Association Schemes and Delay Analysis in Wireless Sensor Networks with Cluster Tree Topology

ASSOCIATION SCHEMES AND DELAY ANALYSIS IN WIRELESS SENSOR NETWORKS WITH CLUSTER TREE TOPOLOGY

BY

WENJUAN LIU, B. Eng

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

Master of Applied Science (2011) McMaster University

(Electrical & Computer Engineering) Hamilton, Ontario, Canada

TITLE: Association Schemes and Delay Analysis in Wireless Sensor Networks with Cluster Tree Topology

AUTHOR: Wenjuan Liu

B. Eng

Beijing JiaoTong University, Beijing, China

SUPERVISOR: Dr. Dongmei Zhao

NUMBER OF PAGES: xiv, 74

This thesis is dedicated to my parents.

Abstract

Ubiquitous wireless sensor networks (WSNs) are expected to play an important role in the future society for various applications. As a result, carefully managing the network resources to improve the network performance becomes a hot research topic. In this thesis, we study the performance of WSNs with a cluster tree topology, where all the cluster heads (CHs) form a tree topology. The sensor nodes transmit data to their directly associated CHs, which forward the traffic to the sink through other CHs in the cluster tree.

We first study the associations between sensor nodes and the cluster heads (CHs). In a WSN where there is a strong overlapping coverage area between the CHs, associating the sensor nodes to different CHs may result in different network performance. As the sensor node associations affect the traffic load within each cluster and that between the clusters, timeline of the CHs should be allocated accordingly. We formulate three optimization problems by jointly considering the sensor node associations and CH timeline allocations. The objectives are maximizing the throughput per sensor node, balancing the energy consumption among the CHs, and maximizing the network level throughput, respectively. Corresponding to each of the objectives, a heuristic association scheme is designed and the timeline allocations of the CHs are calculated. Numerical results based on computer simulation demonstrate that the proposed schemes achieve close-to-optimum performance.

In the second part of the thesis we study the end-to-end transmission delay for traffic at different levels of a WSN with the cluster tree topology. The end-to-end delay includes both local transmission delay between the sensor nodes and their directly associated CHs and inter-CH transmission delay between the forwarding CHs along the path to the sink. Given the timeline allocations of each CH for local and inter-cluster traffic transmissions, we find the distribution of the local traffic transmission delay and that of the inter-CH transmission delay. Based on these results, we then derive the distribution of the end-to-end transmission delay and the packet drop rate due to excessive delay. The results provide important guidelines for allocating the CH time resources in order to achieve certain delay or packet drop rate performance. By appropriately allocating the CH time resources, it is possible that traffic traversing more hops to the sink experiences better delay performance than that traversing a fewer number of hops.

Acknowledgements

In the first place, I would like to express my deep appreciation to my supervisor, Dr. Dongmei Zhao, for her expert guidance, insightful discussions and patient encouragement throughout my research and thesis writing. Besides, I would like to thank the committee: Dr. Terence Todd and Dr. Jiankang Zhang for their instructive suggestions. Moreover, I do appreciate Dr. Gang Zhu for continuously providing helpful advices for my research and career planning. Last but not least, I am also grateful to my colleagues in the Wireless Network Laboratory, Bin Wang, Zhongliang Liang, Shan Feng, Yang Yang, Ghada Badawy, Mohammad Sheikh Zefreh, Siamak Shakeri, Abdulla Hammad, Hanan Hassanein, and all my dear friends at McMaster Univeristy, for their help and support during my study.

Notations

N_i	Total number of sensor nodes associated with CH i
N_{ij}	Number of sensor nodes in overlapping area of CHs i and j
N_{ii}	Number of sensor nodes located in area covered by CH i only
N_{total}	Total number of sensor nodes
n_{ij}	Number of sensor nodes located in overlapping area of CHs i and j and associated to CH i
$T_{L,i}$	Duration of local data receiving time of CH i
$T_{R,i}$	Duration of inter-CH receiving time of CH i
$T_{T,i}$	Duration of inter-CH transmitting time of CH i
T_{SF}	Duration of a superframe
η	Throughput per sensor node
S_i	Throughput of CH i per unit time during the CAP period
E_i	Energy consumption of CH i in each SF
$E_{\max,i}$	Maximum allowed energy consumption of CH i during an SF
P_r	Power that a CH consumes for receiving data
$P_{T,i}$	Power that CH i consumes for transmitting data
$\mathcal{C}_i$	Set of child CHs of CH i
$\mathcal{P}_i$	Set of the CHs along the path from CH i to the sink (sink is not included)

$\mathcal{B}_i$	Set of neighboring CHs of CH i
d_{ij}	Distance between CHs i and j
T_p	Time for transmitting one packet
$K_{C,i}$	$T_{L,i}/T_p$
$A_{C,i}$	Number of generated packets in one SF
$A_{t,i}$	Number of packets arriving in the same SF as the tagged packet and earlier than the tagged packet
$X_{C,i}$	Number of local packets waiting for transmissions at the beginning of $T_{L,i}$
$Y_{C,i}$	Number of successfully received packets during local receiving time in one SF
$D_{C,i}$	Local transmission delay in cluster i
$Y_{F,i}$	Number of received packets at CH i during inter-CH receiving time in one SF
$Z_{F,i}$	Total number of packets collected during $T_{C,i}$ and $T_{R,i}$ in one SF
$X_{B,j}$	Total number of packets buffered at all child CHs of CH j at the end of $T_{R,j}$
Z_j	Total number of packets received by CH j in one SF
$K_{F,j}$	$T_{R,j}/T_p$
$D_{U,i}$	Inter-CH transmission delay from CH i to its parent CH
$D_{e2e,i}$	End-to-end transmission delay of traffic originated from cluster i
$D_{\max}$	Maximum tolerable end-to-end delay
$P_{dr,i}$	Packet drop rate at cluster i
λ	Average arrival rate of local traffic

Abbreviations

WSN	Wireless Sensor Network
MEMS	Micro-Electro-Mechanical Systems
VLSI	Very Large Scale Integration
WPAN	Wireless Personal Area Network
MAC	Media Access Control
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
TDMA	Time Division Multiple Access
QoS	Quality of Service
CH	Cluster Head
SF	Superframe
GTS	Guaranteed Time Slot
CAP	Contention Access Period
CFP	Contention Free Period
FFD	Full Function Device
RFD	Reduced Function Device

Contents

List of Figures

Chapter 1

Introduction

This chapter gives a brief introduction about the wireless sensor networks (WSNs), followed by the motivation of this thesis work and a summary of the related works.

1.1 A Brief Introduction to WSNs

1.1.1 Background

A wireless sensor network typically consists of a number of sensor nodes which can sense, measure, and gather information from the environment. With the development in Micro-Electro-Mechanical Systems and Very Large Scale Integrated circuits technology, ubiquitous wireless sensor networks have gained worldwide attention and are expected to play an important role in the future society, such as for military applications, health care, traffic control, home automation, industrial process and environmental monitoring, etc [1].

The research on WSNs started in 1993, when the Wireless Integrated Network Sensors (WINS) program was initiated at the University of California, Los Angeles. The low-power

Wireless Integrated Microsensors (LWIM) program started in 1995 is another project sponsored by Defense Advanced Research Projects Agency (DARPA) [2], and in 2003 the National Science Foundation (NSF) began to widely support the research on WSNs. Recently, WSNs are an active research area in the fields of computer science and telecommunications.

1.1.2 Network Structure

A WSN can be infrastructure-based or ad-hoc. In an infrastructure-based WSN, some nodes form a relative static infrastructure and are responsible for relaying traffic for other nodes. Smart Dust [3] is an example of the network in this mode. While in an ad-hoc network, sensor nodes communicate with each other via one or multiple hops without an infrastructure. As the deployment of an ad-hoc network is less costly, the ad-hoc mode is adopted by many WSN applications. However, managing the network resources and quality of service provisioning in such networks is more difficult than in an infrastructure-based network. Some WSNs may combine the infrastructure-based and ad-hoc modes, where clusters of sensor nodes are inter-connected via some central access points referred to as cluster heads (CHs), and all the data packets collected by the sensor nodes are eventually forwarded by the CHs to the sinks.

The size of a WSN is determined by many factors such as the communication range and coverage requirement, and may vary from a couple of nodes to thousands of nodes. In a typical WSN, each sensor node is equipped with a sensor that measures a certain physical quantity, a microprocessor, a power supply, a radio transmitter and receiver. A sensor node may vary in size and cost, depending on the functional requirements and the resources of the individual sensor nodes such as computational speed, memory and energy [4].

1.1.3 Standards and Specifications

The IEEE 802.15.4 and ZigBee are among the predominant standards and specifications for wireless sensor network communications. The IEEE 802.15.4 standard specifies the physical and medium access control (MAC) layers for low-rate, low-power and flexible wireless personal area networks (WPANs) [5]. The IEEE 802.15.4 standard defines two types of network nodes: full function devices (FFDs) and reduced function devices (RFDs). An FFD can communicate with other FFDs or RFDs, while an RFD can only communicate with an FFD. In a WSN with a cluster topology, the entire network is partitioned into small areas, and the most capable device in a cluster is selected as the cluster head. In a typical WSN deployment, the CHs are FFDs and form the backbone of the network, while a large number of sensors distributed in the network are RFDs.

The ZigBee specification is maintained and published by ZigBee Alliance, which is an association of companies working together to enable low-power, low-cost WPANs based on the IEEE 802.15.4 standards [6]. There are three types of ZigBee devices: ZigBee Coordinator, ZigBee Router and ZigBee End Device. In each network, there is only one ZigBee Coordinator, which is the most capable device in the network. ZigBee Routers can communicate with each other and forward traffic from other devices, while ZigBee End Devices are only allowed to transmit data to the ZigBee Coordinator or ZigBee Routers. According to their functional capabilities, the ZigBee Coordinator and ZigBee Routers are FFDs defined by the IEEE 802.15.4 standards, and the ZigBee End Devices usually are RFDs which are less expensive to manufacture.

There are both contention-based and contention-free transmission periods specified by the IEEE 802.15.4 standards. The standards allow two types of mechanisms for accessing the contention access period (CAP): a slotted CSMA/CA used in the beacon enabled

network, and an unslotted CSMA/CA used in the non-beacon enabled network. For the former, the CHs transmit periodic beacons to other network devices, and a superframe is defined to be the period between two successive beacons. In a non-beacon enabled network, the CHs typically have their receivers continuously active to receive the data from other network devices. Compared to the contention-based transmissions, the contention free transmissions provide much higher efficiency and lower latency. The contention free period (CFP) is composed of several guaranteed time slots (GTSs), and each GTS is assigned to the communication from or to a particular device in the network.

It should be noted that a lot of WSNs currently studied in the literature do not follow the IEEE standards. In particular, TDMA-based WSNs have been studied extensively in the literature, and some examples can be found in [7]-[9].

1.1.4 Network Topologies

ZigBee Alliance defines three network topologies above the IEEE 802.15.4 physical and MAC layers, the mesh topology, the star topology and the cluster-tree topology. In the mesh topology, some of the nodes are connected with more than one adjacent node in the network, and packets determine theirs path to the destinations according to the routing algorithm. Due to its cost and complexity, the mesh topology is usually used in the network with a small number of nodes. The star topology is easy to design and implement, all the devices in a star topology network are connected to central nodes, and the failure of each device or connection does not affect the entire network as long as the central nodes function well. The cluster-tree topology can be considered to integrate several small star topology networks together, in which a root node connects with one or multiple lower level nodes, and each node in the network has a specific number of child nodes connected to it in the

hierarchy.

The topology of a WSN affects many network features and qualities, such as capacity, latency and scalability. Both the star and cluster-tree topologies can use beacon frames to synchronize devices to their parent node, and thus minimize power consumption of the devices by intermittent operations.

1.2 Motivations of the Thesis Work

Observing the environments and gathering data are the fundamental functions of a WSN, and energy consumption and traffic delay are two main performance metrics to evaluate the efficiency of this data collection process. In a WSN with specially deployed CHs, association relation between the sensor nodes and the CHs is important as it affects the overall network throughput, energy consumption, and other performance. In a lot of cases, the CHs are placed in random locations, and strong overlapping may exist between their coverage areas so that the sensor nodes can choose to associate to different CHs. The associations, on the other hand, affect the traffic load within each cluster and further affect the traffic load between the CHs. Therefore, the timeline allocations of the CHs and the sensor node associations should be jointly considered in order to provide desired QoS or optimize certain network performance. Although extensive work has been done on sensor node associations, no work is available for jointly considering the sensor node associations and CH resource allocations in order to optimize the network performance.

Providing data transmissions with guaranteed QoS is of great importance in various areas such as health care and environmental monitoring, and a lot of WSN applications increasingly require real-time traffic with strict latency guarantees. For example, in emergency surveillance, data are supposed to be delivered within the maximum tolerable delay.

As a result, meeting real-time QoS requirements is becoming a key issue of WSN designs. The end-to-end delay is not only related to the resource allocations at the local CH, but also that of all the CHs along the end-to-end path to the sink. Developing an analytical model for the distribution of the end-to-end delay can provide important insights on designing the real-time solutions for WSN applications. More importantly, the experienced end-to-end delay can be significantly different for traffic originated from different clusters in a WSN. Appropriately allocating the CH timeline resources can achieve the desired delay performance. Little work has been done along this direction.

1.3 Related Works

In a network where all sensor nodes are homogeneous, the node serving as the CH consumes more energy than other nodes. Dynamic topology formation using protocols such as Low-Energy Adaptive Clustering Hierarchy (LEACH) and its various modifications can balance the energy consumption of the sensor nodes. The basic idea is to dynamically update the network topology so that in every update nodes with more remaining energy serve as the CHs [10]. During the cluster formation phase of LEACH, sensors in the transmission range of more than one CH choose the CHs that require the minimum communication energy. Thus typically the sensors choose the nearest CH. Energy efficient clustering scheme (EECS) [11] is another distributed and load balanced clustering algorithm. According to the cluster formation stage of this algorithm, sensors with more than one potential CH choose the CHs by considering not only the distance to the CH but also the number of hops from that CH to the sink.

Recently, there has been some work, e.g., [12] [13], considering WSNs with heterogeneous devices. A practical deployment for a WSN is to distribute a number of FFDs as

the CHs and a number of RFDs as the sensor nodes, all in random locations. The FFDs then form an infrastructure for the network with a certain topology. Since the cluster-tree topology has better scalability than the star topology and is more suitable for large-scale sensor networks, it is attracting increasingly more attention, e.g., [12]-[14].

It is possible that sensor nodes located in the overlapping area among several clusters so that the sensor nodes can choose to associate to different CHs. In this case, the association relation between sensor nodes and the CHs is important as it affects the network performance. Some work on sensor node associations in IEEE 802.15.4 network can be found in [15]-[18]. The Simple Association Process (SAP) [15] and the Fast Association mechanism (FAM) [16] are introduced respectively in order to decrease the association delay in practical IEEE 802.15.4 networks. In [17], the authors proposed association schemes for pre-determined devices to associate to their desired coordinator if the device is in the transmission range of multiple similar IEEE 802.15.4 networks.

Designing delay-awareness MAC protocols in multi-hop networks is an active research topic in recent years. In [19], the authors propose a Quality of service enhanced Base station Controlled Dynamic Clustering Protocol (QBCDCP) to support video and imaging traffic delivery in WSNs. According to this protocol, both intra-cluster and inter-cluster communications use a Time Division Multiple Access (TDMA) scheme. Analytical results show the trade-off between the end-to-end delivery delay and the lifetime of sensor node. A low latency MAC protocol for WSNs is proposed in [20]. The idea is that the active period of the sensor nodes in each cluster is synchronized and with the same timing offset. This MAC protocol can reduce the latency and solve the collision problem in inter-cluster communications. Besides, the authors of [21] present a novel MAC protocol for delay sensitive data in a multi-hop mesh network, in which the nodes use local mechanisms to

determine an adaptive transmission schedule and manage the end-to-end delay.

Some other research on delay sensitive applications in multi-hop wireless mesh networks has been developed, like [22] and [23]. In [22], the authors derive a lower bound of delay-bound violation probability over multi-hop paths, and analyze the end-to-end average delay and the delay jitter bounds under different traffic conditions as well. Reference [23] presents an approach to finding conflict-free TDMA schedules with the minimum scheduling delay, in which the authors formulate an optimization problem to find a transmission order with the min-max delay in a set of multi-hop paths, and further devise an algorithm on overlay tree topologies and use it with a modified Bellman-Ford algorithm to find minimum delay schedules in polynomial time.

There has been some work focusing on modeling and estimating the delay performance in WSNs, e.g., [24]-[26]. In [24], the authors analyze the average minimum delay for collecting data in networks with different topologies and the impact of several factors such as packet size, transmission range and channel erasure probability on the delay performance. End-to-end delay analysis in WSNs is investigated in [25] [26]. A closed form expression for the end-to-end delay distributions in a WSN employing data aggregation is derived in [25], and a stochastic end-to-end delay bound to predict the performance of a WSN deployment is derived in [26].

1.4 Overview of the Thesis

The remainder of the thesis is organized as follows. In Chapter 2, sensor node associations and CH timeline allocations are jointly studied. Three optimization problems are formulated with the objectives of maximizing the throughput per sensor node, balancing the CH energy consumption, and maximizing the network level throughput, respectively. One

heuristic scheme is then proposed for each of the objectives. Numerical results demonstrate that the proposed heuristic schemes achieve close-to-optimum performance. In Chapter 3, an analytical model is developed to find the distribution of the end-to-end packet transmission delay in a WSN. The distribution of the local transmission delay between the sensors and their directly associated CHs is first derived, followed by the derivation of the distribution of the inter-CH transmissions and that of the end-to-end transmission delay. The analysis is verified by computer simulations. In Chapter 4 we conclude the thesis and list some future research topics.

Chapter 2

Association Schemes in WSNs with Cluster Tree Topology

In a WSN with multiple clusters, a lot of system performance is related to which cluster the sensor nodes are associated to. For example, transmission delay is dependent on the number of hops (or intermediate CHs) in order for the traffic to reach the sink, and the maximum achievable throughput is determined by the congestion conditions in individual clusters, which depend on the number of the sensor nodes in each cluster. In a lot of cases, the CHs are placed in random locations, and strong overlapping may exist between their coverage areas so that the sensor nodes can choose to associate to different CHs. This provides opportunities to optimize different network performance, including throughput and energy consumption. In this chapter we study the relationship between sensor node associations and different network performance. Section 2.1 gives a brief introduction about the system that this work is based on. Three optimization problems are formulated in Section 2.2 with objectives to maximize throughput per sensor node, to balance CH energy consumption, and to maximize network throughput, respectively. Heuristic association schemes are then

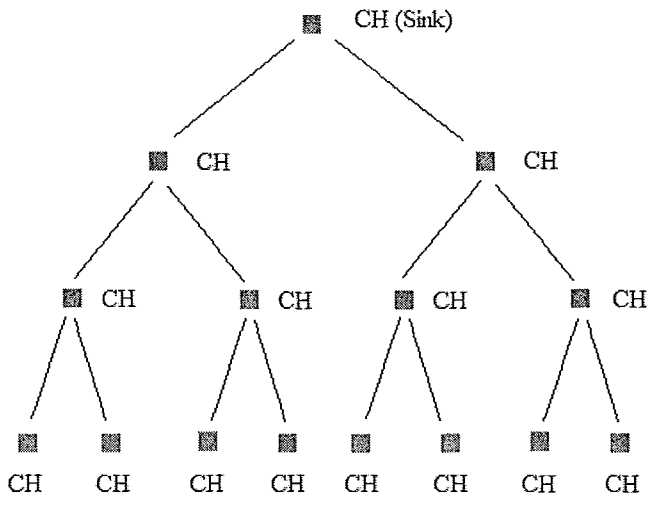

Figure 2.1: Cluster-tree topology

proposed in Section 2.3 for each of the objectives. In Section 2.4 we provide numerical results to demonstrate performance of the proposed schemes. Section 2.5 concludes this chapter.

2.1 System Description

We consider a WSN with the cluster-tree topology as shown in Fig. 2.1. The sink is located at the root of the cluster tree, and all the CHs (including the sink) together form a multi-level wireless backbone. We define local traffic as the traffic between the sensors and their associated CHs. All local traffic should be forwarded by their CHs to higher level CHs (with a smaller number of hops to the sink) and eventually to the sink. Sensors can only communicate with their associated CHs. Each CH periodically broadcasts beacons which the sensor nodes and lower level CHs use to establish the associations with it and acquire

its timeline arrangement. Sensors and lower level CHs receive beacons from the CHs and decide when to transmit.

Sensors scan the beacons and each sensor is associated to one CH. The transmission power of the beacons from each CH determines its coverage area. We assume that a frequency reuse plan is in place and there is no interference between transmissions in different clusters. We focus on the energy consumptions of the CHs, which are powered by batteries and have a limited amount of energy available. The CHs are important for forming the backbone of the network. The energy consumption of the sensor nodes is also an important topic but not studied in this thesis as extensive work has been done in this area in the literature.

At the MAC layer, the timeline of the CH is divided into equal length superframes (SFs), each of which includes a contention access period (CAP) for contention-based transmissions and guaranteed time slots (GTSs) for contention free transmissions. Communications between the CH and its directly associated sensors use CAP, and communications between the CHs use GTSs. We consider cluster i as a typical cluster, and use $T_{L,i}$, $T_{R,i}$ and $T_{T,i}$ to represent the duration of CAP, the inter-CH receiving time for receiving from its child CHs, and the inter-CH transmitting time for transmitting to its parent CHs. When CH i transmits, its parent CH receives at the same channel.

2.2 Problem Formulation

All sensors in the same clusters compete for accessing the CAP time. Let N_i be the total number of sensors associated with CH i, and η be the throughput per sensor node which is represented by the amount of successful transmission time of the sensor node during the CAP period of each SF. In a saturation case, when all sensors always have data to transmit,

we have

$$S_i T_{L,i} = N_i \eta, \tag{2.1}$$

where S_i is the percentage of the successful transmission time during the CAP period. The left-hand side of (2.1) is the total throughput achieved in the CAP interval, and the right-hand side is the total throughput of all sensors during the CAP interval. As the transmissions are contention-based, S_i is a non-linear function of N_i. The analytical relationship between S_i and N_i as well as other MAC layer parameters is derived in [27] and [28]. From (2.1) we can solve $T_{L,i}$ as

$$T_{L,i} = N_i \eta / S_i. \tag{2.2}$$

Since S_i is a function of N_i, $T_{L,i}$ is a function of N_i and η. For a concise presentation in later derivations, we write this function as

$$T_{L,i} \triangleq f(N_i, \eta). \tag{2.3}$$

Equation (2.2) indicates that in order to achieve the required throughput for a number of sensors, the CAP interval should be sufficiently long. This relationship is plotted in Fig. 2.2 based on (2.2) and the formulas in [27] and [28]. The figure shows that the required $T_{L,i}$ increases with N_i for given per sensor node throughput requirement, and it increases more significantly when N_i or η is larger in order to satisfy the increased total throughput requirement and deal with the increased collisions. Consequently, the energy consumption of CH i increases with N_i. Due to the limited available energy consumption at each CH, the number of sensors associated with it should be limited. Given the total available energy of a CH and its desired lifetime, the maximum allowed energy consumption of the CH during an SF is limited, which is defined as $E_{\max,i}$. We consider relative static traffic

in this work, and therefore $E_{\max,i}$ is the same for all the SFs for a given CH. The actual energy consumption of the CH depends on the timeline allocations of the CH. Let P_r and $P_{T,i}$, respectively, represent the power that CH i consumes for receiving and transmitting, then the actual energy consumption of CH i during an SF is given by

$$E_i = T_{L,i}P_r + T_{T,i}P_{T,i} + T_{R,i}P_r, \tag{2.4}$$

where we ignore the energy consumption in the inactive period. The transmission time of CH i should be sufficient in order for it to forward to its parent CH all the traffic that it collects from the directly associated sensors and its child CHs. This is given by

$$T_{T,i} = T_{R,i} + T_{L,i}S_i = T_{R,i} + N_i\eta \tag{2.5}$$

if CH i is not the sink. Let CH 1 be the sink, then $T_{T,1} = 0$. For all CHs except the lowest level CHs (without a child CH) in the cluster tree,

$$T_{R,i} = \textstyle\sum_{j\in\mathcal{C}_i} T_{T,j}, \tag{2.6}$$

where $\mathcal{C}_i$ is a set of the directly connected child CHs of CH i. For the CHs at the lowest level in the cluster tree, $T_{R,i} = 0$.

Next we consider the association relationship between the sensor nodes and the CHs. Given the low cost of the devices, the number of the FFDs and the RFDs may be larger than the minimum necessary for coverage purpose, so that there is a strong overlapping area between the coverage areas of different CHs. In this case, a large number of sensor nodes have choices to associate with different CHs, and the association relation can be done

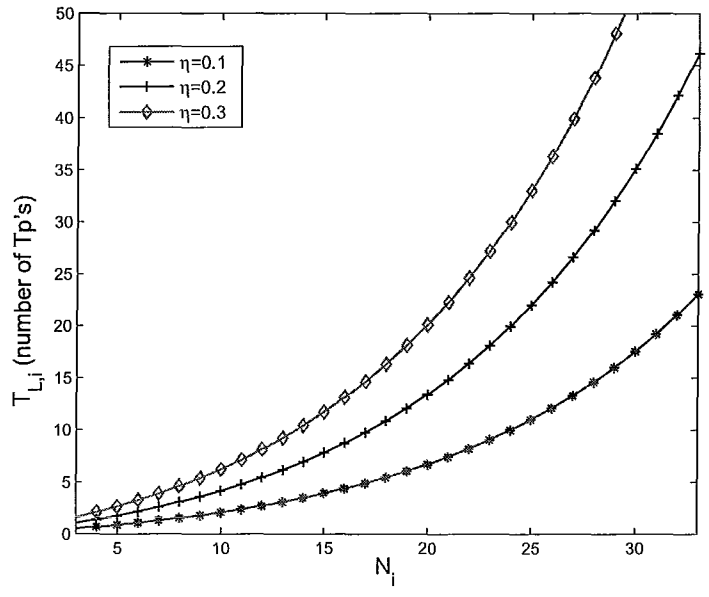

Figure 2.2: Relationship between $T_{L,i}$ and N_i

to optimize the network performance, such as maximizing overall throughput or balancing energy consumptions of the CHs. To simplify the presentation of the optimization problems below, we assume that any sensor can be in the coverage area of at most two CHs. The optimization problems and the association schemes can be extended to a more general case, in which the sensors can be in the coverage area of multiple CHs.

Let N_{ij} be the number of sensors in the overlapping area of CH i and CH j, and $N_{ij} = N_{ji}$. When $i = j$, $N_{ij} = N_{ii}$ represents the number of sensor nodes located in the coverage area of CH i but not in the coverage area of any other CHs, and $N_{total} = \sum_{\text{all } i} \sum_{j \leq i} N_{ij}$ is the total number of sensor nodes. Define n_{ij} as the number of sensors that are located in the overlapping area of CH i and CH j and associated with CH i. We then have $n_{ij} + n_{ji} = N_{ij}$ for $i \neq j$, and $N_i = \sum_j n_{ij}$ is the total number of sensors associated to CH i. For a special case, $n_{ii} = N_{ii}$.

To maximize the per sensor node throughput, an optimization problem can be formulated as follows

$$\mathbf{P1}: \quad \max \eta \tag{2.7}$$

$$\text{s.t.} \quad N_i = N_{ii} + \textstyle\sum_{j \neq i} n_{ij} \tag{2.8}$$

$$n_{ij} + n_{ji} = N_{ij} \tag{2.9}$$

$$E_i \leq E_{\max,i} \tag{2.10}$$

$$T_{L,i} + T_{T,i} + T_{R,i} \leq T_{SF} \tag{2.11}$$

$$0 \leq n_{ij} \tag{2.12}$$

For given η, the energy consumption of the CHs can be balanced through sensor node associations and appropriate CH timeline allocations. The optimization problem is formulated as

$$\mathbf{P2}: \quad \min \max_i E_i \tag{2.13}$$

$$\text{s.t.} \quad N_i = N_{ii} + \textstyle\sum_{j \neq i} n_{ij} \tag{2.14}$$

$$n_{ij} + n_{ji} = N_{ij} \tag{2.15}$$

$$T_{L,i} + T_{T,i} + T_{R,i} \leq T_{SF} \tag{2.16}$$

$$0 \leq n_{ij} \tag{2.17}$$

Let η_i be the per sensor node throughput requirement for the sensor nodes in cluster i.

The following optimization problem is formulated in order to maximize the total throughput in the system:

$$\mathbf{P3}: \quad \max \sum_i N_i \eta_i \tag{2.18}$$

$$\text{s.t.} \quad N_i = N_{ii} + \textstyle\sum_{j \neq i} n_{ij} \tag{2.19}$$

$$n_{ij} + n_{ji} \leq N_{ij} \tag{2.20}$$

$$E_i \leq E_{\max,i} \tag{2.21}$$

$$T_{L,i} + T_{T,i} + T_{R,i} \leq T_{SF} \tag{2.22}$$

$$0 \leq n_{ij} \tag{2.23}$$

When $\eta_i = \eta$ for all i, problem P3 is equivalent to maximizing the total number of sensor nodes that can be served with the required per sensor node throughput.

In all the three optimization problems, the unknowns are n_{ij}'s, $T_{L,i}$'s, $T_{T,i}$'s and $T_{R,i}$'s. Therefore, solving the above problems not only finds number of sensor nodes associated to each CH, but also the timeline allocations of the CHs.

2.3 Proposed Association Schemes

The association schemes work for the network where there are strong overlapping areas between CHs. Sensors located in non-overlapping areas do not have a choice to associate to different CHs. We assume these sensors have associated to their CHs already and their association will not be mentioned below. Before presenting the association schemes, we first introduce the criteria that the CHs use to determine whether they are overloaded or can accept more sensor node association requests.

CH i is overloaded if any of the following two conditions is not satisfied.

$$T_{L,i}P_r + T_{R,i}P_r + T_{T,i}P_{T,i} \leq E_{\max,i}, \tag{2.24}$$

$$T_{L,i} + T_{R,i} + T_{T,i} \leq T_{SF}. \tag{2.25}$$

Define $\mathcal{P}_i$ as a set of the CHs along the path from CH i to the sink (not including the sink). Note that adding one more sensor node to CH i increases not only the traffic load of CH i, but also that of all CHs in $\mathcal{P}_i$. When N_i is increased to $N_i + 1$, the local traffic throughput of CH i is increased by η, its CAP period should be increased to $f(N_i + 1, \eta)$, and its $T_{T,i}$ should be changed to $T_{T,i} + \eta$. Meanwhile, for all $j \in \mathcal{P}_i$, $T_{R,j}$ is increased to $T_{R,j} + \eta$, and $T_{T,j}$ is increased to $T_{T,j} + \eta$. Given that there are N_i sensors associated to CH i currently, CH i can accept one more sensor node only if

$$f(N_i + 1, \eta)P_r + T_{R,i}P_r + (T_{T,i} + \eta)P_{T,i} \leq E_{\max,i}, \tag{2.26}$$

$$f(N_i + 1, \eta) + T_{R,i} + (T_{T,i} + \eta) \leq T_{SF}, \tag{2.27}$$

and

$$T_{L,j}P_r + (T_{R,j} + \eta)P_r + (T_{T,j} + \eta)P_{T,j} \leq E_{\max,j}, \tag{2.28}$$

$$T_{L,j} + (T_{R,j} + \eta) + (T_{T,j} + \eta) \leq T_{SF}, \tag{2.29}$$

for all CH $j \in \mathcal{P}_i$. Therefore, when CH i receives an association request, it should pass the request to all CHs in $\mathcal{P}_i$, and the request can be accepted only if CH i and all the CHs in $\mathcal{P}_i$ can accept it.

For the sink, having one more sensor node associated to it increases N_1 to $N_1 + 1$.

Meanwhile, this reduces its inter-CH receiving time by η, since traffic of this sensor will not be forwarded by another CH. The sink can accept one more sensor node association request if and only if the following two conditions are satisfied:

$$f(N_1+1,\eta)P_r + (T_{R,1}-\eta)P_r \leq E_{\max,1}, \tag{2.30}$$

$$f(N_1+1,\eta) + (T_{R,1}-\eta) \leq T_{SF}. \tag{2.31}$$

Based on the above discussions, we design three sensor association schemes below.

2.3.1 Association Scheme for Maximizing η

The associations are performed in iterations. After the initialization process, each iteration includes two phases. The first phase is for sensor node reassociations among non-sink CHs. The second phase reduces the target η value if it is not achieved in the first phase.

Initialization. i) Each sensor node in the overlapping coverage area of multiple CHs randomly chooses a CH to associate. For a sensor node that is in the overlapping area of the sink and another CH, it associates to the non-sink CH. ii) Each CH finds N_i, the total number of sensor nodes associated to it, based on received association signalling messages. iii) Assuming the sink is the throughput bottleneck, the maximum possible value of η can be found from the following two expressions, whichever gives the smaller value:

$$f(N_1,\eta) + (N_{total}-N_1)\eta = T_{SF} \tag{2.32}$$

$$f(N_1,\eta)P_r + (N_{total}-N_1)\eta P_r = E_{\max,1}. \tag{2.33}$$

Note that $T_{L,1} = f(N_1, \eta)$, and

$$T_{R,1} = \sum_{i \neq 1} T_{L,i} S_i = \sum_{i \neq 1} N_i \eta = (N_{total} - N_1)\eta. \tag{2.34}$$

The sink does the calculation and broadcasts the calculated η to all other CHs. This is the initial target value for η. iv) Upon receiving the η value, each CH calculates its $T_{L,i}$. v) After this, the CHs pass their N_i and $T_{L,i}$ values to their parent CHs. This is first done at the lowest level CHs, and then hop by hop until the sink. During this process, every CH calculates its $T_{R,i}$ and $T_{T,i}$, checks its overload status, and reports the status to the sink hop by hop.

Phase 1: reassociations. If any CH is overloaded, the target η cannot be achieved, and reassociations are performed between neighboring CHs. In this phase, the re-associations are performed between non-sink CHs, since this does not change the load at the sink, and, if successful, the current desired η can be achieved. All CHs broadcast about their overload/underload information. Sensors that are currently associated with an overloaded CH look for a neighboring underloaded CH and send reassociation requests to the CH. If a CH receives multiple re-association requests, it processes them one by one. If a re-association request is rejected, the original CH is notified. The process is completed when no more reassociation can be done.

Phase 2: reducing desired η. If there is still at least one CH overloaded, the desired η is reduced to $\eta(1 - e)$, where $0 < e < 1$ is a small number close to 0, such as $e = 0.02$. The exact value of e depends on the required accuracy. Given the new desired η, the maximum N_1 can be found from (2.32) and (2.33), whichever gives the smaller one. The difference between this calculated value of N_1 and the current N_1 gives the maximum number of sensors that can reassociate to the sink from other CHs. The sink notifies the neighboring

CHs, which inform their associated sensors to reassociate to the sink. Any sensors that are in the coverage area of the sink but are associated to another CH can send a reassociation request. Eventually, the number of re-associations can also be limited by the actual number of sensors in the overlapping areas. The new η is broadcast to other CHs hop by hop, and a new iteration starts. This process is repeated until all sensor nodes find their associated CHs.

2.3.2 Association Scheme for Energy Balancing

This is performed when all sensors have completed the associations and no CH is overloaded. At this time, each CH calculates its energy consumption, and neighboring CHs exchange their energy consumption information. Upon receiving the messages, CH i calculates

$$E_{ave,i} = \frac{E_i + \sum_{k \in \mathcal{B}_i} E_k}{|\mathcal{B}_i| + 1}, \tag{2.35}$$

where $\mathcal{B}_i$ is a set of neighboring CHs of CH i. If E_i is larger than $E_{ave,i}$ by a certain threshold, one sensor currently associated to CH i is informed to reassociate to a neighboring CH with lower energy consumption, if the sensor node is located in the overlapping area of the two CHs. The following algorithm is performed distributively at each CH.

1: Let $\mathcal{B}_i' = \mathcal{B}_i$.
2: **while** $\frac{E_i - E_{ave,i}}{E_{ave,i}} > \delta$ and $\mathcal{B}_i' \neq \emptyset$ **do**
3: Find $k = \arg\min_{j \in \mathcal{B}_i} E_j$
4: **while** $n_{ik} > 0$ **do**
5: One sensor in CH i sends a reassociation request to CH k
6: **if** The request is accepted **then**

7: CH i updates $n_{ik} = n_{ik} - 1$ and CH k updates $n_{ki} = n_{ki} + 1$. N_i and N_k are updated accordingly.

8: CHs i and k update E_i and E_k, respectively, and their timeline allocations.

9: For all $j \in \mathcal{P}_i$, CH j updates its timeline allocations and energy consumptions.

10: **end if**

11: **end while**

12: $\mathcal{B}' = \mathcal{B}' \setminus \{k\}$

13: **end while**

Note that the reassociation process can be time consuming. However, this is not done dynamically, but only at the time when the network is first set up. Once this is done, the association relation between the CHs and the sensors as well as the CH timeline allocations are fixed unless the traffic conditions are changed.

2.3.3 Association Scheme for Maximizing Network Throughput

Initially, all sensors associate to the closest CHs, and sensors in overlapping area of multiple CHs randomly associate to one of the CHs. After this process, each cluster can find N_i, the total number of sensors currently associated to it. Based on N_i and the required throughput, the CH can find the minimum amount of CAP duration as

$$T_{L,i} = f(N_i, \eta_i). \tag{2.36}$$

Meanwhile, each child CH passes the value of N_i (total number of sensor nodes associated to it) to its parent, and this is done from the lowest level CHs to the sink hop by hop. When this process ends, each CH can find the required time for inter-CH transmissions.

Given the intra-cluster traffic load, the maximum amount of time remaining for inter-CH transmissions at CH i is $T_{inter,i} = T_{SF} - T_{L,i}$. Meanwhile, the maximum energy available after collecting the intra-cluster traffic is

$$E_{inter,i} = E_{i,max} - T_{L,i}P_r. \tag{2.37}$$

The total amount of energy consumption for inter-CH transmissions is

$$T_{R,i}P_r + T_{T,i}P_{T,i} = T_{R,i}(P_r + P_{T,i}) + S_i T_{L,i} P_{T,i} \tag{2.38}$$

In order to support the current sensor associations, the following conditions must hold

$$T_{R,i}(P_r + P_{T,i}) + S_i T_{L,i} P_{T,i} \leq E_{inter,i} \tag{2.39}$$

and

$$T_{R,i} + T_{T,i} + T_{L,i} = 2T_{R,i} + (1 + S_i)T_{L,i} \leq T_{SF} \tag{2.40}$$

If both (2.39) and (2.40) are satisfied for all i, then all sensor nodes can be served, the process for throughput maximization is completed. Otherwise, the following reassociation scheme is performed, and the process is repeated at each CH until both the conditions are satisfied.

The CHs that have both (2.39) and (2.40) satisfied with inequality may have extra resources available for accepting more sensor nodes. However, this is subject to the resource availability of the CHs along its path to the sink. Therefore, in order to determine whether each CH is able to accept more sensor nodes, resource availability messages are passed from the sink, hop by hop, to the lowest level CHs.

Adding one extra sensor node to cluster i increases not only the local traffic of CH i, but also the inter-CH traffic of all CHs along its path to the sink. Let $T'_{L,i} = f(N_i + 1, \eta_i)$, and $T'_{R,j} = T_{R,j} - S_i T_{L,i} + S_i T'_{L,i}$. $T'_{L,i}$ is the required CAP time if N_i is increased to $N_i + 1$, and $T'_{R,j}$ is the new inter-CH receiving time for CH j along the path between CH i to the sink. Then CH i can accept one more sensor node only if

$$T_{R,i}(P_r + P_{T,i}) + S_i T'_{L,i} P_{T,i} \leq E_{inter,i} \tag{2.41}$$

$$2T_{R,i} + (1 + S_i)T'_{L,i} \leq T_{SF} \tag{2.42}$$

and

$$T'_{R,j}(P_r + P_{T,j}) + S_j T_{L,j} P_{T,j} \leq E_{inter,j} \tag{2.43}$$

$$2T'_{R,j} + (1 + S_j)T_{L,j} \leq T_{SF} \tag{2.44}$$

for all CH $j \in \mathcal{P}_i$.

The CH broadcasts to its associated sensors, which start listening beacon signals from neighboring CHs. Those located in the overlapping areas of multiple CHs receive beacons from other CHs. There can be different ways for a sensor node to do reassociation, either the current associated CH informs the sensor node to associate to a neighboring CH, or the sensor node looks for a new CH to associate after receiving an overload message from the current CH. In either way, we are concerned about the number of sensors associated to each CH, but not which sensors associated to which CHs.

For a sensor node located in the coverage overlapping area of both CH i and CH k, after receiving an overload message from CH i, it switches to listen to CH k. If CH k is not overloaded, it sends an association message to CH k and waits for a response. Upon

receiving the association message, CH k checks whether it can accept one more sensor node. The sensor then receives an association acknowledgement from CH k and switches to CH k after sending a disassociation message to CH i.

It is also possible that CH i is overloaded and all neighboring CHs are also overloaded. In this case, CH i can either remove sensors from its current associated ones until it is not overloaded, or try to reduce its traffic load for inter-CH communications by informing its directly connected child CHs to reduce their local traffic loads. We consider only the former case in this work.

2.4 Numerical Results

We consider a WSN with the cluster-tree topology as shown in Fig. 2.3, where there are 7 clusters with the root at CH 1. Sensor nodes are not shown in the figure for clarity. The distance between any two directly communicating CHs is 1 unit. The transmission power for CH i to reach its parent CH is given by $P_{T,i} = P_0 d_{ij}^{\alpha}$, where CH j is the parent CH of CH i and d_{ij} is the distance between the two CHs. The values of N_{ii}'s are uniformly distributed between 0 and 10, and that of N_{ij}'s for $i \neq j$ are uniformly distributed between 0 and r which varies during the simulation in order to change the total number of sensor nodes in the network. Default parameters are listed in Table 2.1. We compare the proposed sensor association schemes with the corresponding optimum solutions as well as two other straightforward schemes, referred to as "C1" and "C2". For C1, each sensor node always associates to the closest CH, and for C2, the sensor node always associates to the CH with the smallest number of hops to the sink.

In Figs. 2.4 and 2.5 we compare the proposed association scheme for maximizing η with other association solutions, with Fig. 2.4 showing the average throughput per sensor

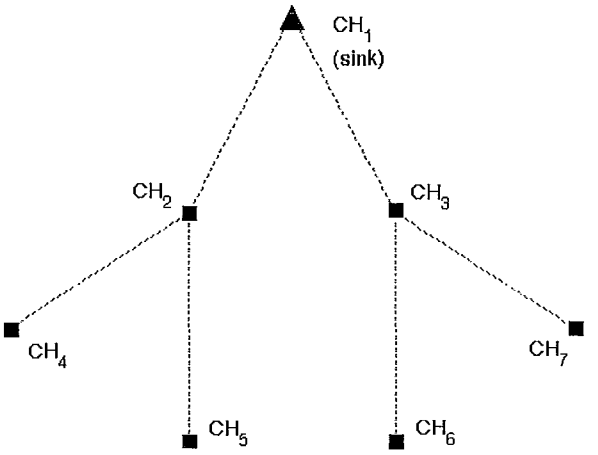

Figure 2.3: CH distribution

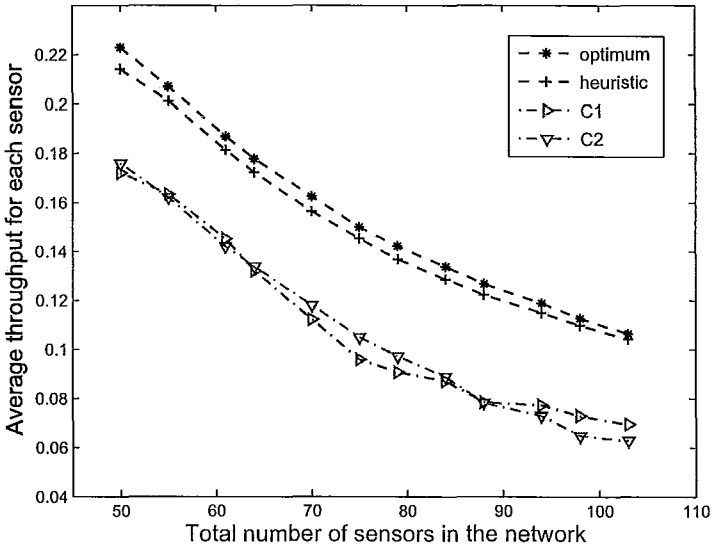

Figure 2.4: Comparison of schemes for maximizing η: per sensor node throughput

Table 2.1: Default Simulation Parameters

Parameter	Value
CH receive power, P_r	30 mW
Normalized CH transmission power, P_0	40 mW
Path loss exponent, α	2
Per node throughput requirement, η	$0.1T_p$/SF
Superframe duration, T_{SF}	$15T_p$
CH Max. energy consumption per SF, $E_{\max,i}$	400 mW $\times$ time slot

node and Fig. 2.5 showing the total throughput in the network. Both figures show that the performance of the proposed scheme is quite close to the optimum results and much better than both C1 and C2.

Meanwhile, we also find that the average throughput for each sensor decreases with the total number of sensors in the network. This is due to the increased number of sensor nodes to share the total available network resources. On the other hand, we find from Fig. 2.5 that the total throughput keeps almost constant using the proposed scheme or the optimum solution, while decreases using either C1 or C2, as the number of sensor nodes increases. This is because both C1 and C2 are unable to resolve the increased collisions during the CAP transmissions, which reduce the channel efficiency. On the other hand, the proposed scheme can adjust the number of associated sensor nodes to each CH so that to best utilize the network resources.

Fig. 2.6 compares performance of the proposed scheme for balancing energy consumption to the optimum solution as well as schemes C1 and C2. First, we find that the proposed scheme achieves energy consumption performance very close to the optimum solution, both in terms of the average energy consumption and the total energy consumption. The figure shows that all the schemes achieve approximately the same average energy consumption.

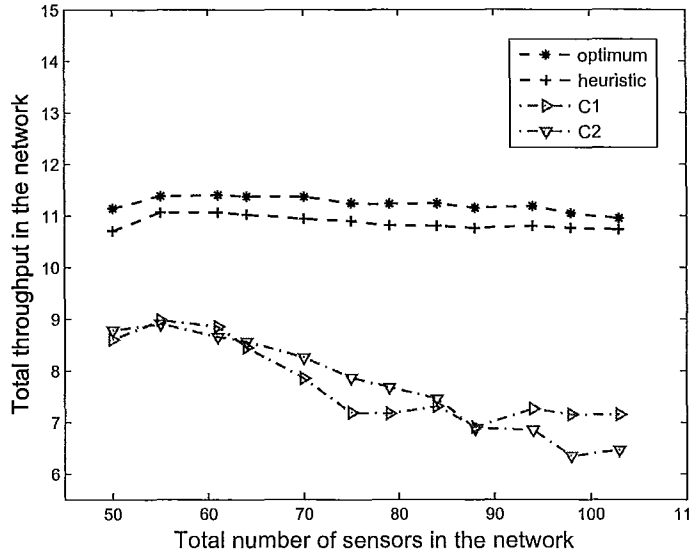

Figure 2.5: Comparison of schemes for maximizing η, total throughput

This is due to that the traffic load is relatively low and collisions during the CAP transmissions are low. Therefore, the amount of network resources is not very much affected by number of sensor nodes associated to different CHs. On the other hand, both C1 and C2 result in much higher maximum energy consumption than the proposed scheme. This is due to that the traffic loads at different CHs are not balanced using C1 or C2. Note that in a practical system, the maximum energy consumption determines the network lifetime. Therefore, using C1 or C2 leads to much shorter network lifetime than the proposed scheme.

In addition, we fix the total number of sensors at $N_{total} = 86$ and vary the percentage (ratio) of sensor nodes in the overlapping areas. Given the total number of sensor nodes, the number of sensor nodes in the overlapping areas affects the performance of the reassociation schemes. When the ratio increases, the energy consumption of the proposed scheme

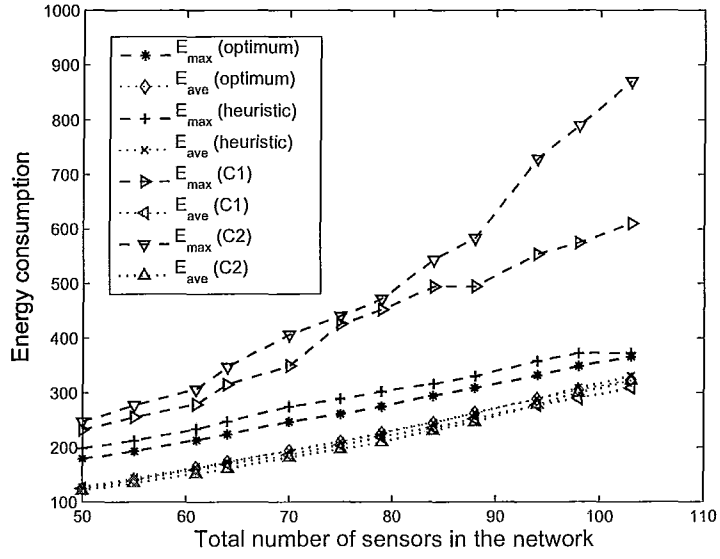

Figure 2.6: Comparison of schemes for CH energy balancing: energy consumption vs. total number of sensor nodes

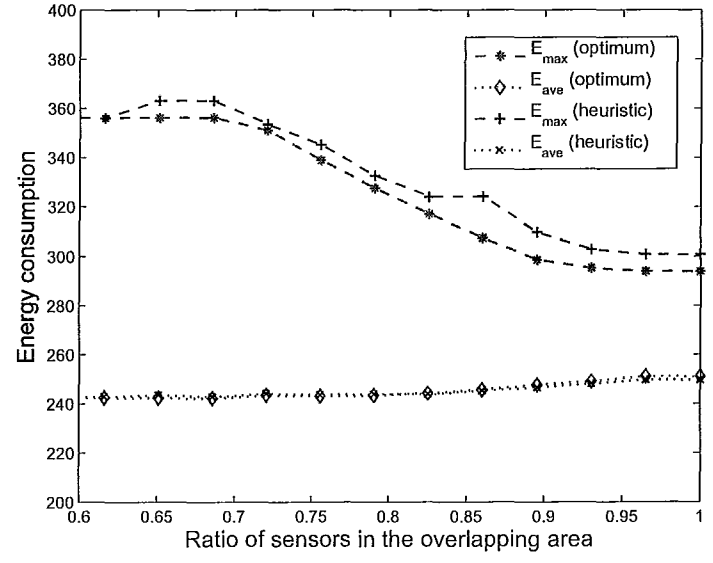

Figure 2.7: Comparison of schemes for CH energy balancing: energy consumption vs. ratio of sensor nodes in overlapping areas

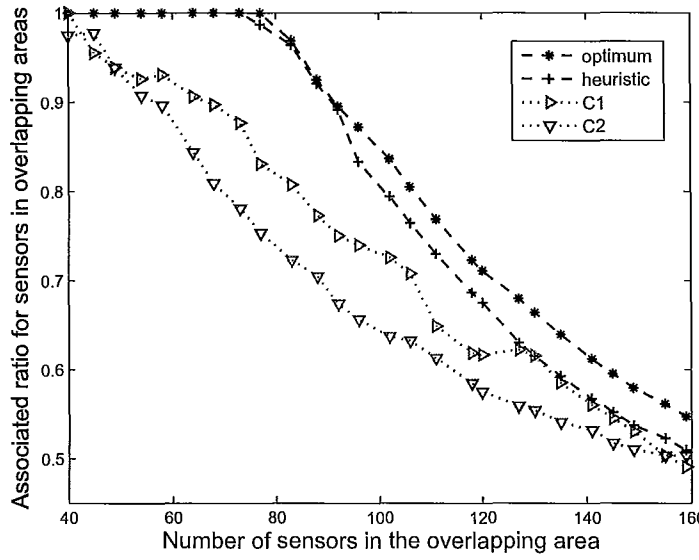

Figure 2.8: Comparison of schemes for maximizing total throughput: the associated ratio in overlapping area

is very close to the optimum solution as shown in Fig. 2.7, in terms of both average energy consumption per CH and total energy consumption of all the CHs. We have also found that as the number of sensor nodes in the overlapping areas increases, the total energy consumption of all the CHs decreases. This is due to that more sensor nodes can reassociate to different CHs, making it possible for better adjusting the energy consumption of the CHs. Meanwhile, the average CH energy consumption increases slightly due to the increased minimum energy consumption.

In Figs. 2.8 and 2.9 we compare our proposed scheme for maximizing the total throughput in the network with other association solutions. In Fig. 2.8, we fix $\sum_{\text{all } i} N_{ii} = 30$, and vary the number of sensors in the overlapping areas. Fig. 2.8 shows the associated ratio of the sensors when the total number of sensors in the overlapping area increases, where the associated ratio r= (the number of the associated sensors in the overlapping area) / (total

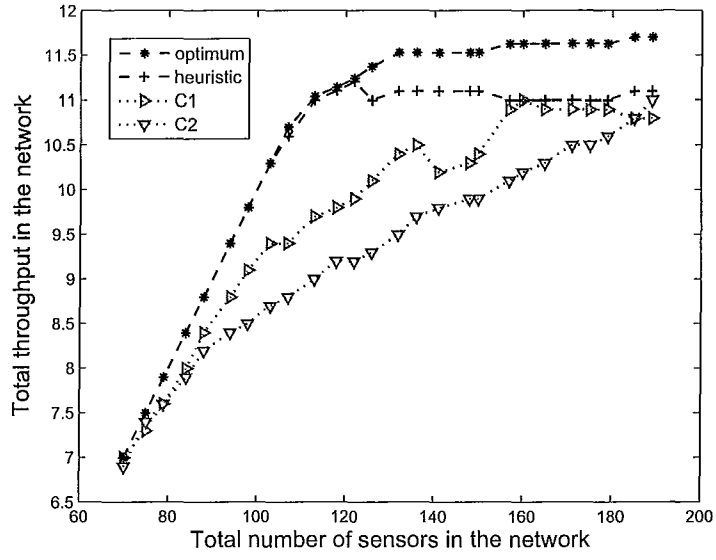

Figure 2.9: Comparison of schemes for maximizing total throughput: total throughput vs total number of sensor nodes

number of sensors in the overlapping area). We notice that if there are not too many sensors in the overlapping area, all these sensors can be associated with a CH. When the number of sensors in the overlapping area increases, especially more than 80, the association ratio of these sensors drops fast.

Fig. 2.9 shows the total throughput in the network versus the total number of the sensors in the network. From this figure we can see that when there are not too many sensors in the network and all the sensors in the overlapping area can be associated, the total throughput is proportional to the total number of sensors in the network, as the throughput is $\sum_i N_i \eta_i$ and $\eta_i = 0.1$ for all CHs. When the number of sensors in the network goes up, the throughput becomes a constant since the total number of sensor nodes that can associate to each CH is upper limited for given η_i requirement.

2.5 Summary

In this chapter, we have formulated three optimum sensor association problems for maximizing per sensor node throughput, balancing CH energy consumption and maximizing network throughput. Heuristic sensor association schemes have been proposed to achieve these objectives. Numerical results have demonstrated that the proposed schemes achieve close-to-optimum performance. In addition, these association schemes achieve much better throughput and energy performance than the straightforward association schemes when the sensor nodes choose to associate to either the closest CH or the one with the smallest number of hops to the sink.

Chapter 3

Delay Analysis in a WSN with Cluster Tree Topology

In a typical wireless sensor network with multiple clusters, the CHs form a relatively static infrastructure with a certain topology, and the sensor nodes select their directly associated CHs. The sensors in the coverage area of a given cluster send their data to the CH, which further forwards the data to the sink through other CHs via one or multiple hops. Therefore, the end-to-end delay includes local transmission delay between the sensors and the associated CHs and inter-cluster delay between different CHs. In general, the end-to-end transmission delay does not only depend on the resource allocations within each cluster, but also that between the CHs. In this chapter we study the delay performance in a WSN with the cluster tree topology and find the relationship between the end-to-end transmission delay and the timeline allocations of the CHs at different levels of the cluster tree. The remainder of the chapter is organized as follows. We first describe the system on which this work is based in Section 3.1. The distribution of the local transmission delay is derived in Section 3.2, and the distribution of the inter-CH transmission delay and the end-to-end

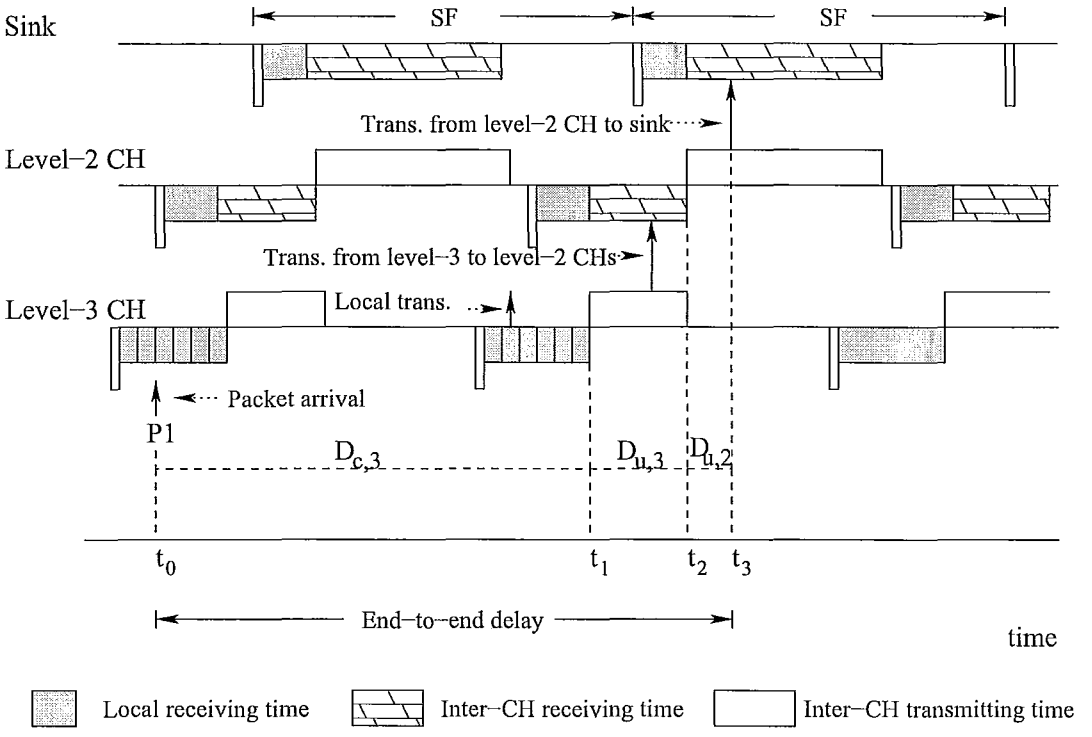

Figure 3.1: Superframe structure and timeline allocations

transmission delay is derived in Section 3.3. Numerical results are demonstrated in Section 3.4, where simulation results are used to verify the analysis. Section 3.5 concludes this chapter.

3.1 System Description

We consider a WSN with the cluster-tree topology, where the sink (which is also a CH) and other CHs form a multi-level wireless backbone. Each CH periodically broadcasts beacons which the sensors and lower level CHs use to establish the associations with it. Sensors (or lower level CHs) also receive the timeline arrangement information from their associated CH (or parent CH) and decide when they can transmit. A number of sensors are associated

to each CH and communicate directly with the CH. The CHs collect data from the sensors and then forward the data through other CHs hop-by-hop to the sink. In a typical WSN, most traffic is from the sensors to the sink, and traffic in the opposite direction is mainly for control signalling, which affects little on the system traffic load and data transmission delay. Therefore, we only consider traffic transmissions from the sensors. We define local traffic as the traffic from the sensors to their directly associated CHs, and forwarded traffic as the traffic between CHs.

As shown in Fig. 3.1, channel time of each CH is divided into equal length superframes. For CH i, each of its superframes is further divided into multiple intervals, the local receiving time with duration $T_{L,i}$ for receiving the local traffic, inter-cluster receiving time with duration of $T_{R,i}$ for receiving traffic from its directly connected child (lower level) CHs, and the inter-cluster transmitting time with duration $T_{T,i}$ for forwarding to the parent CH at the higher level. For a lower level CH, its inter-cluster transmitting time corresponds to the inter-cluster receiving time of its parent CH. The end-to-end delay is defined between the time when the packet is generated to the time when it is received by the sink. Fig. 3.1 illustrates the end-to-end delay of packet P1, where t_0 is the time when the packet is generated, and t_3 is the time when it is received at the sink. It is further divided into local transmission delay and inter-CH transmission delay. Depending on whether the receiver is the sink, the local and inter-CH transmission delay is defined differently. Below we define the local transmission delay and the inter-CH transmission delay.

For sensors directly associated to the sink, their packet tran.smissions are finished as soon as the packets reach the sink. Therefore, its local transmission delay is from the time when it is generated to the time when it is received at the sink. For a non-sink CH i, the local traffic collected by the CH should first be buffered in the CH, and the buffered packets

will be delayed and forwarded to the parent CH during the inter-cluster transmission time. The local transmission delay is defined to include the time from the moment when a packet is generated at the sensor to the starting time of the first inter-cluster transmission time of CH i (the directly associated CH) after the packet is received by the CH. In Fig. 3.1, the local transmission delay for packet P1 is from t_0 to t_1, even the packet is transmitted to the local CH earlier than t_1.

When the parent CH is not the sink, the inter-CH delay of a packet from CH i to CH j is from the start of the first inter-cluster transmitting time of CH i after the packet is received at CH i to the end of the inter-cluster transmitting time of CH i after the packet is received by CH j. In Fig. 3.1, the inter-CH transmission delay for packet P1 from the level-3 CH to the level-2 CH is from t_1 to t_2, even the actual packet transmission time is earlier than t_2. When the parent CH is the sink, the inter-CH transmission delay of a packet from CH i is from the start of the first inter-cluster transmitting time after the packet is received by CH i to the time when the packet is received by the sink. In Fig. 3.1, the inter-CH transmission delay for packet P1 from the level-2 CH to the sink is from t_2 to t_3.

Each CH keeps a buffer to store its received packets. Let T_p be the transmission time for each packet, including all the overhead and the inter-frame time and ACKs. Define $K_{C,i}$ and $K_{F,i}$, respectively, as the normalized values of $T_{L,i}$ and $T_{R,i}$ with respect to T_p, i.e., $K_{C,i} = T_{L,i}/T_p$ and $K_{F,i} = T_{R,i}/T_p$. We consider that both $K_{C,i}$ and $K_{F,i}$ are integers. We consider that $T_{T,i} = T_{R,j}$ for all $i \in C_j$, and all child CHs share the same inter-cluster receiving interval of the parent CH. When one child CH finishes transmitting to the parent CH, the next child CH starts transmitting, and this is done in a round-robin fashion.

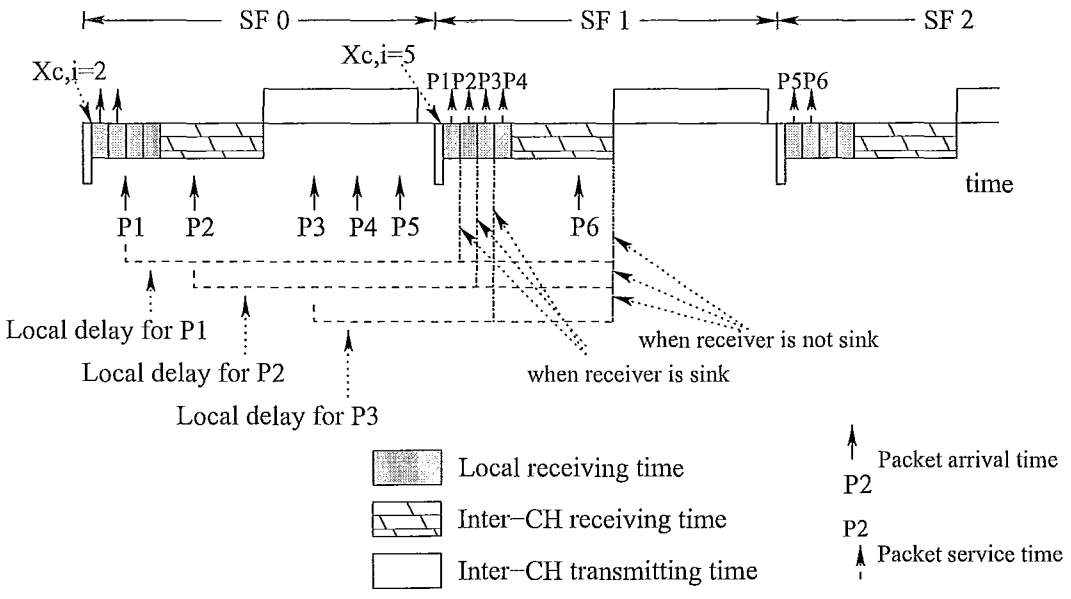

Figure 3.2: Local transmission delay

3.2 Local Transmission Delay Analysis

Each sensor generates data at a random time and stores the data in a buffer. Sensors with packets to be transmitted share the local receiving time of their associated CH on TDMA-basis. At the beginning of each superframe (which is also the beginning of the local receiving time) of CH i, there are a number of minislots for the sensors and lower level CHs to report the number of packets to be transmitted to CH i. When the number of child CHs is small and the local traffic load is low, the mini-slots can share the same timeline as that for packet transmissions. Otherwise, the CH can be equipped with two radios, one dedicated for signalling exchanges and the other for data packet transmissions. In either way, the effect of the reporting process on the packet transmission delay can be neglected. Based on these reports, the CH assigns the local receiving time to the sensors and the inter-CH receiving time to the child CHs. If less than $K_{C,i}$ packets are reported from the sensors,

then all these packets can be transmitted to the CH during the local receiving time in the current superframe. In order for each CH to know at the beginning of a superframe the exact number of local and forwarded packets that it will receive during the superframe, data packets generated after the start of the superframe cannot be transmitted in the local receiving interval of the same superframe. For example, in Fig. 3.2, packet P1 is served in SF 1, even there is available time in the local receiving time of SF 0.

To simplify the analysis, we consider a virtual buffer that stores the data packets generated by all the sensors in the same cluster and then study the average packet transmission delay for the entire cluster. We consider that the aggregate packet arrivals from all the sensors in cluster i follow a Poisson distribution and use $A_{C,i}$ to represent the number of generated packets in one SF. Let $X_{C,i}$ represent the total number of packets in the virtual buffer at the beginning of each SF, and we find that $X_{C,i}$ is a Markov chain. Let x be the value of $X_{C,i}$ at the beginning of the local receiving interval in a given SF, and $x^{'}$ be its value at the beginning of the local receiving interval in the next SF, then we can find the transition probability of $X_{C,i}$ from x to $x^{'}$ based on different cases. First, when $x \leq K_{C,i}$, all the x packets buffered at the beginning of the current SF can be served in the local receiving time of the next SF, and $x^{'}$ is equal to $A_{C,i}$, which is the number of new arrivals during the current SF. This is the example from SF 0 to SF 1 shown in Fig. 3.2. At the beginning of SF 0, $X_{C,i} = 2$. Since $K_{C,i} = 4$, all the buffered 2 packets can be served in the local receiving interval during SF 0. Since 5 packets arrive in SF 0, $A_{C,i} = 5$, and then $X_{C,i} = 5$ at the beginning of SF 1. Second, when $x > K_{C,i}$ in the current SF, only $K_{C,i}$ of the packets can be transmitted to CH i in the next SF and the remaining $x - K_{C,i}$ packets are still be buffered. Then, $x^{'} = x - K_{C,i} + A_{C,i}$. This is the example from SF 1 to SF 2 shown in Fig. 3.2. Since $X_{C,i} = 5$ at the beginning of SF 1, and $K_{C,i} = 4$, only 4 of

the buffered packets can be served in the local receiving interval of SF 1. In addition, one packet (P6) arrives during SF 1. Therefore, $X_{C,i} = 5 - 4 + 1 = 2$ at the beginning of SF 2. Given these observations, the transition probability of $X_{C,i}$ is given by

$$\begin{aligned} Q^{C}_{xx',i} &= \Pr.\{X_{C,i} = x' | X_{C,i} = x\} \\ &= \begin{cases} \Pr.\{A_{C,i} = x'\}, & \text{if } x \leq K_{C,i} \\ \Pr.\{A_{C,i} = x' - x + K_{C,i}\}, & \text{otherwise.} \end{cases} \end{aligned} \tag{3.1}$$

Given the mean of $A_{C,i}$, the steady state distribution of $X_{C,i}$, $\Pi_{X_{C,i}} = \Pr.\{X_{C,i} = x\}$, can be found.

3.2.1 Average Local Transmission Delay

Consider a tagged packet that arrives t ($0 < t \leq T_{SF}$) time after the beginning of a reference SF, for example, SF 0. The packet should first wait for $T_{SF} - t$ until the beginning of the local receiving time in the next SF, SF 1. The total number of buffered packets at the start of SF 1 is given by $(X_{C,i} - K_{C,i})^+ + A_{C,i}$, among which $(X_{C,i} - K_{C,i})^+ + A_{t,i}$ packets arrive before the tagged packet, where $A_{t,i}$ is the number of packets that arrival from the beginning of SF 0 to time t. In addition to the waiting time $T_{SF} - t$, the delay for serving the tagged packet also includes the time for serving all the $(x - K_{C,i})^+ + A_{t,i}$ packets, and the time for transmitting the tagged packet. That is, the total waiting time is $T_{SF} - t$ plus the time for serving the $(x - K_{C,i})^+ + A_{t,i} + 1$ packets. When this number is larger than $K_{C,i}$, it takes more than one SF until the tagged packet is transmitted.

For sensors directly associated to the sink, the end-to-end delay is equal to the local transmission delay. Given that $X_{C,1} = x$ and that a packet arrives t time after the beginning

of a SF, the delay for transmitting the packet is given by

$$D_{C,1}|_{x,t} = (T_{SF} - t) + \beta_1 T_{SF} + \left[(x - K_{C,1})^+ + A_{t,1} + 1 - \beta_1 K_{C,1}\right] T_p, \tag{3.2}$$

where the second and third terms on the right-hand-side of (3.2) together give the total amount of time in order to transmit all the $(x - K_{C,1})^+ + A_{t,1} + 1$ packets,

$$\beta_i = \left\lceil \frac{(x - K_{C,i})^+ + A_{t,i} + 1}{K_{C,i}} \right\rceil - 1 \tag{3.3}$$

for $i = 1$ (the sink) gives the total number of full SFs that the tagged packet should wait before it is transmitted, and $[(x - K_{C,1})^+ + A_{t,1} + 1 - \beta_1 K_{C,1}]$ gives the number of packets transmitted in the same SF as the tagged packet. Consider P3 in Fig. 3.2. It arrives during SF 0, when $X_{C,i} = 2$ ($x = 2$), $K_{C,i} = 4$, and therefore $(x - K_{C,i})^+ = 0$. In addition, 2 packets arrive before P3 during SF 0, that is, $A_{t,i} = 2$. Therefore, $\beta_i = 0$ for P3. Similarly, $\beta_i = 0$ for P1 and P2, and P4. For P5, there are 4 packets that arrive before it during SF 0, and therefore $A_{t,i} = 4$ and $\beta_i = 1$

Given (3.2), the unconditional mean delay is given by

$$\bar{D}_{C,1} = \int_0^{T_{SF}} \sum_{a,x=0}^{\infty} D_{C,1}|_{x,t} \Pr.\{X_{C,1} = x\} \Pr.\{A_{t,1} = a\} \frac{1}{T_{SF}} dt, \tag{3.4}$$

where $\frac{1}{T_{SF}}$ is due to that t is uniformly distributed between 0 and T_{SF}.

For a non-sink cluster i, given that a packet arrives t time after the beginning of a reference SF, the conditional delay is given by

$$D_{C,i}|_{x,t} = (T_{SF} - t) + \beta_i T_{SF} + T_{L,i} + T_{R,i}, \tag{3.5}$$

and the unconditional mean is given by

$$\bar{D}_{C,i} = \int_0^{T_{SF}} \sum_{a=0}^{\infty} \sum_{x=0}^{\infty} D_{C,i}|_{x,t} \Pr.\{X_{C,i} = x\} \Pr.\{A_{t,i} = a\} \frac{1}{T_{SF}} dt. \tag{3.6}$$

3.2.2 Distribution of Local Transmission Delay

We first derive the probability distribution of $D_{C,1}$ and then the distribution of $D_{C,i}$ for $i \neq 1$.

Distribution of $D_{C,1}$. Although the actual delay is a continuous variable, we approximate it by considering a discrete case for simplicity. Later on the numerical results will show that the error caused by such approximation can be neglected, compared to the delay results from computer simulation. Assume that a packet can only arrive at an integer multiple of T_p after the start of each SF. In this way, $D_{C,1}$ can only take values of $n_1 T_{SF} + n_2 T_p$ for non-negative integers n_1 and n_2. Define $\tilde{t} = \frac{t}{T_p}$ and $K_{SF} = \frac{T_{SF}}{T_p}$. We derive the distribution for different values of n_1 and n_2 as follows.

1. When $n_1 = 0$ and $0 < n_2 \leq K_{C,i}$, $D_{C,1}|_{x,t} < T_{SF}$. From (3.2) we can find that $\beta_1 = 0$, and

$$n_2 = K_{SF} - \tilde{t} + \left[(x - K_{C,1})^+ + A_{t,1} + 1\right], \tag{3.7}$$

from which we have

$$A_{t,1}|_{X_{C,1}=x} = \tilde{t} - K_{SF} - (x - K_{C,1})^+ - 1 + n_2 \triangleq \tilde{a}_{t,1}. \tag{3.8}$$

Thus the distribution of $D_{C,1}$ can be found as

$$\Pr.\{D_{C,1} = n_2 T_p\} = \sum_{\tilde{t}=K_{SF}-(n_2-1)}^{K_{SF}} \sum_{x=0}^{\infty} \frac{1}{K_{SF}} \Pr.\{A_{t,1}|_{X_{C,1}=x} = \tilde{a}_{t,1}\} \Pr.\{X_{C,1} = x\} \quad (3.9)$$

where $\frac{1}{K_{SF}}$ is due to the fact that the arrival time of the tagged packet is equally likely distributed between 1 and K_{SF}, and the lower limit of $\tilde{t}$ is to ensure $\tilde{a}_{t,1} \geq 0$.

2. When $n_1 = \beta_1$ and $K_{C,1} + 1 \leq n_2 < K_{SF}$. Since $1 \leq (x - K_{C,1})^+ + A_{t,1}|_{X_{C,1}=x} + 1 - \beta_1 K_{C,1} \leq K_{C,1}$, from (3.2) we can find that

$$n_2 = K_{SF} - \tilde{t} + \left[(x - K_{C,1})^+ + A_{t,1}|_{X_{C,1}=x} + 1 - \beta_1 K_{C,1}\right], \quad (3.10)$$

from which we can find $A_{t,1}|_{X_{C,1}=x}$ as

$$A_{t,1}|_{X_{C,1}=x} = \tilde{t} - K_{SF} - (x - K_{C,1})^+ - 1 + n_1 K_{C,1} + n_2 \triangleq \tilde{a}_{t,2}. \quad (3.11)$$

Thus the distribution of $D_{C,1}$ can be found as

$$\Pr.\{D_{C,1} = n_1 T_{SF} + n_2 T_p\} = \sum_{\tilde{t}=K_{SF}-(n_2-1)}^{K_{SF}-(n_2-K_{C,1})} \sum_{x=0}^{\infty} \frac{1}{K_{SF}} \times \Pr.\{A_{t,1}|_{X_{C,1}=x} = \tilde{a}_{t,2}\} \Pr.\{X_{C,i} = x\}. \quad (3.12)$$

As the value of the third term on the right-hand side of (3.2) is between 0 and $K_{C,1}T_p$, the range constraint of $\tilde{t}$ for the first summation on the right-hand side of (3.12) guarantees that the condition $K_{C,1} + 1 \leq n_2 < K_{SF}$ is satisfied.

3. When $n_1 \geq 1$ and $n_2 = 0$, then

$$(T_{SF} - t) + \beta_1 T_{SF} + \left[(x - K_{C,1})^+ + A_{t,1} + 1 - \beta_1 K_{C,1}\right] T_p = n_1 T_{SF}. \quad (3.13)$$

Then $n_1 = \beta_1 + 1$, and

$$-\tilde{t} + \left[(x - K_{C,1})^+ + A_{t,1}|_{X_{C,1}=x} + 1 - \beta_1 K_{C,1}\right] = 0, \quad (3.14)$$

which is equivalent to

$$A_{t,1}|_{X_{C,1}=x} = \tilde{t} - (x - K_{C,1})^+ - 1 + (n_1 - 1)K_{C,1} \triangleq \tilde{a}_{t,3}. \quad (3.15)$$

The distribution of $D_{C,1}$ can be found as

$$\Pr.\{D_{C,1} = n_1 T_{SF}\} = \sum_{\tilde{t}=1}^{K_{C,1}-1} \sum_{x=0}^{\infty} \Pr.\{A_{t,1}|_{X_{C,1}=x} = \tilde{a}_{t,3}\} \Pr.\{X_{C,1} = x\} \frac{1}{K_{SF}}. \quad (3.16)$$

Since $0 \leq (x - K_{C,1})^+ < K_{C,1}$, the range for $\tilde{t}$ is from 1 to $K_{C,1} - 1$ in order to make n_2 to be zero in (3.14).

4. When $n_1 \geq 1$ and $0 < n_2 \leq K_{C,1}$, we have either

$$\begin{cases} n_1 = \beta_1, \\ n_2 = K_{SF} - \tilde{t} + \left[(x - K_{C,1})^+ + A_{t,1}|_{X_{C,1}=x} + 1 - \beta_1 K_{C,1}\right], \end{cases} \quad (3.17)$$

or

$$\begin{cases} n_1 = \beta_1 + 1, \\ n_2 = -\tilde{t} + \left[(x - K_{C,1})^+ + A_{t,1}|_{X_{C,1}=x} + 1 - \beta_1 K_{C,1}\right]. \end{cases} \tag{3.18}$$

In the first sub-case, we have

$$A_{t,1}|_{X_{C,1}=x} = \tilde{t} - K_{SF} - (x - K_{C,1})^+ - 1 + n_1 K_{C,1} + n_2 \triangleq \tilde{a}_{t,41}. \tag{3.19}$$

In the second sub-case,

$$A_{t,1}|_{X_{C,1}=x} = \tilde{t} - (x - K_{C,1})^+ - 1 + (n_1 - 1)K_{C,1} + n_2 \triangleq \tilde{a}_{t,42}. \tag{3.20}$$

The distribution of $D_{C,1}$ can be found as

$$\begin{aligned} \Pr.\{D_{C,1} &= n_1 T_{SF} + n_2 T_p\} \\ &= \sum_{\tilde{t}=K_{SF}-(n_2-1)}^{K_{SF}} \sum_{x=0}^{\infty} \frac{1}{K_{SF}} \Pr.\{A_{t,1}|_{X_{C,1}=x} = \tilde{a}_{t,41}\} \Pr.\{X_{C,1} = x\} \\ &+ \sum_{\tilde{t}=1}^{K_{C,1}-1-n_2} \sum_{x=0}^{\infty} \frac{1}{K_{SF}} \Pr.\{A_{t,1}|_{X_{C,1}=x} = \tilde{a}_{t,42}\} \Pr.\{X_{C,1} = x\}. \end{aligned} \tag{3.21}$$

The lower and upper limits of $\tilde{t}$ on the right-hand side of (3.21) ensure that condition $0 < n_2 \leq K_{C,1}$ is satisfied in each sub-case.

Distribution of $D_{C,i}$ for $i \neq 1$. After finishing deriving the distribution of $D_{C,1}$, next we derive the distribution of $D_{C,i}$ for $i \neq 1$, which is the local transmission delay for a non-sink cluster. Again, we consider that t can only take an integer multiple of T_p, in which case $D_{C,i}$ takes discrete values. Below we find the probability of $D_{C,i} = n_1 T_{SF} + T_{L,i} + T_{R,i} - n_2 T_p$

for integer values of n_1 and n_2. According to equation (3.5),

$$\begin{aligned}(T_{SF}-t)+\beta_i T_{SF}+T_{L,i}+T_{R,i} =&(K_{SF}-\tilde{t})T_p+\beta_i T_{SF}+T_{L,i}+T_{R,i}\\ =&n_1 T_{SF}+T_{L,i}+T_{R,i}-n_2 T_p.\end{aligned} \tag{3.22}$$

Since $T_{L,i}$ and $T_{R,i}$ are all integer multiples of T_p, we have $n_1 = \beta_i + 1$, and $n_2 = \tilde{t}$ for $0 \leq n_2 < K_{SF}$. Based on (3.3) we have $\frac{(X_{C,i}-K_{C,i})^+ + A_{t,i}+1}{K_{C,i}} \leq \beta_i < \frac{(X_{C,i}-K_{C,i})^+ + A_{t,i}+1}{K_{C,i}} + 1$. Then

$$(n_1-1)K_{C,i} \leq (X_{C,i}-K_{C,i})^+ + A_{t,i} + 1 < n_1 K_{C,i}. \tag{3.23}$$

Given the value of t and $X_{C,i} = x$, the range of $A_{t,i}|_{X_{C,i}=x}$ can be found as

$$(n_1-1)K_{C,i} - (x-K_{C,i})^+ - 1 \leq A_{t,i}|_{X_{C,i}=x} < n_1 K_{C,i} - (x-K_{C,i})^+ - 1. \tag{3.24}$$

Define a_1 and a_2, respectively, as the lower bound and upper bound of $A_{t,i}|_{X_{C,i}=x}$ in (3.24), we have

$$\begin{cases} a_1 = (n_1-1)K_{C,i} - (x-K_{C,i})^+ - 1, \\ a_2 = n_1 K_{C,i} - (x-K_{C,i})^+ - 1. \end{cases} \tag{3.25}$$

Then the distribution of $D_{C,i}$ is given by

$$\begin{aligned}&\Pr.\{D_{C,i} = n_1 T_{SF} + T_{L,i} + T_{R,i} - n_2 T_p\}\\ &= \sum_{x=0}^{b} \sum_{a_1 \leq a < a_2} \Pr.\{A_{t,i}|_{X_{C,i}=x} = a\} \Pr.\{X_{C,i} = x\} \Pr.\{\tilde{t} = n_2\} \qquad (3.26)\\ &= \sum_{x=0}^{b} \sum_{a_1 \leq a < a_2} \Pr.\{A_{t,i}|_{X_{C,i}=x} = a\} \Pr.\{X_{C,i} = x\} \frac{1}{K_{SF}}. \qquad (3.27)\end{aligned}$$

3.3 End-to-End Transmission Delay Analysis

Consider CH i. The total number of packets that it receives during an SF is $Z_{F,i} = Y_{C,i} + Y_{F,i}$, where $Y_{C,i}$ is the number of packets collected from the local sensors and $Y_{F,i}$ is the number of packets from the child CHs. $Y_{C,i} = \min\{X_{C,i}, K_{C,i}\}$, and its distribution can be found as

$$\Pr.\{Y_{C,j} = y\} = \begin{cases} \Pr.\{X_{C,i} \geq K_{C,j}\}, & \text{if } y = K_{C,j} \\ \Pr.\{X_{C,i} = y\}, & \text{if } y < K_{C,i} \\ 0, & \text{otherwise} \end{cases} \tag{3.28}$$

Let $\mathcal{C}_j$ denote a set of the child CHs of CH j, and CH i is in $\mathcal{C}_j$. We consider the aggregate traffic from the CHs in $\mathcal{C}_j$, and study the average transmission delay for all packets from these CHs to CH j. All the packets that the child CHs have collected before the inter-CH receiving time of CH j are then waiting to be transmitted to CH j. The total number of these packets is given by $Z_j = \sum_{i \in \mathcal{C}_j} Z_{F,i}$ and its distribution can be found based on the distribution of $Z_{F,i}$.

In addition, there are also packets that have not been transmitted from previous SFs from the child CHs to CH j, and this is denoted as $X_{B,j}$. Together, the total number of packets that are waiting to be transmitted to CH j at the beginning of its inter-cluster transmission interval is

$$X_{F,j} = X_{B,j} + Z_j. \tag{3.29}$$

We find that $X_{B,j}$ is a Markov chain and its transition probability is given by

$$\begin{aligned} Q^B_{xx'} &= \Pr.\{X_{B,j} = x' | X_{B,j} = x\} \\ &= \begin{cases} \Pr.\{Z_j \leq K_{F,j} - x\}, & \text{if } x' = 0 \\ \Pr.\{Z_j = x' - x + K_{F,j}\}, & \text{otherwise} \end{cases} \end{aligned} \tag{3.30}$$

where $K_{F,j} = T_{R,j}/T_p$. The steady-state probability of $X_{B,j}$ can be found based on the above transition probability and the distribution of Z_j.

Denote the number of successfully transmitted packets during the inter-cluster receiving interval of CH j as $Y_{F,j}$, which is given by $Y_{F,j} = \min\{X_{F,j}, K_{F,j}\}$. Then

$$\Pr.\{Y_{F,j} = y | Z_j = z\} = \begin{cases} \Pr.\{X_{B,j} \geq K_{F,j} - z\}, & \text{if } y = K_{F,j} \\ \Pr.\{X_{B,j} = y - z\}, & \text{if } y < K_{F,j} \\ 0, & \text{o.w.} \end{cases} \tag{3.31}$$

Thus the unconditional distribution of $Y_{F,j}$ is given by

$$\Pr.\{Y_{F,j} = y\} = \sum_{z=0}^{\infty} \Pr.\{Y_{F,j} = y | Z_j = z\} \Pr.\{Z_j = z\}. \tag{3.32}$$

3.3.1 Inter-CH Transmission Delay

In this subsection we consider the single-hop transmission delay between CHs. We consider the case when the receiver (parent CH) is the sink first, and then the case when the receiver (parent CH) is not the sink.

The parent CH is the sink. When $X_{F,j} = x$, the inter-CH delay for each of the x buffered

packets at CH i, $i \in \mathcal{C}_j$, can be found as

$$\begin{aligned}
\text{1st packet,} \quad & D_{U,i} = T_p \\
\text{2nd packet,} \quad & D_{U,i} = 2T_p \\
\vdots \quad & \vdots \\
K_{F,j}\text{'th packet,} \quad & D_{U,i} = K_{F,j}T_p \\
(K_{F,j}+1)\text{'th packet,} \quad & D_{U,i} = T_{SF} + T_p \\
\vdots \quad & \vdots \\
2K_{F,j}\text{'th packet,} \quad & D_{U,i} = T_{SF} + K_{F,j}T_p \\
(2K_{F,j}+1)\text{'th packet,} \quad & D_{U,i} = 2T_{SF} + T_p \\
\vdots \quad & \vdots \\
x\text{'th packet,} \quad & D_{U,i} = \left(\lceil \tfrac{x}{K_{F,j}} \rceil - 1\right) T_{SF} + \left[x - \left(\lceil \tfrac{x}{K_{F,j}} \rceil - 1\right) K_{F,j}\right] T_p
\end{aligned}$$

Let $\alpha_j = \left(\lceil \frac{x}{K_{F,j}} \rceil - 1\right)$, the transmission delay for the x'th packet is $\alpha_j T_{SF} + (x - \alpha_j K_{F,j})\, T_p$. Given $X_{F,j} = x$, the conditional one-hop average delay is given by

$$\bar{D}_{U,i}|_{X_{F,j}=x} = \frac{\sum_{k_1=1}^{\alpha_j} \sum_{k_2=1}^{\min\{K_{F,j}, x-\alpha_j K_{F,j}\}} (k_1 T_{SF} + k_2 T_p)}{x}. \tag{3.33}$$

The unconditional one-hop average delay is then given by

$$\bar{D}_{U,i} = \sum_{x=1}^{\infty} \bar{D}_{U,i}|_{X_{F,j}=x} \Pr.\{X_{F,j} = x\}. \tag{3.34}$$

Given integers $n_1 \geq 0$ and $n_2 \geq 0$, when the number of buffered packets is larger than or equal to $n_1 K_{F,j} + n_2$, there is one packet in the buffer with delay $n_1 T_{SF} + n_2 T_p$.

Therefore, we have

$$\Pr.\{D_{U,i} = n_1 T_{SF} + n_2 T_p\} = \sum_{x=n_1 K_{F,j}+n_2}^{\infty} \frac{\Pr.\{X_{F,j} = x\}}{x}. \tag{3.35}$$

The parent CH is not the sink. When $X_{F,j} = x$, the first $K_{F,j}$ packets in the buffer of CH i, $i \in \mathcal{C}_j$, have delay of $D_{U,i} = T_{R,j}$, and the next $K_{F,j}$ buffered packets will be delayed to the next SF and have delay of $D_{U,i} = T_{SF} + T_{R,j}$. Given the total number of packets $X_{F,j} = x$, and define $\xi_j = \lceil x/K_{F,j} \rceil - 1$, the average one-hop transmission delay can be found as

$$\bar{D}_{U,i}|_{X_{F,j}=x} = \frac{\sum_{k_1=0}^{\xi_j} (k_1 T_{SF} + T_{R,j}) \min\{K_{F,j}, x - (\xi_j - 1)K_{F,j}\}}{x}, \tag{3.36}$$

and the unconditional average one-hop transmission delay is

$$\bar{D}_{U,i} = \sum_{x=1}^{\infty} \bar{D}_{U,i}|_{X_{F,j}=x} \Pr.\{X_{F,j} = x\}. \tag{3.37}$$

When the number of buffered packets is x and $x > n_1 K_{F,j}$, there are $\min\{K_{F,j}, x - n_1 K_{F,j}\}$ packets with inter-CH transmission delay of $T_{R,j} + n_1 T_{SF}$. Therefore, the distribution of the one-hop transmission delay is given as

$$\Pr.\{D_{U,i} = T_{R,j} + n_1 T_{SF}\} = \sum_{x=n_1 K_{F,j}+1}^{\infty} \frac{\min\{K_{F,j}, x - n_1 K_{F,j}\}}{x} \Pr.\{X_{F,j} = x\}. \tag{3.38}$$

3.3.2 End-to-end Delay and Drop Rate

The end-to-end delay includes $D_{C,i}$, which is the delay between the sensors and their associated CHs, and the delay between the CHs along the path to the sink. Therefore,

$$D_{e2e,i} = D_{C,i} + \sum_{k \in \mathcal{P}_i} D_{U,k} \tag{3.39}$$

where $\mathcal{P}_i$ is a set of the CHs along the path from CH i to the sink but does not include the sink. The mean and distribution of $D_{e2e,i}$ can be found based on the mean and distribution of $D_{C,i}$ and $D_{U,k}$'s.

Define $D_{\max}$ as the maximum tolerable end-to-end delay, and the packets with delay larger than $D_{\max}$ are dropped. The drop rate for the packets from the sensors associated with CH i can be found as

$$P_{dr,i} = \Pr.\{D_{e2e,i} > D_{\max}\}. \tag{3.40}$$

3.4 Numerical Results

Consider a 3-level sensor network as shown in Fig. 3.3. The local traffic arrivals in each cluster follow the Poisson process. We consider that the average packet arrival rate is the same for all the clusters and equal to λT_{SF} packets per SF. The duration of an SF is $80T_p$.

3.4.1 Local Transmission Delay

In this subsection we demonstrate the local transmission delay for traffic from the clusters at different levels of the cluster tree.

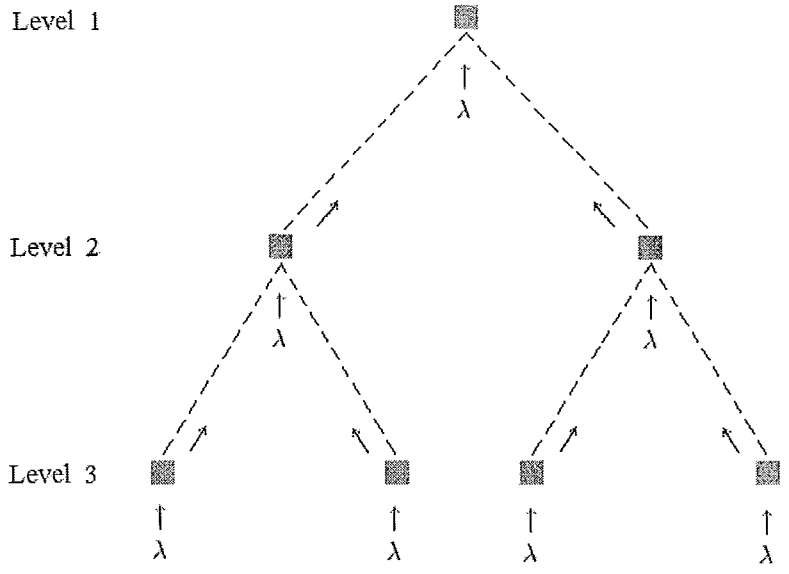

Figure 3.3: Network topology

We first vary $K_{C,3}$ and observe the local transmission delay for the level-3 traffic. Note that $K_{C,3}T_p = T_{L,3}$. Therefore, varying $K_{C,3}$ is equivalent to varying $T_{L,3}$. Fig. 3.4 shows that when $K_{C,3}$ is small, the average local delay can be very large, since the available local transmission time is insufficient for serving all the traffic, and the local traffic service system is unstable. The local transmission delay decreases as $K_{C,3}$ increases. However, when $K_{C,3}$ is sufficiently large, further increasing $K_{C,3}$ has little effect on the local delay.

Fig. 3.5 shows the distribution of local traffic delay for the level-3 CHs. In the figure, we can see that the analytical results match the simulation results very well. A packet has the minimum delay if it is generated at the end of an SF and transmitted to the CH in the next SF. In this case, its local transmission delay is approximately $T_{L,3}(= 8T_p)$. Otherwise, if the packet is generated at the sensor right after the beginning of a SF, it has to wait for a full SF until the start of the following SF, and the minimum local transmission delay for

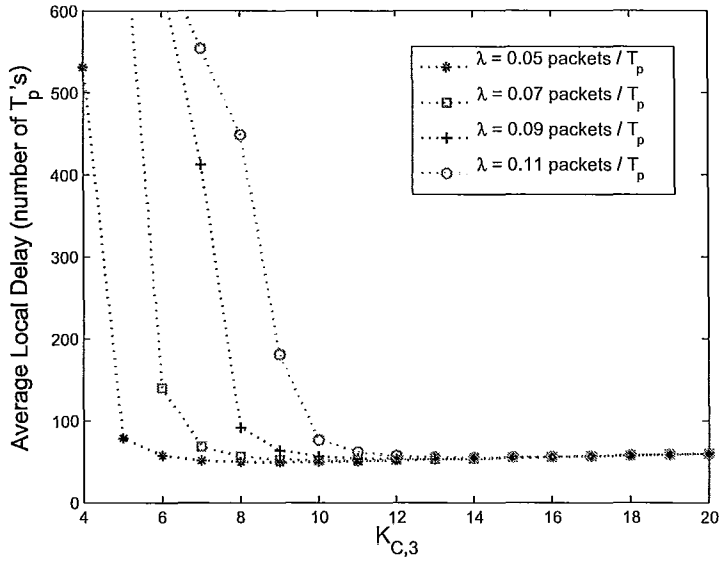

Figure 3.4: Average local traffic delay for level-3 CHs

this packet is $T_{SF} + T_{L,3}(= 88T_p)$. Fig. 3.5 shows that most of the local transmission time is between $8T_p$ and $88T_p$, which indicates that most of the packets can be transmitted to the CH in the following SF after they are generated. Therefore, this represents a light local traffic load case.

The distribution of the local traffic delay of the level-3 traffic is shown in Figs. 3.6 and 3.7 for different values of $K_{C,3}$ and λ values, respectively. Fig. 3.6 shows that for larger $K_{C,3}$, the packets are less likely to be delayed to later SFs. Fig. 3.7 shows that when the traffic load is relatively light, e.g. λ = 0.05 packets/T_p, the local transmission delay for most packets is less than one SF. As the traffic load becomes heavy, more packets experience longer delay. When λ = 0.11 packets/T_p, packet transmissions can experience very long delay because the local traffic service system is close to be overloaded.

Next we show the local traffic delay for traffic in the level-2 clusters. Its distribution

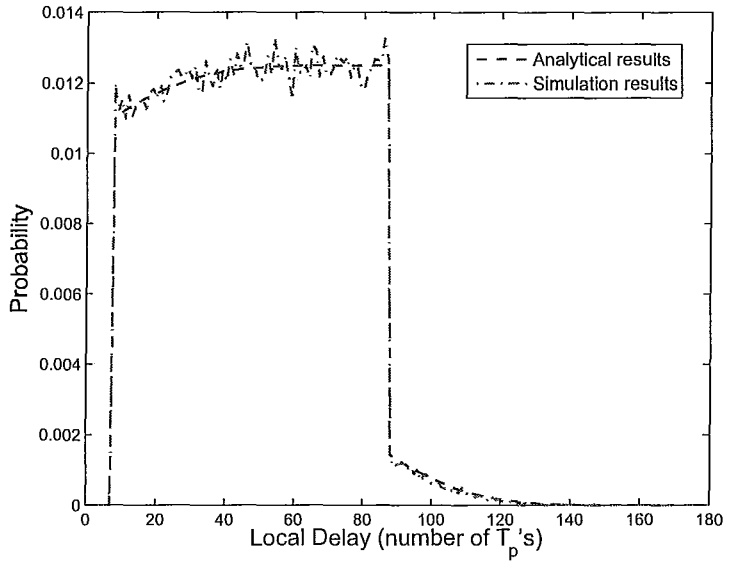

Figure 3.5: Distribution of local traffic delay for level-3 CHs ($K_{C,3}$=8, λ=0.05 packets/T_p)

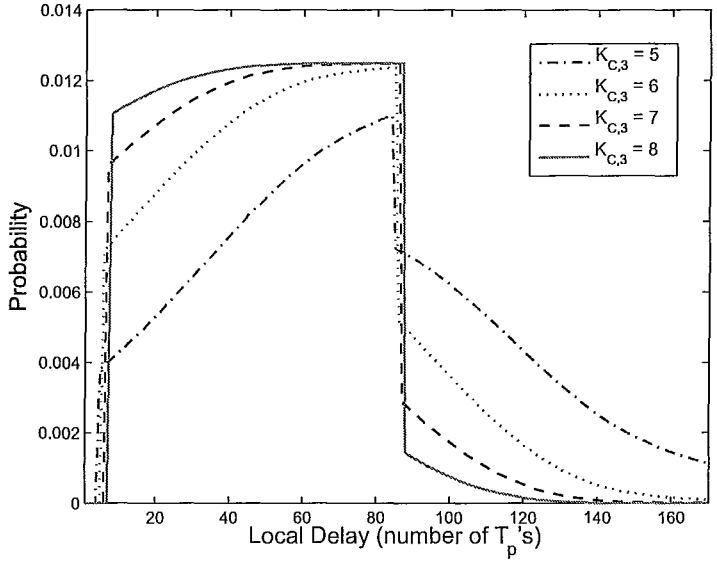

Figure 3.6: Distribution of local traffic delay for level-3 CHs (λ=0.05 packets/T_p)

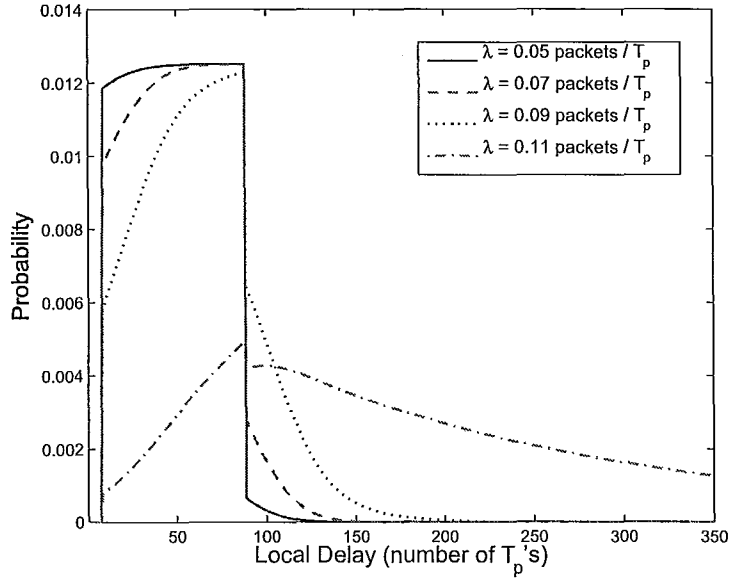

Figure 3.7: Distribution of local traffic delay for level-3 CHs ($K_{C,3}$=9)

is shown in Fig. 3.8. It is seen that the simulation results match the analytical results very well. Comparing Fig. 3.8 and Fig. 3.5 we find that the curves of local transmission delay distribution for traffic at the level-2 and level-3 clusters have a similar shape, since all the clusters have the same distribution of local traffic arrivals and the local traffic service rate, and therefore they have the same distribution of $X_{C,i}$'s and the same distribution of $Y_{C,i}$'s. The reason that the level-2 and level-3 traffic have the different minimum local transmission delay is that $T_{R,i}$ is included in the local transmission delay as shown in (3.6), $T_{R,3} = 0$ for the level-3 CHs, while $T_{R,2} = 16T_p$ in the simulated system.

Fig. 3.9 shows the distribution of local traffic delay at the sink. The distribution curve is similar to those for the non-sink CHs. However, as the local traffic delay at the sink does not include $T_{R,1}$, the minimum delay can be as small as T_p. Figs. 3.10 and 3.11 show the distribution of the local transmission delay for the traffic from the sensors directly

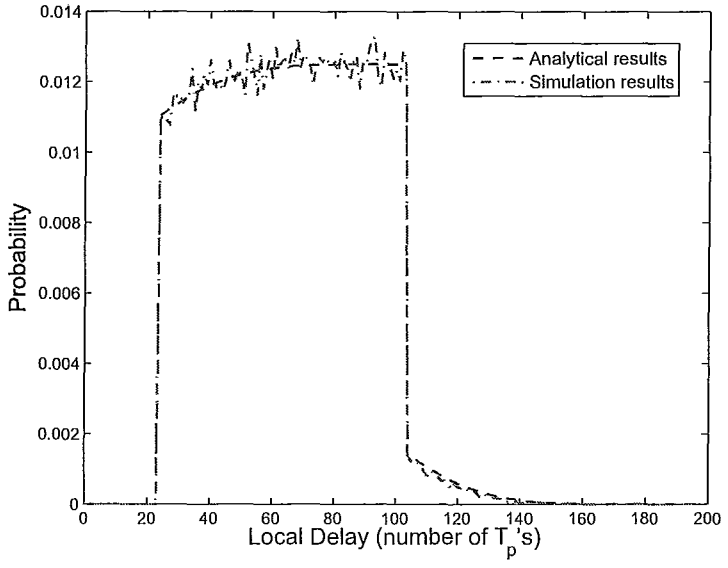

Figure 3.8: Distribution of local traffic delay for level-2 CHs ($K_{C,2}$=8, $K_{F,2}$=16, λ=0.05 packets/T_p)

associated to the sink. We can see that the distribution is very similar to that for the level-3 traffic.

3.4.2 Inter-CH transmission delay

In this subsection we demonstrate the results for single-hop inter-CH transmission delay. Fig. 3.12 shows the average transmission delay from the level-3 CHs to the level-2 CHs when the traffic load is relatively light, and Fig. 3.13 shows the delay when the traffic load is relatively heavy. Fig. 3.12 shows that when $K_{F,2}$ is small (less than 9), the available channel time for the inter-CH traffic is insufficient and the delay is very large. The delay is relatively small when $K_{F,2} \geq 9$. However, increasing $K_{F,2}$ too much will deteriorate the delay performance slightly, since $T_{F,2}$ itself is counted as part of the inter-CH delay. We also

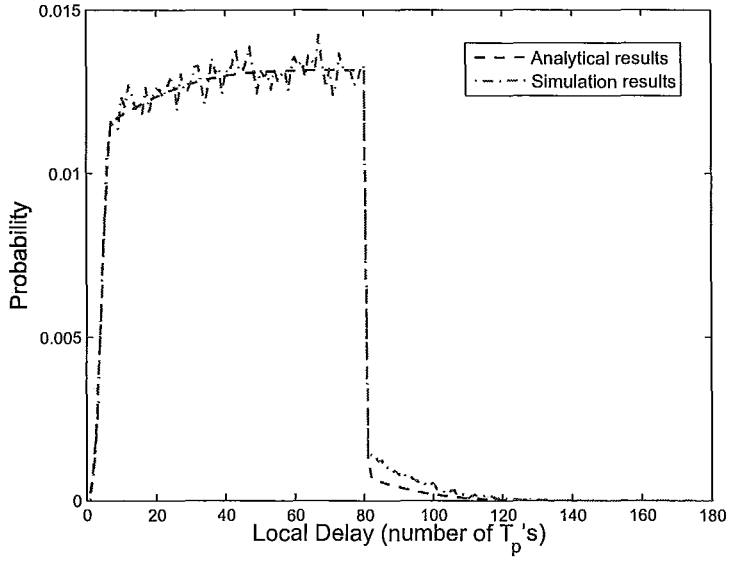

Figure 3.9: Distribution of local traffic delay for the sink ($K_{C,1}$=8, λ=0.05 packets/T_p)

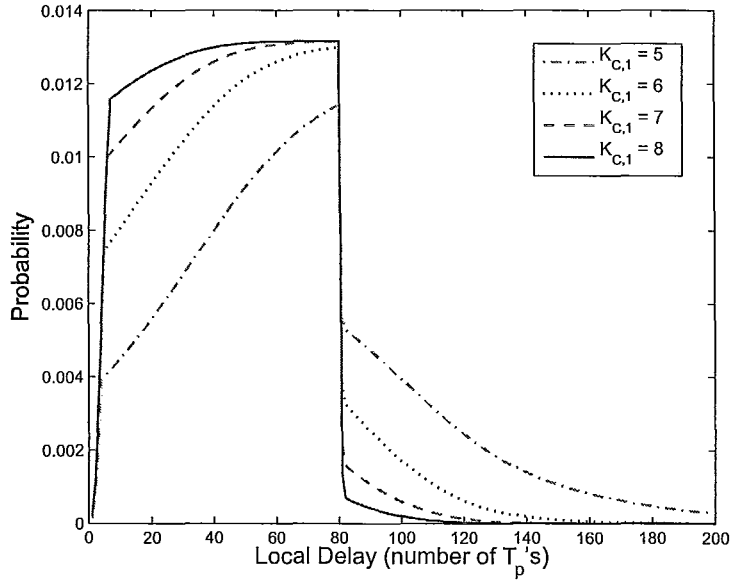

Figure 3.10: Distribution of local traffic delay for the sink (λ=0.05 packets/T_p)

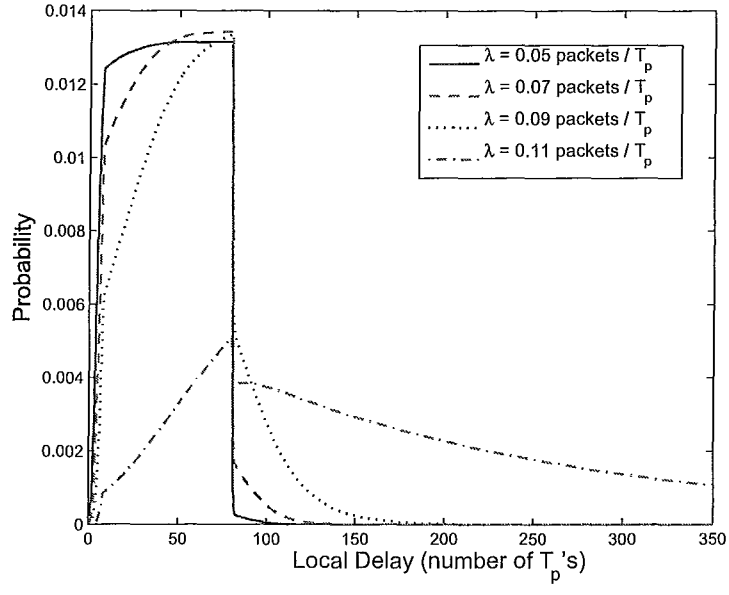

Figure 3.11: Distribution of local traffic delay for the sink ($K_{C,1}$=9)

observe that having smaller $K_{C,3}$ can achieve slightly better inter-CH delay performance, as $K_{C,3}$ limits the amount of local traffic that can be collected by the level-3 CHs, and therefore limits the inter-CH traffic load. Similar observations can be found in Fig. 3.13 for the heavy local traffic load case, except that the required $K_{F,2}$ is much larger in order to keep stable inter-CH transmission delay.

Figs. 3.14-3.16 show the distribution of the inter-CH transmission delay from the level-3 CHs to the level-2 CHs. Fig. 3.14 shows that if $K_{F,2}$ is relatively large, such as 13 or larger, all the packets can be forwarded in the same superframe as the packets are received by the local CH. However, when $K_{F,2}$ is smaller, i.e. $K_{F,2} = 9$ or 11, more packets may be delayed to later superframes. Similar observations can be found in Fig. 3.15 when the traffic load is relatively heavy. When $K_{F,2} = 18$ or larger, all the packets can be delivered in the same superframe as they reach the local CH, thus relatively good inter-CH transmission

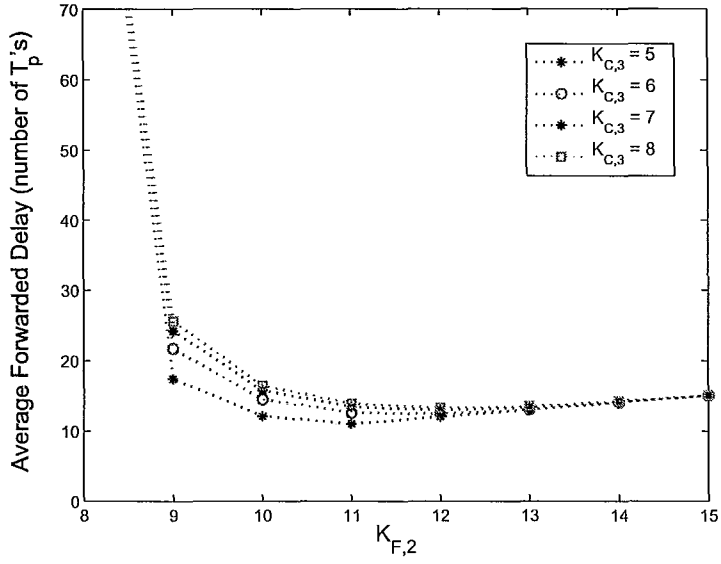

Figure 3.12: Forwarded traffic delay between level-3 and level-2 CHs (λ=0.05 packets/T_p)

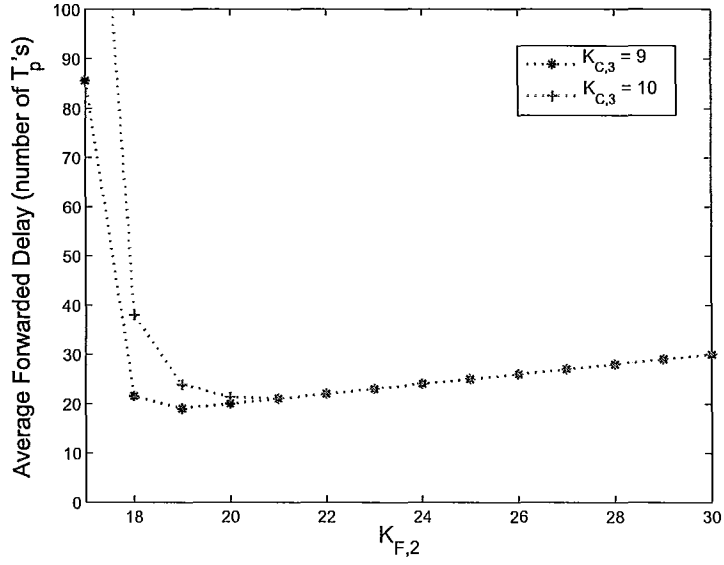

Figure 3.13: Forwarded traffic delay between level-3 and level-2 CHs (λ=0.11 packets/T_p)

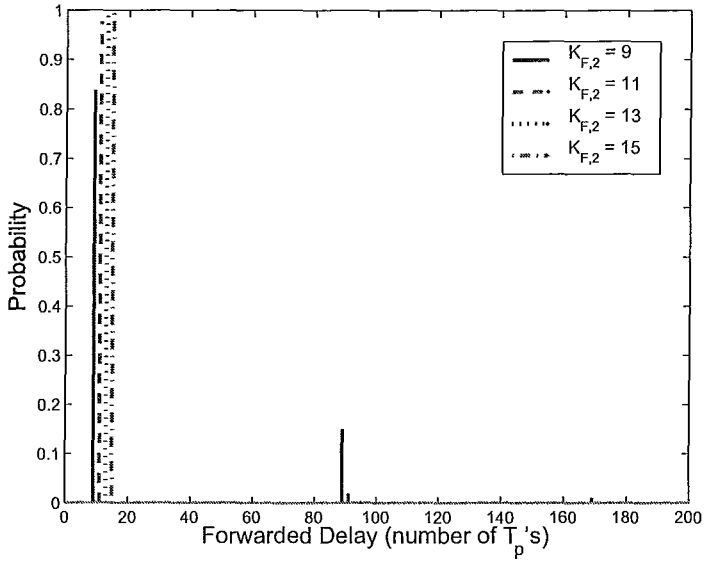

Figure 3.14: Distribution of inter-CH transmission delay between level-3 and level-2 CHs ($K_{C,3}$=9, λ=0.05 packets/T_p)

delay performance can be achieved. When $K_{F,2} = 14$ or smaller, more packets have to be delayed to later superframes. Fig. 3.16 shows that when the traffic load is heavier, more packets are delayed to later superframes.

Next we set $K_{C,1} = K_{C,2} = K_{C,3}$ and $\lambda = 0.11$ packets/T_p for all clusters and collect the average inter-CH transmission delay from the level-2 CHs to the sink. The results are shown in Fig. 3.17, where we observe that the average inter-CH delay for the case with $K_{C,1} = K_{C,2} = K_{C,3} = 9$ and $K_{F,2} = 18$ is the best among all the three parameter settings shown in the figure. This indicates that increasing $K_{C,i}$ and $K_{F,i}$ at the lower level CHs does not always improve the inter-CH transmission delay performance, similar to the observations that we had in Figs. 3.12 and 3.13.

Figs. 3.18-3.20 show the distribution of the inter-CH transmission delay from the level-2 CHs to the sink. Fig. 3.18 shows good match between the simulation and analytical

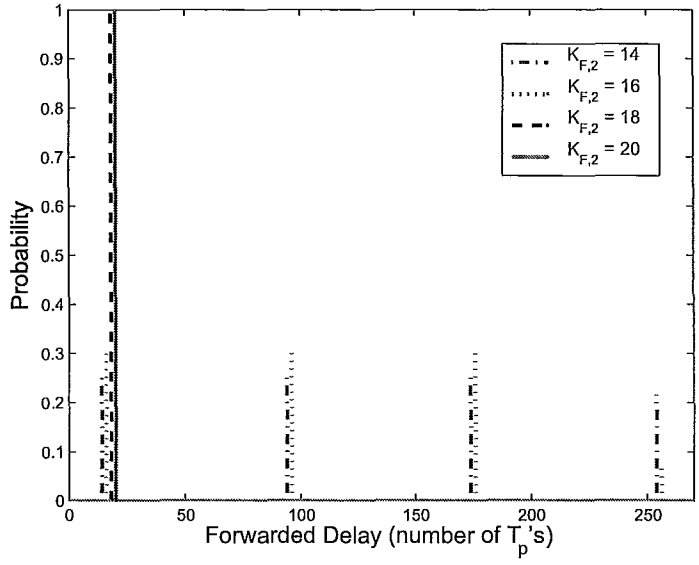

Figure 3.15: Distribution of inter-CH transmission delay between level-3 and level-2 CHs ($K_{C,3}$=9, λ=0.11 packets/T_p)

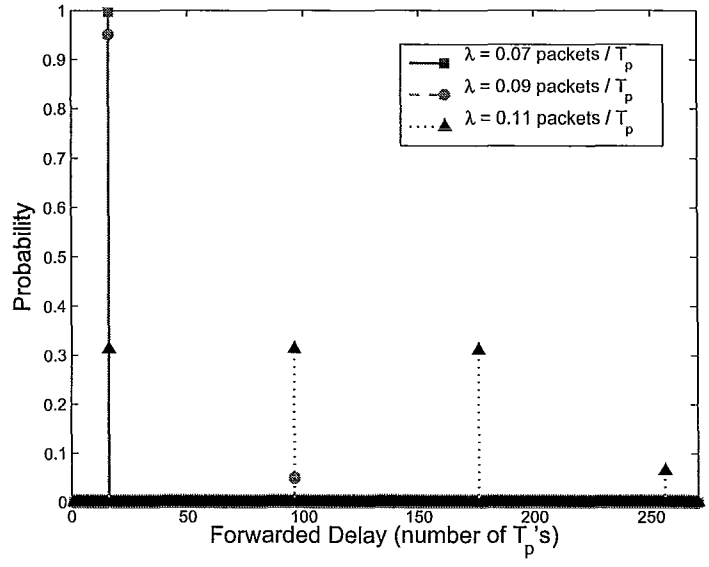

Figure 3.16: Distribution of inter-CH transmission delay between level-3 and level-2 CHs ($K_{C,3}$=9, $K_{F,2}$=16)

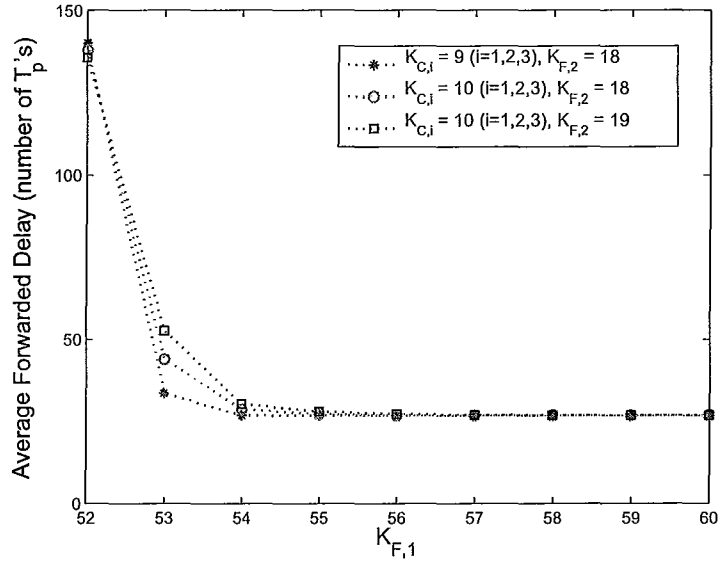

Figure 3.17: Forwarded traffic delay between the level-2 CHs and the sink (λ=0.11 packets/T_p)

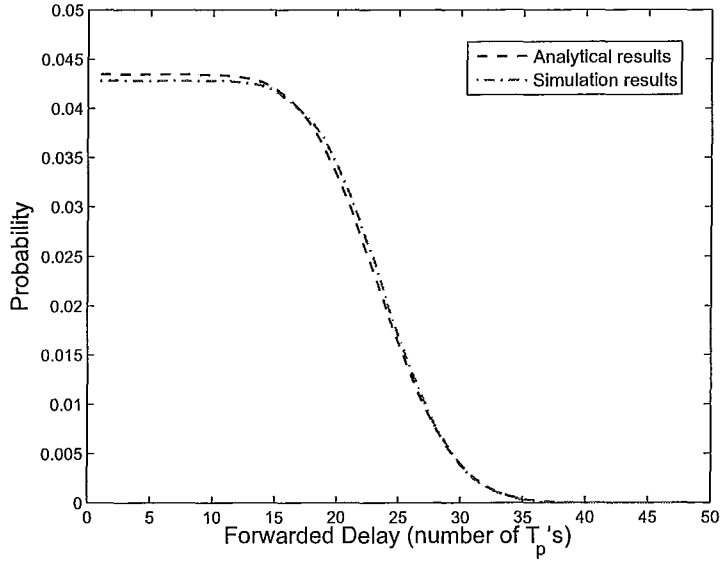

Figure 3.18: Distribution of inter-CH transmission delay between the level-2 CHs and the sink ($K_{C,i} = 9$, $K_{F,2} = 16$, $K_{F,1} = 48$, $\lambda = 0.05$ packets/T_p)

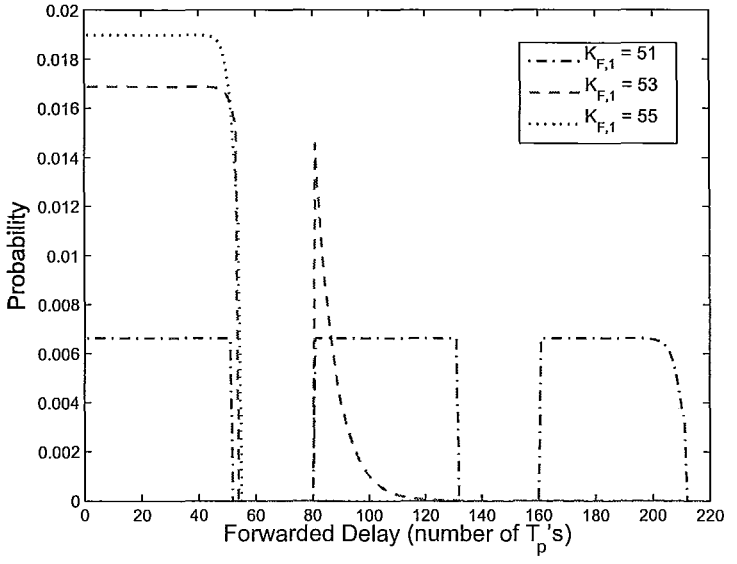

Figure 3.19: Distribution of inter-CH transmission delay between the level-2 CHs and the sink ($K_{C,i}$=9, $K_{F,2}$=18, λ=0.11 packets/T_p)

results. As the receiver is the sink, the minimum delay can be as small as T_p. Fig. 3.19 shows the distribution of the inter-CH delay for different $K_{F,1}$ values when the local traffic load is heavy. It is shown that when $K_{F,1} = 55$, all collected packets from the sensors and lower level CHs can be transmitted to the sink within one SF. When $K_{F,1}$ is smaller, more packets are delayed at the level-2 CHs before they are forwarded to the sink. Fig. 3.20 shows the distribution of the inter-CH transmission delay for different local traffic loads. When λ is larger, more packets are delayed for longer time.

3.4.3 End-to-end Transmission Delay

Fig. 3.21 shows the distribution of the end-to-end delay for the level-3 traffic at different local traffic loads. When the local traffic load is lighter, the end-to-end transmission is

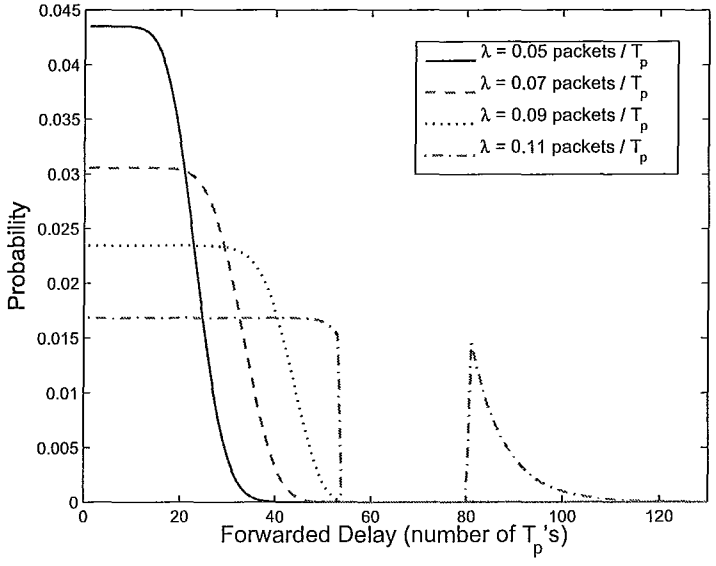

Figure 3.20: Distribution of inter-CH transmission delay between the level-2 CHs and the sink ($K_{C,i}$=9, $K_{F,2}$=18, $K_{F,1}$=53)

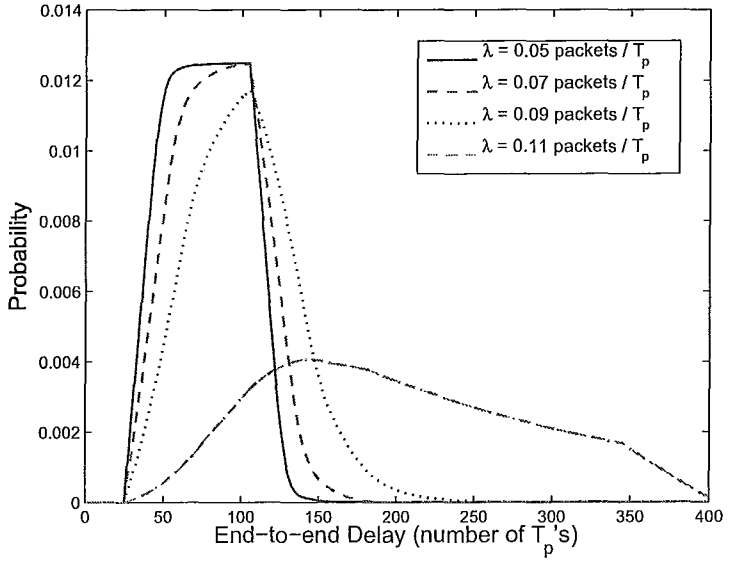

Figure 3.21: Distribution of end-to-end delay for level-3 traffic ($K_{C,i}$ = 9, $K_{F,2}$ = 18, $K_{F,1}$ = 53)

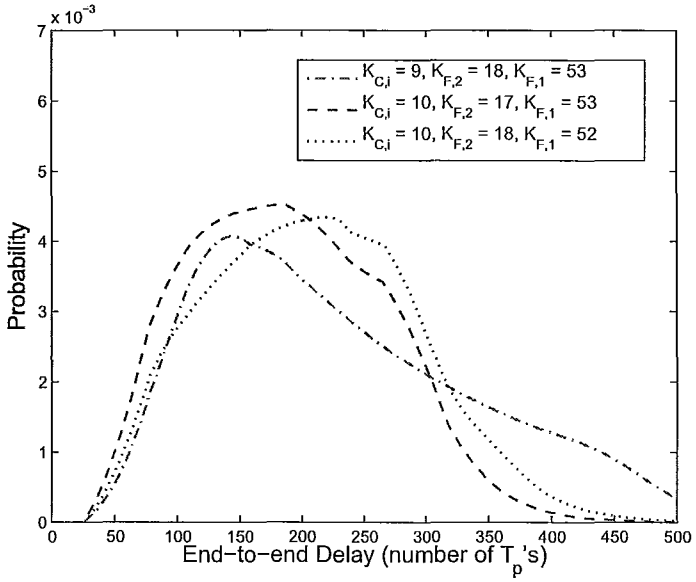

Figure 3.22: Distribution of end-to-end delay for level-3 traffic ($\lambda = 0.11$ packets / T_p)

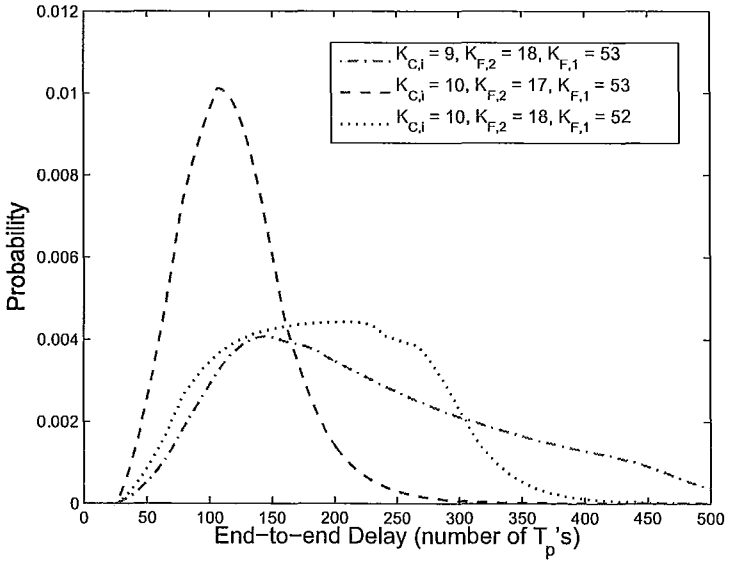

Figure 3.23: Distribution of end-to-end delay for level-2 traffic ($\lambda = 0.11$ packets / T_p)

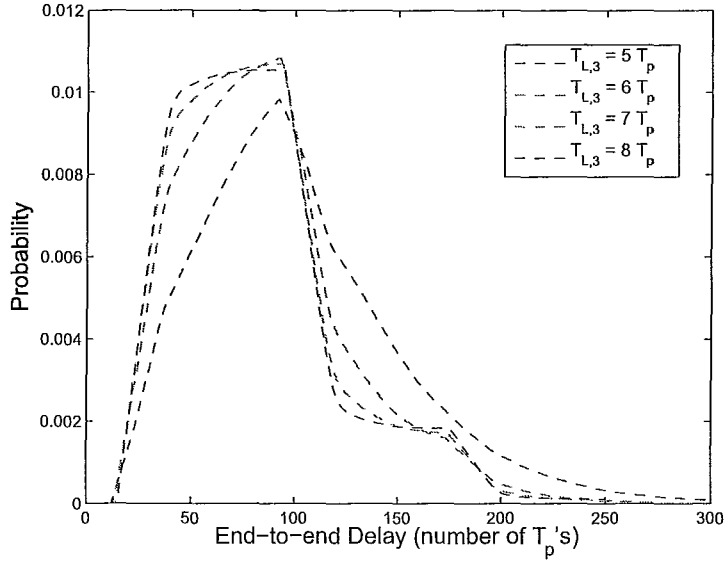

Figure 3.24: Distribution of end-to-end delay for level-3 traffic ($\lambda = 0.05$ packets/T_p)

more likely to be completed in shorter time. When the local traffic load is heavy, such as $\lambda = 0.11$ packets/T_p, most packets are delayed by more than one SF along the end-to-end path. When $\lambda = 0.11$ packets/T_p, the end-to-end delay of the level-3 and level-2 traffic is shown in Figs. 3.22 and 3.23, respectively. From these two figures we can see that the delay distribution is slightly affected by tuning the timeline allocations. For the same three sets of timeline allocations, the level-2 traffic experiences shorter delay, and the same changes in the timeline allocations affect more on the end-to-end delay of the level-2 traffic than on the level-3 traffic. Among the three sets of timeline settings, $K_{C,i} = 10$, $K_{F,2} = 17$ and $K_{F,1} = 53$ achieves the best end-to-end delay performance for both the level-2 and the level-3 traffic.

In Fig. 3.24, we fix the timeline allocations related to the level-2 traffic $T_{L,2} = 5T_p$, $T_{R,2} = 9T_p$ and $T_{T,2} = 50T_p$, then vary the time allocated to the local traffic at the level-3

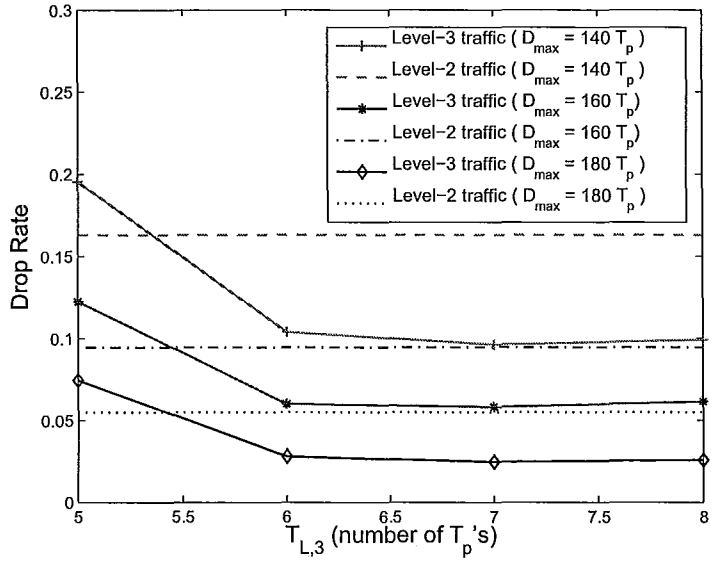

Figure 3.25: Drop rate for level-2 and level-3 traffic ($\lambda = 0.05$ packets/T_p)

CH to observe the distribution of end-to end delay for level-3 traffic. We find that it is possible to improve the end-to-end delay performance by increasing $T_{L,3}$. In Fig.3.25 we keep all these timeline allocations related to the level-2 traffic so that to fix the packet drop rate for the level-2 traffic for the given maximum tolerable delay requirement, we then adjust the time allocated to the local traffic at the level-3 CH. It is seen that when $T_{L,3}$ is relatively small, the drop rate for the level-3 traffic is higher than that for the level-2 traffic. As $T_{L,3}$ increases, the drop rate for the level-3 traffic is below that for the level-2 traffic. At one particular value of $T_{L,3}$, it is possible to achieve an equal drop rate for the traffic at both levels.

The distribution of end-to-end delay for Level-2 and Level-3 traffic can be found in Figs. 3.26 and 3.27, respectively. We can see that the simulation results match the analytical results very well. It is seen that the distributions of the end-to-end delay for traffic at both

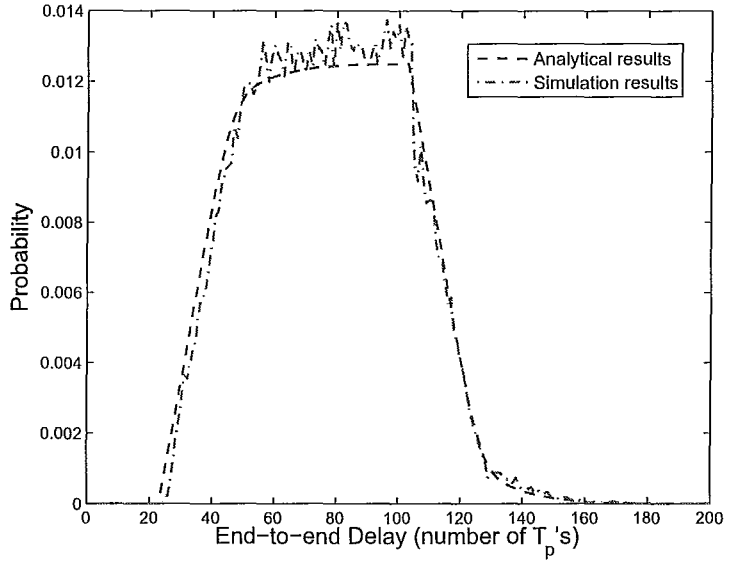

Figure 3.26: Distribution of end-to-end delay for level-2 traffic ($\lambda = 0.05$ packets/T_p, $K_{C,i} = 8$)

levels have almost the same shape and values. This is because the local traffic load is the same at both levels, the local receiving time is the same for the two levels, and therefore, the local transmission delay is the same for both levels. After the packets are received by their respective local CHs, the level-2 packets should wait for a $T_{R,2}$ interval until the start of the inter-CH transmission time to the sink, while the level-3 packets are served during the $T_{R,2}$ interval. Although queueing delay is normally expected for the level-3 packets during this period of inter-CH transmissions, in this example, $T_{R,2} = 18T_p = 2T_{L,3}$, and the available inter-CH time is sufficient for all the level-3 packets to reach the level-2 CH in the same SF as they are received at the level-3 CH.

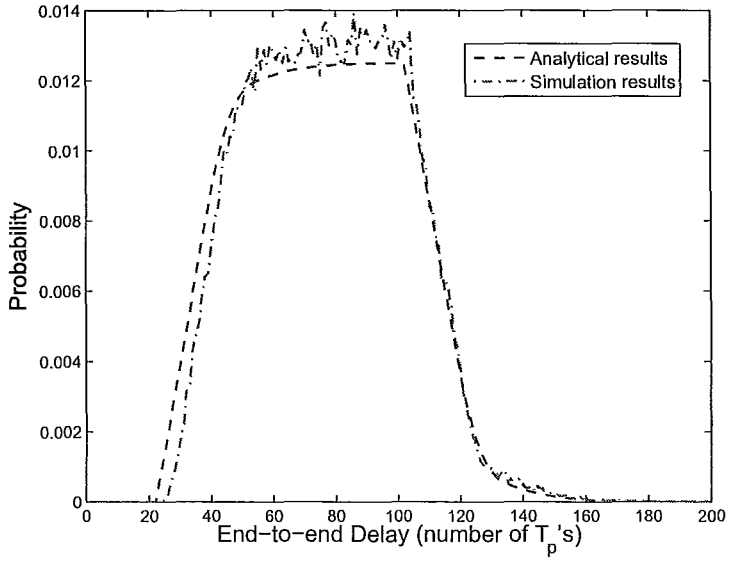

Figure 3.27: Distribution of end-to-end delay for level-3 traffic ($\lambda = 0.05$ packets/T_p, $K_{C,i} = 8$)

3.5 Summary

In this chapter, we analyze the local transmission delay, forwarded delay as well as the end-to-end transmission delay in a WSN with cluster tree topology. The analytical expressions for both average delay and distribution of the delay are derived. Our analytical results match the simulation results very well, and the numerical results indicate that it is possible to arrange the timeline allocations in order to reduce the transmission delay and balance the end-to-end transmission delay for the traffic from different levels in the network.

Chapter 4

Conclusions and Future Work

In this thesis, we have studied the performance of WSNs with the cluster tree topology related to the CH timeline allocations. We have proposed schemes to jointly consider the timeline allocations of the CHs and the sensor node associations so that to optimize different network parameters, including maximizing the throughput of individual sensor nodes, balancing the energy consumption among the CHs, and maximizing the network level throughput. This work can be further extended to include more dynamic traffic loads. In this case, the timeline allocations and the sensor node associations should be dynamically changed, and the overhead due to such dynamic changes should be taken into consideration.

We have also developed an analytical model to find the distribution of the end-to-end transmission delay for traffic at different levels of the cluster tree, and based on this the packet drop rate due to excessive delay can be derived. This provides important guidelines for allocating the CH time resources in order to achieve certain delay or packet drop rate performance. The analytical model also helps find a solution to achieve balanced end-to-end delay or packet drop rate performance among traffic in different clusters. We will design a scheme to find appropriate timeline allocations of the CHs based on the local

traffic load in each cluster to achieve this objective.

Bibliography

[1] J. Yick, B. Mukherjee, D. Ghosal, "Wireless sensor network survey", *Computer Networks*, Aug. 2008, vol. 52, pp. 2292-2330.

[2] G. J. Pottie and W. J. Kaiser, "Wireless Integrated Network Sensors", *Communications of ACM*, May 2000, 43(5).

[3] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Emerging Challenges: Mobile Networking for Smart Dust", *Journal of Communications and Networks*, Sep. 2000, 2(3):188196.

[4] Kay Romer and Mattern Friedemann, "The Design Space of Wireless Sensor Networks", *IEEE Wireless Communications*, Dec. 2004, 11(6): 5461.

[5] "IEEE Std. 802.15.4-2006, IEEE Standard for Local and Metropolitan Area Networks part 15.4, Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)", *IEEE Press*, 2006.

[6] "ZigBee Specification Version 1.0", *ZigBee Alliance*, 2005.

[7] L. Shi and A.O. Fapojuwo, "TDMA Scheduling with Optimized Energy Efficiency

and Minimum Delay in Clustered Wireless Sensor Networks", *IEEE Transactions on Mobile Computing*, vol. 9, no. 7, pp.927-940, July 2010.

[8] C. Shanti and A. Sahoo, "DGRAM: A Delay Guaranteed Routing and MAC Protocol for Wireless Sensor Networks", *IEEE Transactions on Mobile Computing*, vol. 9, no. 10, pp.1407-1423, Oct. 2010.

[9] H. Wang *et al*,, "Network Lifetime Optimization in Wireless Sensor Networks", *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 7, pp.1127-1137, Sept. 2010.

[10] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocols for wireless microsensor networks", *Proc. Hawaaian International Conference on Systems Science*, Jan. 2000.

[11] M. Ye, C. Li, G. Chen, J. Wu, "EECS: an energy efficient clustering scheme in wireless sensor networks", *Proceedings of 24th IEEE International Performance, Computing and Communications Conference*, 2005, pp. 535-540.

[12] Y. Zhou, M. Hart, S. Vadgama, and A. Rouz, "A hierarchical clustering method in wireless ad hoc sensor networks", *Proc. IEEE International Conference on Communications (ICC)*, 2007, pp. 3503-3509.

[13] F. Cuomo, S. D. Luna, E. Cipollone, P. Todorova, and T. Suihko, "Topology formation in IEEE 802.15.4: cluster-tree characterization", *Proc. Sixth Annual IEEE International Conference on Pervasive Computing and Communications*, 2008, pp. 276-281.

[14] J. Cho and S. An, "Optimal beacon scheduling scheme for cluster-tree WPANs", *Proc. The Second International Conference on Sensor Technologies and Applications*, 2008, pp. 160-165.

[15] F. Zhang, F. Wang, B. Dai and Y. Li, "Performance Evaluation of IEEE 802.15.4 Beacon-enabled Association Process", *International Conference on Advanced Information Networking and Applications - Workshops*, 2008, pp. 541-546.

[16] F. Meng and Y. Han, "A New Association Scheme of IEEE 802.15.4 for Real-time Applications", *Wireless Communications, Networking and Mobile Computing*, 2009, pp. 1-5.

[17] P. Seeling and J. B. Starren, "Ad-Hoc Association of Pre-determined ZigBee Devices", *6th Annual International Mobile and Ubiquitous Systems: Networking and Services*, 2009, pp. 1-6.

[18] Cuomo F., Della Luna S. Monaco, U. and Melodia F., "Routing in ZigBee: benefits from exploiting the IEEE 802.15.4 association tree", *IEEE International Conference on Communications (ICC)*, 2007, pp.3271-3276.

[19] A.O. Fapojuwo and A. Cano-Tinoco, "Energy Consumption and Message Delay Analysis of QoS Enhanced Base Station Controlled Dynamic Clustering Protocol for Wireless Sensor Networks", *IEEE Transactions on Wireless Communications*, vol. 8, no. 10, Oct. 2009.

[20] S.S. Choi, "Analysis of Low Latency MAC Protocols for Clustered Sensor Networks", *IEEE Wireless Communications and Networking Conference*, 2008, pp.1894-1898

[21] J. Yackoski and C. Shen, "Managing End-to-End Delay for VoIP Calls in Multi-Hop Wireless Mesh Networks", *INFOCOM IEEE Conference on Computer Communications Workshops*, 2010

[22] J. Chen and Y. Yang, "Multi-hop Delay Performance in Wireless Mesh Networks", *IEEE Global Telecommunications Conference (GLOBECOM)*, 2006, 1

[23] P. Djukic and S. Valaee, "Delay Aware Link Scheduling for Multi-Hop TDMA Wireless Networks", *IEEE/ACM Transactions on Networking*, vol.17, issue: 3, June 2009

[24] C. Florens, M. Sharif and R.J. McEliece, "Random Sensory Networks: A Delay Analysis", *IEEE Transactions on Information Theory*, vol. 55, issue: 4, pp.1650-1664, March 2009.

[25] L. Galluccio and S. Palazzo, "End-to-End Delay and Network Lifetime Analysis in a Wireless Sensor Network Performing Data Aggregation", *IEEE Global Telecommunications Conference (GLOBECOM)*, 2009, pp.1-6

[26] Y. Deng, F. Ren and C. Lin, "An End-to-End Stochastic Delay Bound Analysis in Sensor Networks", *3rd International Conference on Advanced Computer Theory and Engineering*, 2010, vol. 3, pp.370-374

[27] S. Pollin *et al*, "Performance analysis of slotted carrier sense IEEE 802.15.4 medium access layer", *IEEE Transactions on Wireless Communications*, vol. 7, no. 9, pp.3359-3371, Sept. 2008.

[28] X. Ling, Y. Cheng, J. W. Mark, and X. Shen, "A renewal theory based analytical model for the contention access period of IEEE 802.15.4 MAC", *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, June 2008.