The attacker can not use it to decrypt the ciphertext.

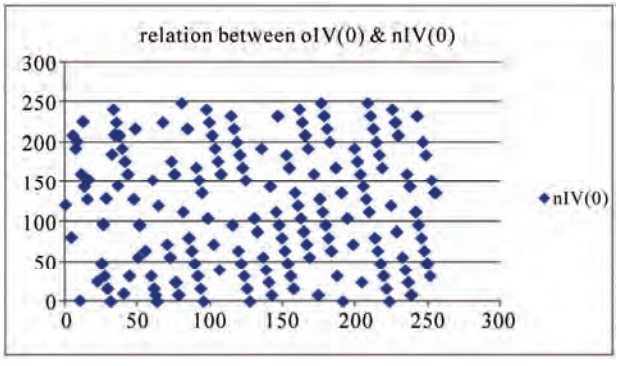


Figure 7. shows the relation between oIV[0] and nIV[0].

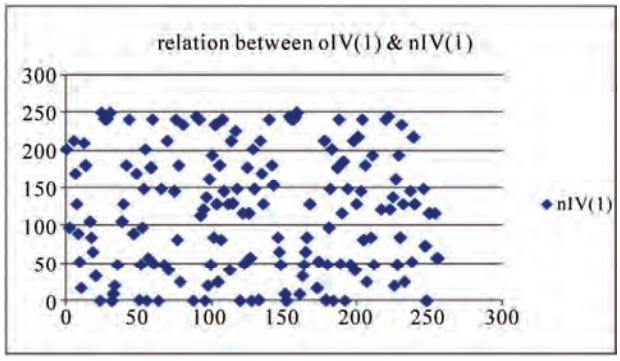


Figure 8. Shows the relation between oIV[1] and nIV[1].

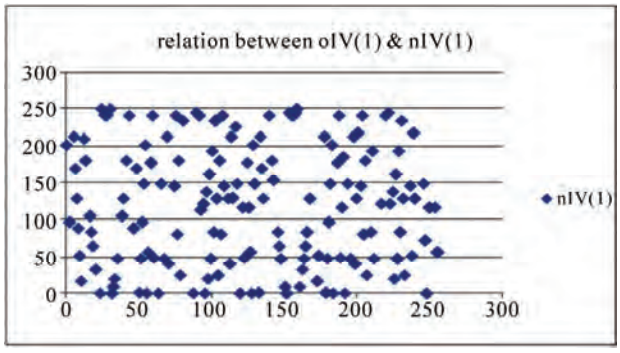


Figure 9. Shows the relation between oIV[2] and nIV[2].

3) Combine the secret K and IV', with K' they will be the seed for RC4 algorithm.

4) Send cipher text and IV.

2.2. Comparison the IV Value Rollover

For a device sending 10000 packets per-second

1) 24-bits, an IV will be reused after 16777216 packets, after 27.9 min. [WEP static secret Key]

2) 48-bit, an IV will be reused after 281474976710656 packets, after 892.6 years. [WPA static secret Key]

3) 24-bits, an IV will be reused after 16777216*39 *13*16 =136096776192 packets, after 5.2 month [Dynamic Session Key].

2.3. Comparison the Bandwidth Overhead

For a device sending 10000 packets per-second:

1) WEP overhead 44byte per-packet, means every second 4.4 Kbytes.

2) WPA overhead 56byte per-packet, means every second 5.6kbyte.

3) Our approach overhead 44byte per-packet means, 4.4 Kbytes.

3. References

[1] "WEP fix using RC4 fast packet keying," February 2002, http://www.rsasecurity.com/rsalabs/technotes/wep-fix.html

[2] S. Jariwala, “Enhancing wireless security with WPA,” April 2004.

[3] A. Jain and S. Karan, “Wireless LAN security”.

[4] M. S. Ahmad and V. Ramachandran, “Cafe latte with a free topping of cracked WEP—retrieving WEP keys from road warriors”.

[5] M. O. Pervaiz, M. Cardei, and J. Wu, “Security in wireless local area networks,” Department of Computer Science & Engine.

[6] D. Kalina, “WAP, WPA, and EAP,” March 2005, http://islab.oregonstate.edu/koc/ece478/05Report/Kalina.doc

[7] T. Takahashi, “WPA passive dictionary attack overview,” http://www.personalwireless.org/tools/WPA-Cracker/WPA_Passive Dictionary_Attack_Overview.pdf

[8] V. Moen, H. Raddum, and K. J. Hole, “Weaknesses in the Temporal Key Hash of WPA,” http://bora.uib.no/bitstream/1956/1901/21/Paper4 Moen.pdf

[9] H. R. Hassan and Y. Challal, “Enhanced WEP: An efficient solution to WEP threats”.

[10] W. Stallings, “Cryptography and network security,” 3rd edition, 1999.

[11] J.-C. Lin, Y.-H. Kao, and C.-W. Yang “Secure enhanced wireless transfer protocol,” First International Conference on Availability, Reliability and Security, 2006.

[12] A. Klein, “Attacks on the RC4 stream cipher,” 27 February 2007.

Wireless Sensor Network, **2010, 2, 243-249**
doi:10.4236/wsn.2010.23033 Published Online March 2010 (http://www.scirp.org/journal/wsn)

Scientific Research

QPSK DS-CDMA System over Rayleigh Channel with a Randomly-Varying Frequency Narrow-Band Interference: Frequency Tracking Analysis

Aloys N. Mvuma
School of Informatics, University of Dodoma, Dodoma, Tanzania
E-mail: {*anmvuma, mvuma*}*@udom.ac.tz*
Received August 25, 2009; *revised September* 7, 2009; *accepted September* 21, 2009

Abstract

This paper analyses frequency tracking characteristics of a complex-coefficient adaptive infinite-impulse response (IIR) notch filter used for suppression of narrow-band interference (NBI) with a randomly-varying frequency in a quadriphase shift keying (QPSK) modulated direct-sequence code-division multiple-access (DS-CDMA) communication system. The QPSK DS-CDMA signals are transmitted over a frequency non-selective Rayleigh fading channel. The analysis is based on a first-order real-coefficient difference equation with respect to steady-state instantaneous frequency tracking error from which a closed-form expression that relates frequency tracking mean square error (MSE) with number of DS-CDMA active users and NBI power is obtained. Closed-form expressions for optimum notch bandwidth coefficient and step size constant that minimize the frequency tracking MSE are also derived. Computer simulations are included to substantiate the accuracy of the analyses.

Keywords: Code-Division Multiple Access (CDMA), Quadriphase Shift Keying (QPSK), Complex Adaptive IIR Notch Filter, Narrow-Band Interference (NBI), Frequency Tracking MSE

1. Introduction

Direct-sequence code-division multiple-access (DS-CDMA) is a preferred multiplexing technique in cellular telecommunications services as it exhibits desired features that are not inherently found in other multiple access techniques, *i.e.*, time-division multiple-access (TDMA) and frequency-division multiple-access (FDMA). Some of superior DS-CDMA features include robustness in multipath fading environment, flexibility in allocation of channels and increased spectral efficiency due to its capability of sharing bandwidth with narrow-band communication systems [1–3]. The bandwidth sharing capability is made possible by the inherent narrow-band interference (NBI) suppression capacity of DS-CDMA due to the processing gain of spread spectrum systems. However, for high levels of NBI power, the NBI suppression capacity of DS-CDMA system can be enhanced by means of signal processing techniques at the receiver. Several methods with varying complexities have been proposed for suppression of NBI in DS-CDMA communication systems [4–9].

In [10], a complex coefficient adaptive notch filter implemented as a constrained infinite-impulse response (IIR) filter with a complex Gauss-Newton adaptation algorithm was proposed. Its application in the suppression of NBI in quadriphase shift keying (QPSK) direct-sequence spread-spectrum (DSSS) communication system was shown to result in a better signal-to-noise ratio (SNR) improvement factor than that achieved by finite-impulse response (FIR) adaptive prediction filter. In [11] a complex coefficient adaptive IIR notch filter with a simplified gradient-based algorithm that does not require any matrix inversion was presented. Analyses of its convergence, steady-state and tracking characteristics were presented in [12–15], respectively. Its application in suppression of NBI in QPSK-DSSS system with fixed unknown frequency was introduced in [13].

This paper investigates frequency tracking characteristics of the complex-coefficient adaptive IIR notch filter in [11] that is used for suppression of NBI with randomly-varying frequency in a synchronous QPSK DS-CDMA communication system communicating over a frequency non-selective Rayleigh fading channel. The analysis is based on the derived first-order real-coeff-

icient difference equation with respect to steady-state instantaneous frequency tracking error from which a closed-form expression for frequency tracking mean square error (MSE) is obtained. In addition, closed-form expressions for optimum notch bandwidth coefficient and step size constant are also derived. Computer simulations are included to substantiate accuracy of the analyses.

This paper is organized as follows. Section 2 presents the system model. Complex coefficient adaptive IIR notch filter is presented in Section 3 whereas frequency tracking is analyzed in Section 4. In Section 5 simulations are presented and discussed before a conclusion in Section 6.

2. System Model

We consider a synchronous DS-CDMA system over frequency-nonselective Rayleigh fading channel using QPSK modulation, quaternary pseudo-noise (PN) spreading m-sequences and the rectangular chip waveform. It is assumed that there are K simultaneously transmitting users. Referring to **Figure 1**, the transmitted signal for the *i*-th user $s_i(t)\big|_{i=1}^{K}$ is expressed as

$$\begin{aligned} s_i(t) &= \sqrt{P} b_i^I(t) c_i^I(t) \cos(\omega_c t + \theta) \\ &\quad + \sqrt{P} b_i^Q(t) c_i^Q(t) \sin(\omega_c t + \theta) \end{aligned} \tag{1}$$

where:

P is the power of the transmitted signal.

$b_i^I(t)$ and $b_i^Q(t)$ are in-phase and quadrature-phase binary data signals, respectively.

$c_i^I(t)$ and $c_i^Q(t)$ are in-phase and quadrature-phase PN spectrum spreading signals, respectively.

ω_c and θ are the modulator carrier frequency and phase, respectively.

We define the rectangular pulse $p_T(t)$ with duration T as

$$p_T(t) = \begin{cases} 1, & 0 \le t < T \\ 0, & \text{elsewhere} \end{cases} \tag{2}$$

Therefore the in-phase and quadrature-phase data and spectrum spreading signals are expressed as

$$b_i^I(t) = \sum_{j=-\infty}^{\infty} b_{i,j}^I p_{T_s}(t - jT_s), \tag{3}$$

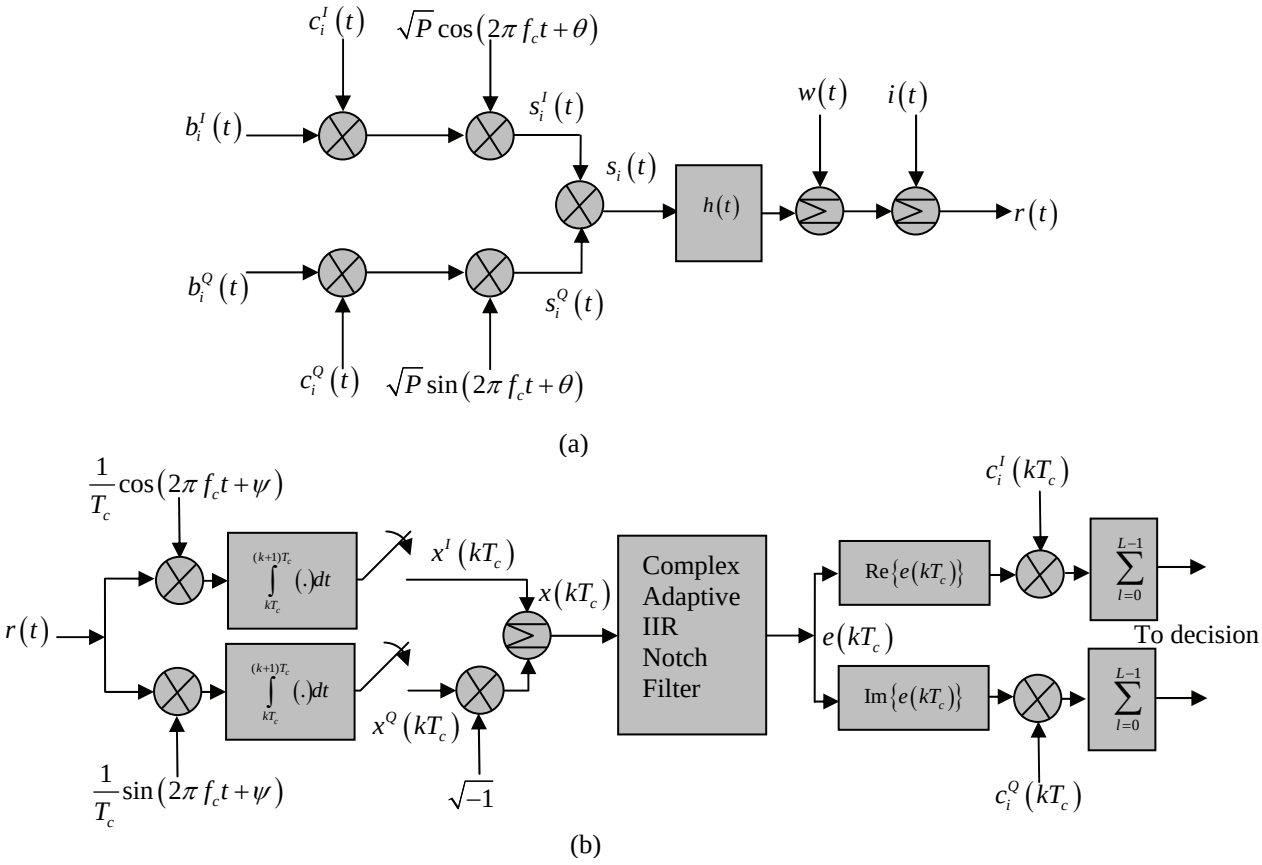


Figure 1. QPSK DS-CDMA communications system model. (a) Transmitter; (b) Receiver.

$$b_i^Q(t)=\sum_{j=-\infty}^{\infty} b_{i,j}^Q p_{T_s}(t-jT_s) \quad (4)$$

$$c_i^I(t)=\sum_{j=-\infty}^{\infty} c_{i,j}^I p_{T_c}(t-jT_c), \quad (5)$$

$$c_i^Q(t)=\sum_{j=-\infty}^{\infty} c_{i,j}^Q p_{T_c}(t-jT_c) \quad (6)$$

where

$b_{i,j}^I \in \{+1,-1\}$ and $b_{i,j}^Q \in \{+1,-1\}$ are identically and independently distributed (IID) random *j*-th data bits of the *i*-th user for the in-phase and quadrature-phase components.

$c_{i,j}^I \in \{+1,-1\}$ and $c_{i,j}^Q \in \{+1,-1\}$ are IID random *j*-th chips for the *i*-th user for the in-phase and quadrature-phase components, respectively.

T_c and T_s are the chip duration and the symbol duration, respectively, where $T_s/T_c = L$ is the number of chips per symbol or processing gain.

The QPSK DS-CDMA signal comprising of signals for all *K* active users is transmitted over a frequency non-selective Rayleigh fading channel with impulse response given by

$$h(t)=\alpha \exp(j\vartheta)\delta(t-\tau) \quad (7)$$

where ϑ is the phase shift with uniform PDF over $[0,\ 2\pi)$, τ is the time delay which is uniformly distributed over $[0,\ T_s)$ and α is the Rayleigh distributed attenuation having a probability density function (PDF) expressed as

$$f_A(\alpha)=\begin{cases}\alpha \exp\left(-\dfrac{\alpha^2}{2}\right), & \alpha \geq 0,\\ 0, & \alpha < 0.\end{cases} \quad (8)$$

The transmitted signal is corrupted with a zero-mean additive white Gaussian noise (AWGN) $w(t)$ with two-sided power spectral density (PSD) $N_0/2$ and a NBI modeled as

$$I(t)=\sqrt{2J}\cos(\omega_c t+\phi(t)) \quad (9)$$

where J is the power of the interference and $\phi(t)$ is the instantaneous phase deviation.

The received signal $r(t)$ at the input of the correlator bank in **Figure 1** is expressed as

$$r(t)=\alpha\sum_{i=1}^{K} s_i(t-\tau)+w(t)+I(t) \quad (10)$$

At time kT_c the samples $x^I(kT_c)$ and $x^Q(kT_c)$ in **Figure 1** can be written as (for simplicity, T_c is normalized to unity)

$$x^I(k)=s^I(k)+n^I(k)+\zeta^I(k) \quad (11)$$

$$x^Q(k)=s^Q(k)+n^Q(k)+\zeta^Q(k) \quad (12)$$

$$s^I(k)=\alpha\sqrt{\frac{P}{4}}\sum_{i=1}^{K} b_{i,k}^I c_{i,k}^I \quad (13)$$

$$s^Q(k)=\alpha\sqrt{\frac{P}{4}}\sum_{i=1}^{K} b_{i,k}^Q c_{i,k}^Q \quad (14)$$

where

$b_{i,k}^I \in \{+1,-1\}$ and $b_{i,k}^Q \in \{+1,-1\}$ are the values of data signals at the *k*-th sampling instant of the *i*-th user for the in-phase and quadrature-phase components.

$c_{i,k}^I \in \{+1,-1\}$ and $c_{i,k}^Q \in \{+1,-1\}$ are the values of the spectrum spreading signals at the *k*-th sampling instant for the *i*-th user for the in-phase and quadrature-phase components, respectively.

$n^I(k)$ and $n^Q(k)$ are independent and uncorrelated random processes with zero mean and variance $\sigma^2=\dfrac{N_0}{4T_c}$.

Assuming $\phi(t)$ to be varying slowly such that it is constant over one chip interval, then NBI components $\zeta^I(k)$ and $\zeta^Q(k)$ in (11) and (12) are expressed as

$$\zeta^I(k)=\sqrt{\frac{J}{2}}\cos\phi(k) \quad (15)$$

$$\zeta^Q(k)=\sqrt{\frac{J}{2}}\sin\phi(k) \quad (16)$$

It can easily be shown that $s^I(k)$ and $s^Q(k)$ are zero-mean uncorrelated random processes each with variance $\sigma_s^2=E[\alpha^2]\dfrac{P}{4}K$.

3. Complex Coefficient Adaptive IIR Notch Filter

Using complex notation, a complex input signal to a complex coefficient adaptive IIR notch filter in **Figure 1** is of the form

$$x(k)=s(k)+\zeta(k)+n(k), \quad (17)$$

where

$$s(k)=s^I(k)+js^Q(k) \quad (18)$$

$$\zeta(k)=\sqrt{\frac{J}{2}}\exp(j\phi(k)) \quad (19)$$

$$\phi(k)=\phi(k-1)+\omega(k), \quad (20)$$

$$n(k) = n^I(k) + jn^Q(k). \quad (21)$$

$\omega(k)$ is the instantaneous frequency that follows a *random walk model* and is expressed as

$$\omega(k) = \omega(k-1) + \beta\nu(k),\ \omega(0) = \omega_0 \quad (22)$$

where $\nu(k)$ is a zero-mean white noise with variance σ_v^2, β is the scaling factor for the frequency drift and is assumed to be small, *i.e.*, $\beta \ll 1$ and ω_0 is the initial frequency.

Transfer function of first-order complex coefficient IIR notch filter for suppression of the NBI in (17) is given by [11].

$$H(z) = \frac{1+\alpha_0}{2}\frac{1-e^{j\alpha_1(k)}z^{-1}}{1-\alpha_0 e^{j\alpha_1(k)}z^{-1}} \quad (23)$$

where α_0 is the notch bandwidth coefficient and α_1 is the notch frequency coefficient. Adaptation algorithm used here to estimate the instantaneous frequency of the NBI is expressed as [11].

$$\alpha_1(k+1) = \alpha_1(k) + \mu\,\mathrm{Re}\left[e(k)\varphi^*(k)\right]. \quad (24)$$

Re[.] denotes real number and * denotes complex conjugate. Here μ is the step size constant, $e(k)$ is the complex notch filter output and $\varphi(k)$ is the gradient signal. Referring to (23), the instantaneous frequency can be estimated by $\hat{\omega}(k)$ where

$$\hat{\omega}(k) = \alpha_1(k). \quad (25)$$

Transfer function from the input $x(k)$ to the gradient signal $\varphi(k)$ in (24) is given by [11].

$$G(z) = \frac{1+\alpha_0}{2}\frac{je^{j\alpha_1(k)}z^{-1}}{1-\alpha_0 e^{j\alpha_1(k)}z^{-1}}. \quad (26)$$

4. Frequency Tracking Error Analysis

In this section, frequency tracking error of the algorithm in (24) is analyzed. We define steady-state tracking error of instantaneous frequency $\delta(k)$ as [14], [15].

$$\delta(k) = \hat{\omega}(k) - \omega(k). \quad (27)$$

Referring to the coefficient adaptation algorithm in (24), it follows from (27) that [15]:

$$\delta(k+1) = \delta(k) - \beta\nu(k) + \mu\,\mathrm{Re}[e(k)\varphi^*(k)], \quad (28)$$

A first-order difference equation with respect to $\delta(k)$ is obtained by using approximations for steady-state signals $e(k)$ and $\varphi(k)$ derived in [14], and is expressed by

$$\delta(k+1) = \left(1 - \frac{J}{2}\frac{\mu}{4\gamma^2}\right)\delta(k) + \mu\xi(k),\ \gamma = \frac{1-\alpha_0}{1+\alpha_0} \quad (29)$$

where the input to the difference equation $\xi(k)$ is given by

$$\xi(k) = \sqrt{\frac{J}{2}}\frac{1}{4\gamma}\left\{n_e(k)e^{-j(\phi(k)+\frac{\pi}{2})} + n_e^*(k)e^{j(\phi(k)+\frac{\pi}{2})}\right\} + \frac{1}{2}\left\{n_e(k)n_\varphi^*(k) + n_e^*(k)n_\varphi(k)\right\} - \frac{\beta\nu(k)}{\mu} \quad (30)$$

Here $n_e(k)$ and $n_\varphi(k)$ are the output of $H(z)$ and $G(z)$, respectively, due to $s(k)+n(k)$. By referring to (29), a first-order real coefficient transfer function $F(z)$ between the input $\xi(k)$ and the output $\delta(k)$ is given by

$$F(z) = \frac{\mu z^{-1}}{1-\left(1-\frac{J}{2}\frac{\mu}{4\gamma^2}\right)z^{-1}}. \quad (31)$$

4.1. Square of Tracking Error Due to Frequency Drift

By referring to (29), (30) and (31), square of tracking error due to the frequency drift is found to be given by

$$\sigma_1^2 = \sigma_v^2\left(\frac{\beta}{\mu}\right)^2\frac{1}{2\pi j}\oint F(z)F(z^{-1})z^{-1}dz = \frac{2}{J}\frac{2\sigma_v^2\beta^2\gamma^2}{\mu}. \quad (32)$$

(32) is obtained by assuming that $\frac{J}{2}\mu \ll 4\gamma^2$ which means the step-size parameter is sufficiently small.

4.2. Frequency Tracking Mean Square Error

Frequency tracking mean square error (MSE) is the sum of σ_1^2 in (32) and frequency error variance σ_2^2 due to $n_e(k)$ and $n_\varphi(k)$. σ_2^2 is obtained by [12,14].

$$\sigma_2^2 \approx \frac{\frac{J}{2}\left(E\left[\alpha^2\right]\frac{P}{2}K + \frac{N_0}{2T_c}\right)\mu^2(1+\alpha_0)^3}{16(1-\alpha_0)^2(1-\alpha_0\varepsilon)(1+\varepsilon)}, \quad (33)$$

$$\varepsilon = 1 - \frac{J}{2}\frac{\mu}{4\gamma^2} \approx 1, \quad (34)$$

The following observations are made from (33) and (34):

1) Frequency tracking error variance is directly proportional to the NBI power J.

2) For small AWGN PSD N_0 frequency tracking error variance is directly proportional to the signal transmit power P and number of active users K.

3) Frequency tracking error variance can be reduced by expanding the notch bandwidth, *i.e.*, reducing the

notch bandwidth coefficient α_0.

4) Frequency tracking error variance varies proportionally with the square of step-size constant.

From (32), (33) and (34), the closed-form expression for the frequency tracking MSE can be expressed as

$$MSE = \frac{2}{J}\frac{2\sigma_v^2\beta^2\gamma^2}{\mu} + \frac{\frac{J}{2}\left(E\left[\alpha^2\right]\frac{P}{2}K + \frac{N_0}{2T_c}\right)\mu^2}{32\gamma^3}. \quad (35)$$

4.3. Optimum Step Size and Notch Bandwidth Coefficients

From (35), the optimum step-size μ_{opt} that corresponds to minimum frequency tracking MSE is derived by equating the first derivative of MSE in (35) with respect to μ to zero. This yield:

$$\mu_{opt} = 2\left(\frac{4\sigma_v^2\beta^2\gamma^5}{\left(\frac{J}{2}\right)^2\left(E\left[\alpha^2\right]\frac{P}{2}K + \frac{N_0}{2T_c}\right)}\right)^{\frac{1}{3}}. \quad (36)$$

Similarly, equating to zero the first derivative of MSE in (35) with respect to γ we obtain:

$$\alpha_{0opt} = \frac{1-\gamma_{opt}}{1+\gamma_{opt}}. \quad (37)$$

where:

$$\gamma_{opt} = \frac{1}{2}\left(\frac{3\left(\frac{J}{2}\right)^2\left(E\left[\alpha^2\right]\frac{P}{2}K + \frac{N_0}{2T_c}\right)\mu^3}{4\sigma_v^2\beta^2}\right)^{\frac{1}{5}} \quad (38)$$

5. Simulation and Discussions

In this section, computer simulation results are compared with analytical values obtained in (35), (36), (37) and (38) to substantiate the accuracy of the proposed analytical method. All simulated results were obtained by averaging over 50 independent computer runs for $k = 2500$ to 14000 with J = 2.0, and P = 0.1, *i.e.*, interference-to-signal power ratio (ISR) of 13.0 dB. The values for β, σ_v^2, and ω_0 were set to 10^{-3}, 0.2, and 0.2π, respectively, and L=255. Here the AWGN PSD N_0 is ignored as it is assumed to be much smaller than the NBI power J.

In **Figure 2**, simulated results for the frequency tracking MSE are plotted and compared with theoretical values obtained from (29) with the DS-CDMA system number of active users K as a parameter. Here the step size constant μ was set to 10^{-4}. For small values of α_0, the figure shows a decrease in the frequency tracking MSE which increases after α_0 exceeds the optimal notch bandwidth coefficient α_{opt} whose value depends on K. Close agreement between analytical and simulation results is clearly demonstrated by the figure.

Figure 3 shows plots of theoretical values of the frequency tracking MSE obtained from (35) and simulated results with the number of active DS CDMA users K as a parameter for $\alpha_0 = 0.90$. It is observed from the figure that the frequency tracking MSE decreases for values of μ less than the optimal step size μ_{opt} which depends on K. For values of μ above μ_{opt}, the frequency tracking MSE increases steadily. Similarly, close agree-

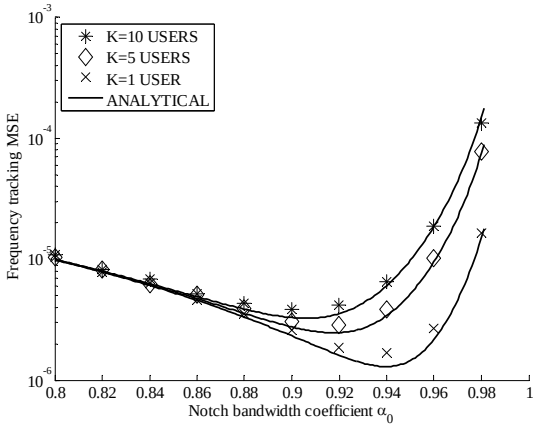


Figure 2. Frequency tracking MSE with number of DS-CDMA users *K* as a parameter.

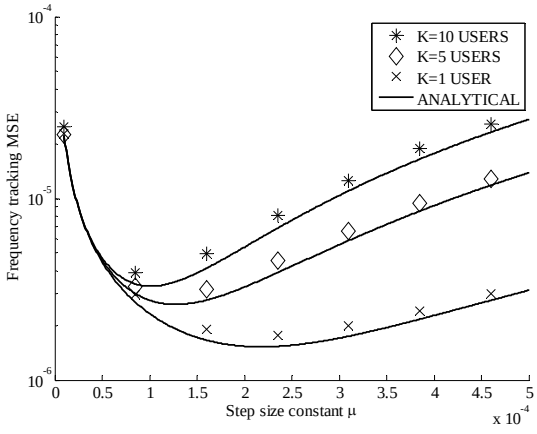


Figure 3. Frequency tracking MSE with number of DS-CDMA active users *K* as a parameter.

ment between analytical values and simulation results is clearly shown by the figure.

Figure 4 shows simulated results and theoretical values for the optimum notch bandwidth coefficient $\alpha_{0_{opt}}$ plotted against the number of DS-CDMA active users *K* for $\mu = 10^{-4}$. Similarly, the figure shows a decrease in $\alpha_{0_{opt}}$ with the increase in *K* as predicted by (37) and (38). There is a close agreement between simulated results and theoretical values as validated by the figure.

Simulated results and theoretical values obtained from (36) for the optimum step-size constant μ_{opt} versus number of active DS-CDMA users *K* are plotted in **Figure 5** with $\alpha_0 = 0.9$. The figure shows a decrease in μ_{opt} with the increase in *K* as anticipated by (30). Close agreement between simulated results and theoretical values is clearly demonstrated by the figure.

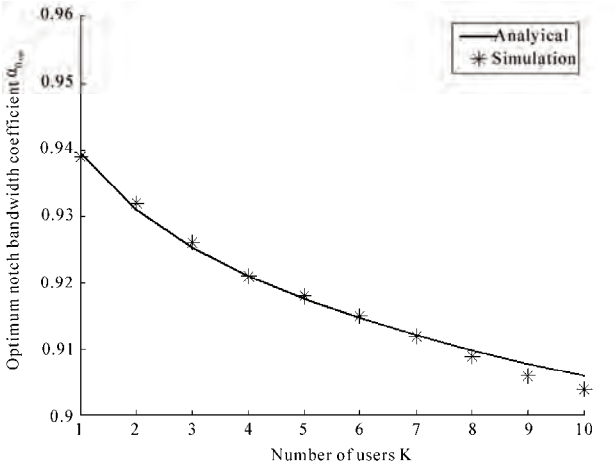


Figure 4. Optimum step-size constant versus number of active users *K*.

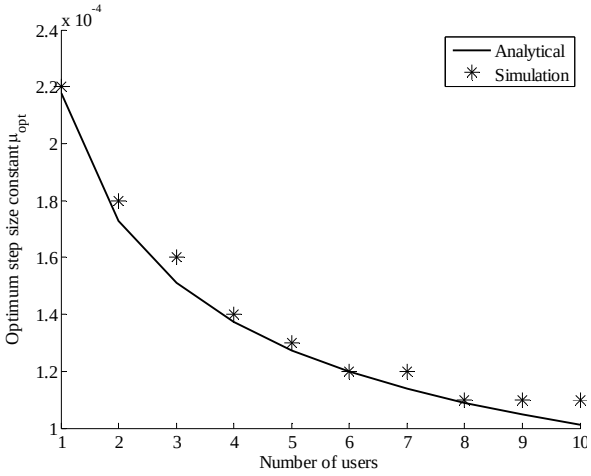


Figure 5. Optimum notch bandwidth coefficient versus the number of DS-CDMA active users *K*.

6. Conclusions

Frequency tracking characteristics of the complex-coefficient adaptive IIR notch filter for suppression of NBI with randomly-varying frequency in a DS-CDMA communication system over a Rayleigh fading channel were investigated in this paper. Derived closed-form expressions for frequency tracking MSE and optimum step size and notch bandwidth coefficient have revealed a need for proper setting of adaptation algorithm and IIR notch filter parameters to minimize frequency tracking MSE. Moreover, computer simulation results have demonstrated the accuracy of the analytical approach. In the future, probability of bit error of the DS-CDMA system with NBI suppression complex adaptive IIR notch filter will be investigated.

7. References

[1] W. C. Y. Lee, "Overview of cellular CDMA," IEEE Transactions on Vehicular Technology, Vol. 40, pp. 291–301, May 1991.

[2] R. L. Pickholtz, D. L. Schilling, and L. B. Milstein, "Theory of spread spectrum communications: A tutorial," IEEE Transactions on Communications, Vol. COMM-30, pp. 855–884, May 1982.

[3] L. B. Milstein, *et al.*, "On the feasibility of a CDMA overlay for personal communications networks," IEEE Journal on Selected Areas in Communications, Vol. 10, pp. 655–668, May 1992.

[4] H. V. Poor and L. A. Rusch, "Narrowband interference suppression in spread spectrum CDMA," IEEE Communications Magazine, Vol. 1, pp. 14–27, August 1994.

[5] J. Wang and L. N. Miltein, "Adaptive LMS filters for cellular CDMA overlay situations," IEEE Journal on Selected Areas in Communications, Vol. 14, pp. 1548–1559, October 1996.

[6] M. Lops, G. Ricci, and A. M. Tulino, "Narrowband interference suppression in multiuser CDMA systems," IEEE Transactions on Communications, Vol. 46, pp. 1163–1175, September 1998.

[7] S. Buzzi, M. Lops, and A. M. Tulino, "Time-varying narrow-band interference rejection in asynchronous multiuser DS/CDMA systems over frequency-selective fading channels," IEEE Transactions on Communications, Vol. 47, pp. 1523–1535, October 1999.

[8] W. S. Hou, L. M. Chen and B. S. Chen, "Adaptive narrowband interference rejection in DS-CDMA systems: A scheme of parallel interference cancellers," IEEE Journal on Selected Areas in Communications, Vol. 19, pp. 1103–1114, June 2001.

[9] V. Krishnamurthy, G. Yin, and S. Singh, "Adaptive step-size algorithms for blind interference suppression in DS/CDMA systems," IEEE Transactions on Signal Processing, Vol. 49, pp. 190–201, January 2001.

[10] S. C. Pei and C. C. Tseng, "Complex adaptive IIR notch filter algorithm and its applications," IEEE Transactions on Circuits and Systems, Vol. CAS-41, No. 2, pp. 158–163, February 1994.

[11] S. Nishimura and H. Y. Jiang, "Simplified realization of cascaded adaptive notch filters with complex coefficients," Proceedings of International Symposium on Circuits and Systems, Vol. 5, pp. 269–272, 1998.

[12] H. Y. Jiang, S. Nishimura, and T. Hinamoto, "Convergence analysis of complex adaptive IIR notch filters for the detection of single sinusoid," IEICE Transactions Fundamentals, Vol. E82-A, No. 6, pp. 912–915, June 1999.

[13] H. Y. Jiang, S. Nishimura, and T. Hinamoto, "Steady-state analysis of complex adaptive IIR notch filter and its application to QPSK communication systems," IEICE Transactions Fundamentals, Vol. E85-A, No. 5, pp. 1088–1085, May 2002.

[14] A. Mvuma, S. Nishimura, and T. Hinamoto, "Tracking analysis of complex adaptive IIR notch filter for a linear chirp signal," IEICE Transactions Fundamentals, Vol. E92-A, pp. 1526–1529, June 2009.

[15] A. Mvuma, S. Nishimura, and T. Hinamoto, "Complex coefficient adaptive IIR notch filter tracking characteristics," Proceedings of Midwest Symposium on Circuits and Systems, pp. 640–643, August 2009.

Wireless Sensor Network, **2010, 2, 250-253**
doi:10.4236/wsn.2010.23034 Published Online March 2010 (http://www.scirp.org/journal/wsn)

Study on the Wireless Heat Meters

Rujian Ma, Xueyan Yang, Xiao Wang, Enping Zhang
School of Control Science and Engineering, University of Jinan, Jinan, China
E-mail: *rjma@ujn.edu.cn*
Received August 16, 2009; *revised September* 21, 2009; *accepted October* 19, 2009

Abstract

The design of a new type heat meters based on RFID technology are presented in this paper. By use of RFID technology in heat meters, the data can be exchanged between heat meters and heat supplying department by RF cards. The information can be transmitted in a non-contact way. In this way, the purpose of automatic identification can be achieved. The experimental study of the heat meters is also performed in the paper. The results show that the new type of heat meters can meet the demands of users. Compared with the ordinary heat meters, the new type of heat meters have the advantages of small in meter volume, high accuracy, no impact of water quality and good reliability.

Keywords: Heat Meters, Design, Experimental Study, RFID

1. Introduction

Radio frequency identification (RFID) technology is a non-contact automatic identification technology [1]. Based on the wireless data communication technology, the RFID technology has been widely used for wireless remote meter-reading technology and provides a sufficient technology foundation. At the present, the central heating supply in the urban area of China, waste of energy and unreasonable charges by the heating area, these problems have restricted the further development of central heating. In the implementation of structural reform in the urban heating, heat meters are the key equipment in heating metering and charging control system, which can achieve the metering of urban central heating household and energy charges in accordance with the implementation of heat, and this has gradually become a new hot industry. To make the study of intelligent heat meters based on RFID technology, we consulted the International Organization of Legal Metrology OIML-R75 rules and the European standards EN1434 heat meters which are most advanced in the world, also we draw lessons from European advanced technology and experience of similar products, these heat meters can realize not only for the centralized hot household heat metering, but also the non-contact measurement of data transmission and the control of clearing transactions. As a result, the heat meters provide the possibility of intelligent management in cities and a firm material foundation for the digital communities and cities, which can bring high social and economic benefits.

Heat meters are the terminal equipments of the heat metering management system, as shown in **Figure 1**. They exchange data with the department of heat metering by radio frequency (RF) cards. As an important aspect of RFID technology, the RF cards achieve non-contact transmission of information by using radio frequency signals through space coupling. Furthermore, they can achieve the purpose of automatic identification through the transmission of information.

In heat metering management system, each user gets a new RF card when installed new equipment. Before using heat meters, users must open accounts in the department and the management system establishes detailed information for each user. After opening accounts, users buy heat according to their actual needs of the heat required and bring about closer ties between RF card and their own heat meters. From the information security point of view, in the card all user information, heat data calorie and other management information is transmitted

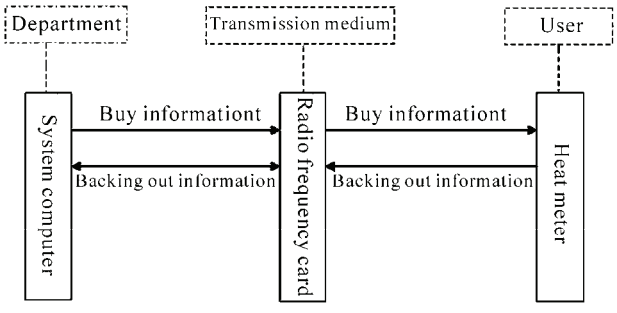


Figure 1. Diagram of the heat supplying management system.

in the form of encryption, heat meters do the corresponding decryption work when they accept. At the same time, heat meters transmit the user's use of heat, the remaining heat value, the corresponding record in history and heat meters state information to heat supplying department by RF card. Heat supplying department analyses and compares the historical data they collected, and obtains a large number of heating data, these data provide strong data support for the department to make decisions.

2. Design of Heat Meters

Heat meters consist of two major parts, the RF card and smart heat metes.

The Mifare One of Philips chips (referred to M1 card) is a kind of commonly used RF card. The parameters of the card are: 8K in storage capacity, 10 years of data retention, 100,000 cycles for rewriting, unlimited time for reading, and no battery needed. The card has its own antenna, encryption control logic circuit and communication logic circuit, inter-national common DES and cross-RES secrecy algorithm is adopted in the communication between the card and reader, its confidentiality performance is very high, with characteristics of fast anti-collision, high reliability in data communication, anti-strong electrical interference, moisture-proof, waterproof, convenience, fast and so on [2].

The smart heat metes contains a microprocessor, a mechanical rotor-type non-magnetic flow sensor, a pair of temperature sensor, a voltage monitor and radio frequency modules, and so on. The overall structure of the heat meters is shown in **Figure 2**.

The microprocessor uses singlechip of MSP430FW425 type for the wireless long-distance transmission calorimeter. Its internal composite signal processor is key microprocessor of integrating monitor, superior performance, reliable and cost-effective [3]. The current sensor is mechanical rotor type multi-striation flowmeter which is improved to comply with the request of non-magnetic current signal acquisition calibration module and measurement characteristic. It is characterized by a small wheel rotor installed. Signal damping board is light in weight, wide flow measurement, and high accuracy of flow measurement and applicable to poor quality water etc. [4]. A pair of temperature sensors are used for measuring temperature and signal of difference in temperature about heat exchanger inlet and outlet, the precision of the sensors less than 0.1℃ [5]. Integrating monitor is hardcore. On one hand, it collects and calibrates the flow rate signal, temperature and signal of difference in temperature integration demonstration quantity of heat of non-magnetic flowmeter. On the other hand, it uses RF transport module to transports data.

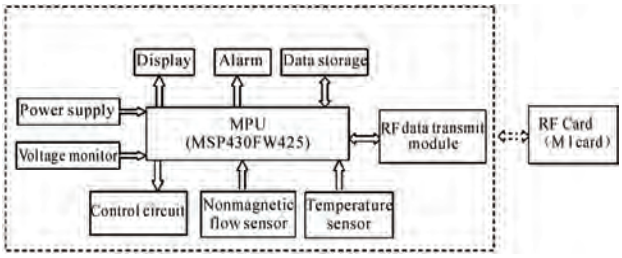


Figure 2. Overall structure diagram of heat meters.

Control software of heat meters are designed based on the modularizaion method, including radio frequency read-write module, non-magnetic flowmeter scan module, flow rate module and so on. Each part functionally independent designed, effectively to guarantee the reliability of the scale of intelligent hot work. The table with applicability and flexibility and humanized design concept are full considered in the design, heat meters based on RFID technology has superior intelligence and fully meets the market demand. Flow chart of the main program and the card interrupt program are shown in **Figures 3** and **4**.

3. Experimental Study

Three heat meters of DN20 type are used for the prototype test. The test includes accelerated wear test, metrological characteristic test and data communication test.

The accelerated wear test requires that the meter is continuously running 300 hours under the maximum flow rates. The maximum flow rate of this meter is 5000 kg/h. All the three meters are in good working situation after the test.

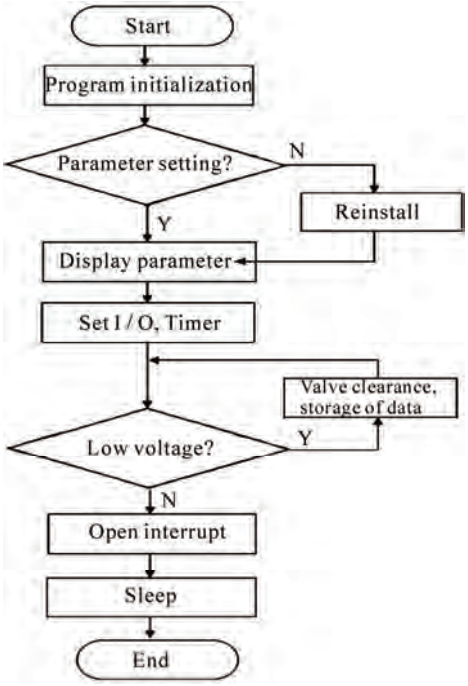


Figure 3. Flow chart of the software program.

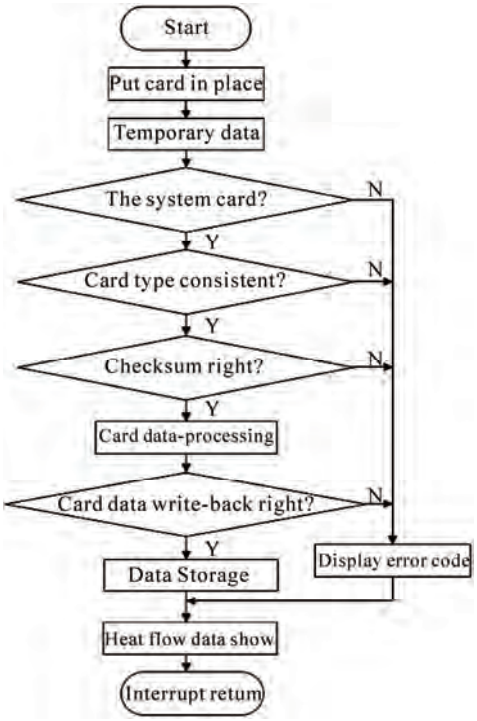


Figure 4. Flow chart of the card interrupt program.

Metrological characteristic test includes temperature test and flow rate test. The measuring points for the temperature test are 10, 50 and 90 degrees Centigrade, representing the low, medium and high temperatures respectively. The measurement results of temperatures are listed in **Table 1**. It can be seen from the table that all the data is less than the permit error of level 2 meters (3.5%). The measurement of flow rates is performed in three temperature points representing the low, medium and high temperatures. The results are shown in **Figures 5, 6** and **7**. It can be seen from the figures that all the data is less than the permit error of level 2 meters.

The test items of data communication test include Distance of data reading & writing, resetting data, data summation, data transfer and replacement of cards and meters. The measurement results of the data communication test are listed in **Table 2**. Compared with the performance required of the meters (also listed in the table) indicates that all the performance indexes are satisfied with the requirement of the meters.

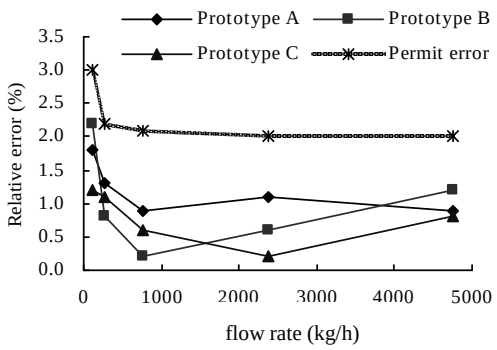


Figure 5. Measurement errors of flow rates at 10 degree Centigrade.

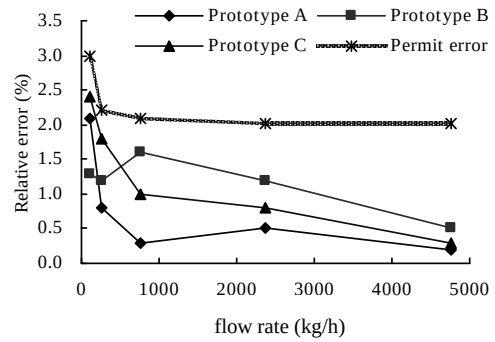


Figure 6. Measurement errors of flow rates at 50 degree Centigrade.

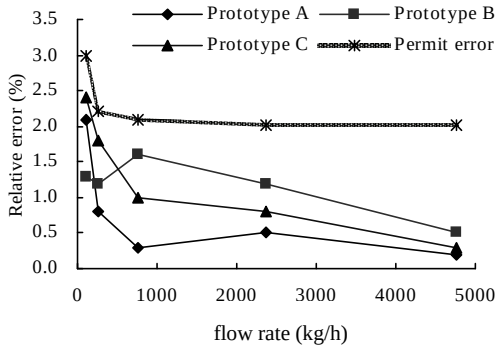


Figure 7. Measurement errors of flow rates at 90 degree Centigrade.

Table 1. Measurement results of temperatures.

Temperature Test points (℃)	Maximum errors (%)		
	Prototype A	Prototype B	Prototype C
10, 50 and 90	1.17	1.92	0.79

Table 2. Results of data communication tests.

Test items	Performance required	Prototype numbers		
		No. A	No. B	No. C
Distance of data reading & writing	≥2.0 cm	2.2 cm	2.4 cm	2.3 cm
Resetting data	The valve is closed after data resetting	Normal	Normal	Normal
Data summation	The valve will be open after data summation	Normal	Normal	Normal
Data transfer	Will be work normally after the data transfer	Normal	Normal	Normal
Replacement of cards and meters	Will be work normally after the replacement	Normal	Normal	Normal

4. Conclusions

A new type of heat meter is designed based on RFID technology in this paper. This type of heat meters can provide centralized heat metering, residence intelligence, management digitization and information system management with new plan. The experimental study is also presented in the paper. The results indicate that the heat meters have high accuracy in measurement, no impact of water quality, good reliability and low power consumption.

5. Acknowledgements

This project is supported by Research Foundation of University of Jinan (XKY0806) and Key Subject Research Foundation of Shandong Province.

6. References

[1] Y. Shen, "Application of RFID technology in management of college teaching," Instrumentation Analysis Monitoring, No. 1, 2008.

[2] X. Yi, "Application of Mifare one RF card in prepayment electric meter," Process Automation Instrumentation, Vol. 28, No. 8, 2007.

[3] S. Guo, "Design of RFID reader based on MCUMSP430," Journal of Jilin University, Vol. 28, No. 2, 2007.

[4] X. Yang, "Design of digital water meter based on MSP430FW425 MCU scan IF module," Computer Engineering and Applications, Vol. 42, No. 18,42, 2006.

[5] X. Wei, "MSP430 singlechip interface technology and system design examples," Beijing University of Aeronautics & Astronautics Press, 2002.

Wireless Sensor Network, **2010, 2, 254-263**
doi:10.4236/wsn.2010.23035 Published Online March 2010 (http://www.scirp.org/journal/wsn)

ACTIVE-A Real Time Commit Protocol

Udai Shanker, Nikhil Agarwal, Shalabh Kumar Tiwari, Praphull Goel, Praveen Srivastava
Department of Computer Science and Engineering, M. M. M Engineering College, Gorakhpur, India
E-mail: {*udaigkp, nikhilmmmec, shalabhmmmec, goelpraphul, praveen*047}*@gmail.com*
Received November 19, 2009; *revised November* 30, 2009; *accepted December* 15, 2009

Abstract

Many existing real time commit protocols try to improve system performance by allowing a committing cohort to lend its data to an executing cohort, thus reducing data inaccessibility. They block the borrower from sending WORKDONE/PREPARED message and restrict them from lending data so that transaction abort chain is limited to one. Thus, transaction execution time increases. This paper proposes a modified real time commit protocol for distributed real time database systems (DRTDBS), Allow Commit Dependent and in Time borrowers for Incredible Value added data lending without extended abort chain (ACTIVE), where borrower cohorts are categorized as commit and abort dependent. Further, the commit dependent borrowers can lend data to executing cohorts with still limiting the transaction abort chain to one only and reducing the data inaccessibility. Also, an incoming executing cohort having borrowing factor greater than one can only borrow the dirty data items from lender. This minimizes the fruitless borrowing by the cohort. The performance of ACTIVE is compared with PROMPT, 2SC and SWIFT protocols for both main memory resident and disk resident databases with and without communication delay. Simulation results show that the proposed protocol improves the system performance up to 4% as transaction miss percentage.

Keywords: Distributed Real Time Database System, Commit Protocol, Conflict Resolution, Dependency, Lender, Borrower

1. Introduction

Maintenance of transaction's ACID semantics are the well known complexities that real time distributed database have to contend with when operating on the distributed data. Several important factors contribute to the difficulty in meeting transaction deadlines in DRTDBS. Data conflicts are one of the most important factors amongst the transactions. Two kinds of conflicts between transactions [1,2] arise. One occurs between executing transactions, which is resolved by a concurrency control protocol to ensure distributed transaction serializability; the other occurs between executing-committing transactions, which is resolved by a commit protocol to ensure distributed transaction atomicity. Limited work is done in case of executing-committing conflicts.

Database systems are currently being used as backbone to thousands of applications, which have very high demands for availability and fast real-time responses. A large part of the workload consists of read, write-blind and updates transactions against DRTDBS which these applications generate. Business transactions using these applications in the absence of real time could lead to financial devastations and in worst case cause injuries or deaths. Examples include telecommunication systems, trading systems, online gaming, sensor networks etc. Typically, a sensor network consists of a number of sensors (both wired and wireless) which report on the status of some real-world conditions. The conditions include sound, motion, temperature, pressure and moisture, velocity etc. The sensors send their data to a central system that makes decisions based both on current and past inputs. To enable the networks to make better decisions, both the number of sensors and the frequency of updates should be increased. Thus, sensor networks must be able to tolerate an increasing load. For applications such as Chemical Plant Control, Multi Point Fuel Injection System (MPFI), Video Conferencing, Missile Guidance System etc., data is needed in real-time, and must be extremely reliable and available as any unavailability or extra delay could result in heavy loss. Many applications listed above using DRTDBS require distributed transaction to be executed at more than one site. To maintain consistency, a commit protocol ensures that either all the effects of the transaction persist or none of them persist. Failure of site or communication link and loss of messages do not hamper the transaction processing. Commit

protocols must ensure that little overheads are laid upon transactions during processing. Hence, the need of developing a new commit protocol for better performance of DRTDBS arises.

The two phase commit protocol (2PC) referred to as the Presumed Nothing 2PC protocol (PrN) is the most commonly used protocol in the study of DDBS [3–5]. It ensures that sufficient information is force-written on the stable storage to reach a consistent global decision about the transaction. A number of 2PC variants [6] commit protocols have been proposed and can be classified into following four groups [7].

1) Presumed Abort/Presumed Commit
2) One Phase
3) Group Commit
4) Pre Commit/Optimistic

Presumed commit (PC) and presumed abort (PA) [8] are based on 2PC. Soparkar *et al*. [9] have proposed a protocol that allows individual site to unilaterally commit. Gupta *et al*. [10,11] proposed optimistic commit protocol and its variant. Enhancement has been made in PROMPT commit protocol [12], which allows executing transactions to borrow data in a controlled manner only from the healthy transactions in their commit phase. However, it does not consider the type of dependencies between two transactions. The impact of buffer space and admission control is also not studied. In case of sequential transaction execution model, the borrower is blocked for sending the WORKDONE message and the next cohort can not be activated at other site for its execution. It will be held up till the lender completes. If its sibling is activated at another site anyway, the cohort at this new site will not get the result of previous site because previous cohort has been blocked from sending the WORKDONE message due to being borrower [13]. In shadow PROMPT, a cohort forks off a replica of the transaction, called a shadow, without considering the type of dependency whenever it borrows a data page.

Lam *et al.* [1] proposed deadline-driven conflict resolution (DDCR) protocol which integrates concurrency control and transaction commitment protocol for firm real time transactions. DDCR resolves different transaction conflicts by maintaining three copies of each modified data item (before, after and further) according to the dependency relationship between the lock requester and the lock holder. This not only creates additional workload on the systems but also has priority inversion problem. The serializability of the schedule is ensured by checking the before set and the after set when a transaction wants to enter the decision phase. The protocol aims to reduce the impact of a committing transaction on the executing transaction which depends on it. The conflict resolution in DDCR is divided into two parts (a) resolving conflicts at the conflict time; and (b) reversing the commit dependency when a transaction, which depends on a committing transaction, wants to enter the decision phase and its deadline is approaching.

If data conflict occurs between the executing and committing transactions, system's performance will be affected. Pang Chung-leung and Lam K. Y. [2] proposed an enhancement in DDCR called the DDCR with similarity (DDCR-S) to resolve the executing- committing conflicts in DRTDBS with mixed requirements of criticality and consistency in transactions. In DDCR-S, conflicts involving transactions with looser consistency requirement and the notion of similarity are adopted so that a higher degree of concurrency can be achieved and at the same time the consistency requirements of the transactions can still be met. The simulation results show that the use of DDCR-S can significantly improve the overall system performance as compared with the original DDCR approach.

Based on PROMPT and DDCR protocols, B. Qin and Y. Liu [14] proposed double space commit (2SC) protocol. They analyzed and categorized all kind of dependencies that may occur due to data access conflicts between the transactions into two types commit dependency and abort dependency. The 2SC protocol allows a non-healthy transaction to lend its held data to the transactions in its commit dependency set. When the prepared transaction aborts, only the transactions in its abort dependency set are aborted and the transactions in its commit dependency set execute as normal. These two properties of the 2SC reduce the data inaccessibility and the priority inversion that is inherent in distributed real-time commit processing. 2SC protocol uses blind write model. Extensive simulation experiments have been performed to compare the performance of 2SC with that of other protocols such as PROMPT and DDCR. The simulation results show that 2SC has the best performance. Furthermore, it is easy to incorporate it in any commit protocol.

Ramamritham *et al*. [15] have given three common types of constraints for the execution history of concurrent transactions. The paper [16] extends the constraints and gives a fourth type of constraint. The weak commit dependency and abort dependency between transactions, because of data access conflicts, are analyzed. Based on the analysis, an optimistic commit protocol Two-Level Commit (2LC) is proposed, which is specially designed for the distributed real time domain. It allows transactions to optimistically access the locked data in a controlled manner, which reduces the data inaccessibility and priority inversion inherent and undesirable in DRTDBS. Furthermore, if the prepared transaction is aborted, the transactions in its weak commit dependency set will execute as normal according to 2LC. Extensive simulation experiments have been performed to compare the performance of 2LC with that of the base protocols PROMPT and DDCR. The simulation results show that 2LC is effective in reducing the number of missed transaction deadlines. Furthermore, it is easy to be incorporated with the existing concurrency control protocols.

Udai Shanker *et al.* [17] proposed SWIFT protocol. In SWIFT, the execution phase of a cohort is divided into two parts, locking phase and processing phase and then, in place of WORKDONE message, WORKSTARTED message is sent just before the start of processing phase of the cohort. Further, the borrower is allowed to send WORKSTARTED message, if it is only commit dependent on other cohorts instead of being blocked as opposed to PROMPT. This reduces the time needed for commit processing and is free from cascaded aborts. However, SWIFT commit protocol is beneficial only if the database is main memory resident. Based on the SWIFT protocol, Dependency Sensitive Shadow SWIFT (DSS-SWIFT) protocol [18] was proposed, where the cohort forks off a replica of itself called a shadow, whenever it borrows dirty value of a data item, and if, the created dependency is abort type as compared to creating shadow in all cases of dependency in Shadow PROMPT. Also the health factor of cohort is used for permitting to use dirty value of lender rather than health factor of transaction as whole.

Here, our modified work proposes a commit protocol ACTIVE, which allows lender to lend data to healthy borrowers. Also, commit dependent cohorts are further allowed to lend data. This protocol is beneficial both for main memory and disk resident databases.

The remainder of this paper is organized as follows. Section 2 introduces the distributed real time database system model. Section 3 discusses the type of dependencies created between conflicting cohorts. Section 4 states the condition for fruitful borrowing. Section 5 presents ACTIVE commit protocol and its pseudo code. Section 6 discusses the simulation results and Section 7 gives an outline for future research directions. Finally, Section 8 concludes the paper.

2. Distributed Real Time Database System Model

The common model for DRTDBS is given below in **Figure 1**. The structure of our simulation model including the description of its various components such as system model, database model, network model, cohort execution model, locking mechanism and the model assumptions is given below [17,18]. At each site, two types of transactions are generated: global transactions and local transactions. Each global transaction consists of m cohorts, where m is less than or equal to the number of database sites N_{site}. We use the same model for local and global transactions. Each local transaction has a coordinator and a single cohort both executing at the same site. Each transaction consists of N_{oper} number of database operations. Each operation requires locking of data items and then processing.

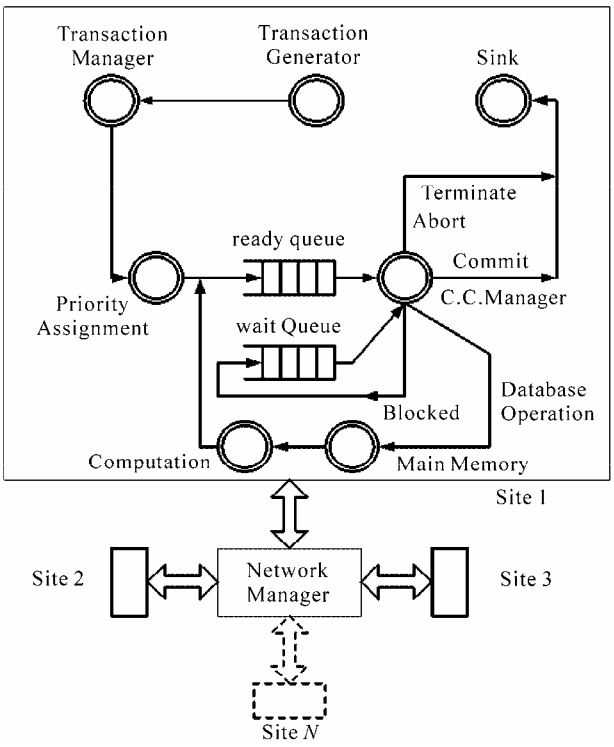


Figure 1. Distributed real-time database system model.

2.1. System Model

Each site consists of a transaction generator, a transaction manager, a concurrency controller, a CPU, a ready queue, a local database, a communication interface, a sink and a wait queue. The transaction generator is responsible for creating the transactions independent to the other sites using Poisson distribution with the given inter-arrival time. The transaction manager generates cohorts on remote site on behalf of the coordinator. Before a cohort performs any operation on a data item, it has to go through the concurrency controller to obtain a lock on that data item. If the request is denied, the cohort is placed in the wait queue. The waiting cohort is awakened when the requested lock is released and all other locks are available. After getting all locks, the cohort accesses the memory and performs computation on the data items. Finally, the cohort commits/aborts and releases all the locks that it is holding. The sink component of the model is responsible for gathering the statistics for the committed or terminated transactions.

2.2. Database Model

The database is modeled as a collection of data items that are uniformly distributed across all the sites. Transactions make requests for the data items and concurrency control is implemented at the data item level. No replication of data items at various sites is considered here.

2.3. Network Model

A communication network interconnects the sites. There is no global shared memory in the system. All sites communicate via messages exchange over the communication network. The network manager models the behavior of the communications network.

2.4. Cohort Execution Model

In our work, we have considered the parallel execution of cohorts. The coordinator of the transaction spawns all cohorts together by sending messages to remote sites to activate them, lists all operations to be executed at that site and then cohorts may start execution at the same time in parallel. The assumption here is that a cohort does not have to read from its sibling and operations performed by one cohort during its execution are independent of the results of the operations performed by other cohorts at some other sites. In other words, the sibling cohorts do not require any information from each other to share.

2.5. Locking Mechanism

The main technique used to control concurrent execution of transactions is based on the concept of locking data items. A lock is a variable associated with a data item that describes the status of the item with respect to possible operations that can be applied to it. Generally there is one lock for each data item in the database. Locks are means for synchronizing the access of concurrent transactions to the database items. A transaction is said to follow the two phase locking protocol if all locking operations precede the first unlock operation in the transaction. There is a number of variations of the two phase locking (2PL) such as static two phase locking (S2PL) and dynamic two phase locking (D2PL). The static 2PL (S2PL) requires a transaction to lock all needed data items before the transaction begins execution, by predeclaring it's read-set and write-set. If any of the predeclared data item can not be locked, the transaction does not lock any items; instead, it waits until all data items are available for locking.

2.6. Model Assumptions

We assume that the transactions are firm real time transactions. The model assumptions are listed below.

1) Processing of a transaction requires the use of CPU and data items located at local or remote site.

2) Arrival of the transactions at a site is independent of the arrivals at other sites and uses Poisson distribution.

3) Each transaction pre-declares its read-set (set of data items that the transaction will only read) and update-set (set of data items that the transaction will update).

4) The cohorts are executed in parallel.

5) A lending transaction cannot lend the same data item in read/update mode to more than one cohort to avoid cascaded abort.

6) The communication delay considered is either 0ms or 100ms to study the impact of the network delay on the system.

7) A distributed real time transaction is said to commit, if the coordinator has reached to commit decision before the expiry of the deadline at its site. This definition applies irrespective of whether cohorts have also received and recorded the commit decision by the deadlines [19].

8) Each cohort makes read and update accesses.

9) S2PL-HP is used for locking the data items.

10) Studies have been made for both main memory resident and disk resident database.

11) In case of disk resident database, buffer space is sufficiently large to allow the retention of data updates until commit time.

12) The updating of data items is made in transaction own memory rather than in place updating

3. Types of Dependencies

Sharing of data items in conflicting modes creates dependencies among conflicting transactions and constraints their commit order. We assume that a cohort requests an update lock if it wants to update a data item x. The prepared cohorts called as lenders lend uncommitted data to concurrently executing transactions known as borrower. Here, the borrower is further divided into two categories.

1) Commit Dependent Borrower

2) Abort Dependent Borrower

Therefore, modified definitions of dependencies used in this paper are given below:

Commit dependency (CD). If a transaction T_2 updates a data items read by another transaction T_1, a commit dependency is created from T_2 to T_1. Here, T_2 is called as commit dependent borrower and is not allowed to commit until T_1 commits.

Abort dependency (AD). If T_2 reads/updates an uncommitted data item updated by T_1, an abort dependency is created from T_2 to T_1. Here, T_2 is called as abort dependent borrower. T_2 aborts, if T_1 aborts and T_2 is not allowed to commit before T_1.

Each transaction/cohort T_i, that lends its data while in prepared state to an executing transaction/cohort T_j, maintains two sets.

1) Commit Dependency Set CDS (T_i): set of commit dependent borrower T_j, that are borrowed dirty data from lender T_i.

2) Abort Dependency Set ADS (T_i): the set of abort

dependent borrower T_j that are borrowed dirty data from lender T_i.

These dependencies are required to maintain the ACID properties of the transaction. Each lender is associated with a health factor defined as follows:

HF (health factor) = TimeLeft/MinTime

where TimeLeft is the time left until the transaction's deadline, and MinTime is the minimum time required for commit processing. The health factor is computed at the time when the coordinator is ready to send the YES-VOTE messages. MinHF is the threshold that allows the data held by committing transaction to be accessed. The variable MinHF is the key factor to influence the performance of the protocol. In our experiments, we have taken MinHF as 1.2, the value of MinHF used in PROMPT [12].

4. Condition for Fruitful Borrowing

If the deadline of an executing cohort is lesser than a committing cohort's commit time, then borrowing of locked data item is useless. Therefore, for fruitful borrowing, an incoming executing cohort having borrowing factor (BF) greater than a threshold value must be permitted to borrow the dirty data items from lender. The borrowing factor can be computed as given below.

Let us consider that transaction/cohort T_i that lends its data items while in prepared state to an executing transaction/cohort T_j, Here, T_i's voting phase is over and has entered in decision phase. The commit time (C_i) of T_i is the mean time required for the decision phase. It includes the time for sending the final commit message to the participating cohorts, the time for writing the final decision into stable storage, the time for permanently updating the data items for write operations and the time needed for releasing the locks [1]. *T*he minimum transaction response time (R_j) of T_j can be calculated as given below [17].

The deadline of a transaction is controlled by the runtime estimate of a transaction and the parameter slack factor, which is the mean of an exponential distribution of slack time. We allocate deadlines to arriving transacttions using the method given below. The deadlines of transactions (both global and local) are calculated based on their expected execution times [1,17].

The deadline (D_j) of transaction (T_j) is defined as:

$$D_j = A_j + SF * R_j$$

where, A_j is the arrival time of transaction (T_j) at a site; SF is the slack factor; R_j is the minimum transaction response time. As cohorts are executing in parallel, the R_j can be calculated as:

$$R_j = R_p + R_c$$

where, R_p, the total processing time during execution phase and commitment phase, and R_c, the communication delay during execution phase and commitment phase are given as below. For global transaction

R_p = max. $((2T_{\text{lock}} + T_{\text{process}})N_{\text{oper local}}, (2T_{\text{lock}} + T_{\text{process}})N_{\text{oper remote}})$

$R_c = N_{\text{comm}} T_{\text{com}}$

For local transaction

$R_p = (2T_{\text{lock}} + T_{\text{process}})N_{\text{oper local}}$

$R_c = 0$

Where, T_{lock} is the time required to lock/unlock a data item; T_{process} is the time to process a data item (assuming read operation takes same amount of time as write operation); N_{comm} is no. of messages; T_{com} is communication delay *i.e.* the constant time estimated for a message going from one site to another; $N_{\text{oper local}}$ is the number of local operations; $N_{\text{oper remote}}$ is maximum number of remote operations taken over by all cohorts. If T_2 is abort dependent on T_1.

The BF can be the ratio of $(SF*R_j - T_{com})/C_i$. For Fruitful Borrowing.

$$(SF*R_j - T_{com})/C_i > 1$$

5. ACTIVE Commit Protocol

5.1. Type of Dependency in Different Cases of Data Conflicts

Let T_1 be a transaction/cohort with identifier id_1 and holding a lock on data item *x*, and let T_2 be another transaction/cohort with id_2 requesting the same data item x, two situation may result depending on the status of T_1.

Situation 1: T_1 is a prepared cohort called as lender, lending uncommitted data. When data conflicts occur, there are three possible cases.

Case 1: Read-update conflict.

If T_2 requests an update-lock and it's BF > 1 while T_1 is holding a read-lock a commit dependency is defined from T_2 to T_1. First, the transaction id of T_2 is added to the CDS (T_1). Then T_2 acquires the update-lock

Case2: Update-Update conflict.

If T_2 requests an update-lock and it's BF > 1 while T_1 is holding an update-lock and HF(T_1) ≥ MinHF an abort dependency is defined from T_2 to T_1. The transaction id of T_2 is added to ADS (T_1), and T_2 acquires the update-lock; otherwise, T_2 is blocked.

Case 3: Update-Read conflict.

If T_2 requests a read-lock and it's BF > 1 while T_1 is holding an update-lock and HF(T_1) ≥ MinHF, an abort dependency is defined from T_2 to T_1. The transaction id of T_2 is added to ADS (T_1), and T_2 acquires the read-lock; otherwise, T_2 is blocked.

Situation 2: T_2 is commit dependent borrower going to become a lender simultaneously by lending its uncommitted data to an incoming transaction/cohort T_3. When

data conflicts occur, there are two possible cases of conflict.

Case 1: Update-Update conflict.

If T_3 requests an update-lock and it's BF > 1 while T_2 is holding an update-lock and HF(T_2) ≥ MinHF an abort dependency is defined from T_3 to T_2. The transaction id of T_3 is added to ADS (T_2), and T_3 acquires the update-lock; otherwise, T_3 is blocked.

Case 2: Update-Read conflict.

If T_3 requests a read-lock and it's BF > 1 while T_2 is holding an update-lock and HF(T_2) ≥ MinHF an abort dependency is defined from T_3 to T_2. The transaction id of T_3 is added to ADS (T_2), and T_3 acquires the read-lock; otherwise, T_3 is blocked.

In our Protocol, locks on data item are granted to those borrowers whose BF > 1. On the basis of the data conflicts discussed above, the access of data items in conflicting mode are processed by lock manager as follows.

If (T_1 is an independent lender)
{ If ((T_2 (BF > 1) CD T_1)
{ CDS (T_1) = CDS (T_1) ∪ {T_2};
T_2 is granted Update lock.
}
else if ((T_2(BF > 1) AD T_1) AND (HF(T_1) ≥ MinHF))
{ ADS (T_1) = ADS (T_1) ∪ {T_2};
T_2 is granted the requested lock (read or Update).
}
else T_2 will be blocked;
}
else if(T_1 is commit dependent borrower)
{ If ((T_2 (BF > 1) AD T_1) AND (HF (T_1) ≥ MinHF))
{ ADS (T_1) = ADS (T_1) ∪ {T_2};
T_2 is granted the requested lock (read or Update)
}
else T_2 will be blocked; }
else T_1 is not allowed to lend data

5.2. Mechanics of Interaction between Lender and Borrower Cohorts

Let us consider three transactions T_1, T_2 and T_3. T_2 has borrowed the data locked by T_1 being an independent lender. If T_2 is commit dependent borrower and has become a lender simultaneously by lending uncommitted data to an incoming transaction/cohort T_3, the following three scenarios are possible:

***Scenario 1*: T_1 receives decision before T_2 is going to start processing phase after getting all locks.**

T_1 is an independent lender

If the global decision is to commit, T_1 commits.

1) All cohorts in ADS (T_1) and CDS (T_1) will execute as usual and the sets ADS (T_1) and CDS (T_1) are deleted.

2) If the global decision is to abort, T_1 aborts. The cohorts in the dependency sets of T_1 will execute as follows:

All cohorts in ADS (T_1) will be aborted;

All cohorts in CDS (T_1) will execute as usual;

Sets ADS (T_1) and CDS (T_1) are deleted.

If, there is another commit dependent cohort T_2 on T_1 that has lent its dirty data to another cohort T_3, processing will be done as follows.

If T_1 aborts or commits, there are two possibilities with T_2.

1) It has either received the Yes-Vote message from its coordinator, or

2) It is still in wait state for the same.

In case of possibility 1, T_2's Yes-Response Message will be send to its coordinator. Type of Dependency between T_2 and T_3 will be governed by either Situation 1 Case 2 or Situation 1 Case 3 already discussed earlier depending on whether T_3 is update or read. In case of possibility 2, T_3 can not commit until T_2 terminates (*i.e.* commits or aborts) [21].

***Scenario 2*: T_2 is going to start processing phase after getting all locks before T_1 receives global decision.**

If T_1 is independent lender, T_2 is allowed to send a WORKSTARTED message to its coordinator, if it is commit dependent only; otherwise it is blocked from sending the WORKSTARTED message. It has to wait until

1) Alternative 1: either T_1 receives its global decisions, or.

2) Alternative 2: its own deadline expires, whichever, occurs earlier.

In Alternative 1, the system will execute as in the Scenario 1. In Alternative 2, T_2 will be killed and will be removed from the dependency set of T_1.

If, there is another cohort T_3 has borrowed dirty data from commit dependent borrower T_2, T_3 can not commit until T_2 terminates (*i.e.* commits or aborts).

***Scenario 3*: T_2 aborts before T_1 receives decision**

In this situation, T_2's and T_3's updates are undone and T_2 will be removed from the dependency set of T_1. T_2 and T_3 will be killed and removed from the system.

5.3. Algorithm

On the basis of above discussions, the complete pseudo code of the protocol is given below. if (T_1 (an independent lender) receives global decision before, T_2 is going to start processing phase after getting all locks)
{
ONE: if (T_1's global decision is to commit)
{
T_1 enters in the decision phase;

All cohorts in ADS (T_1) and CDS (T_1) will execute as usual;
Delete set of ADS (T_1) and CDS (T_1);
}
else //T_1's global decision is to abort
{
$T1$ aborts;
The cohorts in CDS (T_1) will execute as usual;
Transaction in ADS (T_1) will be aborted;
Delete sets of ADS (T_1) and CDS (T_1);
}
if (T_2 commit dependent on T_1 has lent its dirty data to another cohort T_3)
{
if (T_2 has received the Yes-Vote message from its coordinator) // T_1 has aborted or committed
{
T_2 send its Yes-Response Message its coordinator
T_2 & T_3 execute as two normal cohorts with T_3 dependent on T_1
}
else // T_2 is still waiting for Yes-Vote message
{
T_3 can not commit until T_2 terminates
}
}
}
else if (T_2 is going to start processing phase after getting all locks before, T_1(an independent lender) receives global decision)
{
check type of dependencies;
if (T_2's dependency is commit only)
{
T_2 sends WORKSTARTED message;
}
else
{
T_2 is blocked for sending WORKSTARTED message;
while (! (T_1 receive global decision OR T_2 misses deadline))
{
if (T_2 misses deadline)
{
Undo computation of T_2;
Abort T_2;
Delete T_2 from CDS (T_1) & ADS (T_1);
}
else GoTo ONE;
}
}
if (T_3 has borrowed dirty data item from commit dependent borrower T_2)
T_3 cannot commit until T_2 terminates
}
else //T_2 is aborted by higher transaction before,
// T_1 receives decision
{
Undo computation of T_2 & T_3;
Abort T_2 & T_3;
Delete T_2 from CDS (T_1) & ADS (T_1);
}

5.4. Main Contributions

1) Borrower cohorts are categorized as commit or abort dependent.

2) Commit dependent borrowers are allowed to further lend data, thereby reducing the data inaccessibility. The length of cascaded abort chain is same as in the case of PROMPT.

3) Borrowing factor is computed for each borrower, so that lock on requested dirty data items can be granted to borrowers in fruitful way only.

To maintain consistency of database, cohort sends the YES-VOTE in response to its coordinator's VOTE-REQ message only when its dependencies are removed and it has finished its processing.

6. Performance Measures and Evaluation

The default values of different parameters for simulation experiments are same as in [17]. The concurrency control scheme used is static two phase locking with higher priority. Miss Percentage (MP) is the primary performance measure used in the experiments and is defined as the percentage of input transactions that the system is unable to complete on or before their deadlines [19].

Since there were no practical benchmark programs available in the market or with research communities to evaluate the performance of protocols and algorithms, an event driven based simulator was written in C language. In our simulation, a small database (200 data items) is used to create high data contention environment. For each set of experiment, the final results are calculated as an average of 10 independents runs. In each run, 100000 transactions are initiated.

6.1. Simulation Results

Simulation was done for both the main memory resident and the disk resident databases at communication delay of 0ms and 100ms. We compared ACTIVE with PROMPT, 2SC and SWIFT in this experiment. **Figure 2** to **Figure 6** show the Miss Percent behavior under normal and heavy load conditions with/without communication delay. **Figure 2** and **Figure 3** deal with main memory based database system while rest of the figures deal with disk resident database system. In these graphs, we

first observe that there is a noticeable difference between the performances of the various commit protocols throughout loading range. This is due to careful lending of data to a borrower. Here, commit dependent borrowers has also lent data to executing cohort reducing the data inaccessibility. At the same time, the transaction abort chain restricts to one only. Also, an incoming executing cohort having borrowing factor greater than one has been permitted only to borrow the dirty data items from lender. Thus, the work done by the borrowing cohort is never wasted because of better borrowing choice. This minimizes the fruitless borrowing by the cohort. In general the number of transaction being committed are more than number of aborted transactions in real life situations. In this way, we can increase some more parallelism in the distributed system. The ACTIVE commit protocol provides a performance that is significantly better than other commit protocols.

Main Memory Resident Database

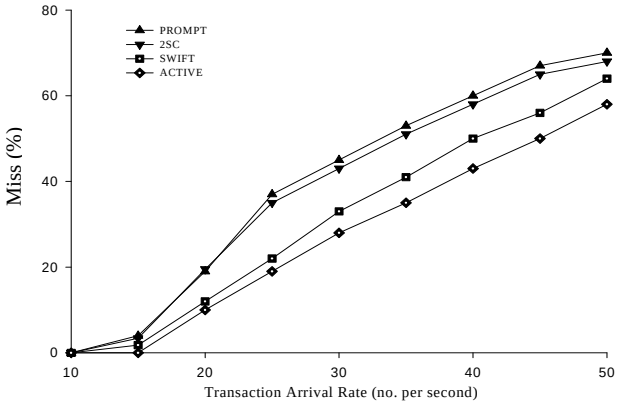


Figure 2. Miss % with (RC + DC) at communication delay 0ms normal and heavy load.

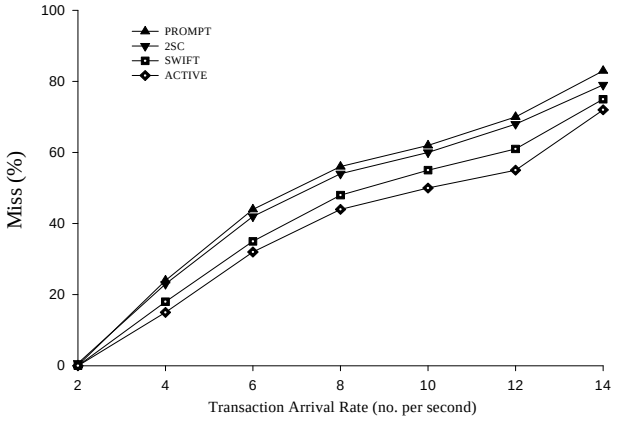


Figure 3. Miss % with (RC + DC) at communication delay 100ms normal and heavy load.

Disk Resident Database System

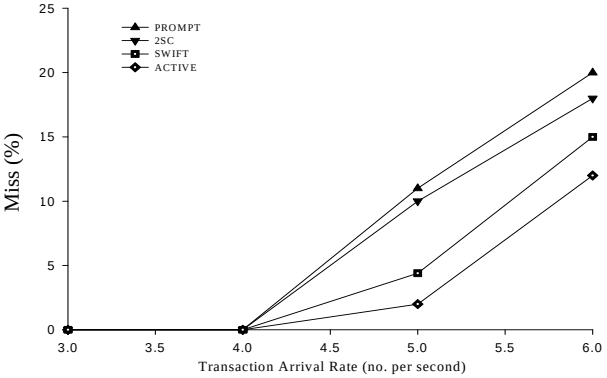


Figure 4. Miss % with (RC + DC) at communication delay 0ms normal load.

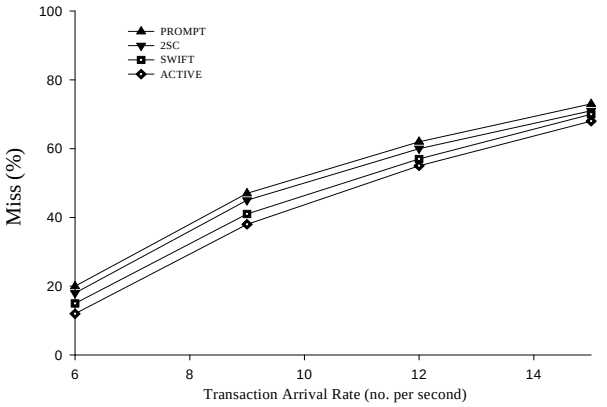


Figure 5. Miss % with (RC + DC) at communication delay 0ms heavy load.

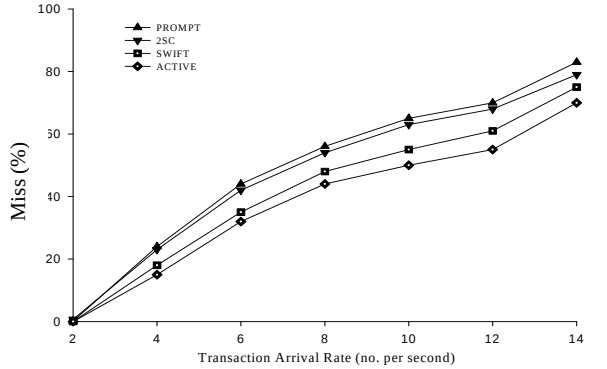


Figure 6. Miss % with (RC + DC) at communication delay 100ms normal and heavy load.

7. Future Research Directions

Following are some suggestions to extend this work [13,20].

1) Despite much research in the area of wireless sensor networks in recent years, the programming of sensor

nodes is still time-consuming and tedious. A new paradigm which seems to be qualified to simplify the programming of sensor networks is the Service Oriented Architecture. The composition of simple services to more complex ones can be a convenient way to design applications. To enable the sophisticated techniques known from service oriented architectures like replication and migration of services, a transaction model for sensor networks is required. The paper [22] studied the applicability of the standard commit protocols Two Phase Commit (2PC) and Transaction Commit on Timeout and show in experiments with real sensor nodes that 2PC can enable consistent service migration in wireless sensor networks. This study can further be extended with this protocol.

2) Our performance studies are based on the assumption that there is no replication. Hence, a study of relative performance of the topic discussed here deserves a further look under assumption of replicated sensor databases.

3) The work can be extended for Mobile DRTDBS, pear-to-pear database systems, grid database systems etc. where memory space, power and communication bandwidth are bottleneck. There is a need to design various protocols for different purposes that may suit to the specific need of hand held devices.

4) Biomedical Informatics is quickly evolving into a research field that encompasses the use of all kinds of biomedical information, from genetic and proteomic data to image data associated with particular patients in clinical settings. Biomedical Informatics comprises the fields of Bioinformatics (e.g., genomics and proteomics) and Medical Informatics (e.g., medical image analysis), and deals with issues related to the access to information in medicine, the analysis of genomics data, security, interoperability and integration of data-intensive biomedical applications. Main issues in this field is provision of large computing power such that researchers have access to high performance distributed computational resources for computationally demanding data analysis, e.g., medical image processing and simulation of medical treatment or surgery and large storage capacity and distributed databases for efficient retrieval, annotation and archiving of biomedical data. What is missing today is full integration of methods and technologies to enhance all phases of biomedical informatics and health care, including research, diagnosis, prognosis, etc. and dissemination of such methods in the clinical practice, whenever they are developed, deployed and maintained. Hence it is another topic of research interest.

8. Conclusions

This paper proposes a new commit protocol ACTIVE for DRTDBS, where the commit dependent borrower is allowed to lend its data to an incoming cohort. It reduces the data inaccessibility by allowing the commit dependent borrowers to further act as lenders. Also, to increase the benefit of borrowing, an incoming cohort having borrowing factor greater than one can borrow the dirty data items from a lender. This reduces the fruitless borrowing. To ensure non-violation of ACID properties, checking of the removal of cohort's dependency is required before sending the Yes-Vote message. The performance of ACTIVE is compared with other protocols for both main memory resident and disk resident databases with and without communication delay. Simulation results show that the proposed protocol improves the system performance.

As future work, it is desirable to study the performance of the proposed protocol in real environment.

9. References

[1] Y. Lam, C-L. Pang, S. H. Son, and J. Cao, "Resolving executing-committing conflicts in distributed real-time database systems," Computer Journal, Vol. 42, No. 8, pp. 674–692, 1999.

[2] C.-L. Pang and K. Y. Lam, "On using similarity for resolving conflicts at commit in mixed distributed real-time databases," Proceedings of the 5th International Conference on Real-Time Computing Systems and Applications, 1998.

[3] G. K. Attaluri and K. Salem, "The presumed-either two-phase commit protocol," IEEE Transactions on Knowledge and Data Engineering, Vol. 14, No. 5, pp. 1190–1196, 2002.

[4] J. Gray and A. Reuter, "Transaction processing: Concepts and technique," Morgan Kaufman, San Mateo, California, 1993.

[5] J. Gray, "Notes on database operating systems," Operating Systems: An Advanced Course, Lecture Notes in Computer Science, Springer Verlag, Vol. 60, pp. 397–405, 1978.

[6] P. Misikangas, "2PL and its variants," Seminar on Real-Time Systems, Department of Computer Science, University of Helsinki, 1997.

[7] I. Lee and Y. H. Yeom, "A single phase distributed commit protocol for main memory database systems," 16th International Parallel & Distributed Processing Symposium (IPDPS), Ft. Lauderdale, Florida, USA, 2002.

[8] C. Mohan, B. Lindsay, and R. Obermarck, "Transaction management in the R* distributed database management system," ACM transaction on Database Systems, Vol. 11, No. 4, 1986.

[9] N. Soparkar, E. Levy, H. F. Korth, and A. Silberschatz, "Adaptive commitment for real-time distributed transaction," Technical Report TR-92–15, Department of Computer Science, University of Texax, Austinm, 1992.

[10] R. Gupta, J. R. Haritsa, and K. Ramamritham, "More optimism about real-time distributed commit processing," Technical Report TR–97-Database System Lab, Super-

computer Education and Research Centre, Indian Institute of Science, Bangalore, India, 1997.

[11] R. Gupta, J. R. Haritsa, K. Ramamritham, and S. Seshadri, "Commit processing in distributed real time database systems," Proceedings of Real-Time Systems Symposium, IEEE Computer Society Press, Washington DC, San Francisco, 1996.

[12] J. R. Haritsa, K. Ramamritham, and R. Gupta, "The PROMPT real time commit protocol," IEEE Transaction on Parallel and Distributed Systems, Vol. 11, No. 2, pp. 160–181, 2000.

[13] U. Shanker, M. Misra, and A. K. Sarje, "Distributed real time database systems: Background and literature review," International Journal of Distributed and Parallel Databases, Springer Verlag, Vol. 23, No. 2, pp. 127–149, 2008.

[14] B. Qin and Y. Liu, "High performance distributed real time commit protocol," Journal of Systems and Software, Elsevier Science Inc., pp. 1–8, 2003.

[15] K. Ramamritham and P. K. Chrysanthis, "A taxonomy of correctness criteria in data-base applications," Journal of the Very Large Data Bases, Vol. 5, pp. 85–97, 1996.

[16] B. Qin, Y. Liu, and J. Yang, "A commit strategy for distributed real-time transaction," Journal of Computer Science and Technology, Vol. 18, No. 5, pp. 626–631, 2003.

[17] U. Shanker, M. Misra and A. K. Sarje, "SWIFT - a new real time commit protocol," International Journal of Distributed and Parallel Databases, Springer Verlag, Vol. 20, No. 1, pp. 29–56, 2006.

[18] U. Shanker, M. Misra, A. K. Sarje and R. Shisondia, "Dependency sensitive shadow SWIFT," Proceedings of the 10th International Database Applications and Engineering Symposium, Delhi, India, pp. 373–276, 2006.

[19] O. Ulusoy, "Concurrency control in real time database systems," PhD Thesis, Department of Computer Science, University of Illinois Urbana-Champaign, 1992.

[20] U. Shanker, M. Misra, and A. K. Sarje, "Hard real-time distributed database systems: Future directions," Communication Network, Department of Electronics & Computer Engineering, Indian Institute of Technology Roorkee, India, pp. 172–177, 2001.

[21] D. Agrawal, A. El Abbadi, R. Jeffers and L. Lin, "Ordered shared locks for real-time databases," International Journals of Very Large Data Bases, Vol. 4, No. 1, pp. 87-126, January 1995.

[22] Christoph Reinke, Nils Hoeller, Jana Neumann, Sven Groppe, Volker Linnemann and Martin Lipphardt, "Integrating standardized transaction protocols in service-oriented wireless sensor networks," Proceedings of the 2009 ACM symposium on Applied Computing, Honolulu, Hawaii, pp. 2202-2203, 2009

Wireless Sensor Network, **2010, 2, 264-266**
doi:10.4236/wsn.2010.23036 Published Online March 2010 (http://www.scirp.org/journal/wsn)

Scientific Research

Self Umpiring System for Security in Wireless Mobile Ad Hoc Network

Ayyaswamy Kathirvel, Rengaramanujam Srinivasan
Assistant Professor, B.S.A. Crescent Engineering College, Chennai, India
Professor, B.S.A. Crescent Engineering College, Chennai, India
E-mail: {*kathir, drsrs*}*@crescentcollege.org*
Received December 22, 2009; *revised January* 6, 2010; *accepted January* 8, 2010

Abstract

A wireless mobile ad hoc network (MANET) is a self-created self-organized and self-administering set of nodes connected via wireless links without the aid of any fixed infrastructure or centralized administrator. In this paper we propose a solution of self-umpiring system that provides security for routing and data forwarding operations. In our system each node in the path from source to destination has dual roles to perform: packet forwarding and umpiring. In the umpiring role, each node in the path closely monitors the behavior of its succeeding node and if any misbehavior is noticed immediately flags off the guilty node. The umpiring system proposed is sufficiently general and can be applied to any networking protocol. For demonstration, we have implemented the self-umpiring system by modifying the popular AODV protocol. Simulation studies show that the proposed system increases throughput by 166.9% with an increase in communication overhead of 13.3% as compared to plain AODV, when 40% of the nodes are malicious and are roaming with a mobility of 20 m/s.

Keywords: MANET, Security, AODV, Self-Umpiring System

1. Introduction

A wireless mobile ad hoc network (MANET) is a self-created self-organized and self-administering set of nodes connected via wireless links without the aid of any fixed infrastructure or centralized administrator. Each node moves and operates in a distributed peer-to-peer mode, generating independent data and acting as a router to provide multi-hop communication. MANET is ideally suited for potential applications in civil and military environments, such as responses to hurricane, earthquake, tsunami, terrorism and battlefield conditions. Security is an important aspect in such mission critical applications.

In this paper we tackle the problem of securing the network layer operations from malicious nodes. Malicious nodes may disrupt routing algorithms by transmitting a false hop count; they may drop packets, route the packets through unintended routes and so on. Our work rests on the foundations of two excellent systems already proposed: the twin systems of watchdog and pathrater [1] and SCAN [2].

Our self-umpiring system has been strongly influenced by the above two schemes. In our system all the active nodes have dual roles just as in watchdog; we also exploit promiscuous hearing functionality as done by both SCAN and watchdog. We have adopted the token concept from SCAN. However we have dropped partially the pathrater functionality. We believe link reliability assessment of pathrater may not be correct; a proper reliability metric for path assessment should consider the direction and velocity of movement of active nodes. Having dropped the link reliability factor from the pathrater, the only other functionality that remains is avoidance of malicious nodes. We achieve the avoidance of malicious nodes by a system of tokens, which is similar to the ones used in SCAN. Token is a pass or validity certificate enabling a node to participate in the network. It contains two fields: nodeID and status bit; nodeID is considered to be immutable. Initially the status bit of all participating nodes is set as 0 indicating "green flag" with freedom to participate in all network operations. It is assumed that a node cannot change its own status bit. When an umpiring node finds its succeeding node misbehaving it sends a M-Error message to the source and malicious node's status bit is changed using M-Flag message (set to 1 indicating "red flag"). With "red flag" on the culprit node is prevented from participating in the network.

The rest of the paper is organized as follows: Section 2 provides an overview of Self_USS models. Section 3 presents simulation results; Section 4 gives the related work and Section 5 gives the conclusions

2. Self-Umpiring System Security Model: Self_Uss

In the self-umpiring system each node is issued with a token at the inception. The token consists of two fields: NodeID and status. NodeID is assumed to be unique and deemed to be beyond manipulation; status is a single bit flag. Initially the status bit is preset to zero indicating a green flag. The token with green flag is a permit issued to each node, which confers it the freedom to participate in all network activities. Each node in order to participate in any network activity, say Route Request RREQ, has to announce its token. If status bit is "1" indicating "red flag" protocol does not allow the node to participate in any network activity. The working of the self-umpiring system is explained with reference to **Figure 1**.

In the self-umpiring system all the nodes have dual roles—packet forwarding and umpiring. In the forward path during data forwarding, each node monitors the performance of immediate next node. That way, node A can tell correctly whether B is forwarding the packet sent by it, by promiscuously hearing B's transmissions. Similarly during reply process RREP, C can verify whether B is unicasting the route reply RREP and whether the hop count given by B is correct. Thus during forward path A is the umpire for B and C is the umpire for B during reverse path operations.

When a node is found to be misbehaving—say dropping packets, the corresponding umpire immediately sends a M-ERROR message to the source and the status bit of guilty node is set to "1"—red flag using M-Flag message. In order to correctly correlate the overheard messages an additional field next_hop has been introduced in all routing messages as done in SCAN [2]. Though there are several kinds of misbehavior that could be captured by promiscuous hearing we are focusing only on two types of malicious actions: dropping packets and transmitting false hop count.

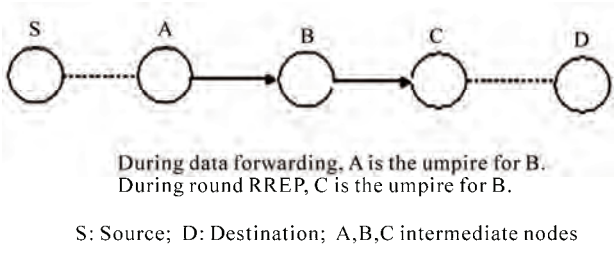


Figure 1. Self umpiring system model.

3. Simulation and Results

We use a simulation model based on QualNet 4.5 in our evaluation [3]. Our performance evaluations are based on the simulations of 100 wireless mobile nodes that form a wireless ad hoc network over a rectangular (1500 × 600 m) flat space. The MAC layer protocol used in the simulations was the Distributed Coordination Function (DCF) of IEEE 802.11 [4]. The performance setting parameters are given in **Table 1**.

Before the simulation we randomly selected a certain fraction, ranging from 0% to 40% of the network population as malicious nodes. We considered only two attacks —modifying the hop count and dropping packets. Each flow did not change its source and destination for the lifetime of a simulation run.

3.1. Throughput

In the world of MANET, packet delivery ratio has been accepted as a standard measure of throughput. Packet delivery ratio is nothing but a ratio between the numbers of packets received by the destinations to the number of packets sent by the sources.

From packet delivery ratio the following conclusions can be drawn:

1) In general packet delivery ratio decreases as mobility and percentage of malicious nodes increase.

2) We observe that the same results are obtained with Self_USS also. With zero percentage malicious nodes, self-umpiring system and plain AODV have almost identical performances.

3) We find similar increase in throughput at all other combinations of malicious node percentages and mobility values, with self-umpiring system.

From the above results we conclude that self-umpiring system leads to a substantial improvement over plain AODV, from the point of view of throughput.

3.2. Failure to Deduct (False Negatives) Probability

False Negatives Probability can be defined as:

False Negatives Probability = number of malicious nodes left undetected/total number of malicious nodes.

Table 1. Parameter settings.

Simulation Time	1500 seconds
Propagation model	Two-ray Ground Reflection
Transmission range	250 m
Bandwidth	2 Mbps
Movement model	Random way point
Maximum speed	0-20 m/s
Pause time	0 seconds
Traffic type	CBR(UDP)
Payload size	512 bytes
Number of flows	10/20

The above definition requires some elaboration. We can think of two groups of malicious nodes that are left undetected. In the first group are those nodes, which never played a part in the network operation; they were probably traveling along the boundaries and never had a chance to participate in the network activity.

The second groups of malicious nodes are those that played a role as a forwarding node, but went undetected. Clearly our umpiring system is responsible only for the second group. The first group of nodes is similar to reserve players in the sidelines and clearly any umpire cannot show red flag and march off players in the sidelines. Appropriately we have done the failure to detect probability calculation taking into consideration only those nodes, which took part in the network activity. Other researchers adopt the same approach also. The results are similar that of SCAN [2].

3.3. False Accusation (False Positives) Probability

This is the probability of wrongly booking innocent nodes. We find false positive probability increases with increasing percentage of malicious nodes and increased mobility. The values vary between 0 to 10% and are similar to the patterns obtained for SCAN [2].

3.4. Communication Overhead

Communication overhead can be evaluated based on the number of transmissions of control messages like RREQ, RREP, RERR in the case of plain AODV and in addition M_ERROR, M-Flag messages in the self umpiring system. RREQ are to be decimated to the entire network, where as RREP messages are unicasts.

From communication overhead following inferences can be drawn:

1) The communication overhead increases with increasing percentage of malicious nodes and mobility for both plain AODV and Self_USS.

2) Further we find that when there is no malicious nodes (0% malicious nodes) the nodes in their umpiring role have very few message packets to send and the communication overheads for plain AODV and Self_USS are nearly same.

4. Related Works

The key distribution center (KDC) architecture is the main stream in wired network because KDC has so many merits: efficient key management, including key generation, storage, distribution and updating. The lack of trusted third party (TTPs) key management scheme is a big problem in mobile ad hoc network [5–7].

All the above schemes only try to protect the system from the attacker, but not bother about quarantining attackers. The twin systems of *watchdog* and *pathrater* [1] not only detect the mischievous nodes but also prevent their further participation in the network. SCAN [2] also has similar action, but is more comprehensive, in the sense not only packet dropping but also other misbehaviors like giving wrong hop count are covered. Our self-USS is an extension of the above two works.

Routeguard [7] is similar to *pathrater and* is run by each node. Routeguard introduces more detailed and natural classification system that rates each node into one of the five classes: *fresh, member, unstable, suspect* or *malicious.* Accordingly each node is treated differently.

5. Conclusions

A self-umpiring system for security for wireless mobile ad hoc network has been proposed. Simulation studies show that the proposed system increases throughput by 166.9% with an increase in communication overhead of 13.3% as compared to plain AODV, when 40% of the nodes are malicious and are roaming with a mobility of 20 m/s. Research work is in progress.

6. Acknowledgements

We express our thanks to Prof. V. M. Periasamy, the Register and Prof. K. M. Mehata, the Head, Department of CSE, B. S. A. Crescent Engineering College Chennai, Tamilnadu, India for the encouraging environment provided.

7. References

[1] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking, Boston, USA, pp. 255–265, 6-11 August, 2000.

[2] H. Yang, J. Shu, X. Meng and S. Lu, "SCAN: Self-organized network-layer security in Mobile ad hoc networks," IEEE Journals on Selected Areas in Communications, Vol. 24, No. 2, February 2006.

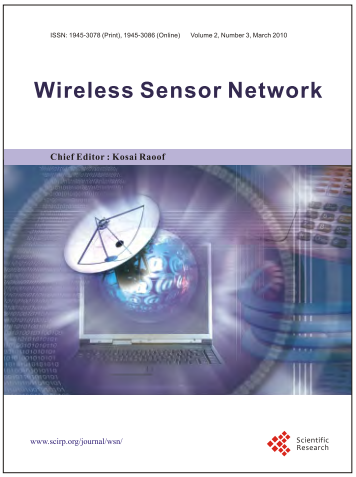
[3] 3. Scalable Networks Technologies: QualNet simulator 4.5, http://www.scalable-networks.com/

[4] IEEE 802.11g. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, August, 1999.

[5] M. A. Azer, S. M. El-Kassas, and M. S. El-Soudani, "Certification and revocation schemes in ad hoc networks survey and challenges," Proceeding of IEEE International Conference on Systems and Networks Communications, 2007.

[6] J. Kong, P. Zerfos, H. Luo, S. Lu, and L. Zhang, "Providing robust and ubiquitous security support for MANET," Proceeding of International Conference on Network Protocols, pp. 251–260, 2001.

[7] N. Nasser and Y. Chen, "Enhanced intrusion detection system for discovering malicious nodes in mobile ad hoc networks," Proceeding of International Conference on Communications, pp. 1154–1159, 2007.



Wireless Sensor Network (WSN)

Call For Papers

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

ISSN 1945-3078 (Print) ISSN 1945-3086 (Online)

WSN is an international refereed journal dedicated to the latest advancement of wireless sensor network and applications. The goal of this journal is to keep a record of the state-of-the-art research and promote the research work in these areas.

Editor-in-Chief

Dr. Kosai Raoof , GIPSA LAB, University of Joseph Fourier, Grenoble, France

Subject Coverage

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

- Network Architecture and Protocols
- Self-Organization and Synchronization
- Quality of Service
- Data Processing, Storage and Management
- Network Planning, Provisioning and Deployment
- Integration with Other System
- Software Platforms and Development Tools
- Routing and Data Dissemination
- Energy Conservation and Management
- Security and Privacy
- Developments and Applications
- Network Simulation and Platforms

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 WSN.

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. Authors are responsible for having their papers checked for style and grammar prior to submission to WSN. Papers may be rejected if the language is not satisfactory. For more details about the submissions, please access the website.

Website and E-Mail

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

Email: wsn@scirp.org

Wireless Sensor Network, 2010, 2, 199-266
Published Online March 2010 in SciRP (http://www.SciRP.org/journal/wsn/).

TABLE OF CONTENTS

Volume 2 Number 3 **March 2010**

9771945307008 08