

行政院國家科學委員會補助專題研究計畫 ☐ 成果報告
☒ 期中進度報告

設計與實作無線微型感測網路攻擊偵測、防禦與高信賴度的
安全傳輸機制

計畫類別：☒ 個別型計畫 ☐ 整合型計畫

計畫編號：NSC 96 — 2221 —E —009 — 139 —MY3

執行期間：2008 年 8 月 1 日至 2009 年 7 月 31 日

計畫主持人：謝續平教授

共同主持人：

計畫參與人員：王繼偉、宋穎昌、施汎勳、張書綸、朱信儒、游釗俊

成果報告類型(依經費核定清單規定繳交)：☐ 精簡報告 ☒ 完整報告

本成果報告包括以下應繳交之附件：

- ☐ 赴國外出差或研習心得報告一份
- ☐ 赴大陸地區出差或研習心得報告一份
- ☐ 出席國際學術會議心得報告及發表之論文各一份
- ☐ 國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、
列管計畫及下列情形者外，得立即公開查詢

☐ 涉及專利或其他智慧財產權，☐ 一年 ☐ 二年後可公開查詢

執行單位：

中 華 民 國 98 年 5 月 30 日

摘要

本成果報告為三年期計畫的第二年報告。第一年研究著重於無線微型感測網路攻擊偵測，已經在前一年繳交報告，探討“無線感測網路中以統計模型為基礎的異常行為偵測 (Detecting Anomalous Behaviors in WSNs with Statistical Learning Model)”。本年度重點在於高信賴度的安全傳輸機制，我們將探討“適用於無線微型感測網路之具容錯性與攻擊抵抗性之多重路徑資訊傳送機制 (A Fault-Tolerant, Attack-Resistant Multipath Information Delivery Scheme for Wireless Sensor Networks)”。

對於大多數無線微型感測網路的應用而言，資訊傳輸的可信賴程度與其安全性為兩項主要的考量因素。由於感測節點通常佈建於無人照料或敵人充斥的環境中，並且感測到的資訊必須經由節點間本質即不穩定的無線連結傳輸至後端基地台，如何正確、安全且即時地傳送感測到的資訊實為一大挑戰。為了解決此項議題，我們提出了一個基於封包切割概念的多重路徑資訊傳送機制，FAMIDS。在此機制之下，資料來源節點將原始封包利用增加冗位元的方式切割成數個小封包，並且利用多重路徑安全地傳送該封包。即使部分的小封包接收錯誤，後端基地台仍能重建原始封包。

除此之外，我們建立了分析 FAMIDS 所能達到的傳輸可信賴程度之分析模型。針對不同的網路條件，我們研究該分析模型之表現情形，並且提出能找出最佳封包切割方式的有效率演算法，以達到最佳的傳輸可信賴程度。相較於傳統之暴力法，我們提出之演算法能有效降低每次封包傳輸之計算複雜度。

關鍵詞：無線微型感測網路、容錯性、攻擊抵抗性、多重路徑資訊傳送機制

A Fault-Tolerant, Attack-Resistant Multipath Information Delivery Scheme for Wireless Sensor Networks

Abstract

Transmission reliability and information confidentiality are two of the major concerns for most applications of wireless sensor networks. Since sensor nodes are usually deployed in unattended or adversarial environments and sensed information is transmitted through inherent instable wireless links among sensor nodes to the base station, how to deliver sensed information correctly, securely and instantaneously becomes a challenging problem. To deal with this issue, we propose an enhanced multipath information delivery scheme FAMIDS based on the concept of packet splitting. With the proposed scheme, the source node splits the original data packet into several sub-packets by adding redundancy, and securely transmits the sub-packets through multiple paths. The base station can reconstruct the original data packet even if part of sub-packets is incorrectly received.

In addition, we develop an analytical model for evaluating communication reliability achieved by our scheme and study the behavior of the analytical form under different network conditions. For each network model, we propose an algorithm for finding the optimal way of packet splitting with highest communication reliability. Compared with the traditional brute-force approach, the proposed algorithms efficiently reduce the computational complexity for each packet transmission.

Keywords: Wireless Sensor Networks Security, Fault-Tolerance, Attack-Resistant, Multipath Information Delivery

1. Introduction

Recently, wireless sensor networks have been widely used for various applications, such as habitat monitoring (i.e. animals, vehicle, or weather) and target detection (i.e. fire, earthquake, or enemies), etc. In such applications, thousands of sensor nodes with restricted computing resources are deployed in the environment being sensed or monitored. The information collected by sensor nodes is transmitted in a hop-by-hop manner between sensor nodes to the base station for the sake of decision making or emergency awareness. Among most applications of sensor networks, transmission reliability and information confidentiality are two critical concerns for system design. Since the sensor nodes are usually deployed in unattended or adversarial environments such as a battlefield, and the data packets are transmitted through inherent unreliable wireless links between sensor nodes, how to guarantee efficient, reliable and instantaneous information delivery becomes a big challenge [11].

To address the issue of communication reliability, many schemes have been proposed. Most schemes achieve end-to-end reliability by single path routing with explicit acknowledgement and retransmission. However, the high packet loss rate resulted from instability of sensor nodes and wireless links can prove troublesome for such mechanism, thus only limited reliability is provided. In addition, multipath routing scheme with three different strategies of packet transmission have been designed for increasing network reliability:

- 1) Selective forwarding: The source node sends the packet to the base station through one randomly chosen path from the multiple paths.
- 2) Packet replication: The source node duplicates the packet and sends each copy through one of the multiple paths.
- 3) Packet splitting: The source node splits the packet into several sub-packets by erasure codes or forward error correcting codes and sends the sub-packets through the

multiple paths simultaneously. Since the packet splitting mechanisms add redundancy to original packet, the base station can reconstruct the original packet with part of the sub-packets.

It is noted that there is a tradeoff between the network traffic load and the achieved reliability. Due to the fault-tolerant and load-balancing nature of packet splitting mechanisms, previous research [3] has showed that multipath routing with packet splitting strategy achieve better efficiency regarding both traffic overhead and reliability level. However, all the previous schemes and reliability analysis of multipath routing with packet splitting approach can not be directly applied to applications of general wireless sensor networks due to the assumptions on the network topology and the characteristics of communication channel. Besides, the packet splitting approach is susceptible to various attacks, such as pollution attack and replay attack. The previous schemes do not provide countermeasures for resisting such attacks.

In this paper, an enhanced multipath information delivery scheme suitable (FAMIDS) for applications of general wireless sensor networks is proposed considering both the reliability and security issue. FAMIDS is built upon any disjoint multipath routing in wireless sensor networks [1],[6][7]. For the fault-tolerant characteristic of FAMIDS, we modify existing packet splitting scheme so as to meet the requirements for wireless sensor networks. As far as attack-resistant characteristic is concerned, we incorporate some security mechanism to resist the above-mentioned attacks. Our scheme's responsibility is to choose the optimal packet allocation that maximizes the probability of successful reception. We develop an analytical model for evaluating the communication reliability achieved by FAMIDS and study the behavior of the analytical form under different network conditions. The analysis result is applicable for all multipath routing with packet splitting schemes in similar wireless networks.

The rest of this paper is organized as follows. Section 2 reviews the related research on multipath routing with packet splitting schemes in the literature. Section 3 explains the detailed

operations of FAMIDS. Section 4 proposes reliability and security analysis of FAMIDS, and section 5 discusses the future works and concludes this paper.

2. Related work

In this paper, we consider the problem of disjoint multipath routing with packet splitting schemes in wireless sensor networks. The idea has been widely used in both wired and wireless networks, where the sender split the original packet into several sub-packets and transmits them separately, and the receiver does not need to receive all sub-packets to rebuild the original packet, and thus provide communication reliability. Two well-known schemes used for packet splitting are information dispersal scheme [9] and diversity coding [2]. We categorized our related work according to the packet splitting approach they utilized.

Information Dispersal Scheme The information dispersal scheme (IDS) was first developed as a method for secure and reliable storage of information. The idea of IDS is that it breaks a file F of length L bits into n pieces, each of length L/m , by a series of matrix multiplication operations, where $m \leq n$, so that every m pieces suffice for reconstructing F . The detailed operations of file splitting and restoration of (m, n) IDS are summarized as follows:

- 1) F can be viewed as a string of characters, $F = (b_1, b_2, \dots, b_L)$. For simplicity, assume that L is a multiple of m . Break F into strings of length m :

$$F = (b_1, b_2, \dots, b_m)(b_{m+1}, b_{m+2}, \dots, b_{2m}) \dots (b_{L-m+1}, b_{L-m+2}, \dots, b_L) = S_1, S_2, \dots, S_{L/m}$$

- 2) Find a $n \times m$ matrix $\underline{A}$ such that all its $m \times m$ sub-matrices are linear independent.
- 3) F is transformed or stored into n pieces denoted by $F_1, F_2, \dots, F_n$. The pieces are calculated by

$$\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{bmatrix}_{n \times \frac{L}{m}} = [\underline{A}]_{n \times m} \cdot \begin{bmatrix} S_1 & S_2 & \dots & S_{\frac{L}{m}} \end{bmatrix}_{m \times \frac{L}{m}}$$

4) Any m pieces can reconstruct the original packet with the inverse matrix $[A]^{-1}$. We denote the m pieces by F_i , $1 \leq i \leq m$. The original data packet can be reconstructed by

$$\begin{bmatrix} S_1 & S_2 & \cdots & S_{\frac{L}{m}} \end{bmatrix}_{m \times \frac{L}{m}} = [A]_{m \times m}^{-1} \cdot \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}_{m \times \frac{L}{m}}$$

There are two important observations from IDS. First, the length of F_i is independent of n . The total number of characters produced by IDS is $(n/m) \times L$, where $r = n/m$ denotes as the information expansion ratio of IDS. With a large information expansion ratio, IDS incurs a larger storage overhead. Second, both the communication ends must have the knowledge of the matrix $\underline{A}$ to successfully encode and decode the shared information. Besides, the matrix A should be shared secretly in order to guarantee the information confidentiality.

IDS has been widely used for fault-tolerant communications in various networks, such as hyper-cube [19] and omega networks [20]. Sun *et al* [12] investigated the performance of IDS used to support fault-tolerant distributed file servers in a distributed system. In [10], the authors analyzed the performance of IDS for fault-tolerant parallel wireless communications over a burst-error channel. In [12] and [10], the authors both proposed an algorithm to find the candidate sets of (m, n) which achieves the optimal reliability.

Moreover, IDS has also been applied to mobile ad hoc networks. In [8], the authors proposed a secure message transmission (SMT) protocol in mobile ad hoc networks which incorporate IDS to adapt the operation of message transmission to be effective even in a highly adverse environment. However, security association (SA) is needed in SMT to negotiate the secret matrix $\underline{A}$ every transmission, while such communication pattern is not suitable for applications in wireless sensor networks. Besides, the authors do not provide methods to find the optimal (m, n) IDS with highest reliability.

Diversity Coding Another approach used for fault-tolerant information sharing is diversity coding [2]. The concept of diversity coding is similar with information dispersal scheme, which is proposed in order to achieve self-healing and fault tolerance in digital communication networks. In [13] and [14], Tsirigos et. al. analyzed the performance of diversity coding upon disjoint multipath routing in mobile ad hoc networks. In these papers, a method is proposed to find the set of paths that maximize the probability of reconstructing the original packet at the destination if the information expansion ratio and the success probability of each path is given.

However, the problem of diversity coding is that it works under the assumption that link failures are treated as an erasure channel problem, which means the successful transmission of packets is in an all-or-nothing manner, independent of packet size. Nevertheless, the assumption on the characteristic of communication channel is not rational in wireless sensor networks. Since link failures are influenced by the bit error rate (BER) of the sensing environment, the success probability of each path should be different when the packet length changes or more packets are sent. For the above reasons, the previous schemes and analysis based on multipath routing with diversity coding in other networks can not be directly applied in wireless sensor networks. As a result, in FAMIDS, we choose IDS as our packet splitting mechanism.

3. Proposed scheme

In this section, we explain the proposed information delivery scheme, FAMIDS. The operations of FAMIDS can be categorized into two parts: path establishment and packet transmission. To make the usage of FAMIDS effective, several assumptions on the network environment and system requirement should be made. Section 3.1 states the assumptions, while the next sections present the detailed operations of FAMIDS.

3.1. Network and Security Assumptions

We describe our assumptions regarding the sensor network scenarios in which FAMIDS can be

used, followed by the notations will be used in the next section.

Network Assumption Generally, we assume that the sensor network consists of many resource-constrained sensor nodes and one powerful base station. The sensor nodes are randomly and densely scattered in the sensing area and are relative stationary since deployment. We assume that the radio frequency (RF) module of each sensor node can provide parameters to indicate the quality of wireless links, such as radio signal strength indicator (RSSI), or link quality indicator (LQI). All the other computational and communication capabilities of sensor nodes are similar to the current generation sensor nodes, e.g. the Berkeley MICA2 motes. The base station, acting as a network monitor and coordinator, is assumed to be a laptop-class device with sufficient storage space, power supply and computation capability. Most network traffic is from source nodes to the base station, or in a reverse direction.

Security Assumption First of all, we assume sensor nodes trust themselves. Once a sensor node is compromised, all the confidential information it holds will be known to the attacker. In addition, since the wireless communication is inherent insecure, we assume an adversary can eavesdrop on all traffic, inject packets, or replay packets. However, we assume that the base station will not be compromised. Finally, to secure packet transmission, we assume some existent secure key establishment and management protocols, such as LEAP [18], is incorporated into the system. We assumed that at least the following keys are securely established:

- 1) Individual key: Every node shares a unique individual key with the base station. Assume the individual key of node u is K_u , node u can generate the encryption key and authentication key by K_u and a pseudo random function f :

$$K_{encr} = f_{K_u}(0), K_{mac} = f_{K_u}(1)$$

- 2) Pairwise key: Every node shares a unique pairwise key with each of its immediate neighbors.
- 3) Cluster key: Every node shares a key with all its neighbors.

In FAMIDS, we only consider the security of data packet transmission. We assume the security of all keying mechanisms is guaranteed by the key establishment and management protocol, and the security of path establishment and maintenance is assured by the disjoint multipath routing protocol.

Notations We list below the notations which appear in the explanations of FAMIDS.

u, v (in lower case)	principals such as communicating nodes
$\{f_k\}$	a family of pseudo-random functions
K_{encr}	The encryption key of the source node
K_{mac}	The authentication key of the source node
$K_{u,v}$	The pairwise key of node u and node v
$\{M\}_k$	encrypting message M with key k
$\text{MAC}\{k, M\}$	message authentication code of message M using key k .
n	the degree of information dispersal
m	[see (m,n) WIDS]
η_2	$\lfloor n/2 \rfloor$
(m, n) FAMIDS	a FAMIDS which breaks a packet into n pieces and transmits them simultaneously through n disjoint paths such that any m pieces received suffice for reconstructing the original packet; $1 \leq m \leq n$
N	maximum number of available paths
L	the packet length of original data packet
Path_i	The i_{th} path among the N paths
$\text{Node}_{i,j}$	The j_{th} intermediate node of Path_i
$\text{Link}_{i,j}$	The j_{th} wireless link of Path_i
Hop_i	The hop count of Path_i

Table 2-1 Notations

3.2. Path Establishment

Path establishment consists of three main phases: path discovery phase, path evaluation phase and path maintenance phase. We explain the operations of each phase below.

Path discovery phase Path discovery phase is initiated by a source node that has some sensed information for the base station and no valid routing entry in its routing table. The source node

adopts disjoint multipath routing protocol to construct multiple paths to the base station. We assume that any disjoint multipath routing protocols for wireless sensor networks can be adopted to acquire these paths. After path discovery finished, N disjoint paths are established and considered as ideally independent, which means they have no intermediate nodes in common.

After path discovery phase finished, all the involved sensor nodes will possess the up-to-date routing information. The intermediate nodes will set up both the forward path (the source node to the base station) and the reverse path (the base station to the source node).

The following figure shows a simplified example of path discovery phase. Figure 2-1(a) shows the topology of the sample network, which consists of one base station and fifteen sensor nodes. Each node has a unique ID and in this example, the base station ID is 0. The dashed lines represent the communication range of the sensor nodes. Node 14 and Node 11 successively initiate path discovery phase to acquire three disjoint paths to the base station. Figure 2-1(b) and Figure 2-1 (c) show the paths constructed by Node 14 and Node 11 after the path discovery phase finished, respectively. The updated routing information is then stored in the routing table of each involved node. Table 2-2 and Table 2-3 show the related entries of routing table of Node 8 (an intermediate node) and Node 11 (a source node), respectively.

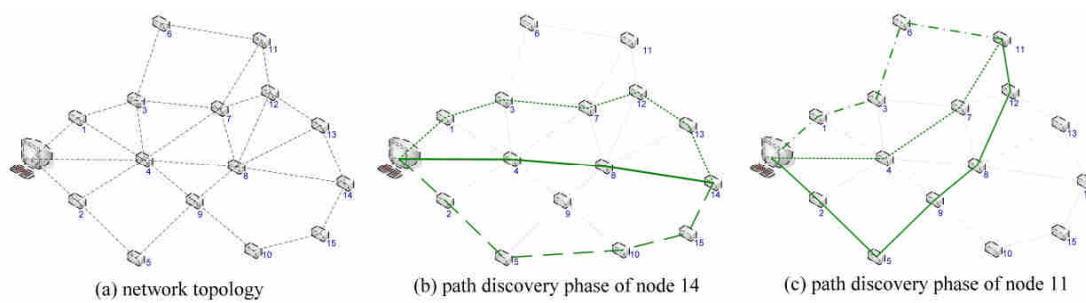


Figure 2-1 Path discovery phase

Destination ID	0	0	14	11
Source ID	14	11	0	0
Hop Count	1	2	2	4
Next Hop ID	4	9	14	12

Table 2-2 Routing table of Node 8

Destination ID	0	0	0
Hop Count	3	4	6
Next Hop ID	7	6	12

Table 2-3 Routing table of Node 11

Path evaluation phase The source node must have the knowledge of successful transmission probability of the paths to perform FAMIDS. It is noted that the established paths have different hop counts and also different successful transmission probability. There are several ways to acquire the successful transmission probability of each path. One viable approach is to utilize the link quality parameter implemented by RF module (e.g. LQI, RSSI) to indicate the successful transmission probability of wireless links. One viable approach is that the base station periodically broadcasts a link quality estimation packet. Each intermediate node receiving the packet appends the bit error rate indicated by the LQI parameter to the packet and retransmits the packet to source nodes through the reverse paths. The source nodes can update the success probability of the paths according to the received packet.

Another mechanism can cooperate with the link quality estimation mechanism to acquire better knowledge of successful transmission probability of each path. The base station that acts as a network coordinator keeps monitoring the network traffic and keeps tracking of the number of successfully transmitted packets of all paths. Compared with the results of the two mechanisms, the base station can possibly detect node compromise.

Path maintenance phase After multiple paths are established, the multipath routing protocol is responsible for maintaining the availability of multiple paths. Many multipath routing protocols provides mechanisms to detect node failures, such as link-layer ACK, and periodical HELLO message broadcast, etc. If a node failure is detected by its upstream neighbor, the neighbor may send packets to notify the affected source nodes. The notified source nodes may decide whether to re-initiate the path discovery phase according to the remaining number of available paths.

3.3. Packet Transmission

After the path establishment phase is finished, the source node will construct N paths to the base station, and simultaneously obtain the successful transmission probability of the paths. The source node is then able to transmit the data packets using (m, n) FAMIDS. We categorize the operations of packet transmission into three phases, packet generation phase, packet splitting phase and packet transmission phase, and explain separately.

Packet generation phase In order to provide secure communication between the source node and the base station, some security mechanisms are needed before packet splitting. Assume the original information sensed by the source node is M . The source node performs some encryption and authentication operations on M , and the message becomes:

$$D = \{M \parallel \text{MAC}(K_{mac}, M \parallel S)\}_{K_{encr}}$$

In the modified message D , S represents the sequence number shared secretly between the source node and the base station which indicates the order of the transmitted packets. The maintenance and synchronization of the sequence number of the two ends can be achieved by some existing security mechanism, such as SPINS. Without loss of generality, we assume the length of the modified message to be L bits.

Packet splitting phase In FAMIDS, a uniform allocation approach is taken, which means all paths used for transmission are assigned with one subpacket, in order to have the advantage of load balance. Based on the successful transmission probability of available multiple paths, the source node decides the optimal (m, n) FAMIDS with the highest communication reliability, where $1 \leq n \leq N, 1 \leq m \leq n$. Compared with the brute-force approach that examining all the (m, n) pairs taken by other schemes, in FAMIDS the base station only needs to consider the communication reliability of the reduced candidate set of (m, n) pairs. In the analysis section, an efficient algorithm for determining the candidate set of (m, n) pairs is proposed.

After determining the optimal (m, n) pair, the source node splits the modified message D of L bits into n sub-packet $D_i, 1 \leq i \leq n$, each of length L/m . In order to solve the sharing problem of the secret matrix A between the source node and the base station, we attach the row A_i of matrix A in front of the sub-packet D_i , where $1 \leq i \leq n$, rather than negotiating the matrix A in advance. Algorithm 3-1 shows the basic packet splitting algorithm, where the $M_i, 1 \leq i \leq n$, computed from the last step are the final sub-packets the source node transmits to the base station. The total computational overhead of the basic algorithm is $O(m^2)$.

There are two important observations of FAMIDS in wireless sensor networks. First, the original data packets are of equal size. Second, the maximum number of available paths, i.e. N , is determined before deployment. Benefit from the observations, the base station preloads each sensor node with sufficient pre-computed sets of rows of matrix A with a sufficient large prime p , and each transmission the source node forms the matrix A by the first m elements of randomly choosing n rows from the pre-computed set. Earlier research has discussed on the appropriate multipath level for wireless sensor networks and indicated that excessive multipath level does not benefit the data transmission because of the maintenance overhead. As a result, the degree of information dispersal, i.e. n , should be relative small in wireless sensor networks; it is efficient to trade storage for computation. With the modified packet splitting algorithm 3-2, the source node only needs to attach the first element r_i in front of the sub-packet D_i to notify the base station of secret matrix A , and thus saving communication bandwidth.

Algorithm 3-1: Basic_Paket_Splitting (D, m, n)

1. Choose a prime p that $p > 2^m - 1$
2. Randomly choose n different elements $\in Z_p$, denoted by $r_1, r_2, \dots, r_n$
3. Computes the rows of matrix A $a_i = [1, r_1, r_1^2, \dots, r_1^{m-1}]$, $1 \leq i \leq n$
4. Splits D into L/m pieces, each piece of length m . (if L is not a multiple of m , pad 0 to the last piece), denoted by $S_1, S_2, \dots, S_{L/m}$.

5. Computes the sub-packets D_i , $1 \leq i \leq n$ by $D_i = \left[a_i S_1, a_i S_2, \dots, a_i S_{L/m} \right]$
6. Append a_i to D_i , $M_i = a_i \parallel D_i$, $1 \leq i \leq n$.

Algorithm 3-2: Modified_Paket_Splitting (D, m, n)

1. Randomly choose n rows from the pre-computed set of matrix A , each row takes the first m elements, i.e. $a_i = [1, r_1, r_1^2, \dots, r_1^{m-1}]$, $1 \leq i \leq n$
2. Splits D into L/m pieces, each piece of length m . (if L is not a multiple of m , pad 0 to the last piece), denoted by $S_1, S_2, \dots, S_{L/m}$.
3. Computes the sub-packets D_i , $1 \leq i \leq n$ by $D_i = \left[a_i S_1, a_i S_2, \dots, a_i S_{L/m} \right]$
4. Append r_i to D_i , $M_i = r_i \parallel D_i$, $1 \leq i \leq n$.

Packet transmission phase The intermediate nodes all keep a specific sequence number of its upstream and downstream neighbors to indicate the number of packets that are successfully received or transmitted by them. For example, $S_{u,v}$ represents the number of successful transmitted packets from node u to node v . Like the sequence number S , the value of each sequence number between two neighbors is maintained by SPINS. After packet splitting phase is finished, the source node transmits the sub-packets M_i 's through the multiple disjoint paths to the base station. The following equations show the transmission process of the sub-packets:

$$\text{Source Node} \rightarrow \text{Node}_{i,1}: M_i \parallel \left\{ S_{\text{source}, \text{node}_{i,1}} \right\}_{K_{\text{source}, \text{node}_{i,1}}}, \quad 1 \leq i \leq N$$

$$\text{Node}_{i,j} \rightarrow \text{Node}_{i,j+1}: M_i \parallel \left\{ S_{\text{node}_{i,j}, \text{node}_{i,j+1}} \right\}_{K_{\text{node}_{i,j}, \text{node}_{i,j+1}}}, \quad 1 \leq j \leq \text{Hop}_i - 1$$

The first equation shows the transmission operations of the source node, while the second one represents the transmission operations of the intermediate nodes. Each intermediate node first verifies the received sub-packet by the encrypted sequence number in the rear of the packet. If a valid sub-packet is recognized, the intermediate node replaces the second part of the sub-packet and re-transmits the sub-packet to the downstream neighbor. The whole transmission process continues until the base station, i.e. $\text{Node}_{i, \text{Hop}_i}$, receives the sub-packet. If enough sub-packets are correctly received by the base station, the base station can reconstruct the original information based on the secret matrix A and the encryption and authentication keys shared with the source

node. Explicit acknowledgment and retransmissions can also be incorporated to enhance the system reliability if needed.

In the following section, we analyze the communication reliability and information security of FAMIDS and some efficient algorithm are proposed to determine the optimal set of (m, n) with highest communication reliability under different network conditions.

4. Analysis

In this section, we analyze the performance of FAMIDS for two aspects: communication reliability and information security. Regarding reliability, we begin with some preliminary knowledge. The communication reliability of FAMIDS, $R(m, n)$ is then derived as an analytical form. In the next sections, we analyze the behavior analytical form under different network conditions and propose separately efficient ways for finding the optimal set of (m, n) with highest communication reliability. In addition, as far as security is concerned, we analyze the security level achieved by FAMIDS for the following concerns: information confidentiality, message integrity and attack resistance. We detail the countermeasures of FAMIDS to resist above-mentioned security problems.

4.1. Reliability Analysis

In this section, we analyze the optimal communication reliability that FAMIDS achieves under four different network conditions: simple case, ideal case, general case and special case. The following analysis is based on the hardware specification of current generation sensor nodes. In addition, since the multipath level of a general sensor network should be relative small, a maximum multipath level of 10 is discussed in the following analysis. The proof of the theorems and corollaries derived in this section is provided in the appendix.

4.1.1. Preliminary

We would like to start our reliability analysis with some preliminary knowledge about the characteristics of wireless channel and the influence of packet splitting.

- **Characteristics of wireless channel**

Table 2-4 summarizes the hardware specification of the radio frequency modules embedded in the Berkeley MICA series nodes [15]. With different signal modulation techniques, the relationship between signal-to-noise ratio (SNR) and bit error rate (BER) varies, as shown in Figure 2-2. In the equation of BER, E_b is defined as the ratio of energy per bit and N_0 represents the spectral noise density. It is easy to find that the increase in SNR will cause a decrease in BER. We can also observe that, under the same sensing environment, binary phase shift keying (BPSK) has better performance.

For a wireless link between two sensor nodes like Figure 2-3(a), the link success probability that a packet of length L bits can successful transmit from one node to the other is

$$P_l = (1 - P_b)^L \quad (1)$$

In equation (1), P_b represents the BER of the wireless link. Figure 2-3(b) gives us an example to illustrate the relationship between BER and P_l . In this example, we assume a packet of 28 bytes is transmitted through a BPSK channel between two nodes. We can observe from the graph that the link success probability drops drastically to 0.1 when BER approaches to 0.01.

Sensor Node	RF Chip	Signal Modulation Techniques	Bit error rate
MICA	TR1000	Amplitude Shift Keying (ASK)	$P_b = Q(\sqrt{E_b / N_0})$
MICA2	CC1000	Frequency Shift Keying (FSK)	$P_b = Q(\sqrt{E_b / N_0})$
MICAZ	CC2420	Binary Phase Shift Keying (BPSK)	$P_b = Q(\sqrt{2E_b / N_0})$

Table 2-4 The Spec. of MICA series sensor nodes

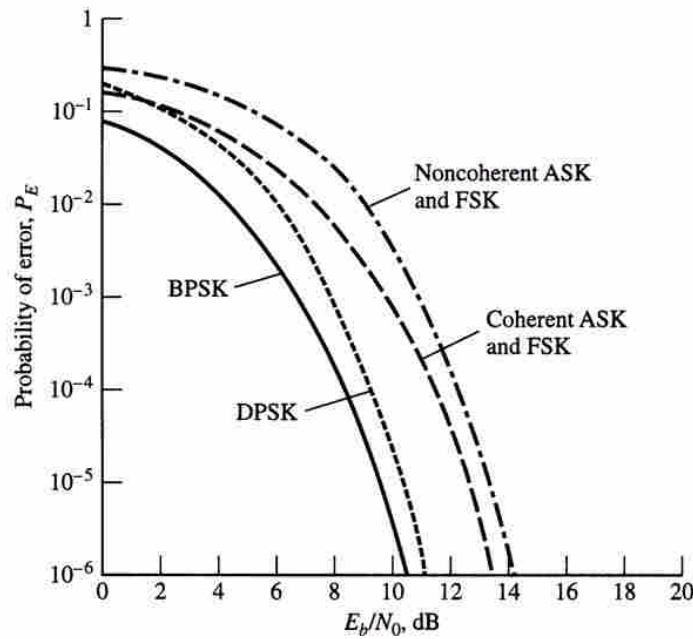


Figure 2-2 Relationship between BER and SNR

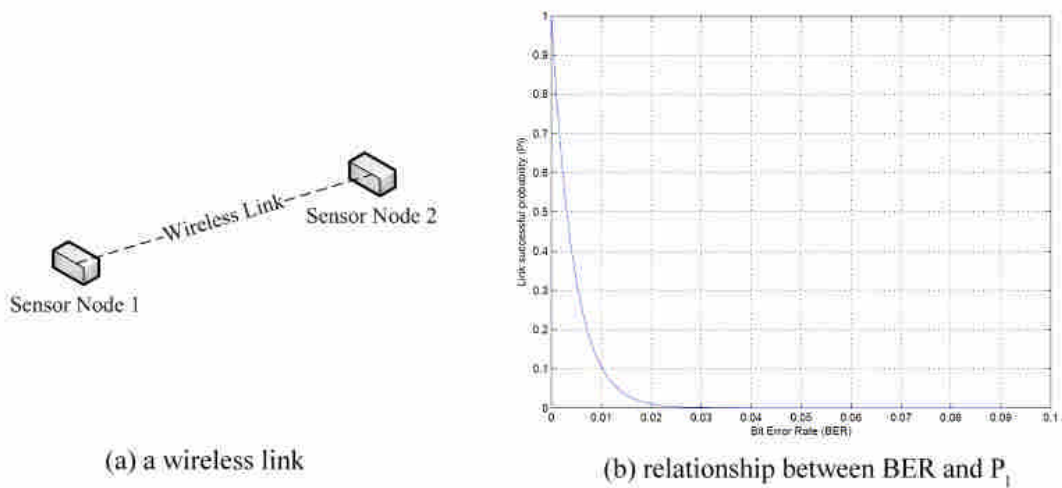


Figure 2-3 An example of link success probability

- The influence of packet splitting on link success probability

Based on the results derived in the previous section, here we would like to explain the influence of packet splitting on link success probability. The link success probability of transmitting the original packet of length L bits is

$$P_l = (1 - P_b)^L \quad (2)$$

If we split the original packet into several sub-packets, each of size $L/2$ bits, the link success probability becomes

$$P_l' = (1 - P_b)^{\frac{L}{2}} = \sqrt{P_l} \quad (3)$$

Similarly, if we split the original packet into several sub-packets, each of size L/m bits, $m \geq 1$, the link success probability afterwards turns into

$$P_l' = (1 - P_b)^{L/m} = \sqrt[m]{P_l} \quad (4)$$

Because of $0 < P_l < 1$ and $m \geq 1$, the link success probability after packet transmitting is greater or equal to the original link success probability. Figure 2-4 shows the link success probability with or without packet splitting approach. Each line in the graph represents the link success probability of different packet size. Two conclusions we can draw from Figure 2-4. First, the link success probability is greatly improved when the original P_l is small. In other words, if the BER of the wireless link is bigger, we can achieve larger link success probability gain resulted from packet splitting. Second, the probability gain from packet length L to $L/2$ is the biggest. The two conclusions will dominate the results of our reliability analysis.

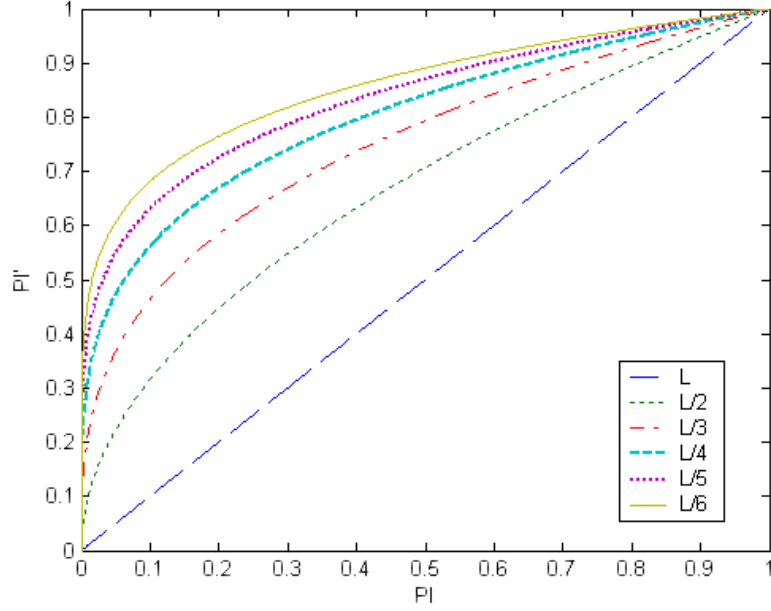


Figure 2-4 The effect of packet splitting

- **The path success probability**

To facilitate the following analysis, we would like to explain the formation of path success probability first. For a path of H hops, there are $H-1$ intermediate nodes and H links between nodes. Figure 2-5 plots a path of 4 hops from a source node to the base station. The path success probability is composed of the success probabilities of the links and the intermediate nodes. It is noted that the path success probability is influenced by the packet length. Assume $P_{l,i}$ and $P_{n,i}$ represent the success probability of the i_{th} link and the i_{th} intermediate node when transmitting a L -bit packet respectively, the path success probability is

$$P_s = \prod_{i=1}^H P_{l,i} \times \prod_{i=1}^{H-1} P_{n,i} = \prod_{i=1}^H (1 - P_{b,i})^L \times \prod_{i=1}^{H-1} P_{n,i} = P_L \times P_N \quad (5)$$

For simplicity, we use P_L and P_N to represent the cumulative link success probability and cumulative node success probability. We will use the definition in the following analysis interchangeably. When (m, n) FAMIDS is applied, the original L -bit packet is fragmented into n sub-packets of length L/m bits. As a result, the path success probability when transmitting a sub-packet becomes

$$P_s' = \prod_{i=1}^H P_{l,i}' \times \prod_{i=1}^{H-1} P_{n,i} = \prod_{i=1}^H (1 - P_{b,i})^{\frac{L}{m}} \times \prod_{i=1}^{H-1} P_{n,i} = \sqrt[m]{P_L} \times P_N \quad (6)$$

Compared with the two equations, we have found that the success probability of intermediate nodes is independent of packet size. The reason is that node success probability is influenced by energy depletion of sensor nodes, while the BER of sensing environment dominates the link success probability. When performing analysis, we usually ignore the problem of energy depletion and assume the involved nodes have efficient energy for packet transmission. As a result, we can simplify the equations as:

$$P_s = \prod_{i=1}^H P_{l,i} = \prod_{i=1}^H (1 - P_{b,i})^L \quad (7)$$

$$P_s' = \prod_{i=1}^H P_{l,i}' = \prod_{i=1}^H (1 - P_{b,i})^{\frac{L}{m}} = \sqrt[m]{P_s} \quad (8)$$

In the following section, we will derive the analytical form of the communication reliability of (m, n) FAMIDS, i.e. $R(m, n)$, based on the above two equations.

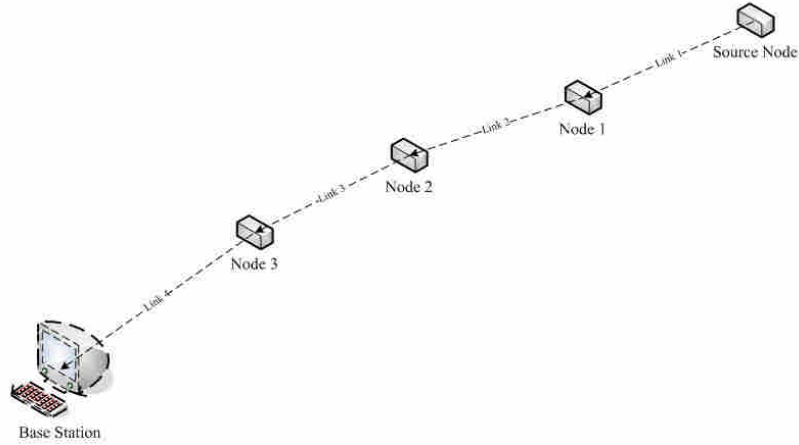


Figure 2-5 A multi-hop path

4.1.2. The Communication Reliability $R(m, n)$

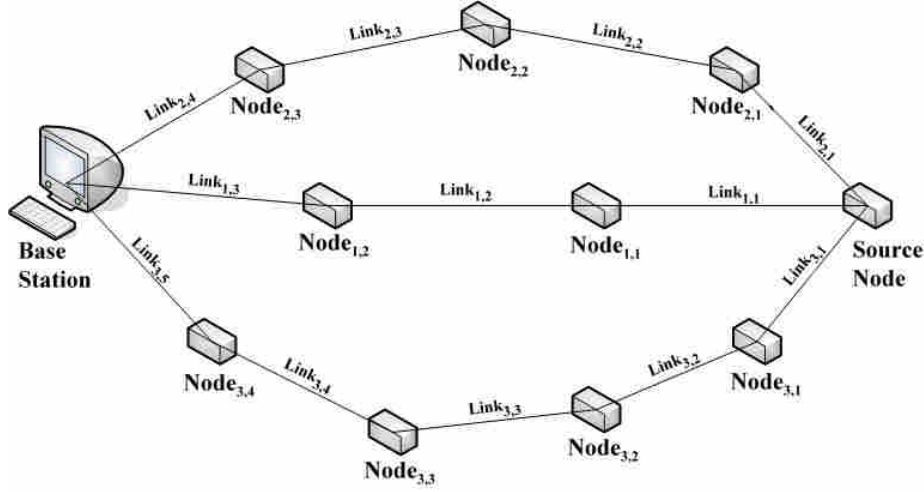


Figure 2-6 The wireless sensor network model

In our network model, we assume that N disjoint paths are available for data packet transmission from a source node to the base station. Figure 2-6 plots an example of $N=3$, where the $\text{Node}_{i,j}$ and $\text{Link}_{i,j}$ means the j th intermediate node and the j th link of path i , respectively. If the source node decides to transmit the original data packet of length L bits by (m, n) FAMIDS, the packet is fragmented into n sub-packets of length L/m , where $1 \leq n \leq N$, $1 \leq m \leq n$. It is noted that a decision to use n paths means that these paths will be the first n ones with the highest success probability, in order to achieve the optimal reliability. The success probabilities of the n paths when transmitting the original packet of length L bits are organized in the probability vector $\underline{P}_s = [P_{s,i}]$, $1 \leq i \leq n$. For easy to explain, we order the paths from the best one to the worst one, i.e. $P_{s,i} \geq P_{s,i+1}$. When (m, n) FAMIDS is applied, the size of sub-packet reduces to L/m and the probability vector turns into $\underline{P}'_s = [\sqrt[m]{P_{s,i}}]$, $1 \leq i \leq n$. Given the fact that uniform allocation is adopted, we define a state vector $\underline{s} = [s_i]$, $1 \leq i \leq n$ to indicate the number of sub-packets that actually reach the base station through the n paths. There are only two possible values for s_i , i.e. 0 and 1. If $s_i=0$, the number of successful transmitted sub-packets through path i is 0. On the other hand, $s_i=1$ indicates the assigned sub-packet of path i actually reaches the base station. The associated probabilities are:

$$P_r\{s_i = 1\} = \sqrt[m]{P_{s,i}} \quad (10)$$

$$\Pr\{s_i = 0\} = 1 - \sqrt[m]{P_{s,i}} \quad (11)$$

The communication reliability of (m, n) FAMIDS, which denotes as $R(m, n)$, is that the base station can reconstruct the original data packet by the successfully received sub-packets. Since the base station can rebuild the original data packet only if more than m correctly received sub-packets, we can define $R(m, n)$ as

$$R(m, n) = P_r \left\{ \sum_{i=1}^n s_i \geq m \right\} \quad (12)$$

- **Formula of $R(m, n)$**

It is easy to understand that there are total 2^n possible states of $\underline{s}$. The probability that the n paths are in state $\underline{s}$ can be calculated as

$$p(\underline{s}) = \prod_{i=1}^n \left(\sqrt[m]{P_{s,i}} \right)^{s_i} \times \left(1 - \sqrt[m]{P_{s,i}} \right)^{1-s_i} \quad (13)$$

We classified all possible states of $\underline{s}$ into $n+1$ classes according to the number of successfully transmitted paths, which denotes as S_k .

$$S_k = \left\{ \underline{s} : \sum_{i=1}^n s_i = k \right\}, \quad 0 \leq k \leq n \quad (14)$$

The size of class S_k is C_k^n . Given the classification of all possible states, we can rewrite the communication reliability $R(m, n)$ as

$$R(m, n) = \sum_{k=m}^n \sum_{\underline{s} \in S_k} p(\underline{s}) \quad (15)$$

Based on the formula of $R(m, n)$, we analyze the behavior of $R(m, n)$ under four different cases of network in the following sections. For each network condition, we derive an efficient way or algorithm for determining the optimal (m, n) with highest communication reliability.

4.1.3. The Simple Model

First of all, the simple model to be analyzed is illustrated as Figure 2-7. N disjoint paths are

constructed from source node to base station for packet transmission. In the simple model, we assume that the influence of packet splitting on path success probability is so trivial that we can ignore the probability gain. That is,

$$\underline{P_s'} = [\sqrt[m]{P_{s,i}}] \cong [P_{s,i}] = \underline{P_s} \quad (16)$$

This is the case that the BER of the sensing environment is relative small and thus the probability gain of link success probability achieved by FAMIDS can be ignored. In other word, the path success probability is independent of packet length.

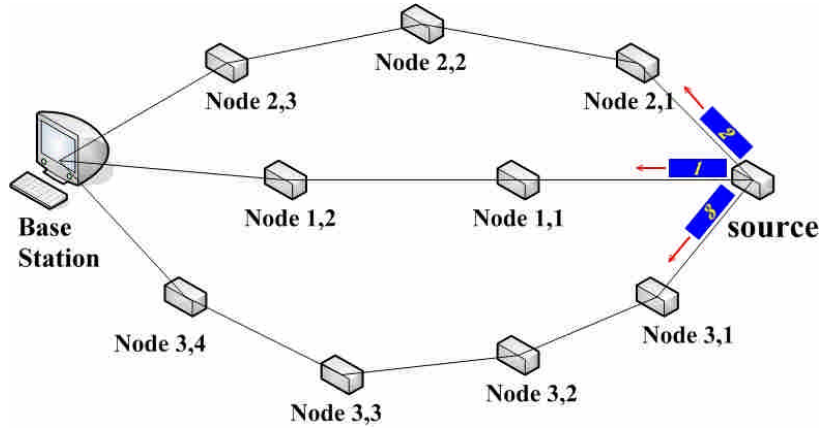


Figure 2-7 The simple model

Assume that the source node uses (m, n) FAMIDS for packet transmission, $1 \leq n \leq N, 1 \leq m \leq n$. Based on the definition of the simple model, the probability that n paths are in state $\underline{s}$ turns into

$$p(\underline{s}) = \prod_{i=1}^n (P_{s,i})^{s_i} \times (1 - P_{s,i})^{1-s_i} \quad (17)$$

As a result, the communication reliability of the simple model can be derived as

$$R(m, n) = \sum_{k=m}^n \sum_{\underline{s} \in S_k} \prod_{i=1}^n (P_{s,i})^{s_i} \times (1 - P_{s,i})^{1-s_i} \quad (18)$$

Theorem 2-1 to Theorem 2-3 discuss the relationship between $R(m, n)$'s and help us find the optimal set of (m, n) with the optimal communication reliability.

Theorem 2-1: $R(m, m) > R(m+1, m+1)$, for $1 \leq m \leq N-1$

Theorem 2-1 concludes that among all FAMIDS that information expansion ratio is 1, the

communication reliability of (1, 1) FAMIDS is maximum.

Theorem 2-2: $R(m, n) < R(m, n+1)$, for $1 \leq n \leq N-1, 1 \leq m \leq n$

For example, $R(2,3) - R(2,2) = P_{s,3} \times [P_{s,1} \times (1 - P_{s,2}) + (1 - P_{s,1}) \times P_{s,2}]$, and

$R(3,4) - R(3,3) = P_{s,4} \times [P_{s,1} \times P_{s,2} \times (1 - P_{s,3}) + P_{s,1} \times (1 - P_{s,2}) \times P_{s,3} + (1 - P_{s,1}) \times P_{s,2} \times P_{s,3}]$.We

observe that every term in the examples is positive, so the difference of communication reliability is also positive. Theorem 2-2 suggests that the communication reliability can be improved if more subpackets are sent and the same number of subpackets is needed for rebuild the original packet.

Theorem 2-3: $R(m, n) > R(m+1, n)$, for $1 \leq m \leq n-1$

Theorem 2-3 tells us that if the same number of paths is used, the communication reliability increases as fewer subpackets are needed for rebuilding the original information.

Although the previous three theorems do not discuss all the relationship between all possible pairs of (m, n), the results is sufficient for us to find the optimal (m, n) FAMIDS in the simple network model. For example, if there are total 5 disjoint paths available for the source node, 15 combinations of (m, n) MIDS can be selected. We summarize the results derived from the three theorems and thus we can get the optimal solution:

- | | |
|----|---|
| 1) | $R(1, 1) > R(2, 2) > R(3, 3) > R(4, 4) > R(5, 5)$ |
| 2) | $R(1,5) > R(1,4) > R(1,3) > R(1,2) > R(1,1)$ |
| 3) | $R(2,5) > R(2,4) > R(2,3) > R(2,2)$ |
| 4) | $R(3,5) > R(3,4) > R(3,3)$ |
| 5) | $R(4,5) > R(4,4)$ |
| 6) | $R(1,5) > R(2,5) > R(3,5) > R(4,5) > R(5,5)$ |
| 7) | $R(1,4) > R(2,4) > R(3,4) > R(4,4)$ |
| 8) | $R(1,3) > R(2,3) > R(3,3)$ |
| 9) | $R(1,2) > R(2,2)$ |

From the results, we can conclude that (1,5) IDS is the optimal solution when there are totally 5 paths available. Corollary 1 is thus established.

Corollary 1: If the path success probability is independent of packet size , (1, N)

FAMIDS is always the optimal solution.

In this simple network case, since the link failure rate is negligible, the probability gain achieved by FAMIDS is limited. In other words, packet splitting does not benefit the communication reliability. As a result, splitting the original data packet with the maximum information expansion ratio, i.e. (1, N) FAMIDS, always achieves the optimal communication reliability. Algorithm 4-1 summarizes the analysis results of this section and shows how to find the optimal m value under simple network model. In the next section, we consider an idealized network case that the influence of packet splitting is taken into consideration.

Algorithm 4-1: Finding Optimal m for (m, N) FAMIDS under Simple Model

Input: N: number of total available paths

Output: m: the optimal value of m, such that (m, N) FAMIDS achieves the optimal communication reliability.

1. The optimal m value is 1.
2. If the optimal m cannot satisfy the information expansion ratio requirement, then choose m=2, m=3, ..., m=N respectively, until the information expansion ratio is satisfied.
3. Output m.

4.1.4. The Ideal Model

In the section, we analyze the behavior of $R(m, n)$ under an idealized network. The network topology is the same as the simple model, except that each path has the same success probability where the problem of node failure is still ignored. That is,

$$P_{s,i}=P_s, \quad \forall 1 \leq i \leq n \quad (19)$$

As a result, each state $\underline{s}$ in the same class S_k has the same probability, which can be derived

as

$$p(\underline{s}) = \left(\sqrt[m]{P_s}\right)^k \times \left(1 - \sqrt[m]{P_s}\right)^{n-k} \quad \text{if } \underline{s} \in S_k \quad (20)$$

Therefore, $R(m, n)$ can be expressed as

$$R(m, n) = \sum_{k=m}^n \sum_{\underline{s} \in S_k} p(\underline{s}) = \sum_{k=m}^n C_k^n \left(\sqrt[m]{P_s}\right)^k \times \left(1 - \sqrt[m]{P_s}\right)^{n-k} \quad (21)$$

The behavior of the communication reliability $R(m, n)$ under ideal case is equivalent to the analysis results provided in [10]. We summarize the behavior of $R(m, n)$ as follows:

- 1) When using total available paths, one gets the highest communication reliability.
- 2) If N paths are available for packet transmission, the optimal (m, n) FAMIDS is in the candidate set $\{(i, N) \text{ FAMIDS} \mid 1 \leq i \leq \eta_2\}$. Figure 2-8 gives an example to illustrate the relationships between $R(1, N)$, $R(2, N)$, ..., and $R(\eta_2, N)$. We can observe that a particular FAMIDS does not always give better reliability than another. According to the value of P_s and the critical probability of (i, N) FAMIDS and $(i+1, N)$ FAMIDS, one can determine the (m, N) FAMIDS with highest communication reliability. Table 2-5 summarizes the intersections of $R(m, n)$ for $m=4..10$. Based on Table 2-5, Algorithm 4-2 can determine the optimal (m, N) FAMIDS with optimal communication reliability under ideal network model.
- 3) If the upper bound on information expansion ratio is given, i.e. u , the candidate FAMIDS set

$C_{u, N}$ is reduced to

$$\begin{aligned} & \{(i, i * u) \text{ FAMIDS} \mid 1 \leq i \leq \eta_u\} \cup \{(\eta_u + j, N) \text{ FAMIDS} \mid 1 \leq j \leq N - \eta_u \\ & \text{and } \eta_u + j \leq \eta_2\} \quad \text{when } u \geq 2, \end{aligned} \quad (22)$$

or

$$\{(i, i * u) \text{ FAMIDS} \mid 1 \leq i \leq \eta_u\} \cup (\eta_u + 1, N) \text{ FAMIDS} \quad \text{when } u \leq 2. \quad (23)$$

The η_u and η_2 in (22) and (23) means the value of $\lfloor N/u \rfloor$ and $\lfloor N/2 \rfloor$, respectively. The source node only needs to check the communication reliability of the candidate set to determine the optimal set.

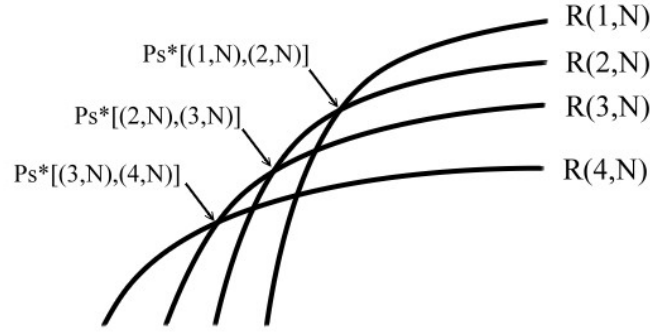


Figure 2-8 Relationships between R(1,N) to R(4,N)

j	$P_s^*[(1,j),(2,j)]$	$P_s^*[(2,j),(3,j)]$	$P_s^*[(3,j),(4,j)]$	$P_s^*[(4,j),(5,j)]$
4	0.271837			
5	0.579696			
6	0.768788	0.022815		
7	0.874048	0.132044		
8	0.931467	0.294322	0.001058	
9	0.962716	0.457744	0.016607	
10	0.979738	0.597068	0.065319	0.000032471

Table 2-5 The intersections: $P_s^*[(i,j),(i+1,j)]$

Algorithm 4-2: Finding Optimal m for (m, N) FAMIDS under Ideal Model

Input: N: number of total available paths

P_s : the path success probability

Output: m: the optimal value of m, such that (m, N) FAMIDS achieves the optimal communication reliability.

1. Look up Table 4-2 to determine the region that P_s belongs to.
2. Search the optimal m value as follows.
 - If $P_s^*[(1, N), (2, N)] < P_s < 1$, then the optimal value of m is 1.
 - If $P_s^*[(i+1, N), (i+2, N)] < P_s < P_s^*[(i, N), (i+1, N)]$, then the optimal value of m is i+1.
 - If $0 < P_s < P_s^*[(\eta_2 - 1, N), (\eta_2, N)]$, then the optimal value of m is η_2 .
3. If the optimal m cannot satisfy the information expansion ratio requirement, then choose $m=m+1$, $m=m+2$..., $m=N$ respectively, until the information expansion ratio is satisfied.
4. Output m.

In the following section, a more general network model will be analyzed where the

restriction on the path success probability under ideal network model is cancelled.

4.1.5. The General Model

Based on the results of the ideal model, we extend our analysis to a more realistic case, which multiple paths have different success probability. Still, a uniform sub-packet allocation is assumed and the problem of node failures is ignored. To facilitate our following analysis, we begin with some more definitions. We define a function $F_k(n)$ to represent the sum of probabilities that all states $\underline{s}$ that belongs to S_k when n paths are being used for packet transmission

$$F_k(n) = \sum_{\underline{s} \in S_k} p(\underline{s}), \quad 0 \leq k \leq n \quad (24)$$

In addition, we define the sums of function F from class k to class n

$$F_k'(n) = \sum_{i=k}^n F_i(n) = \sum_{j=k}^n \sqrt[m]{P_{s,j}} \cdot F_{k-1}(j-1) \quad (25)$$

The equivalence of the two sums of equation (19) is proved in [13]. As a result, the expression for $R(m, n)$ found in (15) can be rewritten as

$$R(m, n) = F_m'(n) \quad (26)$$

The following theorem can explain the relationships between $R(m, n)$ with a fixed value of m for $1 \leq n \leq N$.

Theorem 2-4: $R(m, n) < R(m, n+1)$, for $1 \leq m \leq n$

Based on Theorem 2-4, we can conclude that one can achieve the highest communication reliability when all available paths are used. If total N paths are available for transmission, the candidate FAMIDS set can be reduced to $\{(m, N) \text{ FAMIDS} \mid 1 \leq m \leq N\}$. This conclusion is the same as the ideal case. However, because of the variety of path success probabilities, the influence of packet splitting for each path is different. As a result, even if the number of available paths and the average success probability of the paths are the same, the optimal set of (m, N) with

highest communication reliability may be different. For example, assumed that 6 disjoint paths are available for transmission and average success probability is 0.75. If the probability vector $\underline{P_s}$ is [0.8, 0.8, 0.75, 0.75, 0.75, 0.65], (2, 6) FAMIDS achieves the highest communication reliability. On the other hand, (1, 6) FAMIDS becomes the optimal solution when $\underline{P_s}$ is [0.85, 0.8, 0.8, 0.7, 0.7, 0.65]. The variety of the optimal solutions results from the variety of path success probabilities.

However, resorting to simulation, we have observed that there are still some combinations between the results of the ideal model and the general model. First, the candidate FAMIDS set of both network models ranges from (1, N) FAMIDS to (η_2, N) FAMIDS, where $\eta_2 = \lfloor N/2 \rfloor$. In other words, (η_2+1, N) FAMIDS to (N, N) FAMIDS can not be the optimal solution under both ideal model and general model. Second, according to the average success probability of paths $\overline{P_s}$ and the critical probabilities of (m, n) FAMIDS found in ideal case, one can determine a more exact candidate FAMIDS set, which denotes as C_N :

- 1) If $P_s^*[(1, N), (2, N)] < \overline{P_s} < 1$, $C_N = \{(m, N) \text{ FAMIDS} \mid m=1\}$.
- 2) If $P_s^*[(i+1, N), (i+2, N)] < \overline{P_s} < P_s^*[(i, N), (i+1, N)]$, $C_N = \{(m, N) \text{ FAMIDS} \mid m=1..i+1\}$.
- 3) If $0 < \overline{P_s} < P_s^*[(\eta_2-1, N), (\eta_2, N)]$, $C_N = \{(m, N) \text{ FAMIDS} \mid m=1..\eta_2\}$.

Compared with the brute-force approach taken by other schemes that examining all the possible pairs of (m, n), where $1 \leq n \leq N, 1 \leq m \leq n$, in FAMIDS the source nodes only need to check the (m, n) pairs in the reduced candidate set C_N . Algorithm 4-3 shows how to find the optimal value of m. The algorithm reduces the complexity of finding the candidate set of (m, n) from $O(N^2)$ to $O(N)$.

Algorithm 4-3: Finding Optimal m for (m, N) FAMIDS under General Model

Input N : the number of available paths

$\underline{P}_s$: the probability vector

Output: m : the optimal value of m , such that (m, N) FAMIDS achieves the optimal communication reliability.

1. Calculate the average success probability $\overline{P}_s = \sum_{i=1}^N P_{s,i} / N$.
2. Look up Table 4-2 to determine the region that $\overline{P}_s$ belongs to.
3. Determine the reduced candidate FAMIDS set C_N as follows.
 - If $P_s^*[(1, N), (2, N)] < \overline{P}_s < 1$, $C_N = \{(m, N) \text{ FAMIDS} \mid m=1\}$.
 - If $P_s^*[(i+1, N), (i+2, N)] < \overline{P}_s < P_s^*[(i, N), (i+1, N)]$, $C_N = \{(m, N) \text{ FAMIDS} \mid m=1..i+1\}$.
 - If $0 < \overline{P}_s < P_s^*[(\eta_2 - 1, N), (\eta_2, N)]$, $C_N = \{(m, N) \text{ FAMIDS} \mid m=1..\eta_2\}$.
4. Calculate the communication reliability of (m, N) FAMIDS in C_N . Assume that (m', N) FAMIDS achieves the optimal reliability.
5. If the optimal $m=m'$ cannot satisfy the information expansion ratio requirement, then choose $m=m'+1$, $m=m'+2$, ..., $m=N$ respectively, until the information expansion ratio is satisfied.
6. Output m .

4.1.6. The Special Model

The previous analysis of the general network model is based on the assumption that node failures can be ignored. However, in a realistic network, the situation is usually more complicated with the existence of node compromise. In order to realize the effect of node failures, we illustrate a special example of network environment to explain the influence of node failure over $R(m, n)$. Figure 2-9 shows the special network model, where N disjoint paths with identical number of hops and success probability are considered. All wireless links and intermediate nodes have the same success probability respectively, which are denoted as P_l and P_n . That is, the influence ratio of the two factors is fixed under special network model.

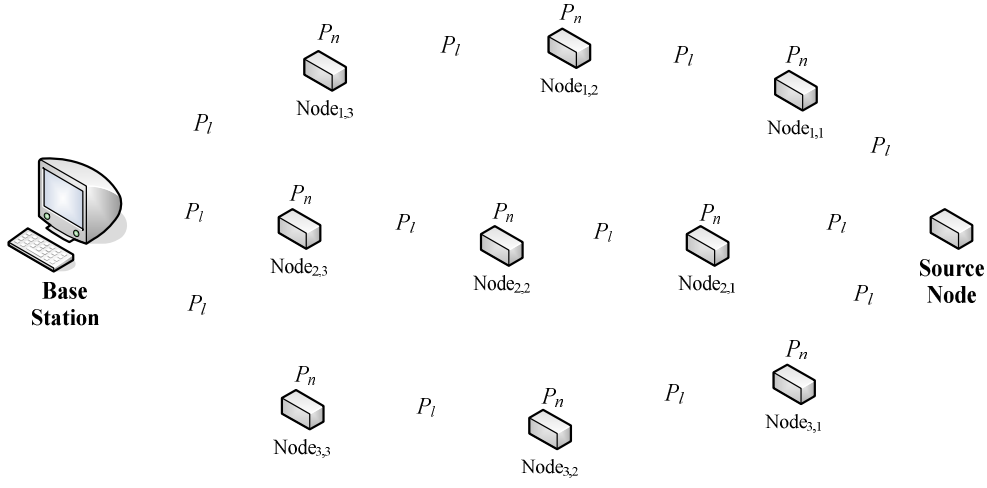


Figure 2-9 The special network case

Based on the assumptions of the network environment, the original path success probability is the same as equation (5). When (m, n) FAMIDS is applied, the path success probability of transmitting a sub-packet of L/m bits can be expressed as equation (6). Using (m, n) FAMIDS, the base station can reconstruct the original information if more than m sub-packets are correctly received. As a result, the communication reliability $R(m, n)$ can be modified as equation (27):

$$R(m, n) = \sum_{i=m}^n C_i^n \cdot \left(\left(\sqrt[m]{P_l} \right)^H \cdot P_n^{H-1} \right)^i \cdot \left(1 - \left(\sqrt[m]{P_l} \right)^H \cdot P_n^{H-1} \right)^{n-i} \quad (27)$$

On account of the definition of cumulative node and link success probability, equation (27) can be rewritten as

$$R(m, n) = \sum_{i=m}^n C_i^n \cdot \left(\sqrt[m]{P_L} \cdot P_N \right)^i \cdot \left(1 - \sqrt[m]{P_L} \cdot P_N \right)^{n-i} \quad (28)$$

Figure 2-10 illustrates an example of the relationship between the original P_s and the communication reliability after $(m, 6)$ FAMIDS is applied. In this example, the number of hops is 3 and the node success probability P_n is 0.9. As shown in Figure 2-10, we can observe that the communication reliability is greatly improved when the original path success probability is relative small because of the increase in the link success probability. In order to determine the optimal (m, n) set with the highest communication reliability under different P_s , some fundamental theorems regarding the relationship between $R(m, n)$'s are derived.

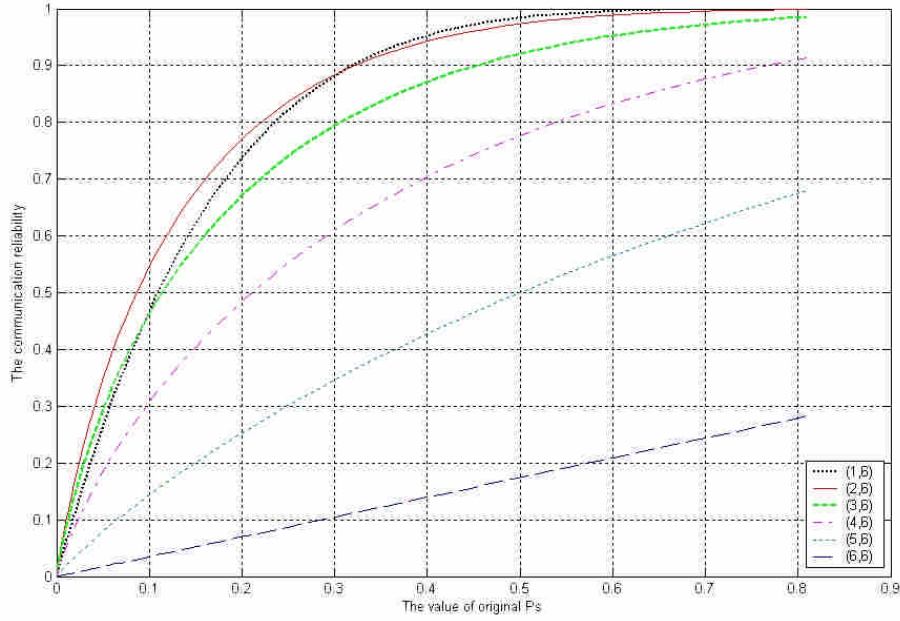


Figure 2-10 Relationship between P_s and $(m,6)$ MIDS

• Fundamental Theorems

Here are some fundamental theorems that the results help us to determine the optimal value of (m, n) set with highest communication reliability.

Theorem 2-5: $R(m, n)$ is strictly increasing for $0 < P_s < 1$.

Theorem 2-5 implies that the communication reliability achieved by a specific (m, n) FAMIDS is proportional to P_s . When the successful transmission probability of each path increases, the communication reliability is also improved.

Theorem 2-6: if $0 < P_s < 1$, then $R(m, n) < R(m, n+1)$.

Theorem 2-6 suggests the communication reliability can be improved if more pieces of sub-packets are sent and the same number of sub-packets is needed to reconstruct the original data packet. Figure 2-11 shows the distribution of the reliability curves of $(1, n)$ MIDS, where $1 \leq n \leq 6$, and each path has 3 hops with $P_n=0.9$. The results implied in Figure 2-11 are

consistent with Theorem 2-6.

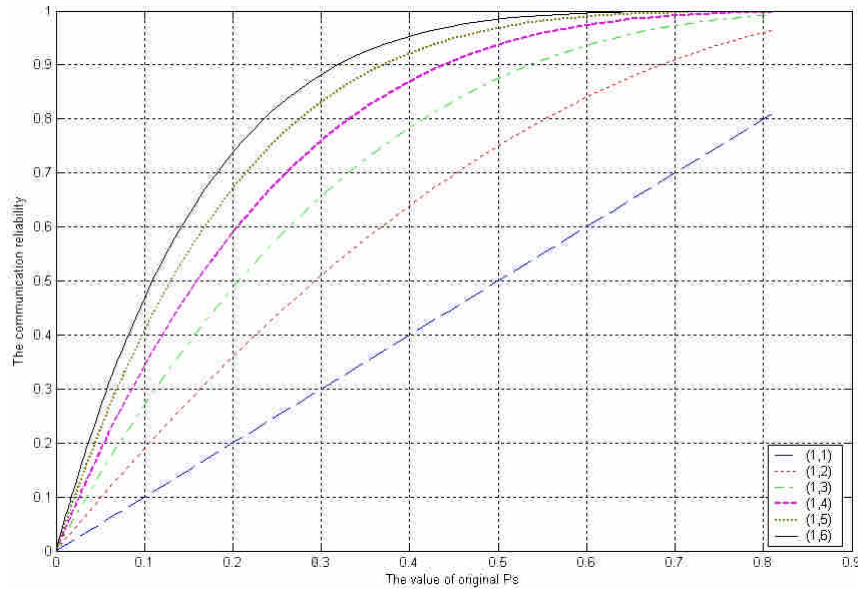


Figure 2-11 Relationship between the P_s and $(1, n)$ MIDS

Theorem 2-7: $R(m, n) > R(m+1, n)$ when $P_s \rightarrow 1^-$

From Theorem 2-7, we can observe that the fewer the number of sub-packets needed to reconstruct the original data packet, the higher the communication reliability when the message dispersal degree n is fixed and P_s is approaching to 1. The only likelihood that P_s is approaching to 1 is both P_1 and P_n are approaching to 1. As a result, we can infer Corollary 2 from Theorem 2-7.

Corollary 2: $R(1, n)$ is the optimal of $R(i, n)$ when $P_s \rightarrow 1^-$, for $1 \leq i \leq n$.

Under the condition that P_s is approaching to 1, it means that both the node and link failure rate are relative small and thus the effect of packet splitting is so trivial that can be neglected. As a consequence, the maximum information expansion ratio achieves the optimal reliability under such network condition. The result is also consistent with the simple model. Compared with the case when $P_s \rightarrow 1^-$, the case when $P_s \rightarrow 0^+$ is more complicated. Theorem 2-8 & Theorem 2-9 give the explanation of such case.

Theorem 2-8: $R(\eta, n)$ is the optimal of $R(i, n)$ when $P_1 \rightarrow 0+$, for $1 \leq i \leq n$, where

$$\eta = \left\lceil \frac{P_N \cdot n - 1}{P_N + 1} \right\rceil$$

The η value derived by Theorem 2-8 is a critical number of (m, n) FAMIDS. When link success probability is approaching to zero, i.e. SNR is extremely small, the optimal (m, n) set is mainly determined by node success probability. From Theorem 2-8, we can guarantee the correctness of Corollary 3 & Corollary 4.

Corollary 3: $R(1, n)$ is the optimal of $R(i, n)$ when $P_1 \rightarrow 0+$ and $P_n \rightarrow 0+$, for $1 \leq i \leq n$.

Corollary 3 conforms to our intuition that the source node should maximize the information expansion ratio when the network is in an exceptionally poor condition. This case happens when the sensor network is deployed in an extremely noisy environment and full of compromised or out-of-battery nodes.

Corollary 4: $R(\eta_2, n)$ is the optimal of $R(i, n)$ when $P_1 \rightarrow 0+$ and $P_n \rightarrow 1-$, for $1 \leq i \leq n$, where

$$\eta_2 = \lfloor n/2 \rfloor.$$

The result of Corollary 4 is similar with the analysis in Tsai's paper [10] where the sender and receiver share n parallel wireless communication channels with no intermediate nodes. Because of the high node success probability, the communication reliability is mainly constrained by the unreliable links.

Figure 2-12 shows the relationship between P_N and the optimal (m, n) FAMIDS when $P_1 \rightarrow 0+$ based on Theorem 2-8, for $n = 1 \sim 10$. It is noted that the only optimal solution is $(1, n)$ when number of paths $n \leq 3$. With the decrease in P_N , the effect of packet splitting also decreases.

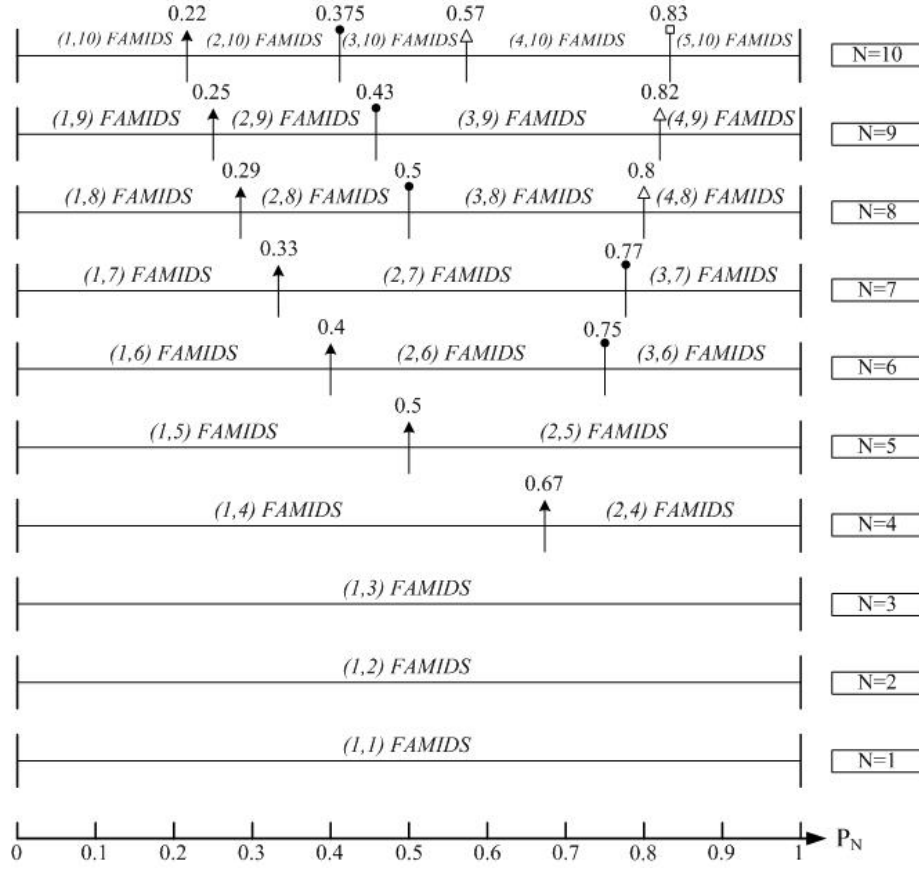


Figure 2-12 Relationship between optimal $R(m, n)$ and P_N

Theorem 2-9: $R(1, n)$ is the optimal of $R(i, n)$ when $P_N \rightarrow 0+$, for $1 \leq i \leq n$.

From Theorem 2-9, we can conclude that in the circumstance of numerous invalid intermediate nodes, the source node should split the original data packet with maximum information expansion ratio. The result of Theorem 2-9 coincides with the result of Theorem 2-8.

Theorem 2-10: If $m \geq \eta$, then $R(m, n) > R(m+1, n)$.

We can conclude from Theorem 2-10 that the curves of $R(\eta, n)$, $R(\eta+1, n)$, ..., $R(n, n)$ do not have intersections. $R(\eta, n)$ would be the optimal solution if $m \geq \eta$. The next theorem discusses the relationship between $R(m, n)$ if $m \leq \eta$.

Theorem 2-11: If $2 \leq m \leq \eta$, then there exists exactly one critical probability P_s^* such

$$\text{that } \begin{cases} R(m, n) > R(m-1, n) & \text{when } 0 < P_s < P_s^* \\ R(m, n) = R(m-1, n) & \text{when } P_s = P_s^* \\ R(m, n) < R(m-1, n) & \text{when } P_s^* < P_s < 1 \end{cases}, \text{ where } \eta = \left\lceil \frac{P_N \cdot n - 1}{P_N + 1} \right\rceil$$

As we can see in Theorem 2-11, in the interval $m \leq \eta$, a particular FAMIDS does not always achieve better communication reliability than another. A FAMIDS can give better reliability in a range of P_s , but worse reliability in the other range. Besides, it is noted that the value of η depends on the cumulative node success probability of each path, P_N . With the decrease in P_N , the value of η also reduces, i.e. fewer pairs of FAMIDS have intersections.

Figure 2-13 shows the relationship between the $R(m, n)$ under different P_N , for $n = 8$, $m = 1 \sim 4$. In this example, each path has three hops as Figure 2-13. Figure 2-13(a) shows the curves of $R(m, n)$ when $P_n=0.9$, i.e. $\eta = 4$. As we can see, the curves of $R(1,8) \sim R(4, 8)$ have intersections. On the other hand, Figure 2-13(b) represents the curves of $R(m, n)$ when $P_n=0.75$, i.e. $\eta = 3$, such that only $R(1,8) \sim R(3, 8)$ have intersections. When P_n decreases to 0.65, i.e. $\eta = 2$, only the first two curves, $R(1,8)$ and $R(2,8)$, have intersections, as shown in Figure 2-13(c). Finally, we can obviously see that the curves of $R(m, n)$ do not have intersections if $\eta = 1$, as shown in Figure 2-13(d).

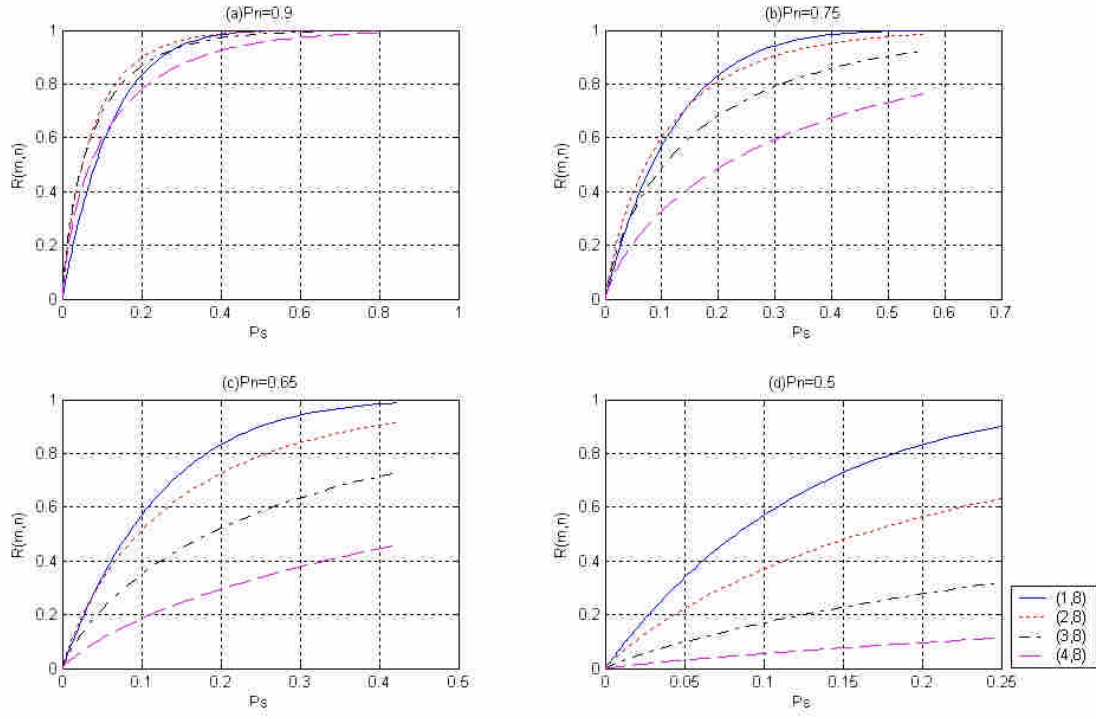


Figure 2-13 $R(m, n)$ under different P_n

Theorem 2-12: For fixed $n \geq 6$ & $P_N > \frac{2}{n-1}$, if $2 \leq m \leq \eta$, then

$$P_s^*[(m-1, n), (m, n)] > P_s^*[(m, n), (m+1, n)], \text{ where } \eta = \left\lceil \frac{P_N \cdot n - 1}{P_N + 1} \right\rceil$$

Theorem 2-12 proves that the intersections of $P_s^*[(m+i-1, n), (m+i, n)]$ have increasing order as i increases for fixed n . The following Figure 2-14 shows the relationship between $P_s^*[i, n] \equiv P_s^*[(i, n), (i+1, n)]$ and P_N for $n = 4 \sim 9$. For example, if P_N is equal to 0.9 and the number of available paths is 9, the following conclusion can be drawn:

- 1) $R(1, 9) > R(2, 9)$ if $P_s > P_s^*[1, 9] = 0.68$
- 2) $R(2, 9) > R(3, 9)$ if $P_s > P_s^*[2, 9] = 0.2$
- 3) $R(3, 9) > R(4, 9)$ if $P_s > P_s^*[3, 9] = 0.0028$

We provide the approximation function of each critical probability for different number of available paths in the appendix I to K. The approximation functions can assist the source node with the determination of the optimal (m, n) FAMIDS with highest communication reliability.

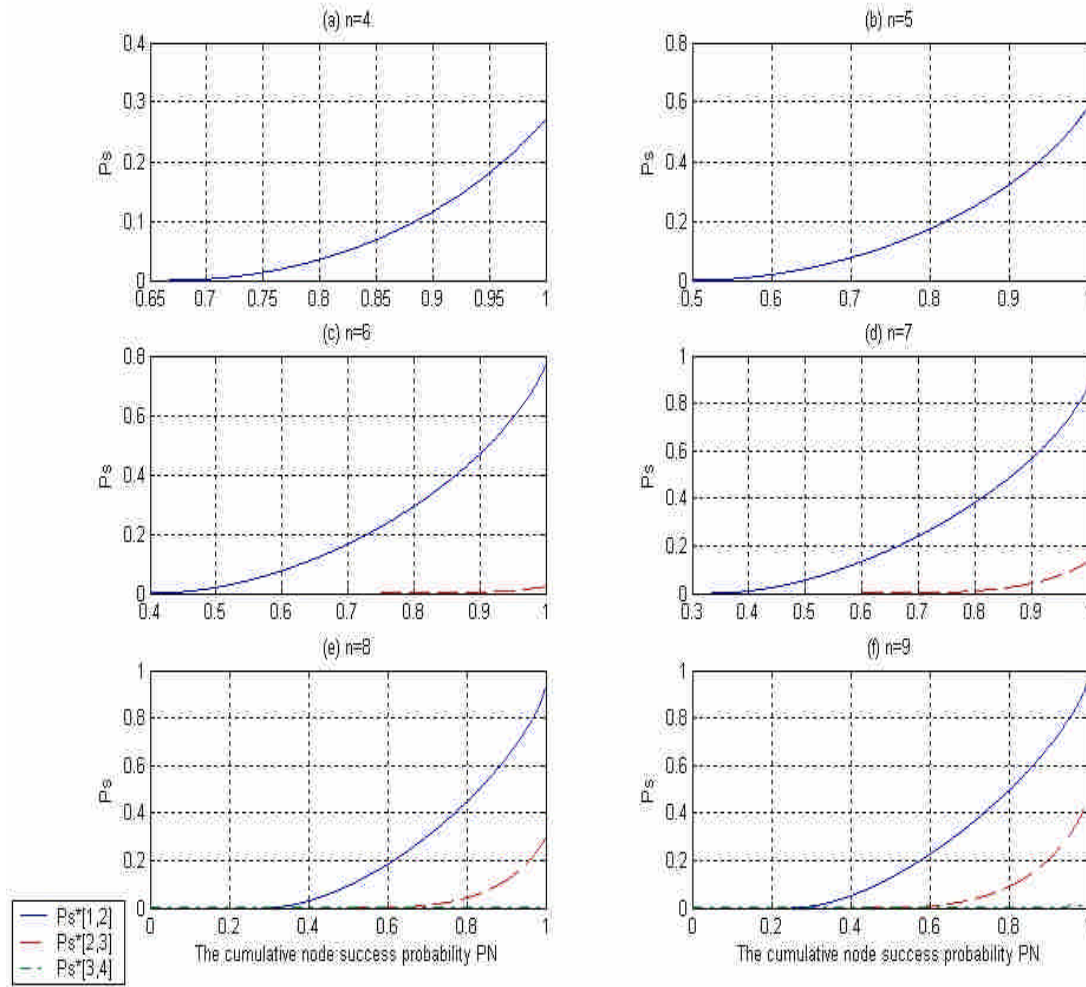


Figure 2-14 Relationship between P_s^* and P_N

• Optimal solution discussion

Based on the fundamental theorems derived in this section, we can determine the optimal set of (m, n) FAMIDS with highest communication reliability under special network condition. First of all, according to Theorem 2-6, we can infer that one achieves the maximum communication reliability when total available paths are used for packet transmission. This conclusion holds for each kind of network. Second, about the relationship between all (m, N) FAMIDS, we have found that any particular FAMIDS does not always achieves better reliability than another. According to the knowledge on P_s and P_N , one can determine the optimal value of m . Algorithm 4-4 finds the optimal value of m under special network model that achieves the optimal reliability.

Algorithm 4-4: Finding Optimal m for (m, N) FAMIDS under Special Model

Input N : the number of available paths

P_s : the path success probability

P_N : the cumulative node success probability

Output: m : the optimal value of m , such that (m, N) FAMIDS achieves the optimal communication reliability.

1. Calculate the critical number $\eta = \left\lceil \frac{P_N \cdot N - 1}{P_N + 1} \right\rceil$.
2. Determine the critical probabilities $P_s^*[(m, N), (m+1, N)]$ for $m=1..\eta-1$ by P_N and the approximation functions.
3. Search the optimal m as follows.
 - If $P_s^*[(1, N), (2, N)] < P_s < 1$, then the optimal value of m is 1.
 - If $P_s^*[(i+1, N), (i+2, N)] < P_s < P_s^*[(i, N), (i+1, N)]$, then the optimal value of m is $i+1$.
 - If $0 < P_s < P_s^*[(\eta-1, N), (\eta, N)]$, then the optimal value of m is η .
4. If the optimal m cannot satisfy the information expansion ratio requirement, then choose $m=m+1$, $m=m+2$, ..., $m=N$ respectively, until the information expansion ratio is satisfied.
5. Output m .

4.2.Security Analysis

In this section, we analyze the security level achieved by FAMIDS. We assume that attacks on sensor network routing in the literature have been tackled by the multipath routing protocols. In our analysis, we put our emphasis on attacks regarding information security. We describe our analysis of FAMIDS as these security concerns: information confidentiality, message integrity, and attack resistance.

4.2.1.Information Confidentiality

The idea of FAMIDS is originated from the concept of information dispersal scheme that the original information is ruffled and evenly distributed into several sub-packets. Since we

incorporate the encryption and authentication procedures into the packet generation phase, an adversary still can not obtain the original information even if enough sub-packets are collected.

4.2.2. Message Integrity

Message authentication code is utilized to ensure the integrity of sub-packets. By comparing the message authentication code decoded from sub-packets and calculated from the information bits on the fly, the base station can easily identify the integrity of the received packets.

4.2.3. Attack Resistance

Traditional packet splitting schemes are susceptible to several attacks, in order to thwart the operation of rebuilding the original information at the base station. We detail our countermeasures to such attacks as follows.

- **Pollution attack**

In this attack, an adversary injects forged sub-packets or modifies the context of sub-packets. Traditional packet splitting schemes provide no mechanism for receivers to examine the validity of received sub-packets, and thus receivers fail to reconstruct the original information. In FAMIDS, the source node encrypts the original information and authentication code with the secret key shared with the base station, therefore forged sub-packets can be detected by the base station in advance of information reconstruction.

- **Replay attack**

An attacker buffers the correctly received sub-packets from the upstream neighbors, and retransmits the sub-packets afterwards. The sub-packets have not been modified by the attacker

and therefore are deemed to be valid from the confidentiality and integrity examination. As a consequence, the receiver will be disturbed by the obsolete sub-packets and reconstruct inconsistent information with the original one generated by the source node. In FAMIDS, we avoid this kind of attack by the sequence number stored in both communicating end. The base station can verify the freshness of a sub-packet by the decoded sequence number. Besides, intermediate nodes can also confirm the incoming sub-packets by the sequence number shared with upstream neighbors.

5. Conclusion

In this paper, we propose an enhanced information delivery scheme, FAMIDS, suitable for general applications of resource-restricted wireless sensor networks. Since FAMIDS adopts the concept of packet splitting schemes, and incorporate lightweight security mechanisms, it simultaneously gains the advantages of fault tolerance and information security. We also propose detailed analysis of the communication reliability achieved by FAMIDS. The communication reliability is first derived as an analytical form $R(m, n)$ and is analyzed under four different network models: the simple model, the ideal model, the general model and the special model. For each network model, we derive fundamental theorems to help us understand the behavior of $R(m, n)$ and propose an algorithm separately to determine the optimal (m, n) set with highest communication reliability. From the analysis results of each network model, we can draw some important conclusions:

- 1) For any kind of network, one achieves the maximum communication reliability only when total available paths are used for packet transmission.
- 2) Transmitting the original packet with the maximum information expansion ratio can not always achieve the best reliability. The reason is that packet splitting causes a drastic gain of link success probability when the BER of the sensing circumstance is relative high. The communication reliability is greatly benefited from packet splitting under such condition.
- 3) The optimal solution with the highest communication reliability for total N paths can only be (m, N) FAMIDS, $1 \leq m \leq \eta_2$, where $\eta_2 = \lfloor n/2 \rfloor$. Under the circumstance of low link failure rate or high node failure rate, the advantages achieved by packet splitting are then diminished, and thus the candidate set can be further reduced.

However, there is still no efficient way in the literature to estimate the node failure rate accurately, especially when the adversaries exist. Besides, the base station can make use of the incorrectly received sub-packets to detect the situation of node compromise, and provide a more

secure network environment. We leave these challenges as our future work. In addition, although packet splitting mechanism achieves the advantage of fault tolerance, it increases the probability of network congestion simultaneously. Many schemes in the literature consider the problem of network congestion in various networks [4],[5],[16],[17]. We can further combine FAMIDS with some congestion control scheme to achieve better utilization of network bandwidth.

Appendix

A. Proof of Theorem 4.1

$$R(m, m) - R(m + 1, m + 1) = P_{s,1} \times P_{s,2} \times \cdots \times P_{s,m} \times (1 - P_{s,(m+1)}) > 0$$

$$\therefore R(m, m) > R(m + 1, m + 1) \text{ for } 1 \leq m \leq N - 1$$

B. Proof of Theorem 4.2

$$R(m, n+1) - R(m, n) = P_{s(n+1)} \times P_r \{ \text{any } m-1 \text{ paths success among } P_1 \sim P_n \} > 0$$

$$\therefore R(m, n) < R(m, n + 1) \text{ for } 1 \leq n \leq N - 1, 1 \leq m \leq n$$

C. Proof of Theorem 4.3

By the definition of $R(m, n)$,

$$R(m, n) - R(m + 1, n) = P_r \{ m \text{ paths success} \} > 0 \rightarrow R(m, n) > R(m + 1, n)$$

D. Proof of Theorem 4.4

$$\therefore R(m, n) = \sum_{j=m}^n \sqrt[m]{P_{s,j}} \cdot F_{m-1}(j-1)$$

$$R(m, n+1) = \sum_{j=m}^{n+1} \sqrt[m]{P_{s,j}} \cdot F_{m-1}(j-1)$$

$$\therefore R(m, n+1) - R(m, n) = \sqrt[m]{P_{s,n+1}} \cdot F_{m-1}(n) > 0$$

$$\rightarrow R(m, n+1) > R(m, n)$$

E. Proof of Theorem 4.5

The first derivative of $R(m, n)$ on P_s is

$$R'(m, n) = C_m^n \cdot m \cdot P_l^{(1-m/m)} \cdot (P_l^{(1-m/m)} P_s)^{m-1} (1 - P_l^{(1-m/m)} P_s)^{n-m} > 0, \text{ for } 0 < P_s < 1$$

$\therefore R(m, n)$ is strictly increasing for $0 < P_s < 1$.

F. Proof of Theorem 4.7

Let $\gamma(P_s) \equiv R(m, n) - R(m+1, n)$.

$$\begin{aligned} \gamma(P_s) &= C_0^n \cdot \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n\right)^n + C_1^n \cdot \left(P_l^{\frac{1}{m+1}} \cdot P_n\right) \cdot \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n\right)^{n-1} + \dots \\ &+ C_m^n \cdot \left(P_l^{\frac{1}{m+1}} \cdot P_n\right)^m \cdot \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n\right)^{n-m} - C_0^n \cdot \left(1 - P_l^{\frac{1}{m}} \cdot P_n\right)^n - \\ &C_1^n \cdot \left(P_l^{\frac{1}{m}} \cdot P_n\right) \cdot \left(1 - P_l^{\frac{1}{m}} \cdot P_n\right)^{n-1} - \dots - C_{m-1}^n \cdot \left(P_l^{\frac{1}{m}} \cdot P_n\right)^{m-1} \cdot \left(1 - P_l^{\frac{1}{m}} \cdot P_n\right)^{n-m+1} \end{aligned}$$

It is known that

$$\lim_{P_s \rightarrow 1^-} \left(P_l^{\frac{1}{m}} \cdot P_n\right) = \lim_{P_l \rightarrow 1^-, P_n \rightarrow 1^-} \left(P_l^{\frac{1}{m}} \cdot P_n\right) = \lim_{P_s \rightarrow 1^-} \left(P_l^{\frac{1}{m+1}} \cdot P_n\right) = \lim_{P_l \rightarrow 1^-, P_n \rightarrow 1^-} \left(P_l^{\frac{1}{m+1}} \cdot P_n\right) = 1$$

$$\begin{aligned} \lim_{P_s \rightarrow 1^-} \left(1 - P_l^{\frac{1}{m}} \cdot P_n\right)^k &= \lim_{P_l \rightarrow 1^-, P_n \rightarrow 1^-} \left(1 - P_l^{\frac{1}{m}} \cdot P_n\right)^k \\ &= \lim_{P_s \rightarrow 1^-} \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n\right)^k = \lim_{P_l \rightarrow 1^-, P_n \rightarrow 1^-} \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n\right)^k = 0, \forall k \in \mathbb{N} \end{aligned}$$

$$\begin{aligned} \lim_{P_s \rightarrow 1^-} \left[\frac{\gamma(P_s)}{\left(1 - P_l^{\frac{1}{m+1}} \cdot P_n\right)^{n-m}} \right] &= \lim_{P_l \rightarrow 1^-, P_n \rightarrow 1^-} \left[\frac{\gamma(P_s)}{\left(1 - P_l^{\frac{1}{m+1}} \cdot P_n\right)^{n-m}} \right] \\ &= \lim_{P_l \rightarrow 1^-, P_n \rightarrow 1^-} \left[C_m^n \cdot \left(P_l^{\frac{1}{m+1}} \cdot P_n\right)^m \right] = C_m^n > 0 \end{aligned}$$

$\therefore \gamma(P_s) > 0$, i.e., $R(m, n) > R(m+1, n)$ when $P_s \rightarrow 1$

G. Proof of Theorem 4.8

Let $\gamma(P_s) \equiv R(m, n) - R(m+1, n)$.

We use another expression of $\gamma(P_s)$ here

$$\begin{aligned} \gamma(P_s) = & C_m^n \cdot \left(P_l^{\frac{1}{m}} \cdot P_n \right)^m \left(1 - P_l^{\frac{1}{m}} \cdot P_n \right)^{n-m} + C_{m+1}^n \cdot \left(P_l^{\frac{1}{m}} \cdot P_n \right)^{m+1} \cdot \left(1 - P_l^{\frac{1}{m}} \cdot P_n \right)^{n-m-1} \\ & + \dots + C_n^n \cdot \left(P_l^{\frac{1}{m}} \cdot P_n \right)^n - C_{m+1}^n \cdot \left(P_l^{\frac{1}{m+1}} \cdot P_n \right)^{m+1} \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n \right)^{n-m-1} - \\ & C_{m+1}^n \cdot \left(P_l^{\frac{1}{m+1}} \cdot P_n \right)^{m+2} \cdot \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n \right)^{n-m-2} - \dots - C_n^n \cdot \left(P_l^{\frac{1}{m+1}} \cdot P_n \right)^n \end{aligned}$$

When $P_l \rightarrow 0^-$, we can conclude that

$$\begin{aligned} \lim_{P_l \rightarrow 0^+} \left(P_l^{\frac{1}{m}} \cdot P_n \right) &= \lim_{P_l \rightarrow 0^+} \left(P_l^{\frac{1}{m+1}} \cdot P_n \right) = 0 \\ \lim_{P_l \rightarrow 0^+} \left(1 - P_l^{\frac{1}{m}} \cdot P_n \right)^k &= \lim_{P_l \rightarrow 0^+} \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n \right)^k = 1, \forall k \in \mathbb{N} \end{aligned}$$

$$\therefore \lim_{P_l \rightarrow 0^+} \left[\frac{\gamma(P_s)}{\left(P_l^{\frac{1}{m}} \cdot P_n \right)^m} \right] = \frac{n! \cdot [(m+1) - P_n(n-m)]}{(m+1)!(n-m)!}$$

$$\text{If } (m+1) - P_n(n-m) > 0, \text{ then } \lim_{P_l \rightarrow 0^+} \left[\frac{\gamma(P_s)}{\left(P_l^{\frac{1}{m}} \cdot P_n \right)^m} \right] > 0.$$

$$\therefore \lim_{P_l \rightarrow 0^+} \left[\frac{\gamma(P_s)}{\left(P_l^{\frac{1}{m}} \cdot P_n \right)^m} \right] > 0 \text{ if } m > \frac{P_n \cdot n - 1}{P_n + 1} \Rightarrow R\left(\left\lceil \frac{P_n \cdot n - 1}{P_n + 1} \right\rceil, n\right) \text{ is the optimal of } R(i, n) \text{ when}$$

$P_l \rightarrow 0^+$, where $1 \leq i \leq n$

Here we analyze two special cases of P_n as follows: (**Corollary 3** & **Corollary 4**)

$$1) \quad 0 \leq P_n \leq \frac{2}{n-1}$$

$$\left\lceil \frac{P_n \cdot n - 1}{P_n + 1} \right\rceil = 1 \Rightarrow R(1, n) \text{ is the optimal solution.}$$

$$2) \quad \frac{n}{n+2} \leq P_n \leq 1$$

$$\frac{n-2}{2} \leq \frac{P_n \cdot n - 1}{P_n + 1} \leq \frac{n-1}{2} \rightarrow \left\lceil \frac{P_n \cdot n - 1}{P_n + 1} \right\rceil = \left\lfloor \frac{n}{2} \right\rfloor$$

$$\Rightarrow R\left(\left\lfloor \frac{n}{2} \right\rfloor, n\right) \text{ is the optimal solution.}$$

H. Proof of Theorem 4.9

Let $\gamma(P_s) \equiv R(m, n) - R(m+1, n)$.

We use the same definition of $\gamma(P_s)$ as Theorem 4.4.

For $P_n \rightarrow 0+$, we have known that

$$\begin{aligned} \lim_{P_n \rightarrow 0+} \left(P_l^{\frac{1}{m}} \cdot P_n \right) &= \lim_{P_n \rightarrow 0+} \left(P_l^{\frac{1}{m+1}} \cdot P_n \right) = 0 \\ \lim_{P_n \rightarrow 0+} \left(1 - P_l^{\frac{1}{m}} \cdot P_n \right)^k &= \lim_{P_n \rightarrow 0+} \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n \right)^k = 1, \forall k \in \mathbb{N} \end{aligned}$$

$$\begin{aligned} &\lim_{P_n \rightarrow 0+} \left[\frac{\gamma(P_s)}{\left(P_l^{\frac{1}{m+1}} \cdot P_n \right)^{m+1}} \right] \\ &= \lim_{P_n \rightarrow 0+} \left[C_{m+1}^n \cdot \left(1 - P_l^{\frac{1}{m+1}} \cdot P_n \right)^{n-m-1} + C_{m+1}^n \cdot P_l^{\frac{1}{m}} \cdot \left(1 - P_l^{\frac{1}{m}} \cdot P_n \right)^{n-m+1} \right] = C_{m+1}^n \left(1 + P_l^{\frac{1}{m}} \right) > 0 \end{aligned}$$

For all possible P_1 values, $\gamma(P_s) > 0$, i.e. $R(m, n) > R(m+1, n)$

$\therefore R(1, n)$ is the optimal solution when $P_n \rightarrow 0+$

I. The formula of $P_s^*[1, n]$ (x below means P_N)

$$P_s^*[1, 4] = 3.97x^3 - 7.12x^2 + 4.28x - 0.87$$

$$P_s^*[1, 5] = 3.2x^3 - 4.6x^2 + 2.45x - 0.48$$

$$P_s^*[1, 6] = 2.3x^3 - 2.6x^2 + 1.3x - 0.26$$

$$P_s^*[1, 7] = 1.585x^3 - 1.26x^2 + 0.656x - 0.15$$

$$P_s^*[1, 8] = 1.07x^3 - 0.418x^2 + 0.328x - 0.1$$

$$P_s^*[1, 9] = 0.72x^3 + 0.096x^2 + 0.176x - 0.075$$

$$P_s^*[1, 10] = 0.464x^3 - 0.43x^2 + 0.105x - 0.0635$$

J. The formula of $P_s^*[2, n]$ (x below means P_N)

$$P_s^*[2, 6] = 2.61x^3 - 6.25x^2 + 5x - 1.33$$

$$P_s^*[2, 7] = 4.02x^3 - 8.32x^2 + 5.756x - 1.329$$

$$P_s^*[2, 8] = 4.622x^3 - 8.55x^2 + 5.3x - 1.1$$

$$P_s^*[2, 9] = 4.635x^3 - 7.8x^2 + 4.43x - 0.84$$

$$P_s^*[2, 10] = 4.35x^3 - 6.71x^2 + 3.543x - 0.63$$

K. The formula of $P_s^*[3, n]$ (x below means P_N)

$$P_s^*[3, 8] = 0.41x^3 - 1.06x^2 + 0.91x - 0.26$$

$$P_s^*[3, 9] = 1.36x^3 - 3.14x^2 + 2.4x - 0.61$$

$$P_s^*[3, 10] = 2.6x^3 - 5.52x^2 + 3.89x - 0.9$$

References

- [1] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *Wireless Communications, IEEE* [see also *IEEE Personal Communications*], vol. 11, pp. 6-28, 2004.
- [2] E. Ayanoglu, C.-L. I, R. D. Gitlin, and J. E. Mazo, "Diversity coding for transparent self-healing and fault-tolerant communication networks," *Communications, IEEE Transactions on*, vol. 41, pp. 1677-1686, 1993.
- [3] S. Dulman, T. Nieberg, J. Wu, and P. Havinga, "Trade-off between traffic overhead and reliability in multipath routing for wireless sensor networks," presented at *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, 2003.
- [4] C. T. Ee and R. Bajcsy, "Congestion control and fairness for many-to-one routing in sensor networks," in *Proceedings of the 2nd international conference on Embedded networked sensor systems* Baltimore, MD, USA ACM Press, 2004 pp. 148-161
- [5] R. Jain, S. Kalyanaraman, S. Fahmy, R. Goyal, and S.-C. Kim, "Source behavior for ATM ABR traffic management: an explanation," *Communications Magazine, IEEE*, vol. 34, pp. 50-56, 1996.
- [6] M. Y. Kemal Akkaya, "A survey on routing protocols for wireless sensor networks," *Ad Hoc Networks*, Elsevier, pp. 325-349, 2003.
- [7] M. K. Marina and S. R. Das, "On-demand multipath distance vector routing in ad hoc networks," presented at *Network Protocols, 2001. Ninth International Conference on*, 2001.
- [8] Z. J. H. Panagiotis Papadimitratos, "Secure message transmission in mobile ad hoc networks," *Ad Hoc Networks*, Elsevier, pp. 193-209, 2003.
- [9] M. O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance," *J. ACM* vol. 36 pp. 335-348 1989
- [10] S.-P. Shieh, Y.-C. Tsai, and Y.-L. Huang, "Optimal information-dispersal for fault-tolerant communication over a burst-error channel," *Reliability, IEEE Transactions on*, vol. 52, pp. 354-366, 2003.

- [11] J. A. Stankovic, T. E. Abdelzaher, C. Lu, L. Sha, and J. C. Hou, "Real-time communication and coordination in embedded sensor networks," *Proceedings of the IEEE*, vol. 91, pp. 1002-1022, 2003.
- [12] H.-M. Sun and S.-P. Shieh, "Optimal information-dispersal for increasing the reliability of a distributed service," *Reliability, IEEE Transactions on*, vol. 46, pp. 462-472, 1997.
- [13] A. Tsirigos and Z. J. Haas, "Analysis of multipath routing, part 2: mitigation of the effects of frequently changing network topologies," *Wireless Communications, IEEE Transactions on*, vol. 3, pp. 500-511, 2004.
- [14] A. Tsirigos and Z. J. Haas, "Analysis of multipath Routing-Part I: the effect on the packet delivery ratio," *Wireless Communications, IEEE Transactions on*, vol. 3, pp. 138-146, 2004.
- [15] M. A. M. Vieira, C. N. Coelho, Jr., D. C. da Silva, Jr., and J. M. da Mata, "Survey on wireless sensor network devices," presented at *Emerging Technologies and Factory Automation, 2003. Proceedings. ETFA '03. IEEE Conference*, 2003.
- [16] C.-Y. Wan, S. B. Eisenman, and A. T. Campbell, "CODA: congestion detection and avoidance in sensor networks," in *Proceedings of the 1st international conference on Embedded networked sensor systems Los Angeles, California, USA ACM Press*, 2003 pp. 266-279
- [17] A. Woo and D. E. Culler, "A transmission control scheme for media access in sensor networks ," in *Proceedings of the 7th annual international conference on Mobile computing and networking Rome, Italy ACM Press*, 2001 pp. 221-235
- [18] S. Zhu, S. Setia, and S. Jajodia, "LEAP: efficient security mechanisms for large-scale distributed sensor networks," in *Proceedings of the 10th ACM conference on Computer and communications security Washington D.C., USA ACM Press*, 2003 pp. 62-72
- [19] L. Gargono, A. A. Rescigno, and U. Vaccaro, "Fault-tolerant hypercube broadcasting via information dispersal," *Networks*, vol. 23, pp. 271-282, 1993.
- [20] Y. D. Lyuu, "Fast fault-tolerant parallel communication for de Bruijn and digit-exchange networks using information dispersal," *Networks*, vol. 23, pp. 365-378, 1993.
- [21] D. Ganesan, R. Govindan, S. Shenker, D. Estrin, "Highly-Resilient, Energy-Efficient

Multipath Routing in Wireless Sensor Networks," *Networks*, vol. 23, pp. 271-282, 1993.

- [22] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin, "Highly-resilient, energy-efficient multipath routing in wireless sensor networks," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, pp. 11-25, 2001.

計畫成果自評

項目	已 完 成	未 完 成
建構適當的無線微型感測器網路實驗模型	✓	
分析構成網路可信賴度與安全性威脅之因素	✓	
針對感測網路的可信賴度與安全性研擬對應措施	✓	
設計一個封包切割概念的多重路徑資訊傳送機制	✓	
提出最佳效率封包切割方式的演算法	✓	
建立 FAMIDS 傳輸可信賴程度之分析模型	✓	

可供推廣之研發成果資料表

☐ 可申請專利

☐ 可技術移轉

日期：98 年 5 月 30 日

國科會補助計畫	計畫名稱：設計與實作無線微型感測網路攻擊偵測、防禦與高信賴度的安全傳輸機制 計畫主持人：謝續平教授 計畫編號：NSC 96-2221-E-009-139-MY3 學門領域：無線網路安全
技術/創作名稱	適用於無線微型感測網路之具容錯性與攻擊抵抗性之多重路徑資訊傳送機制(A Fault-Tolerant ,Attack-Resistant Multipath Information Delivery Scheme for Wireless Sensor Networks)
發明人/創作人	謝續平、陳厚坤
技術說明	對於大多數無線微型感測網路的應用而言，資訊傳輸的可信賴程度與其安全性為兩項主要的考量因素。由於感測節點通常佈建於無人照料或敵人充斥的環境中，並且感測到的資訊必須經由節點間本質即不穩定的無線連結傳輸至後端基地台，如何正確、安全且即時地傳送感測到的資訊實為一大挑戰。為了解決此項議題，我們提出了一個基於封包切割概念的多重路徑資訊傳送機制，FAMIDS。在此機制之下，資料來源節點將原始封包利用增加冗位元的方式切割成數個小封包，並且利用多重路徑安全地傳送該封包。即使部分的小封包接收錯誤，後端基地台仍能重建原始封包。除此之外，我們建立了分析 FAMIDS 所能達到的傳輸可信賴程度之分析模型。針對不同的網路條件，我們研究該分析模型之表現情形，並且提出能找出最佳封包切割方式的有效率演算法，以達到最佳的傳輸可信賴程度。相較於傳統之暴力法，我們提出之演算法能有效降低每次封包傳輸之計算複雜度。

	Transmission reliability and information confidentiality are two of the major concerns for most applications of wireless sensor networks. Since sensor nodes are usually deployed in unattended or adversarial environments and sensed information is transmitted through inherent instable wireless links among sensor nodes to the base station, how to deliver sensed information correctly, securely and instantaneously becomes a challenging problem. To deal with this issue, we propose an enhanced multipath information delivery scheme FAMIDS based on the concept of packet splitting. With the proposed scheme, the source node splits the original data packet into several sub-packets by adding redundancy, and securely transmits the sub-packets through multiple paths. The base station can reconstruct the original data packet even if part of sub-packets is incorrectly received. In addition, we develop an analytical model for evaluating communication reliability achieved by our scheme and study the behavior of the analytical form under different network conditions. For each network model, we propose an algorithm for finding the optimal way of packet splitting with highest communication reliability. Compared with the traditional brute-force approach, the proposed algorithms efficiently reduce the computational complexity for each packet transmission.
可利用之產業 及 可開發之產品	此機制可以應用在大多數的無線微型感測網路，而無線微型感測網路的應用面相當廣泛，軍事、環境、健康、家庭、商業等應用。而其中這些應用裡面對安全性要求較高者更為適合此方法。
技術特點	一般為了達成可信賴性及安全性的傳輸機制已經是大家越來越重視的需求，此研究結果不僅達到這兩項要求，更進一步的提升了效率，針對不同的網路環境，用各種模組去分析後，提出最有效率的演算法來切割封包，來達成最佳的傳輸可信度，此方法有效的降低每次封包傳輸之計算複雜度。
推廣及運用的價值	無線微型感測網路能夠推廣的價值相當的高，因為應用面相當的廣泛，主要可以提供環境上的資訊，監測目標的狀況，來提供操作者資訊。運用上有監控戰場、自然環境、病人健康、家裡環境、工廠品質監控等。

※ 1. 每項研發成果請填寫一式二份，一份隨成果報告送繳本會，一份送 貴單位研發成果推廣單位（如技術移轉中心）。

※ 2. 本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。

※ 3. 本表若不敷使用，請自行影印使用。