ABSTRACT

Title of thesis: Minimum-energy transmission and effect of network architecture on downlink performance of wireless data networks

Adarsh Sridhar, Master of Science, 2005

Thesis directed by: Professor Anthony Ephremides
Department of Electrical and Computer Engineering

The increasing demand for wireless data services in the recent years has led to traditionally voice and video oriented network architectures such as satellite and cellular systems being adapted for data transmission. Also, since most wireless systems are battery-powered, energy consumption has become an important consideration in the research community.

In this thesis, we first examine the problem of controlling transmission power in order to achieve minimum-energy broadcast packet transmission subject to a minimum QoS requirement. We study the structure of the optimal policy when the transmitter chooses between two power levels, and then extend the results to the continuous-power case.

Subsequently, we examine the effect of network architecture on the downlink performance of the network. Satellite and Cellular systems are compared on the basis of energy, delay and throughput. We then show that having a hybrid network architecture can provide additional throughput benefit, at the cost of energy.

Minimum-energy transmission and effect of network architecture on downlink performance of wireless data networks

by

Adarsh Sridhar

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2005

Advisory Committee:

Professor Anthony Ephremides, Chair/Advisor
Professor Alexander Barg
Professor André Tits

To my parents,

Thank you for everything.

ACKNOWLEDGMENTS

I owe my gratitude to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

First and foremost I'd like to thank my advisor, Professor Anthony Ephremides, for giving me an invaluable opportunity to work with him over the past three years. It has been a pleasure to work with and learn from such an extraordinary individual.

I would also like to thank Professor Alexander Barg and Professor André Tits for agreeing to serve on my thesis committee and for sparing their invaluable time reviewing the manuscript.

I owe my deepest thanks to my family - my mother, father and sister who have always stood by me and guided me through trying times. Words cannot express the gratitude I owe them.

In addition, I would like to thank my roommates - Abbhi, Ashok, Lavanya, Renga, Shyam and Som - for putting up with me these three years. We've had lots of great times together, and I'll miss them.

I would also like to thank Ahmed, Ashwin, Brooke, Anna and Song for helping me put this thesis together and taking the time to give me feedback regarding my presentation.

It is impossible to remember everyone, and I apologize to those I've inadvertently left out.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

1.1 Challenges in wireless systems

Wireless systems have seen an explosive growth in the recent years. For example, in the year 2004, a growth in GSM coverage of over 72% was reported in the United States alone. The number of mobile phone units sold worldwide increased by over 30%, and the leading computer manufacturers in the United States reported a 31% increase in the number of laptops sold last year. This sudden growth in the market has led to a significant interest in the field of wireless networks.

The field of wireless networks is relatively unexplored, as compared to their wired counterparts. A lot of new challenges arise in wireless networks that are not present in wired networks. We shall take a look at some of the issues that arise in this context, and see how they fundamentally differ from the wired case.

1. **Nodes, not links**

   The primary difference in wireless networks comes from the fact that we can no longer talk about links between nodes. This is because the existence of a link depends on the transmit power chosen at each node. Also, the links between nodes are no longer independent of each other, since increasing the transmit power at one node to establish a link causes added interference to the other links in the vicinity, potentially disrupting their performance. As a result of this, most of the existing work on wired networks, which is based on the idea of point-to-point links, is inapplicable to wireless systems.

2. **Broadcast nature of channel**

   The broadcast nature of the channel is the underlying cause of many problems in wireless networks. For example, as discussed earlier, transmission at one node can disrupt transmission at other nodes, owing to the broadcast nature of the channel. However, this broadcast nature can be useful depending on the type of application that is being used. In broadcast and multicast applications, wireless systems offer an inherent performance bonus, owing to the idea of "'Wireless Multicast Advantage"' [15]. The idea of multicast advantage is that a single high-power transmission would be enough to reach a lot of nodes, a phenomenon which is absent in wired networks.

3. **Mobility**

   Mobility is an important consideration in wireless systems, since mobile systems are inherently wireless in nature. Mobility introduces a whole host of new problems because it causes the network topology to change constantly. Also, the link quality between nodes that are connected varies with the distance and location of the nodes. Recently, however, it was proposed that mobility might increase the capacity of wireless networks [11], indicating that mobility may not be completely harmful.

4. **Channel quality**

   A big challenge in the wireless environment is that of channel quality. The wireless channel is orders of magnitude more error-prone than the wired channel. This requires that error-detection and error-correction schemes be more robust, which increases the overhead required. Moreover, the transmitted energy dissipates rapidly with increasing distance in wireless systems, which limits the range of transmissions. In conjunction with the broadcast nature of the channel, such rapid dissipation leads to the hidden-terminal and exposed-terminal problems [3], complicating the MAC layer protocols. Wireless networks are also extremely prone to the problem of fading, wherein a transmission can be almost completely killed by the channel going into an extremely bad state for a short period of time. This problem is not present in wired networks.

5. **Security**

   Wireless networks are inherently less secure than their wired counterparts. The primary reason for this, again, is the broadcast nature of the channel. It is now possible for an eavesdropper to scan the entire traffic on the network without actively being a part of it. This makes detection of such malicious entities extremely difficult. Also, mobility introduces further complications, since it now makes masquerading easier. A good list of papers relating to security in Mobile Ad-Hoc networks can be found online here [37].

6. **Energy efficiency**

   An important idea that has surfaced recently in this context is one of energy conservation. Most wireless systems run on batteries, and since battery technology has not made much progress in the recent years, it becomes important to design systems that run for as long as possible on the limited energy stored in the battery, without significant degradation in the performance. While this is not so critical for networks consisting of rechargeable devices (for example, laptop computers and cellphones), it becomes paramount for cases like sensor networks, since the very operation of the network depends on the energy available in the nodes. However, even for rechargeable devices, energy is still an important consideration. It is now well recognized that the conventional idea of layering does not perform well in wireless networks, and cross-layer interaction is a primary focus of research in the community today. Power control, which is part of the physical layer, is now being used increasingly in conjunction with higher layer functions such as routing and MAC.

1.2 Review of existing literature

In this thesis, we examine two aspects of wireless networks, and see how they affect performance. These aspects are those of cross-layer interaction, involving power control and ARQ, and that of network architecture. We look at the existing literature on these two aspects, and see how our work differs from what has already been done.

1.2.1 Power control

A lot of work has been done on power-control schemes, both in the context of energy efficiency as well as for improving the performance of wireless systems. Most of the work in this area can be classified into the following categories :

1. **CDMA systems**

   Code-Division Multiple Access (CDMA) systems are a form of spread-spectrum communication, where all transmitters in a CDMA system transmit on all available bandwidth all the time. The resource division takes place in the form of nearly-orthogonal codes assigned to each transmitter which modulate the transmitted symbols. Such spread-spectrum technology has a large bandwidth but a low power spectral density, which makes the signal significantly more tolerant to narrowband interference.

   A CDMA receiver receives simultaneous transmissions from multiple sources, and attempts to decode all of them from the same received signal. Depending on the decoding method being used, the performance of the receiver depends critically on the received powers. For example, if a matched-filter detector is being used, it is possible that the nearby transmitter drowns out transmissions of the farther transmitter, if they are both transmitting using the same power, leading to a phenomenon called the *Near-far problem* [31]. On the other hand, the optimum multi-user detector works best if the received powers are quite disparate. In either case, however, power control is an important aspect of CDMA systems. A wealth of literature is dedicated to the subject of equalizing received powers and minimizing interference in CDMA systems (e.g. [5], [26], [35], [27], [14]).

2. **Multicasting in Ad-Hoc Networks**

   As mentioned earlier, the algorithms for multicasting and broadcasting in wireless environments are significantly different from that of wired environments. This is because connectivity in wireless networks is determined by the transmit power chosen, and hence existing link-based algorithms do not function optimally. For example, in wired networks, the problem of minimum-energy broadcasting reduces to finding a minimum-cost spanning tree, which can be done in $O(N^2)$ time. However, the same problem in wireless networks is NP-complete [4], and heuristics are needed. Many such algorithms have been proposed in the literature (see [15], [17], [32], [6]), where power control is used to simultaneously determine the network connectivity and the energy expenditure.

3. **Interference mitigation**

   Since transmissions by a node can affect the transmissions of other nodes in its vicinity, it becomes important to control the transmit powers of all the nodes in the network in such a fashion that they cause minimum interference to each other, while still maintaining transmission quality. Scheduling, for example, is a technique that is commonly used to prevent nodes from interfering with each other. A significant amount of work has been done in this area (for example, [22], [21], [30]).

It is now widely accepted that the concept of layering does not extend itself to wireless networks as it does to wired networks. We cannot divide the functionality of the different layers into independent problems, since the process at the different layers influence each other significantly. Power control, which is a physical-layer function, is used in conjunction with multicast routing (a network-layer function), or scheduling (a MAC-layer function) as shown above. We look at a similar cross-layer interaction, namely that of power control with ARQ in a broadcast network.

1.2.2 Network Architecture

Network architecture has always been a critical factor in determining network performance. For example, in wired networks, topology plays an important role in determining how routing protocols perform. This behavior carries over into wireless networks as well.

Networks can be classified into two types on the basis of their architecture.

1. **Infrastructure-oriented**

   These consist of networks that require certain amount of infrastructure before they can be deployed. Examples of this kind of network are Satellite, Cellular and Cable-based networks. While such networks have the disadvantage of requiring a large initial investment and maintenance, they are inherently simpler to design, because having a fixed structure makes it easier to design and implement protocols that are optimized for that infrastructure.

2. **Infrastructure-less or Ad-Hoc networks**

This is a type of network where there is no existing infrastructure. Nodes that are spread out over a geographical area communicate directly with each other to facilitate communication. While they are quicker to deploy, making them very useful in situations such as search-and-rescue, they are much more complicated to design. Ad-hoc networks are the current focus in the research community.

Until recently, the primary application of infrastructure-based wireless systems was real-time traffic such as voice and video. Currently, data-transfer over wireless links is facing an increasing demand. As a result, existing infrastructure is being adapted to carry data traffic, for example, text and picture messaging in cellular phones. However, there does not appear to be any work in the literature where they compare the performance of these architectures when data traffic is being sent. Also, energy considerations were not taken into account in the existing literature. For example, the work in [25] deals with allocating channels for voice in a dynamic fashion to the mobile nodes, for both satellite and cellular architectures. [36] and [23] consider the performance of LEO satellite networks, but do not have any comparison with cellular/hybrid architectures. A comparison of the IRIDIUM and AMPS systems is done in [13], but this is on the basis of the methods used for call registration, handoff and other aspects.

Hybrid networks is another concept that has been introduced recently in the literature ([16], [34]). Ayyagari and Ephremides [1] studied the performance of the hybrid network in terms of blocking probability for voice calls. Following this, Friedman and Ephremides [9] looked at using a hybrid network intelligently to improve throughput. The idea they formulated was that while the first transmission would be through satellite, the retransmissions are sent through the terrestrial link, which was then shown to have higher throughput. A point-to-point transmission was assumed here. A performance analysis involving blocking probability due to handoffs was studied in [18], and further work was done on planning of such networks in [8] and [12].

1.3 Contribution

In broadcast applications, a single node transmits packets to multiple recipients. However, there are many instances where it is not necessary that each packet reach all the destinations, but it is sufficient that each packet reaches a minimum number of nodes. Examples of this type can be found in military applications, distributed control, and search-and-rescue operations. Flooding is another instance where this may be useful. In all these examples, we look at wireless nodes which are inherently energy-limited. Therefore, we would like to look at minimizing energy expenditure in such a situation. In this thesis, we consider the problem of finding the optimal power-control policy when power control is used with Stop-and-Wait ARQ [3], in order to minimize energy expenditure, subject to a minimum number of nodes receiving each packet. To the best of our knowledge, this has not been looked at before.

We also look at the effect of network architecture and the advantage of hybrid networks in network performance. We choose satellite and cellular systems for comparison, and look at how data traffic behaves when transmitted on these networks. The key difference between our work and existing literature is that these networks have been traditionally designed and used for real-time traffic such as voice and video. While there is earlier work comparing satellite systems with each other, and cellular systems with each other, there do not seem to be such studies comparing the two architectures, especially for data traffic. Moreover, we compare them on the basis of energy expenditure as well, which has not been done earlier. Hybrid networks, which consist of both satellite and cellular components, are also examined, and a trade-off between energy and throughput is established in this case as well.

1.4 Organization of the thesis

Chapter 2 deals with the problem of power control in order to achieve minimum energy expenditure, when the transmitter broadcasts packets subject to certain Quality of Service requirements. We describe the problem in detail, formulate it as a Markov decision process, and study the structure of the optimal policy and the performance of the system as the parameters are varied.

Chapter 3 examines the relative benefits of the satellite and cellular architectures when guaranteed broadcast data transmission is considered. We describe a method of comparing them based on sample-path arguments, and confirm the comparison using simulations.

Chapter 4 extends the network architecture into a hybrid structure which consists of both satellite and cellular sections. We discuss the ramifications of such an extension, and see how this architecture can give a higher throughput than the individual components. We also demonstrate a trade-off between throughput and energy in this case.

Chapter 5 summarizes the work, and looks at directions in which this may be extended.

Chapter 2

Optimum power control for minimum-energy transmission

2.1 Introduction

As we have described earlier, energy is of prime concern in wireless networks. Power control is one of the ideas employed in order to efficiently use the limited stock of energy. Choice of transmission power has many implications in wireless networking, such as interference, success probability, energy, delay and buffer overflow.

The main motivation of existing work on power control was in mitigating the effect of interference in order to increase capacity. A power control strategy to maximize battery lifetime under a QoS constraint on the throughput was outlined in [20]. In this paper they considered a wireless network affected by the interference, and outlined an algorithm to find the optimal transmission power given the instantaneous interference value. The model was memoryless and no error control mechanism was specified. Zorzi et al. [19] considered Go-Back-N ARQ in a wireless link and described an improvement to enhance the energy efficiency. However, the analysis was done for point-to-point transmission. Girici and Ephremides [29] derived a power control scheme for a point-to-point link with Stop-and-wait ARQ, where the transmitter had to choose between two powers at each transmission in order to optimize a weighted combination of energy consumption and buffer overflow.

All the work that has been done on this topic are limited to point-to-point transmissions. There appear to be no results of this type for broadcast transmissions. In this thesis, we look at the problem of optimizing the energy consumption when the transmitter has to reach multiple destinations. Another important idea in our problem is that we do not constrain ourself to have guaranteed delivery; however, we do require that a minimum number of receivers receive each packet. It is important to study the structure of the optimal policy, so that we can gain some insight into the design of the control mechanism. The system model and the assumptions are described in detail below.

2.2 System model and assumptions

The network is taken to be a simple broadcast network, with one transmitter and N receivers. Data is in the form of fixed-length packets, and each packet is intended for all N receivers. Time is assumed to be slotted, and all transmissions take place at the beginning of the slot. For simplicity, we assume that each transmission takes exactly one slot.

2.2.1 Channel model

We model each transmitter-receiver channel as an AWGN channel that is independent of all other receivers, and also independent over time. Assuming BPSK modulation, each receiver can successfully receive a packet transmission with a probability μ independent of all other receivers, where μ is given by

$$\mu = \left[1 - Q\left(\sqrt{\frac{P_{tr}}{N_0}}\right)\right]^L \tag{2.1}$$

where P_{tr} is the transmit power, N_0 is the variance of the AWGN and L is the length of the packet (in bits), and Q is the tail probability for the Standard Normal random variable.

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \tag{2.2}$$

2.2.2 Transmission model

At the beginning of each slot, the transmitter chooses a transmit power P_{tr} from a set Ω of allowed powers, and transmits the packet. Depending on the channel conditions, some of the receivers get the packet correctly, while others do not. Each receiver who receives the packet correctly sends an ACK back to the transmitter, which we shall assume is sent instantaneously and is received error-free. Depending on how many receivers have not yet received the packet, the transmitter retransmits it. This process continues until a preset goal is met for each packet.

2.2.3 ARQ

We use a form of Stop-and-Wait ARQ that is adapted to broadcast systems [10], called *Stop-and-Wait with full-memory.* In this form of ARQ, the transmitter maintains a list of receivers who have received the packet correctly, updating it with every retransmission. The transmitter ignores any feedback from receivers who have received the packet prior to the current slot.

2.2.4 Goal

We consider the situation where it is sufficient for the transmitter to reach any K out of the N receivers, where $K \leq N$. The transmitter discards the packet as soon as K or more receivers receive it, and moves on to the next packet. We assume that the transmitter always has a packet to transmit. Guaranteed delivery is modeled as a special case of this formulation, with $K = N$.

2.2.5 Control Problem

We need to find the optimal choice of transmit power at every transmission, in order to minimize the energy expended per packet transmitted. Initially, we constrain the set of possible powers to take only two values $\Omega = \{P_1, P_2\}$ where $P_1 < P_2$. Later, we extend this to multiple transmit powers, and look at how the system behaves in the case of a fine-grained choice of transmit powers.

The power control policy not only determines the total energy consumption per packet, it also determines the average service time of the packet, which in turn affects the delay. Since we seek to optimize only the total energy consumption, the delay does not enter the problem formulation. Therefore, the state of the system is determined by the current packet alone. However, even considering just the total energy is a non-trivial problem. The reason for this is the following observation.

Using a lower transmission power means less energy used per transmission. However, this also leads to lower probability of successful reception, and hence more retransmissions. Overall, the total energy consumption may be greater than just using a higher power in the first place. However, if the higher power were too high, that would clearly be a waste of resources. Clearly, the choice of transmit power must depend on the number of nodes that are yet to receive the packet, and the set of powers to choose from.

2.3 Control problem as Markov Decision Process

We model the problem described above as an optimal first-passage problem in a Markovian decision process.

2.3.1 State

Define the state x_t of the system as the number of receivers yet to be reached at the beginning of slot t in order to achieve the goal for the packet in service. Therefore, $x_t \in \{0, 1, \ldots, K\}$, with $x_0 = K$. The value of x_t continuously reduces from K to 0, at which time the system is reset to the initial state.

2.3.2 Control

The control u_t denotes the choice of transmit power at time t, where $u_t \in \{1, 2\}$. Control $u_t = 1(2)$ means that transmit power $P_1(P_2)$ is being used at time t. Also, u_t can be either 1 or 2 as long as $x_t > 0$. The state $x_t = 0$ represents the goal having been reached, so we do not associate a control with it.

2.3.3 Stopping time

The stopping time τ is defined to be the first time that $x_t = 0$ is reached for the current packet.

$$\tau = \min_{t>0} \{x_t = 0 | x_0 = K\} \tag{2.3}$$

2.3.4 Control problem

We need to compute the policy $\{u(x),\ x = 1, 2, \ldots, K\}$ which minimizes

$$\sigma = E\left[\sum_{t=0}^{\tau-1} P_{u_t}\right] \tag{2.4}$$

where the cost incurred at every stage is P_{u_t}.

2.3.5 State transitions

If power P_u is used for transmission, call the corresponding probability of successful reception by a single receiver μ_u, where μ_u is given by (2.1).

For $i = 1, 2, ..., K$,

$$\begin{aligned} \text{State } x = i &\implies \text{At least } i \text{ more need to receive packet to reach goal} \\ &\implies K - i \text{ nodes have received packet} \\ &\implies N - K + i \text{ nodes have not yet received the packet} \end{aligned}$$

Therefore, a transition from state i to state j will have the following probability:

$$d_{ij}(u) = \begin{cases} \binom{N-K+i}{i-j} \mu_u^{i-j} (1-\mu_u)^{N-K+j} & i = 1, \ldots, K \quad j = 1, ..., i \\ \sum_{l=i}^{N-K+i} \binom{N-K+i}{l} \mu_u^l (1-\mu_u)^{N-K+i-l} & i = 0, \ldots, K \quad j = 0 \\ 0 & otherwise \end{cases} \tag{2.5}$$

2.4 Existing results for the Optimal first-passage problem

Derman proves in [7] that for a system with non-negative stage costs (as is the case here), a stationary policy exists that minimizes the cost function involved. Also, he outlines an algorithm to compute the optimal policy. This algorithm is described here, and later the implementation of the algorithm with respect to the above problem is discussed.

2.4.1 Computing optimal policy

The optimal policy can be computed using the method of value iteration. In the present context, it is as follows. Let $\{v_0(i), i \in I - \{0\}\}$ be arbitrary (where I is the index set of possible states), and define

$$v_{n+1}(i) = \min_{u \in \Omega} \left\{ P_u + \sum_{j \neq 0} d_{ij}(u) v_n(j) \right\}, \qquad i \in I - \{0\} \tag{2.6}$$

It is proved in [7] that $\lim_{n \to \infty} v_n(i)$ converges to the optimal value $\sigma^*(i)$ independent of $\{v_0(i)\}$. As a corollary to this, we also have that the function $\{\sigma^*(i), i \in I - \{0\}\}$ uniquely satisfies

$$\sigma^*(i) = \min_{u \in \Omega} \left\{ P_u + \sum_{j \neq 0} d_{ij}(u) \sigma^*(j) \right\}, \qquad i \in I - \{0\} \tag{2.7}$$

Thus, we start the method of value iteration with an arbitrary vector $\{v_0(i), i \in I - \{0\}\}$, and iterate it according to (2.6). In the limit, we get equation (2.7) with the optimal policy R^* as that policy determined by those actions which minimize the right hand side of (2.7). In practice, the limit may sometimes not be attained, but a large number of iterations of (2.6) should in most cases yield the optimal policy or a good approximation. The theory behind value iteration and other algorithms is also discussed in detail in [2].

2.5 Results for the two-power case

The value iteration algorithm was applied to the problem of minimizing the energy expenditure to reach K out of N receivers. Simulations were run for various values of K and N, and the following results were observed.

2.5.1 Structure of optimal policy

It was observed that the policy is always of the *threshold* type, i.e., P_1 is used for all states $x_t < N_{th}$, and P_2 for all $x_t \geq N_{th}$, where N_{th} is the threshold that is determined by the parameters of the system. In other words, for any set of parameters (N, K, P_1, P_2), the optimal policy is one where the lower transmit power is used when the system is closer to reaching the goal, and the higher transmit power is used when many receivers need to be reached. Also, the threshold does not depend solely on the difference between P_1 and P_2, but rather on the actual values of P_1 and P_2.

- For low values of P_1, with P_2 only slightly greater than P_1, the optimal policy is the one that uses P_2 always. As P_2 increases, keeping P_1 fixed, the optimal policy favors using P_1 for the cases when the number of nodes remaining is small. This threshold increases with P_2, until P_1 becomes optimal for all states.

0																				$P_2 \rightarrow$
	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		0	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3
			0	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	5	5
				0	1	2	2	2	2	3	3	4	4	4	5	6	6	7	8	8
					0	2	2	3	3	3	4	5	5	6	7	8	9	10	11	12
						0	3	3	4	4	5	5	6	7	8	10	11	13	15	17
							0	3	4	5	5	6	7	8	10	11	13	16	18	∞
								0	4	5	6	7	8	9	11	13	15	18	∞	∞
									0	6	7	8	9	10	12	14	16	∞	∞	∞
										0	8	9	10	11	13	15	17	∞	∞	∞
											0	10	11	13	14	16	∞	∞	∞	∞
												0	12	14	16	18	∞	∞	∞	∞
													0	16	18	∞	∞	∞	∞	∞
														0	∞	∞	∞	∞	∞	∞
															0	∞	∞	∞	∞	∞
																0	∞	∞	∞	∞
																	0	∞	∞	∞
																		0	∞	∞
P_1																			0	∞
$\downarrow$																				0

Figure 2.1: Threshold N_{th} when P_1 and P_2 vary from 0.5 to 10 in steps of 0.5

- For large values of P_1 (and larger values of P_2), the optimal policy was to always use P_1. It appears that increasing P_1 beyond a certain point itself is overkill, and using higher power P_2 is simply a waste.

The results of running the algorithm for the case of $N = 20, K = 18$ (and $L = 16, N_0 = 1$) with powers varying from 0.5 to 10 is shown in figure 2.1. The numbers indicate the threshold at which the transmit power switches from P_1 to P_2, with ∞ indicating that P_2 is never used. 0 is used to indicate that $P_1 = P_2$, so the choice becomes immaterial. As we can see from the figure, for small values of P_1, the higher power P_2 is favored even when there are only a few nodes to be reached. This is because the probability of reaching a single node is too low using P_1. As P_1 increases, the system tends to favor it more and more, with P_2 being used only for the higher states, i.e. ones where a large number of nodes are yet to be reached. Finally, as P_1 becomes too high, P_2 is not used at all.

2.5.2 Service time and comparison of threshold policies

While the above algorithm and results describe the structure of the optimal minimum-energy policy, they do not describe its delay performance. We define the Service time of the packet as the number of transmissions it undergoes until it finally reaches the goal. We would like to see how the average service time of the packet under the optimal-energy policy compares with other policies. This comparison is done in figure 2.2.

It is quite clear that the optimal-service-time policy is one where we always transmit with the higher power. Moreover, we expect the service time to reduce as the threshold moves toward lower values, i.e. the higher power is used for more and more states. In order to see this, we take a closer look at figure 2.2, ignoring the first point (which is the one where the lower power is used exclusively). The energy and service times are plotted against the threshold values in figures 2.3 and 2.4.

2.5.3 Varying the goal

So far, we have examined the structure of the optimal policy and the delay for a fixed goal. We would like to see how the optimal policy and the cost of transmission behave when we change the minimum number of nodes that need to be reached. We illustrate the behavior using figures 2.5 and 2.6 for the energy cost, and figure 2.7

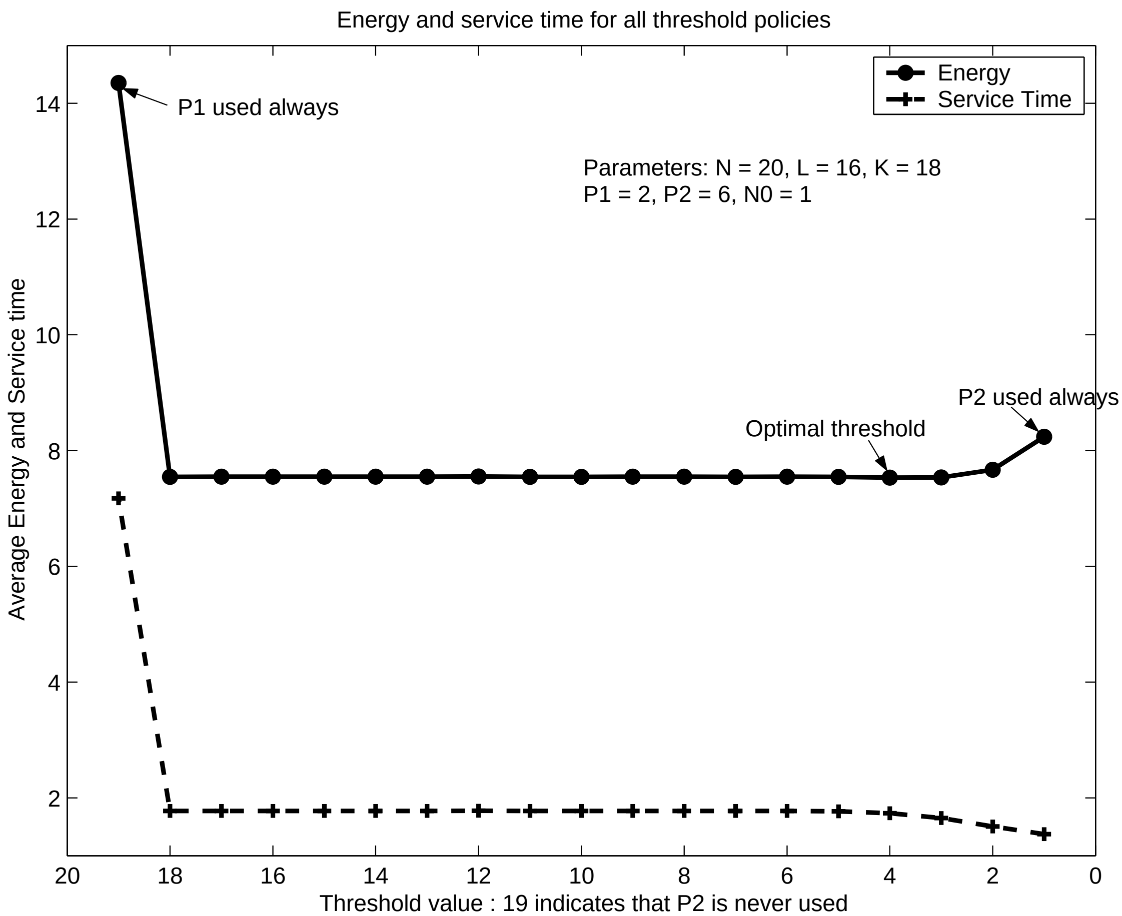

Figure 2.2: Energy and Service time vs. threshold for threshold policies

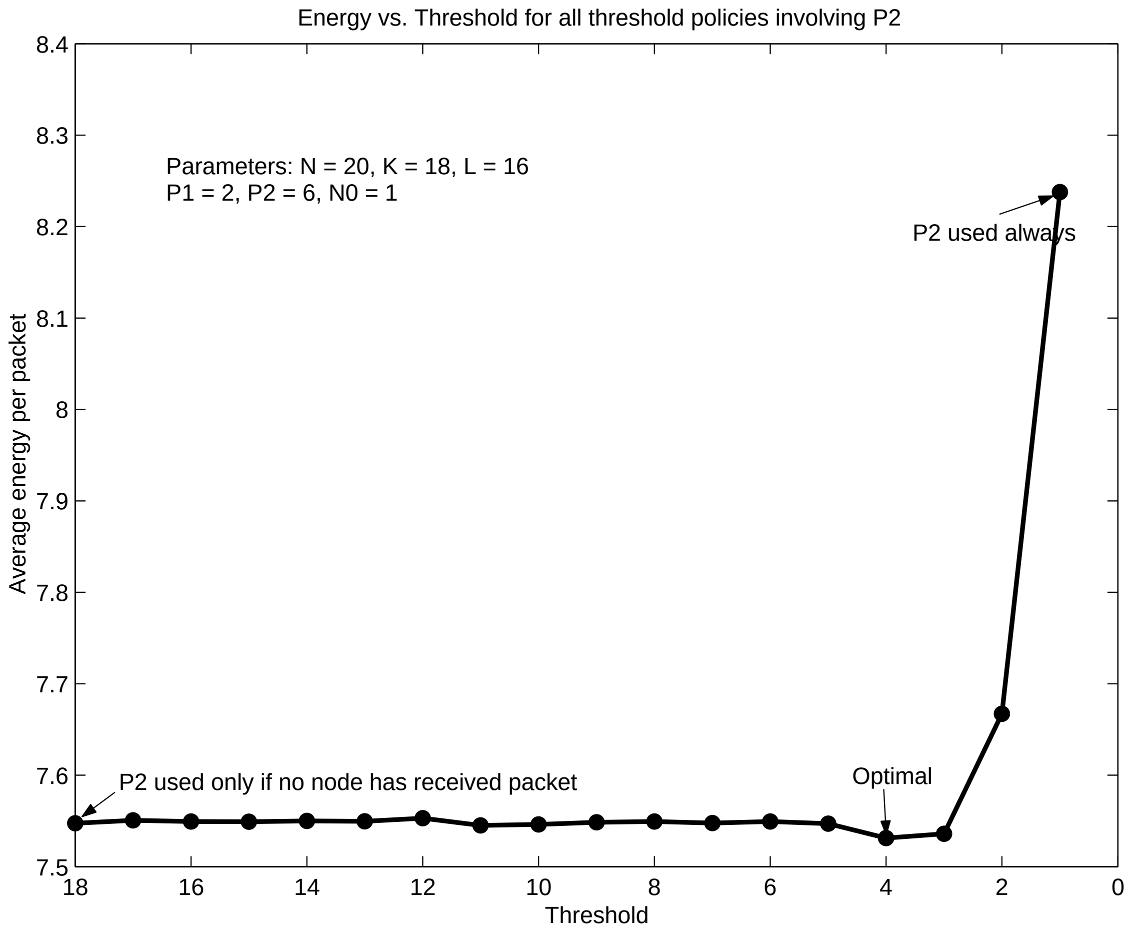

Figure 2.3: Energy vs. Threshold for threshold policies, excluding P_1-always policy

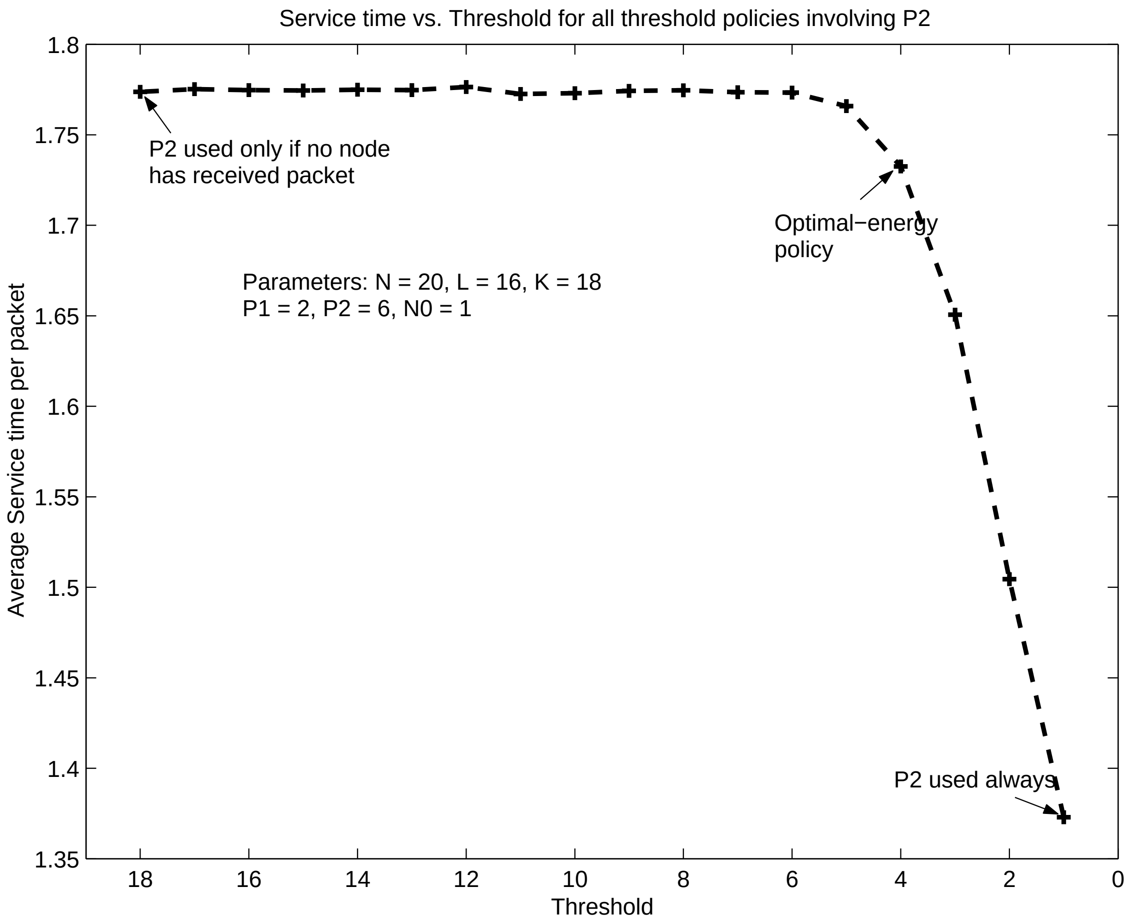

Figure 2.4: Service time vs. Threshold for threshold policies, excluding P_1-always policy

for the service-time performance.

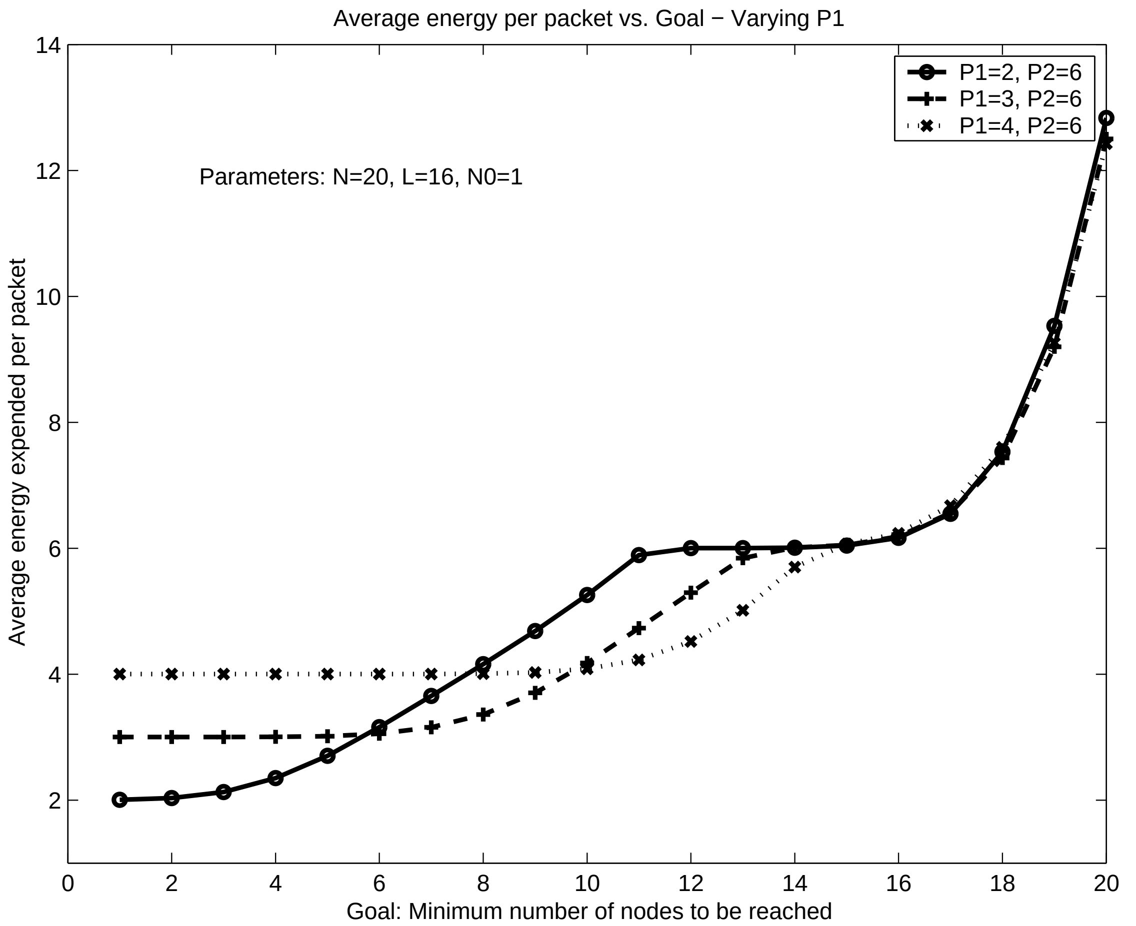

Figure 2.5: Optimal energy per packet vs. Goal - Varying P_1

The following inferences can be made from the figures.

1. As we can see from the figures, for most choices of P_1 and P_2, the curve steadily increases, and then flattens out for a while, before it starts increasing again. The only cases where this does not happen is when the optimal policy (for all goals) is always P_1 or always P_2, which are not interesting cases. Such behavior implies that there exists a range of goals which can all be reached using the same energy cost. Therefore, if we are operating at that energy level,

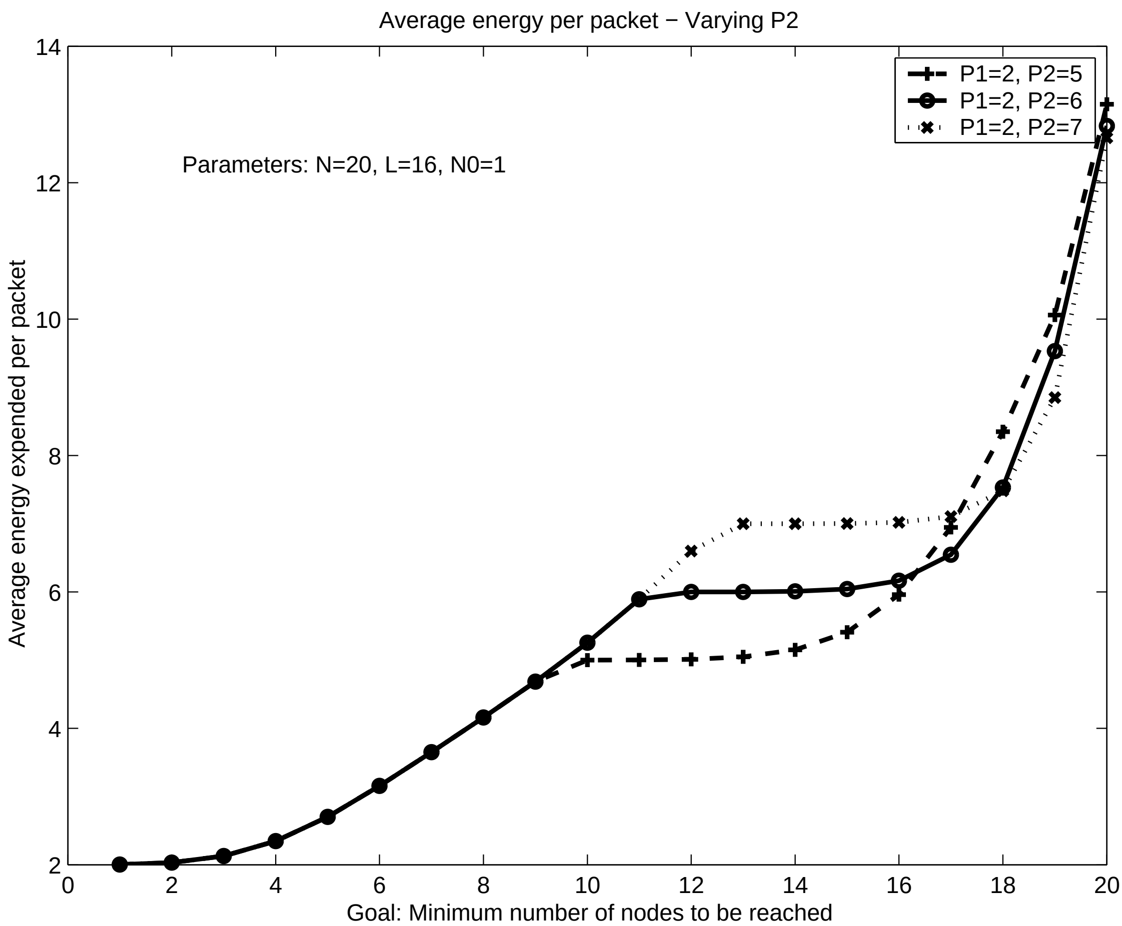

Figure 2.6: Optimal energy per packet vs. Goal - Varying P_2

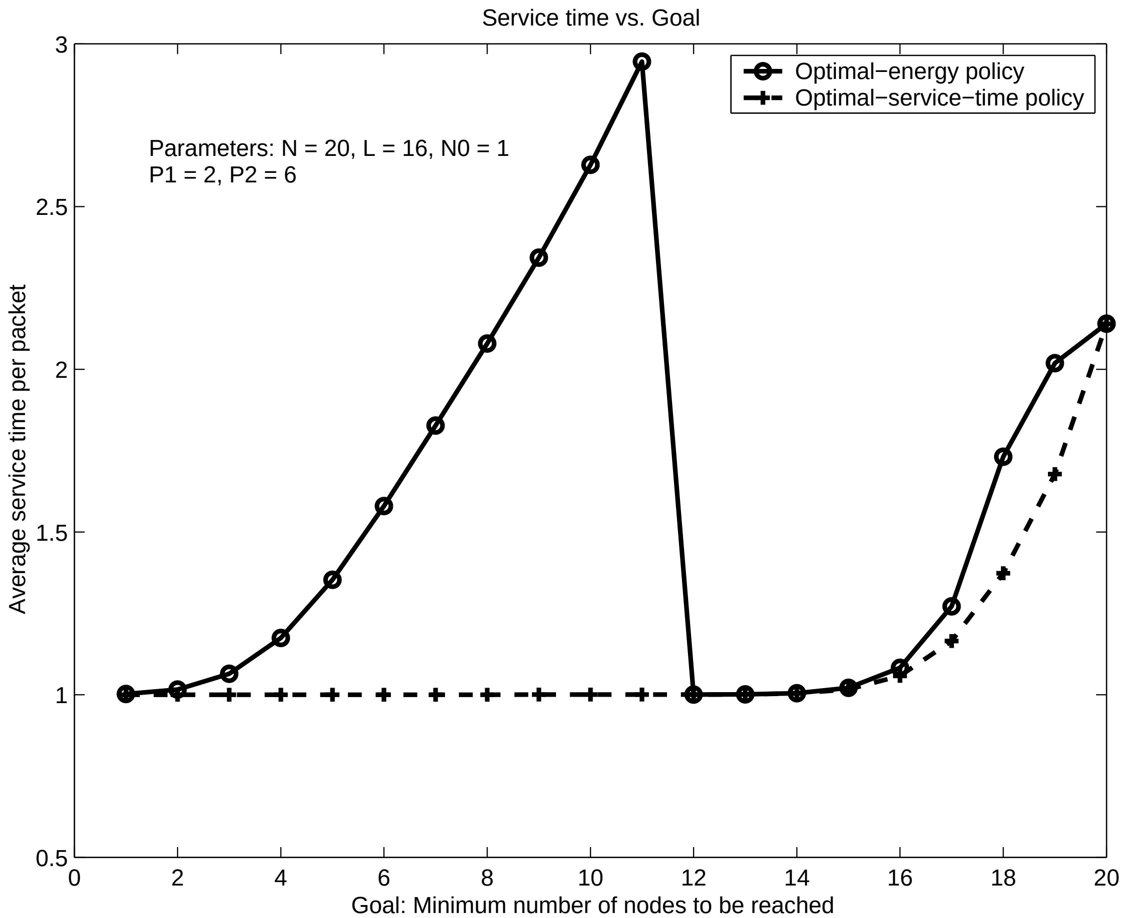

Figure 2.7: Service time vs. Goal, two powers

we could insist on having the higher goal, thus ensuring a higher quality of service for the same energy cost.

2. The width of the flat portion of the curve and the point where it starts depend on the powers chosen. It appears that fixing P_1 and varying P_2 causes the flattening to start later as P_2 increases. Similar behavior is observed when P_2 is held fixed and P_1 varied. An interesting observation can be made with regard to the optimal policy when the curve starts to flatten. We shall discuss this a little later.

3. Note also, that for a fixed P_1 and variable P_2, the energy cost when the goal is small are all the same. This is because the optimal policy in these situations favor using P_1 always. However, as the goal increases, both P_1 and P_2 are used. Here we observe that higher values of P_2 perform worse for intermediate goals, but become better as the goal becomes high. For the other case, i.e. variable P_1 with fixed P_2, the higher goals all have the same performance. This is because P_1 is used only for a few lower states, and its contribution to the net cost is minimal. At low goals, however, higher values of P_1 prove costlier. This is because for the values of P_1 that we considered (for eg, $3, 4, 5$ in the graph), a single transmission is sufficient to reach the low goal. However, if we consider low values of P_1 (for eg, $0.3, 0.4, 0.5$) where multiple transmissions are needed in order to meet the goal, the costs in the low-goal region seem to be reducing with increase in P_1. Therefore, it is difficult to generalize the behavior in this case.

4. We see from the service-time curve (Figure 2.7) that when the optimal-energy policy is used, the service time steadily increases until a certain point, at which it suddenly drops and starts increasing again. The interesting aspect is that the goal at which the service time drops is the same as the point at which the energy curve flattens out. In order to better understand this, we need to look at the optimal policies.

Let us consider the example for which the graphs are shown, i.e. $(N = 20, P_1 = 2, P_2 = 6, N_0 = 1, L = 16)$. The optimal-energy policy for all goals $K \leq 11$ turns out to be the one that always uses P_1. However, when the goal is 12, the optimal policy is one where P_2 is used for the two highest states, i.e. when the packet has just begun transmission, or has reached just one receiver. And it is exactly at this goal (i.e. 12), that the service time drops and the energy curve flattens out. This behavior is seen in all cases - the point at which P_2 enters the optimal policy is the point at which the energy and the delay curves change their behavior (which we shall refer to as the 'critical goal'). For higher goals, P_2 is used for more and more states, i.e. the threshold moves lower. Therefore, looking at the energy curves, we know the optimal policy for all goals less than the critical goal. For goals higher than the critical value, we know that P_2 is being used, but the threshold cannot be seen directly from the energy-goal graph.

2.5.4 Dynamic interpretation of the energy-goal curve

The energy-goal figures (Figures 2.5 and 2.6) are plots of optimal energy expended versus the minimum number of nodes to be reached. They may appear to offer a dynamic interpretation as well, but closer inspection reveals the interpretation is inaccurate. Further explanation of this is best accomplished using an example.

Let us again consider the example that we have used so far. Say, we have $N = 20$ nodes totally, and our goal is to reach at least $K = 13$. We operate according to the optimal policy, and say we reach 3 nodes. That leaves us in state 10. By the Principle of Optimality, we may be tempted to argue that the optimal policy from here on is the same as the optimal policy when starting with a goal of 10. However, a look at the optimal policies for goals 13 and 10 reveal that it is not so. The optimal policies turn out to be :

Goal	Optimal policy
10	P_1 for all states $(1, \ldots, 10)$
13	P_1 for states $(1, \ldots, 8)$ and P_2 for states $(9, \ldots, 13)$.

Therefore, the optimal policy for state 10 is different depending on whether we start from 10 or whether we reach state 10 starting from state 13. The reason for this apparent discrepancy is as follows.

When we start off with a goal of 10, it implies that we have to reach 10 out of 20 nodes, and can ignore 10 nodes. However, when we start off with a goal of 13 and reach 3 nodes, it implies that we now have to reach 10 out of the remaining 17 nodes, and not 10 out of 20. The two problems are not equivalent and therefore the optimal policies are different in the two cases.

2.6 Extension to multiple powers

The previous sections where the transmitter was constrained to choose between two powers gave some insight into the structure of the problem. However, most transmitters capable of power control can usually choose between multiple transmit powers, if not over a continuous range of values. Therefore, we also extend the algorithm and results from the previous sections to a problem where the transmitter has a choice of more than two powers.

The extension of the problem to multiple powers is straightforward. We now need to choose control $u \in \{1, 2, \ldots, N_{pow}\}$, where there are totally N_{pow} powers to choose from, and choosing $u = i$ implies that power P_i will be used for transmission. Also, without loss of generality, $P_1 < P_2 < \ldots < P_{N_{pow}}$. The probability of successful reception, the stopping time and the state transitions are the same as earlier.

2.6.1 Results for multiple-power case

We find that for the multiple-power case, the optimal policy is always of the *separation* type, i.e., it has the structure of using P_1 for states $\{1, ..., n_1\}$, P_2 for states $\{n_1 + 1, ..., n_2\}$, and so on, where $n_1 \leq n_2 \leq \ldots \leq K$.

When the goal is varied, the system exhibits behavior similar to the two-power case, as shown in figure 2.8 where we choose between three powers. The curve flattens at two points (6 and 14), and those points represents the goals at

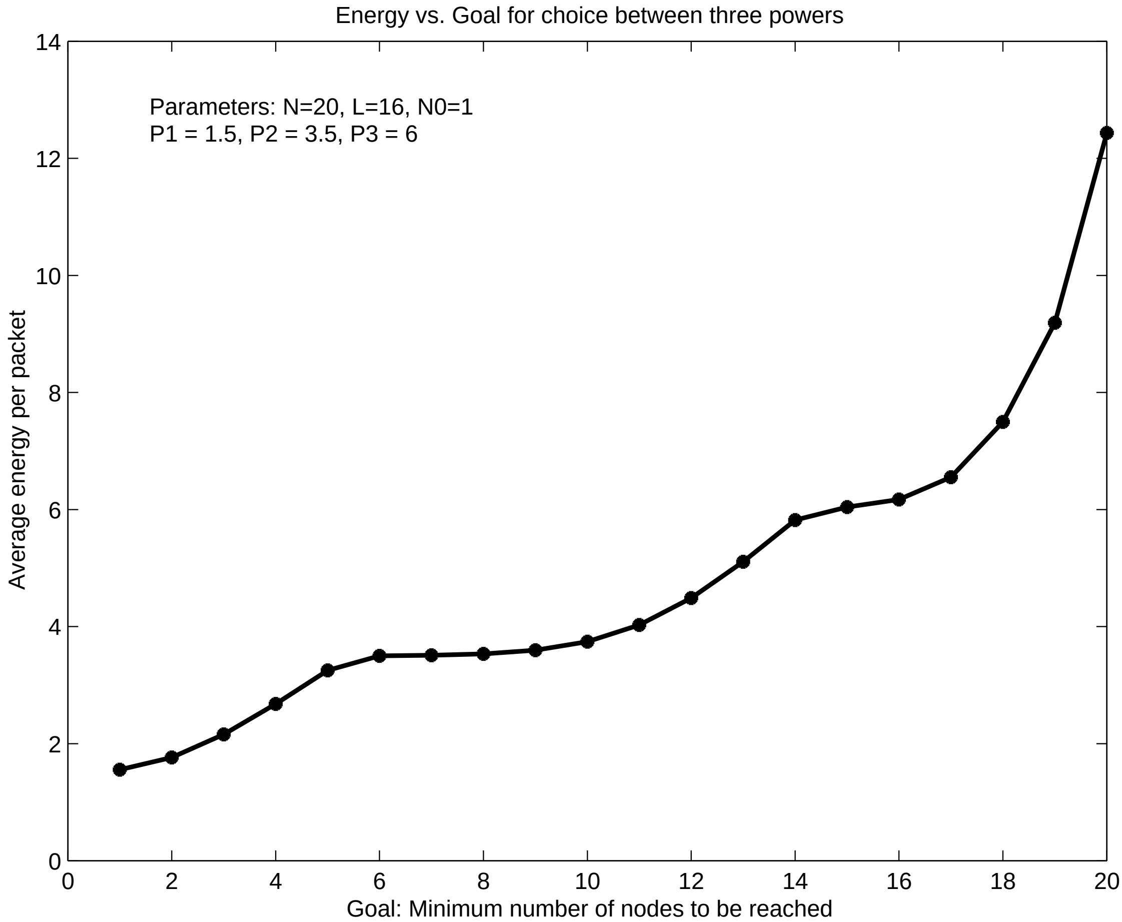

Figure 2.8: Energy vs. Goal, three powers

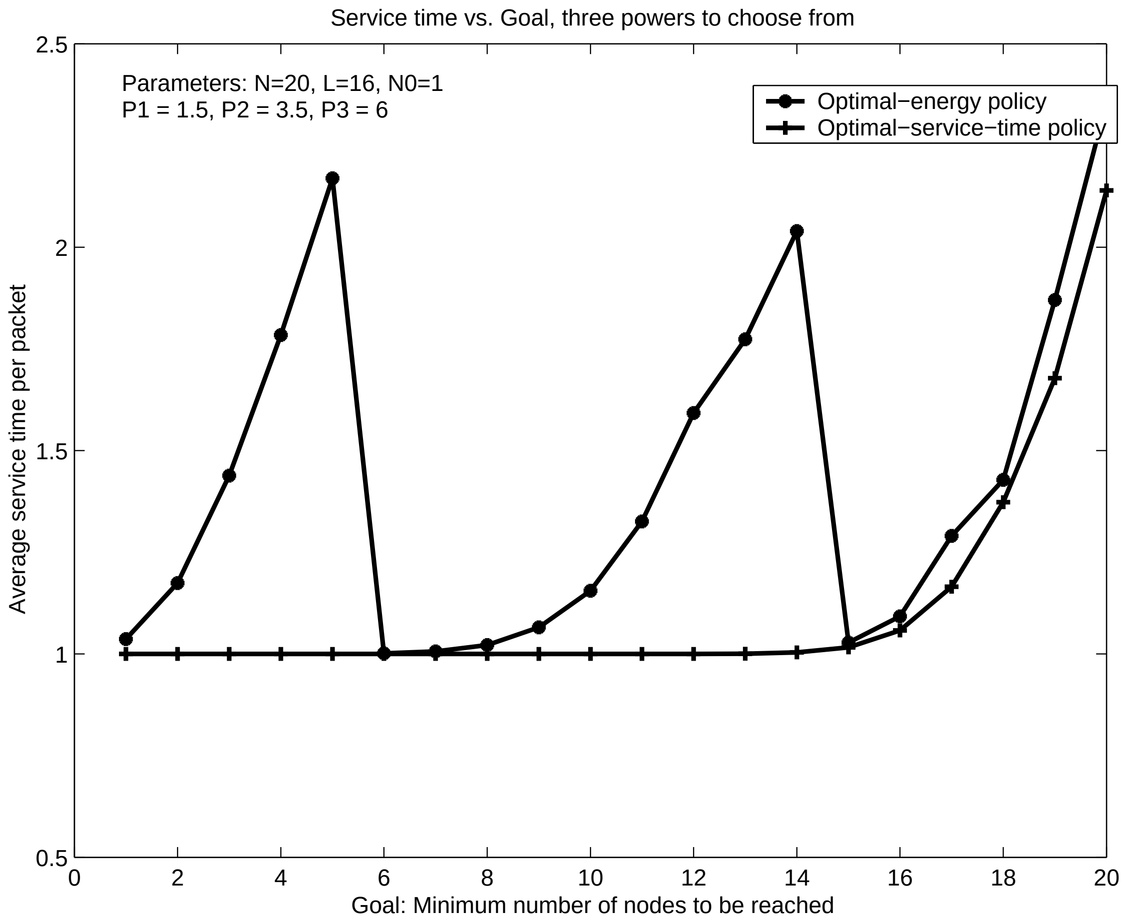

Figure 2.9: Service time vs. Goal, three powers

which P_2 and P_3 enter the optimal policy. The behavior of the service-time curve is also similar to the two-power case, as shown in figure 2.9.

2.6.2 Continuous power control

We would like to extend the problem to approximate continuous power control, i.e., where the transmitter can choose over a range of values. In order to do this, we divide the fixed interval $[P_{min}, P_{max}]$ into $N_{pow} - 1$ equal intervals, where N_{pow} is large. The transmitter has to now optimally choose between N_{pow} powers.

Structure of optimal policy

As we can see from figure 2.10, the optimal choice of transmit power increases with the state (number of nodes to be reached), in a concave fashion. Also, the optimal power for every state increases with the goal, which is also as expected.

Effect of granularity of division on optimal policy

As we can see from figure 2.11, the optimal policy converges as the interval becomes more and more fine-grained. We can see, in this example, that even dividing the interval $[1, 10]$ into 9 equal parts comes close to optimal. There is almost no difference in the policies when dividing the interval $[1, 10]$ into 50 or 500 parts. However, using only two powers is quite wasteful, especially if the higher power is large (as in figure 2.11), since the optimal transmission power does not much go above 4.

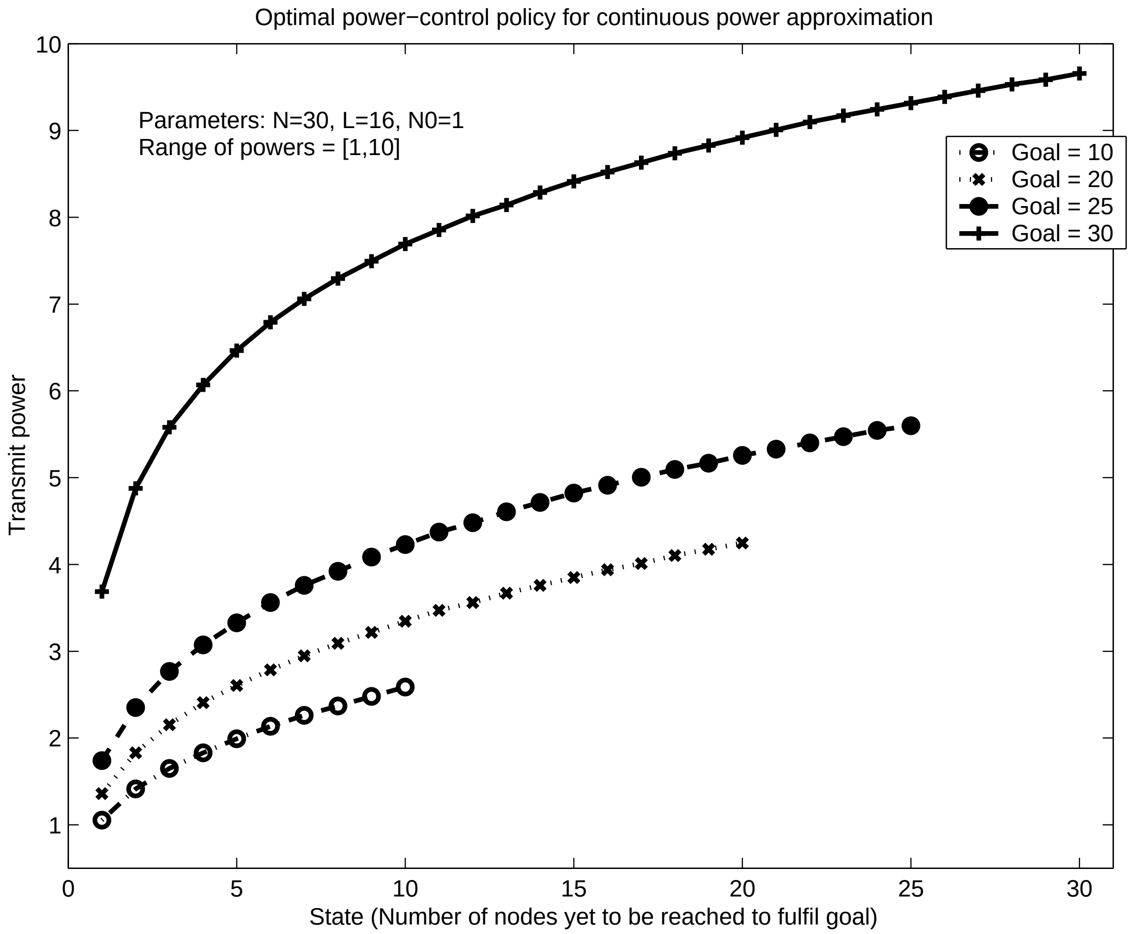

Figure 2.10: Optimal-energy policy, continuous case

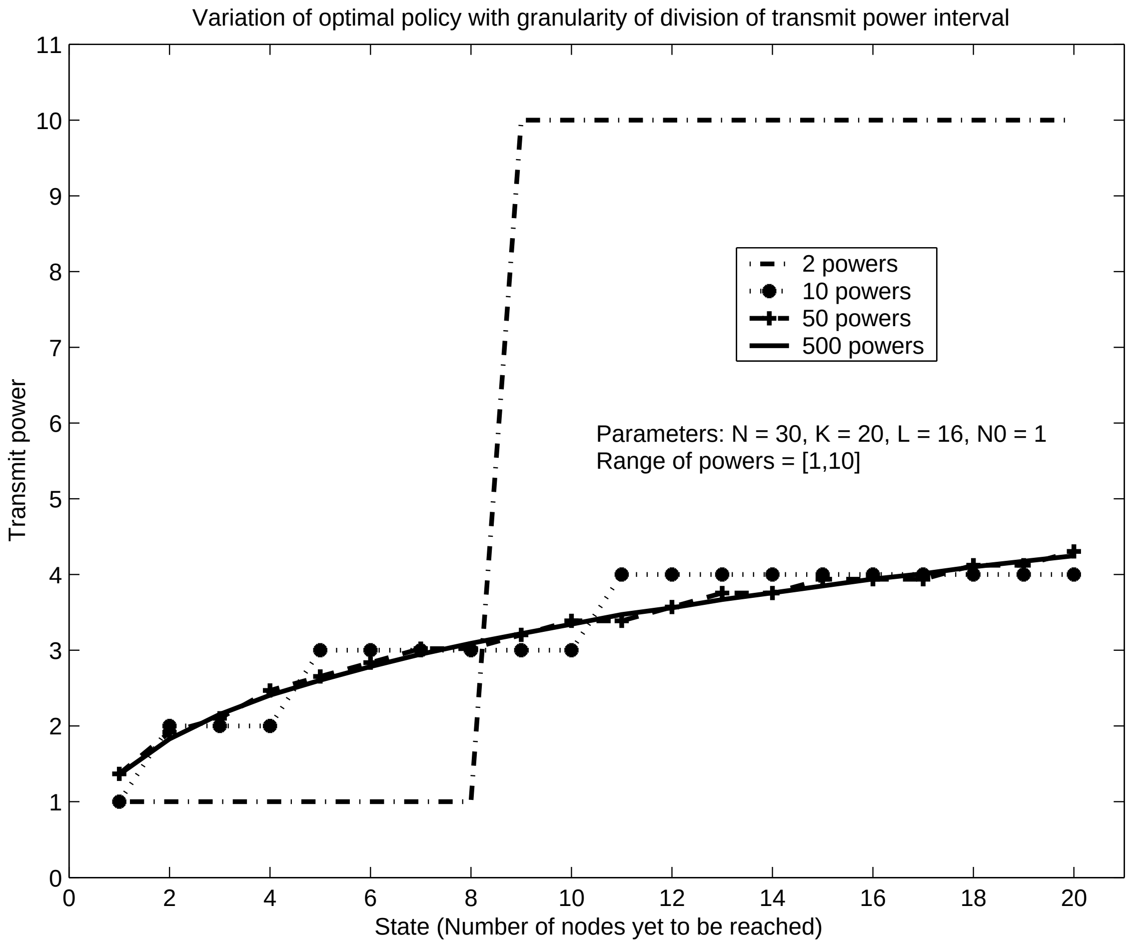

Figure 2.11: Effect of granularity of division of transmit power range

Energy and service time for continuous-power approximation

Figures 2.12 and 2.13 show how the optimum energy and service time vary with the goal. We observe here, that the optimum energy tends to grow linearly with the goal (for small goals), but increases rapidly as the goal gets closer to the total number of nodes. The delay also tends to be roughly constant over most goals, increasing as the goal gets close to the total number of receivers in the network.

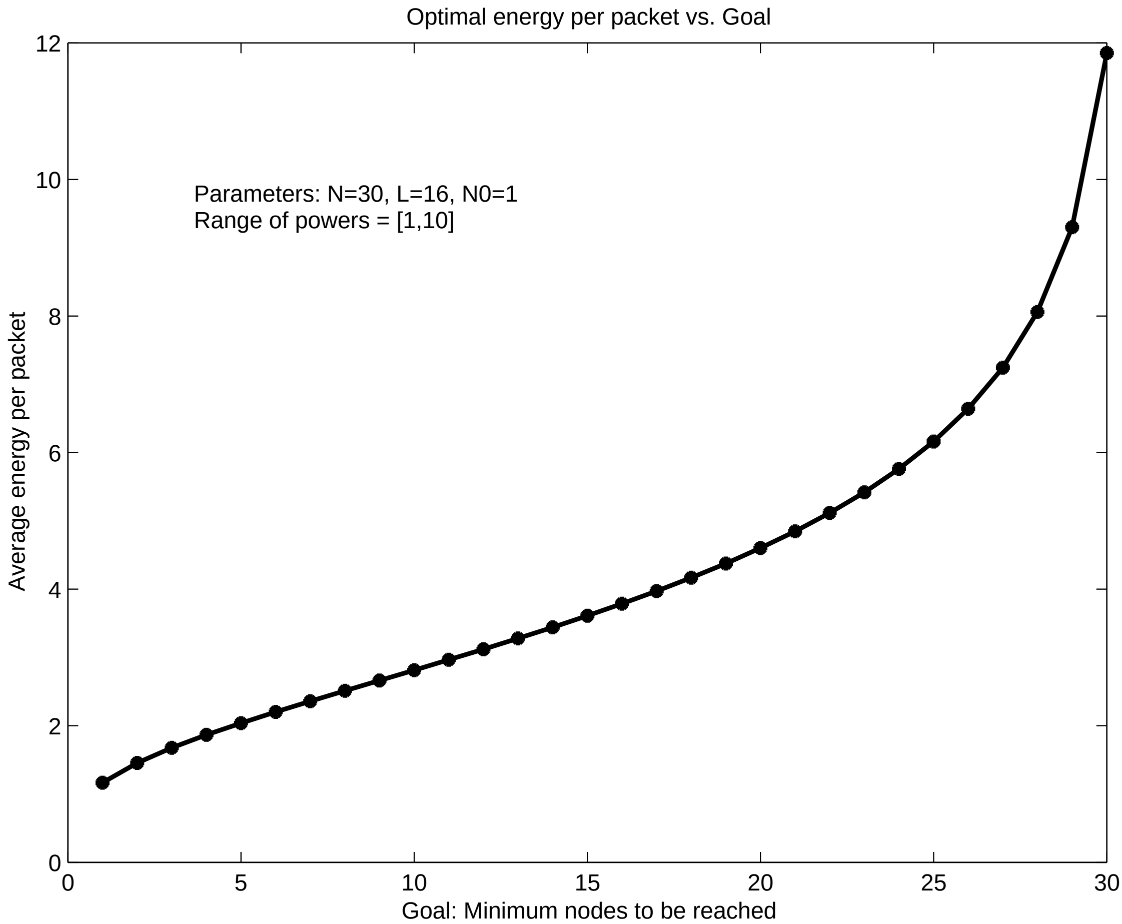

Figure 2.12: Energy vs. Goal, approximation to continuous power control

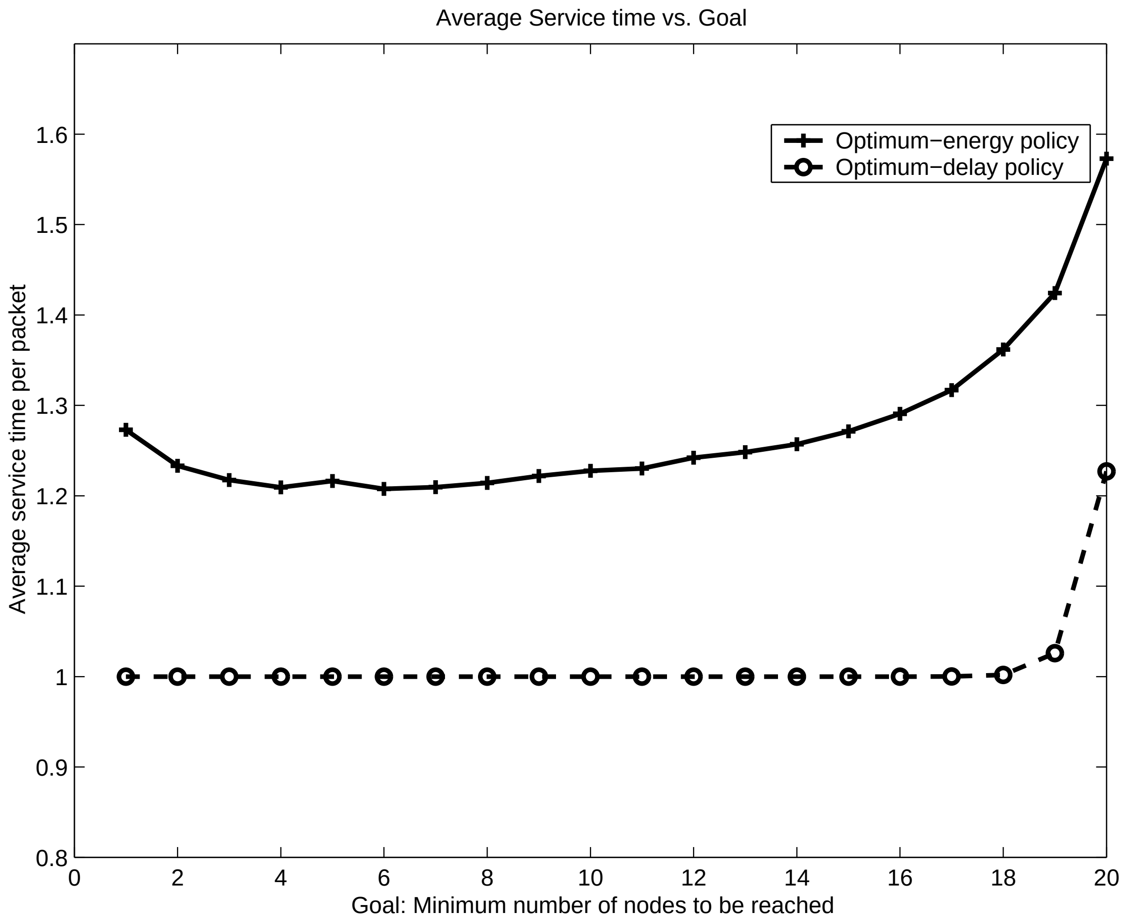

Figure 2.13: Service time vs. Goal for continuous power approximation

2.7 Conclusions

In this chapter, we studied the problem of finding the power-control policy that will give us the minimum energy consumption, when it is sufficient to reach a subset of the total number of receivers. Initially, we considered the case when the transmitter could only choose between two powers, and found that the optimal policy was always of the threshold type. As a result of this, the control mechanism can be implemented using a simple switch. Also, since it does not involve constant swings in transmit power, the electronics can be much simpler to design. Simpler electronics also leads to some power savings in the node, which is an added bonus. We also studied how the threshold varied as functions of the two powers, when all other parameters of the system were fixed.

The variation of the optimum energy with the goal was studied, and it was found that there exist a range of goals for which the energy cost is the same. Therefore, if we are willing to operate at that energy cost, we could insist on having the highest among those goals as the target. The average service time also reduces if we operate in this region. We also found that the goal at which the higher power enters the optimal policy is the same goal at which the energy curve flattens and the service-time curve drops.

This study was extended to the case when the transmitter could choose between more than two powers, and it was found that the optimal policy was now of the separation type. Again, the energy and delay performance was studied as the goal was varied. The results were then further extended to approximate continuous power control. The conclusion that can be drawn from this is that it is expensive (in terms of energy and delay) to insist on guaranteed delivery. Instead, if we can build in some amount of redundancy into the system, by which most (but not all) nodes receive each packet, we will save significantly on the energy and the delay, since they seem to rise sharply as the goal moves closer to the total number of nodes.

Chapter 3

Effect of Network Architecture on Downlink Performance

3.1 Introduction

Network architecture plays an important role in determining the performance of network protocols. In this chapter, we compare the downlink performance of satellite and cellular network architectures, in terms of energy consumption and delay performance, as the parameters of the system are varied.

3.2 Differences between Satellite and Cellular architectures

While both satellite and cellular networks both are infrastructure-based wireless networks, they have some important differences. Some of these differences are discussed in the following section.

1. **Infrastructure cost**

   Both satellite and cellular networks require a significant infrastructure to be set up before they can be deployed. While it is incredibly expensive to design and launch a satellite, the cellular architecture is no less expensive because it involves setting up a large number of base stations.

2. **Integration with existing infrastructure**

   This is a significant advantage that cellular networks have over satellite networks. They can be integrated seamlessly into the existing telephone infrastructure, whereas satellite networks require more infrastructure in the form of earth stations in order to be integrated with the telephone network.

3. **Effect of mobility**

   Mobility is a big issue with cellular networks, with handoffs being an important consideration in network design. Geo-stationary satellites, owing to their fixed position and their great distance, are relatively unaffected by node mobility on the ground. However, Low-Earth Orbit (LEO) satellites have to deal with handoffs, not only because of node mobility, but also because of the satellite's motion in its orbit.

4. **Re-configurability**

   Cellular networks win hands-down on this issue, since it is much easier to upgrade the base stations than it is to perform upgrades on a satellite. Therefore, satellites have to be designed keeping in mind the long-term requirements of the network.

5. **Energy availability**

   Satellites are powered by batteries that are charged by solar cells, while base stations in a cellular network are usually connected to power mains. Therefore, energy is a critical consideration in satellite networks. This, combined with limited processing capability, makes energy-efficient network protocols a very important factor in network design.

6. **Round-trip time**

   Satellites, owing to their great distance, have the associated problem that the round-trip time is comparable to the packet transmission time. This requires special protocols in many cases. For example, CSMA, which is an efficient protocol for small-scale terrestrial networks cannot be used with satellites. This also imposes a minimum delay on packet transmissions, limiting the throughput that can be achieved.

3.3 Problem Statement

We would like to compare the downlink broadcast performance of satellite and cellular network architectures, in terms of energy and delay, as the parameters of the network such as the input rate and the transmit power are varied. The system model that we use, along with the simplifying assumptions that we make, are described in the section below.

3.4 System model

3.4.1 Architecture

Satellite

The network consists of a satellite orbiting at a height h, serving N nodes which lie in the satellite's footprint. In this thesis, we consider the case of LEO satellites, where $h = 1000$km. The nodes on the ground are assumed to be stationary. We assume that the satellite has a single spot-beam.

The issue of handoffs arising due to the movement of the satellite footprint is not addressed. The time period of interest is when the satellite is directly overhead and has connectivity with the area in consideration.

Cellular network

The satellite footprint is also served by a cellular network, consisting of M non-overlapping cells of radius r. We shall make a simplifying assumption that the N nodes are equally distributed among the M base stations ($N >> M$), and are stationary.

3.4.2 Bandwidth

In order to make a fair comparison between the two architectures, we assume that both have the same bandwidth of W bits/second. The satellite can use this bandwidth completely, while for the cellular network it is divided into equal, non-overlapping bands. Because of frequency re-use, and taking the reuse-factor to be 7 as in the GSM case, each base station is therefore allotted bandwidth $W/7$. By this scheme, for any given cell, the cells adjoining it use different frequency bands, and we assume that transmissions over the same band by base stations that are not adjacent to it do not cause any interference.

3.4.3 Data

Data is assumed to be in the form of fixed-length packets of length L bits each. Because of the nature of data traffic, we do not have any strict deadlines on the delay that these packet incur. However, we do require that they be delivered error-free.

3.4.4 Discrete-time model

We assume time to be slotted, and the slot length is chosen to be equal to a packet transmission time when the entire bandwidth W is used for transmission. In other words, slot length $\Delta = L/W$. All transmitters are taken to be slot-synchronous.

3.4.5 Round-trip time

Since we consider LEO satellites (which are at a height of 1000 km), the round trip time is equal to

$$RTT = \frac{2 \times 1000 \text{ km}}{3 \times 10^8 \text{ m/s}} = 6.66 \text{ msec} \tag{3.1}$$

The round-trip time of the satellite is approximated to an integral number of slots. Terrestrial transmissions are assumed to have zero round-trip time.

3.4.6 Type of service

We are interested in the broadcast case, where each packet is intended for all N nodes. Also, we consider the guaranteed-delivery situation, where we require that all N destinations receive the packet correctly.

3.4.7 Queuing

Each of the transmitters (terrestrial and satellite) has a single queue for the packets. The buffer size is assumed to be unlimited in both cases. Packets at the head of the queues are broadcast to all the receivers.

3.4.8 Arrivals

Since we are primarily interested in the downlink performance, we shall assume the packet generation process to be as follows.

Satellite

In each slot, a source generates a packet with a probability λ, independent across slots. Upon generation, it is put into the queue in the satellite, awaiting transmission.

Cellular

In each slot, a central source generates a packet with a probability λ, independent across slots. This packet is then put into the queues at all the base stations. We assume that this process happens instantaneously, so the arrival processes at all M queues is identical.

3.4.9 Channels

The channel between each transmitter-receiver pair is modeled as an AWGN channel with attenuation. As a simplifying assumption, all channels are assumed to be independent of each other, and independent from bit to bit. The amplitude of the received signal y when the transmitted signal is x is given by

$$y = d^{-\gamma} x + w \tag{3.2}$$

where w is white Gaussian noise with variance N_0, and d is the distance between the transmitter and the receiver.

1. Satellite channel

   Considering BPSK modulation, if the satellite transmits a bit with a power P, then the received amplitude is $h^{-\gamma_s}\sqrt{P}$, where h is the height of the satellite and γ_s is the path-loss exponent for the satellite. The received signal power is therefore $h^{-2\gamma_s}P$.

2. Terrestrial channel

   BPSK modulation with transmit power P will result in a received signal power of $r^{-2\gamma_t}P$ and a received amplitude of $r^{-\gamma_t}\sqrt{P}$, where r is the radius of the cell and γ_t is the path-loss exponent for the cellular case. The received power will actually vary with distance from the base station, but we consider the lower bound to apply for all nodes. While we make this assumption for simplicity, we recognize that the actual energy consumption and the end-to-end delay in the cellular architecture will be lower than the values computed here.

3.4.10 Successful reception

For a packet of L bits, assuming BPSK modulation, the entire packet is received successfully with a probability

$$P[success] = \left[1 - Q\left(\frac{a_{recd}}{\sqrt{N_0}}\right)\right]^L \tag{3.3}$$

where a_{recd} is the received amplitude and N_0 is the noise variance.

Therefore, for the satellite and terrestrial channels respectively,

$$p_s = \left[1 - Q\left(h^{-\gamma_s}\sqrt{\frac{P_{tr}}{N_0}}\right)\right]^L \tag{3.4}$$

$$p_t = \left[1 - Q\left(r^{-\gamma_t}\sqrt{\frac{P_{tr}}{N_0}}\right)\right]^L \tag{3.5}$$

where P_{tr} is the transmit power used.

3.4.11 The transmission process

Satellite network

The satellite transmits the packet at the head of the queue with a power P_{tr} for the duration of the packet transmission. Each node receives the packet correctly with a probability p_s described earlier. If the transmission takes A slots and the round-trip time is R slots, then the satellite expends energy for the first A slots, and receives ACKs from all the nodes at the end of $A+R$ slots. If any node has not received the packet correctly, the satellite then retransmits the packet with the same power P_{tr}. We assume that the transmit power does not vary with the state of the system, i.e., no power control scheme is present. When all nodes receive the packet successfully, the packet is removed from the queue and the next packet is served. Acknowledgments are assumed to be instantaneous, and Stop-and-wait ARQ (with full-memory) is followed.

Cellular network

The terrestrial case is similar to the satellite case. All base stations transmit with the same power P_{tr}, which is fixed. If any node that is associated with a base station does not receive the transmitted packet successfully, the base station retransmits the packet with power P_{tr}. Once all nodes in the cell receive the packet correctly, it is removed from the queue and the base station transmits the next packet.

3.5 Quantities of interest

We would like to compare the performance of the two architectures, based on the following criteria. The parameters that are under our control are the packet arrival rate and the power used for transmission. Varying these two, we would like to see how the following quantities behave.

1. **Energy expended per packet**

   This is defined as the total energy expended in transmitting the packet (over all queues). We consider only the transmission energy here, processing and other factors are ignored.

2. **End-to-end delay**

   We define the end-to-end delay of a packet as the time between its arrival instant and the time at which all N receivers receive the packet correctly. In the satellite case, this translates to the instant at which the packet leaves the queue. In the cellular case, however, this implies the time at which the packet leaves the last queue.

3. **Maximum stable throughput**

   This refers to the maximum arrival rate that the queue can sustain, without becoming unstable. This depends on the service time of each packet, which depends on the power used for transmission.

3.6 Mathematical modeling of the network architectures

We model both architectures as systems of queues, fed by a Bernoulli arrival process [33]. For the satellite network, this consists of a single queue into which all packets are put, while for the cellular network we have M queues operating in parallel, but having identical arrival processes. Also, the Bernoulli arrival process enables us to model these systems as discrete-time $Geo/G/1$ queues.

Because the behavior of all queues involved is similar, we can derive results for a generic $Geo/G/1$ queue having a structure similar to the queues in consideration. The results for the satellite and cellular cases can then be obtained by substituting the appropriate parameter values into the generic results.

3.6.1 A Generic $Geo/G/1$ queue

Consider a queue that has a Bernoulli process with rate λ as the input process. This queue serves N_c nodes, repeatedly re-transmitting packets until all N_c nodes receive the packets correctly. The packet transmission time is A slots and the round-trip time is R slots. In order to describe the queue, we need to describe the service-time process.

From the channel model, we see that transmissions from the transmitter to a particular receiver is independent of transmissions to any other receiver. Also, for a given power level P_{tr}, the receiver receives a transmission correctly with probability p, where p is determined by the channel model (Equations 3.4, 3.5). Therefore, for a given packet, node i requires a number of transmissions X_i, where X_i is distributed according to a Geometric distribution with parameter p. In other words,

$$P\left[X_i = r\right] = (1-p)^{r-1}p \quad r = 1, 2, \ldots \quad i = 1, \ldots, N_c \tag{3.6}$$

The packet has to be transmitted atleast X_i times to node i. However, owing to the Wireless Multicast Advantage [15], each transmission reaches all N_c nodes. Therefore, the total number of transmissions that a packet undergoes is equal to the maximum of $\{X_i \; i = 1, ..., N_c\}$. The number of transmissions required for successful reception of the packet can therefore be computed as follows.

$$N_{tr} = \max_{i=1,\ldots,N_c} X_i \qquad X_i \sim Geo(p) \tag{3.7}$$

$$P\left[N_{tr}=n\right]=P\left[\max_{i=1,\ldots,N_c} X_i=n\right]=P\left[all\ X_i\leq n\right]-P\left[all\ X_i<n\right] \tag{3.8}$$

We know that for Geometric distributions,

$$P\left[X_i\leq n\right]=\sum_{j=1}^{n}p(1-p)^{j-1}=1-(1-p)^n \tag{3.9}$$

Combining this with the independence of the X_i, we get the distribution for N_{tr} to be of the following form.

$$P\left[N_{tr}=n\right]=(1-q^n)^{N_c}-(1-q^{n-1})^{N_c}\qquad n=1,2,... \tag{3.10}$$

where $q=1-p$ is the probability that a receiver fails to receive a packet transmission.

PGF of N_{tr}

We would like to compute the Probability Generating Function of the number of transmissions that each packet undergoes, in order to determine the delay experienced by the packets in the queue.

$$\begin{aligned}
F(z)=E[z^{N_{tr}}]&=\sum_{y=1}^{\infty}P[N_{tr}=y]z^y\\
&=\sum_{y=1}^{\infty}\left[(1-q^y)^{N_c}-(1-q^{y-1})^{N_c}\right]z^y\\
&=\sum_{y=1}^{\infty}(1-q^y)^{N_c}z^y-\sum_{y=1}^{\infty}(1-q^{y-1})^{N_c}z^y\\
&=\sum_{y=1}^{\infty}(1-q^y)^{N_c}z^y-\sum_{y=0}^{\infty}(1-q^y)^{N_c}z^y.z\\
&=\sum_{y=0}^{\infty}(1-q^y)^{N_c}z^y(1-z)
\end{aligned} \tag{3.11}$$

$$
\begin{aligned}
\sum_{y=0}^{\infty}(1-q^y)^{N_c}z^y &= \sum_{y=0}^{\infty}\sum_{n=0}^{N_c}\binom{N_c}{n}(-1)^n q^{yn}z^y \\
&= \sum_{n=0}^{N_c}\binom{N_c}{n}(-1)^n\sum_{y=0}^{\infty}(zq^n)^y \\
&= \sum_{n=0}^{N_c}\binom{N_c}{n}(-1)^n\frac{1}{1-zq^n} \qquad for\ \ |z|<1/q^n
\end{aligned} \tag{3.12}
$$

$$
\Longrightarrow\ F(z) = \sum_{n=0}^{N_c}\binom{N_c}{n}(-1)^n\frac{1-z}{1-q^n z} \qquad |z|<1 \tag{3.13}
$$

Simplifying,

$$
F(z) = 1+\sum_{n=1}^{N_c}\binom{N_c}{n}(-1)^n\frac{1-z}{1-q^n z} \qquad |z|<1 \tag{3.14}
$$

From this, we can derive the mean and the second moment of N_{tr}.

$$
\begin{aligned}
E[N_{tr}] &= \lim_{z\to 1}F'(z) = \sum_{n=1}^{N_c}\binom{N_c}{n}\frac{(-1)^{n+1}}{1-q^n} \\
E[N_{tr}^2] &= \sum_{n=1}^{N_c}\binom{N_c}{n}(-1)^{n+1}\frac{1+q^n}{1-q^n}
\end{aligned} \tag{3.15}
$$

Service-time distribution

The service time Y of a packet is equal to the number of transmissions of the packet scaled by the number of slots taken per transmission i.e. $Y = (A+R)N_{tr}$. Therefore, the distribution of the service time (measured in slots), and the PGF are given by

$$
P[Y=y] = \begin{cases} (1-q^n)^{N_c}-(1-q^{n-1})^{N_c} & if\ y=(A+R)n,\ n=1,2,\ldots \\ 0 & otherwise \end{cases} \tag{3.16}
$$

$$
F_y(z) = F(z^{A+R}) = 1+\sum_{n=1}^{N_c}\binom{N_c}{n}(-1)^n\frac{1-z^{A+R}}{1-q^n z^{A+R}} \qquad |z^{(A+R)}|<1 \tag{3.17}
$$

3.7 Application to Satellite and Cellular queues

The results from the above section can now be applied to the particular cases of the satellite and cellular architecture that we are considering. For the satellite architecture, we have $N_c = N, A = 1, p = p_s$, while for each of the queues in the cellular architecture, the parameters are now $N_c = N/M, A = 7, R = 0, p = p_t$. Applying these values to the equations above, we get the following results.

3.7.1 Satellite

For the satellite network, we have the distribution, mean and second moment of the service time to be the following.

$$P[Y = y] = \begin{cases} (1 - q_s^n)^N - (1 - q_s^{n-1})^N & if\ y = (1 + R)n,\ n = 1, 2, \ldots \\ 0 & otherwise \end{cases} \tag{3.18}$$

$$E[Y] = \sum_{n=1}^{N} \binom{N}{n} \frac{(-1)^{n+1}(1 + R)}{1 - q_s^n} \tag{3.19}$$

$$E[Y^2] = \sum_{n=1}^{N} \binom{N}{n} (-1)^{n+1}(1 + R)^2 \frac{1 + q_s^n}{1 - q_s^n} \tag{3.20}$$

3.7.2 Cellular

For each of the queues in the cellular network, the distribution, mean and second moment are as follows.

$$P[Y=y]=\begin{cases}(1-q_t^n)^{N/M}-\left(1-q_t^{n-1}\right)^{N/M} & if\ y=7n,\ n=1,2,\ldots\\ \\ 0 & otherwise\end{cases} \tag{3.21}$$

$$E[Y]=\sum_{n=1}^{N/M}\binom{N/M}{n}\frac{7(-1)^{n+1}}{1-q_t^n} \tag{3.22}$$

$$E[Y^2]=\sum_{n=1}^{N/M}\binom{N/M}{n}(-1)^{n+1}7^2\frac{1+q_t^n}{1-q_t^n} \tag{3.23}$$

Since we assume that the nodes are divided equally between the cells, N/M is an integer.

We are now in a position to compute the average energy and delay of the two architectures, which we shall do in the following section.

3.8 Energy consumption

Before we derive the expressions for the average energy consumption for the two cases, it would be instructive to study the variation of the consumed energy with transmit power, when the transmitter has to reach only a single user.

3.8.1 Energy consumption for single user

Following the analysis done above, it can be quite clearly seen that for a single user, the number of transmissions of a packet is Geometrically distributed with probability p, which depends on the transmit power P_{tr} as described earlier (equation 3.4). Without loss of generality, let us assume that there is no attenuation, and that $N_0 = 1$. The average energy consumption per packet is then given by

$$E_{avg} = N_{avg} E_{trans} \tag{3.24}$$

where N_{avg} is the average number of transmissions, and E_{trans} is the energy expended per transmission. Therefore, the average energy expenditure is given by

$$E_{avg} = \frac{P_{tr}\Delta}{p} = \frac{P_{tr}\Delta}{\left[1 - Q\left(\sqrt{\frac{P_{tr}}{N_0}}\right)\right]^L} \tag{3.25}$$

Assuming $\Delta = 1$, we study the behavior of this curve with P_{tr}. As we can see, for low values of P_{tr}, the denominator rapidly increases to 1, while the numerator does not increase so rapidly. Therefore, the slope of E_{avg} is negative near $P_{tr} = 0$. However, at high P_{tr}, the denominator flattens out close to 1, while the numerator increases linearly. Therefore, the slope of the curve is now positive. This indicates that there exists a point of minimum on the curve. To compute this point of minimum, we take the derivative as shown below.

$$\frac{d}{dP} E_{avg} = \frac{\left[1 - Q\left(\sqrt{\frac{P}{N_0}}\right)\right]^L - P\frac{d}{dP}\left[1 - Q\left(\sqrt{\frac{P}{N_0}}\right)\right]^L}{\left(\left[1 - Q\left(\sqrt{\frac{P}{N_0}}\right)\right]^L\right)^2} \tag{3.26}$$

Equating this to 0 for minimum and substituting $N_0 = 1$,

$$\left[1 - Q\left(\sqrt{P}\right)\right]^L + LP\left[1 - Q\left(\sqrt{P}\right)\right]^{L-1} \frac{d}{dP} Q\left(\sqrt{P}\right) = 0 \tag{3.27}$$

We also know that

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt \tag{3.28}$$

$$\implies \frac{d}{dx} Q(x) = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{3.29}$$

Therefore, the above equation simplifies to

$$\left[1 - Q\left(\sqrt{P}\right)\right] = \frac{L\sqrt{P}}{2\sqrt{2\pi}}e^{-P/2} \tag{3.30}$$

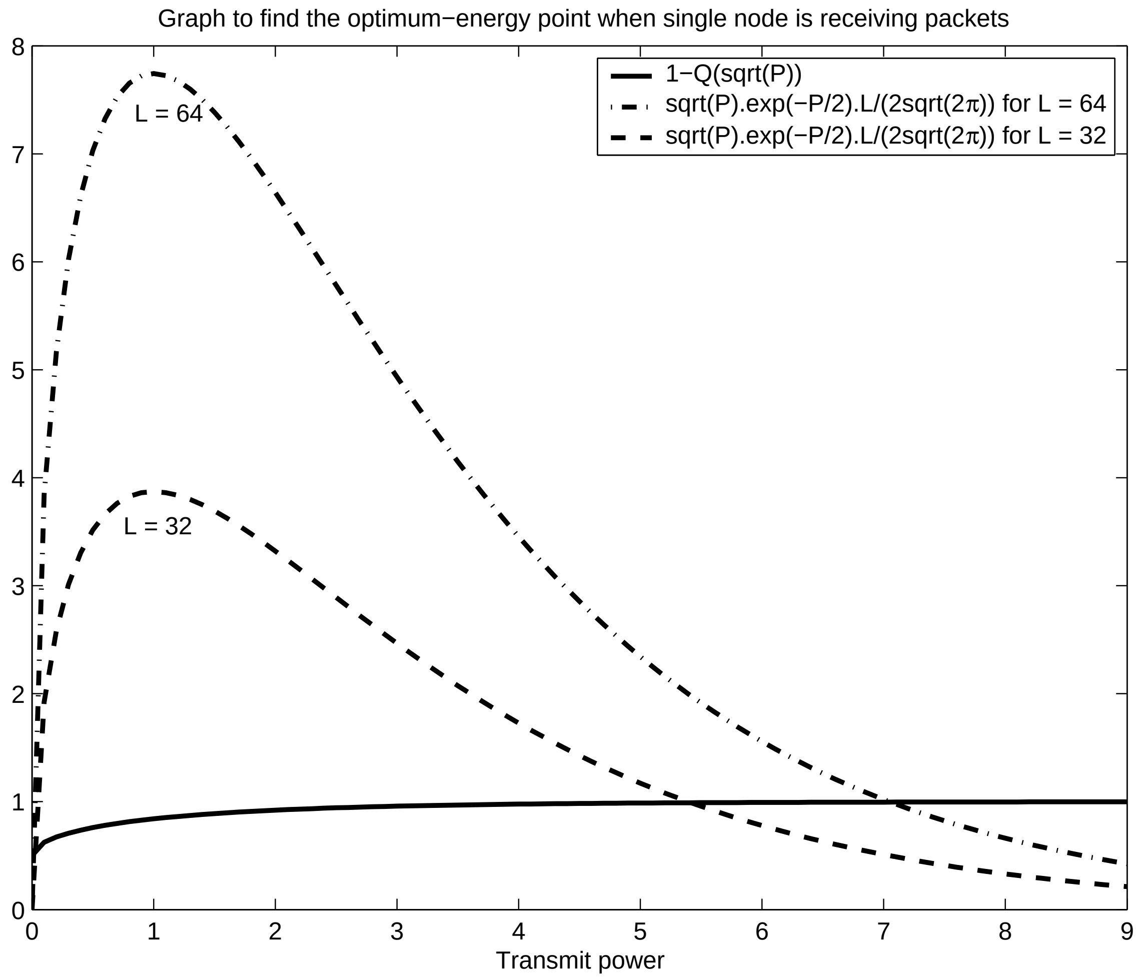

Figure 3.1: Graph to determine optimal-energy point in single-receiver transmission

In order to see the behavior of this equation, we plot both sides of the equation, as shown in Figure 3.1. We can see that there exists a power P which satisfies this equation, and that this P increases with increasing L. Therefore, if the packet length is larger, we require higher transmit power to reach the optimal-energy point for a single receiver.

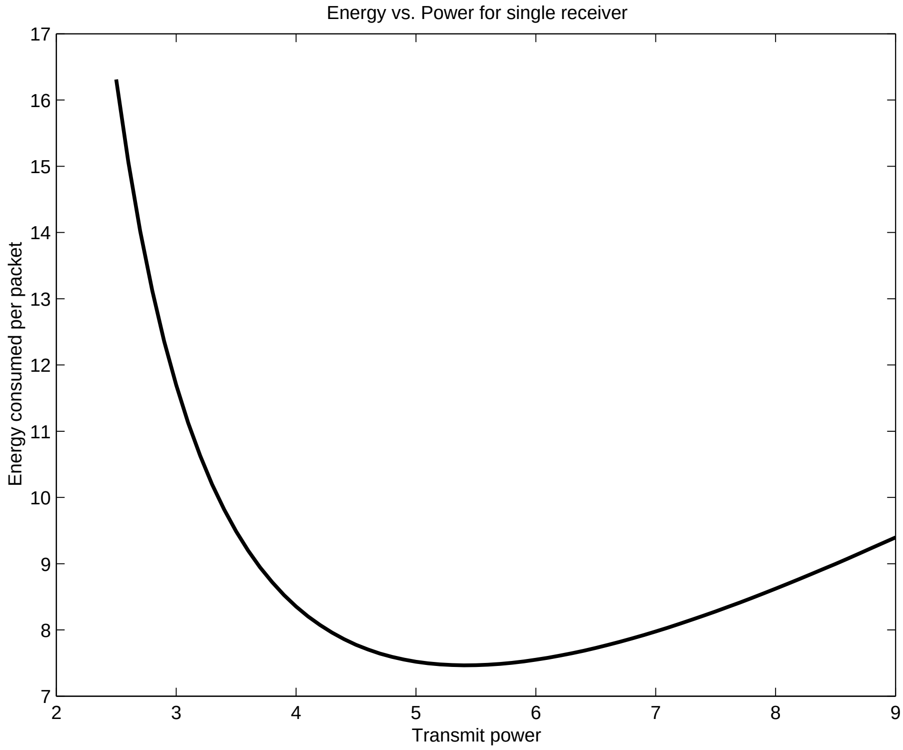

Figure 3.2: Energy vs. Power for single receiver

To confirm this, we also plot the energy consumed versus transmit power, as shown in figure 3.2. As we can see the optimum power matches that from figure 3.1. It would be reasonable for us to expect a similar behavior even when each packet has to reach multiple recipients.

3.8.2 Satellite network

For the satellite case, the energy consumed per packet is equal to the number of times it gets transmitted, times the energy expended per transmission. Therefore, from equation 3.19, we get the energy consumption at the satellite to be equal to

$$\begin{aligned} E_{avg,sat} &= \sum_{n=1}^{N} \binom{N}{n} \frac{(-1)^{n+1} P_{tr} \Delta}{1-(1-p_s)^n} \\ &= \sum_{n=1}^{N} \binom{N}{n} \frac{(-1)^{n+1} P_{tr} \Delta}{1-\left(1-\left[1-Q\left(h^{-\gamma_s}\sqrt{\frac{P_{tr}}{N_0}}\right)\right]^L\right)^n} \end{aligned} \tag{3.31}$$

3.8.3 Cellular network

Using a reasoning similar to the previous case, each base station consumes energy equal to

$$\begin{aligned} E_{single\ queue} &= \sum_{n=1}^{N/M} \binom{N/M}{n} \frac{(-1)^{n+1} P_{tr} 7\Delta}{1-(1-p_t)^n} \\ &= \sum_{n=1}^{N/M} \binom{N/M}{n} \frac{(-1)^{n+1} P_{tr} 7\Delta}{1-\left(1-\left[1-Q\left(r^{-\gamma_t}\sqrt{\frac{P_{tr}}{N_0}}\right)\right]^L\right)^n} \end{aligned} \tag{3.32}$$

The total energy consumed in all M queues, is simply equal to this scaled by M. Therefore,

$$E_{avg,cell} = \sum_{n=1}^{N/M} \binom{N/M}{n} \frac{(-1)^{n+1} 7 M P_{tr} \Delta}{1-\left(1-\left[1-Q\left(r^{-\gamma_t}\sqrt{\frac{P_{tr}}{N_0}}\right)\right]^L\right)^n} \tag{3.33}$$

3.9 Delay comparison

The energy computation is fairly straightforward for both the satellite and the cellular cases. However, the average end-to-end delay is not so straightforward. We will see in the subsequent section that while this can be computed without much trouble for the satellite case, the cellular case is almost impossible to handle.

3.9.1 Satellite case

Since the queue is basically a $Geo/G/1$ queue, we can use the *Pollaczek-Khintchine* formula to compute the average waiting time, following which we can compute the average end-to-end delay. The *Mean Pollaczek-Khintchine* formula is given by

$$W_{avg} = \frac{\lambda E[Y^2]}{2(1 - \lambda E[Y])} \tag{3.34}$$

where Y is the random variable indicating the service time.

The average waiting time can be computed by substituting equations 3.19 and 3.20 in the above equation, and the end-to-end delay is given by the $W_{avg} + E[Y]$. This is a straightforward process, but the end result is not reproduced here, since it is a long and complicated expression that does not give much insight.

3.9.2 Cellular case

It is quite easy to compute the average delay experienced by a packet *in a single queue*, using the same idea as in the satellite case. However, this is not the same as the average end-to-end delay of the packet, since we define it as the time between arrival and the time that the last queue finishes processing the packet. Therefore, the end-to-end delay of the packet is given by the maximum of the delays of the M queues. If the delay in queue i is denoted by D_i, the end-to-end delay is given by $\max_{i=1,\ldots,M} D_i$. Therefore, the average end-to-end delay is given by

$$D_{avg} = E\left[\max_{i=1,\ldots,M} D_i\right] \tag{3.35}$$

However, we still need to determine the distribution of D_i for queue i. Since the queues are statistically identical, it is sufficient to find the distribution of end-to-end delay for a single queue. An expression for the Probability Generating Function of D_i is found in [28], namely,

$$D(z) = \frac{(1-\rho)(1-z)B(z)}{1-\lambda-z+\lambda B(z)} \tag{3.36}$$

where $B(z)$ is the Probability Generating Function of the service time distribution which can be found from equation 3.13 and $\rho = \lambda E[Y]$, where $E[Y]$ is given by equation 3.22.

In principle, we can find the expression for the average end-to-end delay in the cellular case. It involves computing the PGF of D_i using the above equation, then computing the distribution from the PGF, after which the distribution of the maximum should be computed, following which the average can be found. However, in practice, it is very difficult to obtain any closed-form expressions, or even perform the computation numerically. Therefore, we need to resort to alternate means of comparing the two architectures.

3.10 Throughput comparison

Since all queues in consideration are $Geo/G/1$ queues, the criterion for stability in those cases are given by $\rho < 1$, where $\rho = \lambda E[Y]$. We have already found expressions for the average service time, given by equations 3.19 and 3.22. The maximum stable input rate (and hence the maximum stable throughput) is given by the reciprocal of the average service time in both cases.

3.11 Alternate scheme for comparison

Since the earlier comparison did not give us much insight into the problem, we look at an alternate means of comparing the two network architectures, based on sample-path arguments. Before we explain this idea further, it would help to look at the values of some parameters that we will be using, in order to justify the method of comparison.

3.11.1 Parameter values used

As we discussed earlier, we shall be comparing LEO satellite architecture with cellular network. LEO satellites orbit around a height of 1000 km. Also, we assume the radius of the cell to be 10 km, which is a very reasonable value. The path-loss exponent has been determined experimentally, and the values for different environments are given in [24]. For free space propagation, as in the case of satellites, $\gamma_s = 1$ (received power $\alpha\ d^{-2}$), while for the cellular case $\gamma_t = 1.5$ (received power $\alpha\ d^{-3}$), owing to the presence of obstacles and reflections on the path. Thermal noise at the receiver (N_0) has an average value of 10^{-12}. We assume the data rate (bandwidth) to be a 1 Mbps link, and the packet length to be 1024 bits (i.e., 1 Kbit). All these values are standard numbers found in literature. However, they have an interesting relationship that we discuss below.

The packet transmission time is now equal to 1 Kb/1 Mbps = 1 msec. We have seen earlier that the round-trip time is approximately 6 msec. This means that the round-trip time is equal to 6 slots, which makes each transmission 7 slots long. In the cellular case, though the round-trip time is assumed to be zero, each transmission in itself takes 7 slots, since the available bandwidth at each base station is equal to 1/7 of the total bandwidth. Therefore, the time between the start of successive transmissions is the same in both cases.

Also, from the channel model that we assume, for a given transmission power at the satellite, the received power is equal to Ph^{-2}, which is equal to $P.(10^6)^{-2} = P.10^{-12}$. For the cellular network, the received power at the nodes is equal to $Pr^{-3} = P.(10^4)^{-3} = P.10^{-12}$. Therefore, we see that the same transmit power gives us the same received power in both cases. This enables us to compare the two architectures on the basis of the number of transmissions of a packet alone. We now provide an argument by which we show that the total number of transmissions that a packet undergoes is greater in a cellular network than the satellite network, but the end-to-end delay, measured in number of transmission intervals, is greater for the satellite network. Later, we extend this argument to different data rates and packet lengths.

3.11.2 Example to motivate argument

To motivate the idea behind the sample-path argument, let us look at the following argument. Say, we have two cells, both empty to start with. Consider the following sequence of events.

1. At time $t = 0$, one packet arrives, which requires 3 transmissions to complete in cell 1, and 9 transmissions in cell 2.

2. The second packet arrives at the end of the first transmission, and requires 10 and 2 transmissions to reach cells 1 and 2 respectively. Therefore, now, the first queue has 2 more transmissions for the first packet to complete and 10 for the second to complete. The second queue has 8 transmissions for the first packet and 2 for the second packet. Let us consider the service times of both the packets.

If we were using the single-queue mechanism (i.e. satellite architecture), the first packet would take $\max\{9, 3\}$ transmissions to complete, and the second packet would take $\max\{10, 2\}$ transmissions. Therefore, the delay for the first packet would be 9, and for the second packet would be $9 + 10 - 1 = 18$ transmission intervals (the -1 comes because the second packet arrived at the end of the first transmission).

On the other hand, if we were using the multi-queue mechanism, the first packet would finish serving in queue 1 in 3 intervals and in queue 2 in 9 intervals, making the delay for this packet equal to 9, which is the same as in the earlier case. The second packet, however, finishes in queue 1 after 12 intervals, and in queue 2 after 10 intervals, giving it an end-to-end delay of 12 transmission intervals, which is lesser than in the first case.

The enhancement in delay comes from the fact that in a single-queue mechanism, each packet sees the worst-case transmissions for every packet that is ahead of it in the queue, before it starts its own transmission. However, in the other case, the worst-case transmissions get spread out over different queues, which shortens the end-to-end time for each queue. Using this idea, we present a formalized argument, and compare the energy and end-to-end delays of the two architectures.

3.11.3 Sample-path comparison

In this section, we show that the number of transmissions a packet undergoes (and hence the energy consumption) is greater in the cellular architecture than the satellite architecture, but the end-to-end delay (measured in number of transmissions) is lower for the cellular case. Since our parameters are such that the transmit power and the length of a transmission interval is the same under both architectures, this suffices to compare them against each other. We also extrapolate the results obtained to different sets of parameters, where the satellite and the cellular architectures take different number of slots per transmission.

Number of transmissions experienced by a packet

We know that each packet has to reach N users, and that the number of transmissions required to reach user i is denoted as X_i, which is Geometrically distributed with probability p, as discussed earlier.

- For the satellite (single-queue) case, the number of transmissions is equal to

$$N_{trans,sat} = \max_{i=1,\ldots,N} X_i. \tag{3.37}$$

- Each queue in the cellular case serves N/M nodes, i.e., queue i serves nodes $\{(i-1)(N/M)+1, ..., iN/M\}$. Therefore, the number of transmissions experienced by a given packet in queue i is given by

$$r_i = \max_{j=(i-1)(N/M)+1,\ldots,iN/M} X_i. \tag{3.38}$$

Therefore, the total number of transmissions that this packet undergoes in the cellular case is equal to $N_{trans,cell} = \sum_{i=1}^{M} r_i$. In the satellite case, we can see that $N_{trans,sat} = \max_{i=1,\ldots,M} r_i$.

We know that for any set of positive numbers $\{a_i, i = 1, ..., M\}$, the maximum of the M numbers is always less than their sum. This is true for every realization of the X_i. Therefore, we can conclude that the number of transmissions experienced by every packet in the single-queue case is less than that experienced in the multiple-queue case.

Energy consumption

We know that for the satellite case, power is consumed only during one slot; the remaining slots are basically the round-trip time. So, the energy consumed per transmission is one-seventh in the case of the satellite than in the cellular architecture. We have also just seen that satellite architecture leads to fewer transmissions per packet than the cellular case. Therefore, we can conclude that satellite architecture is more energy-efficient than the cellular case.

Delay performance

As we have done in the previous case, we show that the end-to-end delay (in terms of the number of transmissions that a packet sees, from the time it enters the queue(s) to the time it finishes service) is lesser for the cellular case than the satellite case.

Consider a packet that has just arrived and has been put into the queue. Upon arrival, say it sees $K-1$ packets ahead of it. In order to make notation more convenient, we shall label this packet as packet K. We study the cellular case first, as it involves some notation that can be carried over easily to the satellite case.

The state of the system (i.e., the number of transmissions of each packet ahead of it in the queue) that packet K sees can be depicted in the following table. We assume, without loss of generality, that the same packet is being transmitted in all queues when packet K arrives. It is easy to see that if this were not so, the inequality is further strengthened. Therefore, we see that end-to-end delay for a packet in the satellite case can at best be equal to the cellular case.

Packet	Queue 1	Queue 2	...	Queue M
1	$r_{1,1}$	$r_{1,2}$	$\cdots$	$r_{1,M}$
2	$r_{2,1}$	$r_{2,2}$	$\cdots$	$r_{2,M}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
K-1	$r_{K-1,1}$	$r_{K-1,2}$	$\cdots$	$r_{K-1,M}$
K	$r_{K,1}$	$r_{K,2}$	$\cdots$	$r_{K,M}$

Table 3.1: Number of outstanding transmissions seen by packet K upon arrival

Here, $r_{i,j}$ denotes the number of outstanding transmissions of packet i in queue j. Therefore, the total number of transmission that packet K sees in queue i before it completes is given by $\sum_{i=1}^{K} r_{i,j}$. The end-to-end delay is the maximum of the delays in the individual queues, so the net end-to-end delay for this case is equal to

$$D_{cell} = \max_{j=1,\ldots,M} \left[\sum_{i=1}^{K} r_{i,j} \right] \tag{3.39}$$

In the satellite case, with the same realizations of the random variables $r_{i,j}$, we know that the number of transmissions experienced by packet i is the equal to $\max_{j=1,\ldots,M} r_{i,j}$, and therefore the end-to-end delay experienced by packet K under the satellite architecture would be equal to

$$D_{sat} = \sum_{i=1}^{M} \left[\max_{j=1,\ldots,M} r_{i,j} \right] \tag{3.40}$$

Clearly, $D_{cell} \leq D_{sat}$ for any realization of the $r_{i,j}$ (in effect, for any realization of the $X_{i,j}$). Here the D_{cell} and D_{sat} denote the number of transmissions seen in either case, but since in both cases each transmission requires 7 slots, the comparison still remains valid.

Another thing that emerges from the above analysis is that a packet can experience the same delay under both architectures, only under one of the following conditions.

1. It is the first packet in all the queues, i.e., the system is empty when the packet arrives.

2. All packets ahead of it in the queue must have their worst receiver lying in the same base station. Because of the independent nature of the channel, this is very unlikely, especially if the number of packets ahead of it is large.

Implicit in this argument is the assumption that packet K arrives just at the beginning of a transmission. This need not be the case, especially with multi-slot transmissions where arrivals can occur at the beginning of each slot. However, we can easily see that this does not affect the argument. For a packet that arrives during an ongoing transmission, the end-to-end delay in the two cases is equal to $D_{sat} + f_1$ (satellite) and $D_{cell} + f_2$ (cellular), where f_1 and f_2 represent the residual transmission interval when packet K arrives (for the two architectures). Because f_1 and f_2 are fractions, the inequality above for any sample path changes only if $D_{cell} = D_{sat}$. However, if $D_{cell} = D_{sat}$ because the system is empty, then $f_1 = f_2 = 0$, since there is no ongoing transmission when the packet arrives. Therefore, the only sample paths for which the inequality may change are those for which the worst-case transmission for all packets in the queue fall in the same cell, which as we have said before is very unlikely. Consequently, on the average, we conclude that the end-to-end delay in the cellular architecture is lesser than that of the satellite architecture.

We also see that the disparity in the number of transmissions between the two architectures (i.e., equations 3.40 and 3.39) increases with the following conditions :

1. Arrival rate - As the arrival rate increases, the average number of packets already in queue, as seen by an incoming packet gets larger. This increases the differences between the delay in the two architectures.

2. Variance in $r_{i,j}$ - As the variance of $r_{i,j}$ increases, the worst-case number of transmissions becomes much greater than the average case. Since the satellite sees the worst-case number for every packet, this leads to the delay increasing.

3. Number of queues - As the number of queues gets larger, the worst-case gets

spread out further, leading to lower delays in the cellular case as compared to the satellite case.

Extension to other parameter sets

While the above argument is valid for the transmissions seen by an incoming packet, making the actual delay and energy comparisons depend critically on the channel capacity and the packet length. We shall try to extend the above argument to different packet lengths and data rates.

Let us see, for example, what happens when we reduce the packet length by a factor of two (i.e., change it to 512 bits). Because of this, the slot length (defined as a packet transmission time) becomes half of what it was earlier (i.e. 0.5 msec). Since the round-trip time is still 6 msec, the satellite now takes 13 slots between the start of successive transmissions (1 transmit time + 12 round-trip time). The cellular case, on the other hand, still takes only 7 slots, since it still has the same bandwidth. This increase in the inter-transmission interval of the satellite compared to the cellular case only worsens the already poor delay. As far as energy goes, though an inter-transmission interval is longer for the satellite, the satellite is still expending power for only one slot, as compared to 7 in the cellular case. Therefore, the energy comparison also remains the same. Thus we see that reducing the packet length (or in effect, increasing the data rate) does not affect the results obtained.

Doubling the packet length doubles the slot length to 2 msec. Because of this, the satellite now takes 4 slots per transmission (1 transmission + 3 rtt), as compared to the cellular which takes 7 slots of transmission. Again, the energy comparison remains valid (since the number of transmissions is always smaller for the satellite case than for the cellular case, and the satellite consumes power for only one slot compared to 7 in the other case). The delay comparison, however, may change depending on the arrival rate and the transmit power. While we cannot determine which is lesser, we can surely say that as we increase the arrival rate the cellular case will perform better than the satellite case above a certain arrival rate. This is because the disparity between the number of transmissions in the two cases grows with the arrival rate, and above a certain rate this disparity becomes large enough to compensate for the longer packet transmission interval in the cellular case.

3.12 Simulation results

In order to confirm the results obtained above, we run simulations for the various comparisons, which are shown in the following sections. The simulations were written using C++ on Linux, and run on a computer with a Pentium 4 (2.4 GHz) processor and 512 MB RAM. The memory management for the queues was done using the C++ STL container *deque*, and the *drand48()* random number generator was used throughout. The simulation was run for the length of time taken to complete transmission of 2000020 packets starting from an empty queue, and the first twenty packets were ignored while computing the averages.

3.12.1 Energy consumption

The simulation results for the energy consumption of the two architectures are shown in figure 3.3, and the following conclusions can be reached.

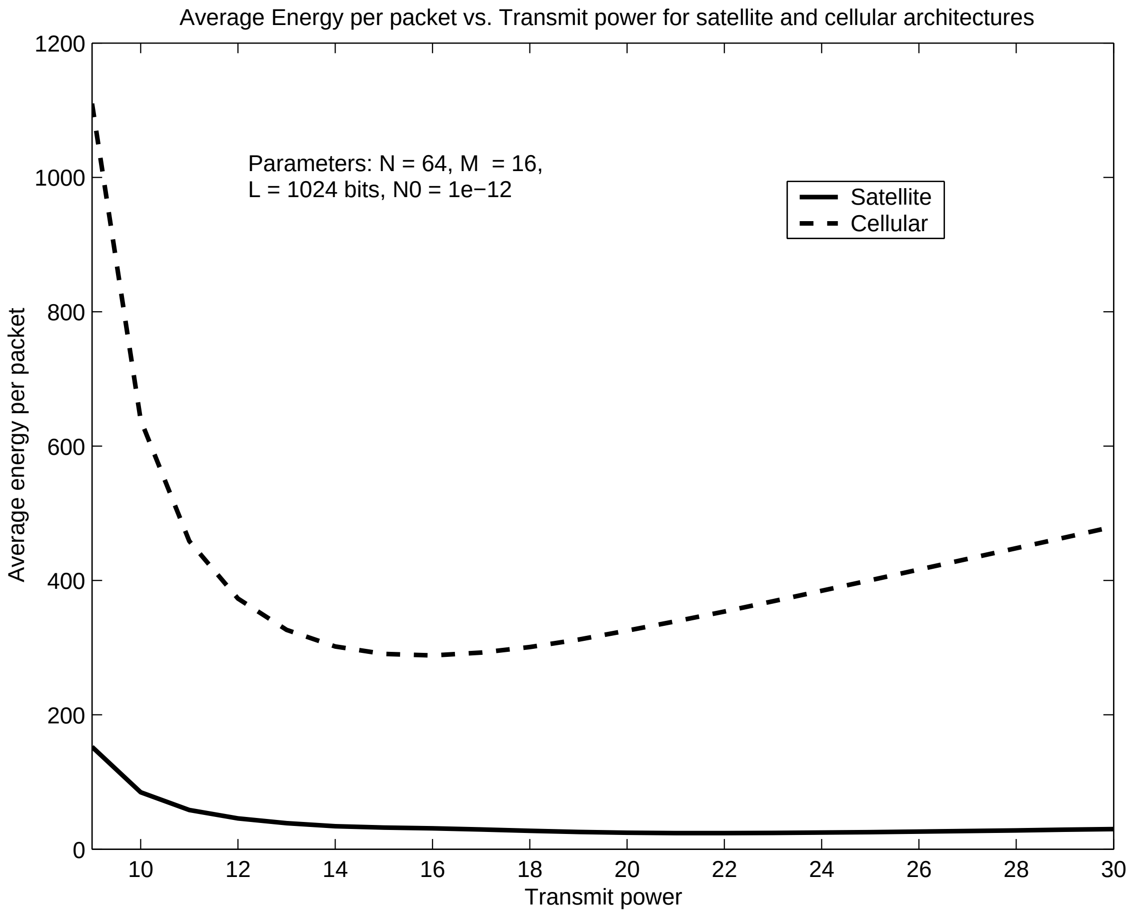

Figure 3.3: Energy vs. Power for Satellite and Cellular architectures

1. The energy consumed per packet is higher in the case of the cellular architecture than the satellite architecture, as we expected.

2. The energy consumed per packet has a minimum point, like in the single-receiver case. The minimum-energy point is not the same for the two cases, however.

3. As we can see from the graph, the average energy consumed in the two cases differ by orders of magnitude. Therefore, we can see that reducing the base station radius (say, by a factor of two) does not affect this comparison, since the difference between the two cases is large enough to overwhelm the gain achievable by shrinking the cell size by a small amount. Consequently, we expect the actual energy consumption in the satellite and cellular architectures (taking node location into consideration) to behave similarly.

3.12.2 Maximum stable throughput

The maximum stable throughput, as discussed earlier, is equal to the reciprocal of the average service time of the packet. For the cellular case, we require all queues to be stable. However, since the queues are statistically indistinguishable, the stability criterion is the same for all of them. Since now each queue serves only part of the total number of customers, we expect that the cellular system has a higher stability threshold. This is verified by the graph in figure 3.4.

We also see from the figure that at high transmit powers, the two systems have the same stable throughput. This is because at such high powers, the probability of correct reception is almost equal to 1. Therefore, each packet completes in approximately one transmission, irrespective of whether the satellite or cellular architectures is being used. Thus, the maximum throughput for the two cases is the same at high powers.

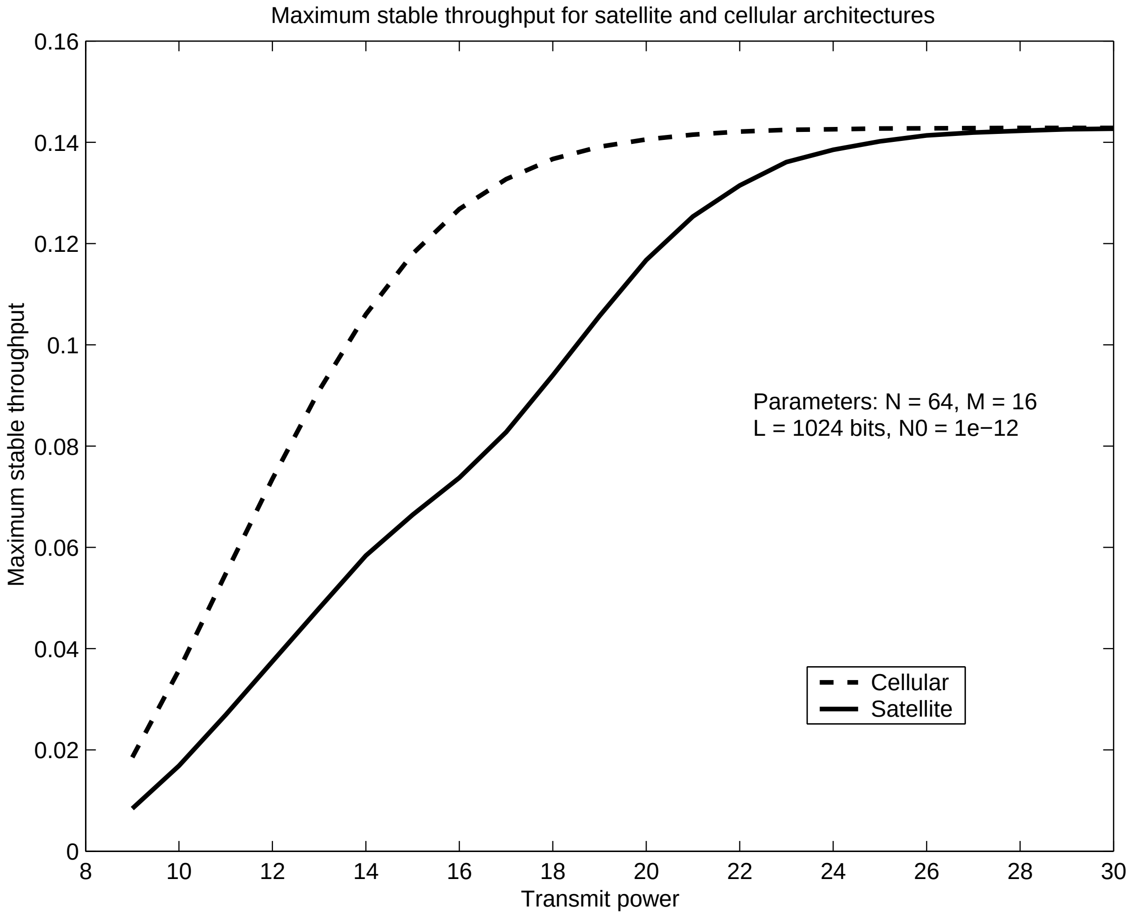

Figure 3.4: Maximum stable throughput vs. Power for both architectures

3.12.3 Delay performance

Again, from our arguments in the previous sections, we expect that the delay performance of the satellite case is worse than that of the cellular case. We verify that this is actually the case, from figures 3.5 and 3.6.

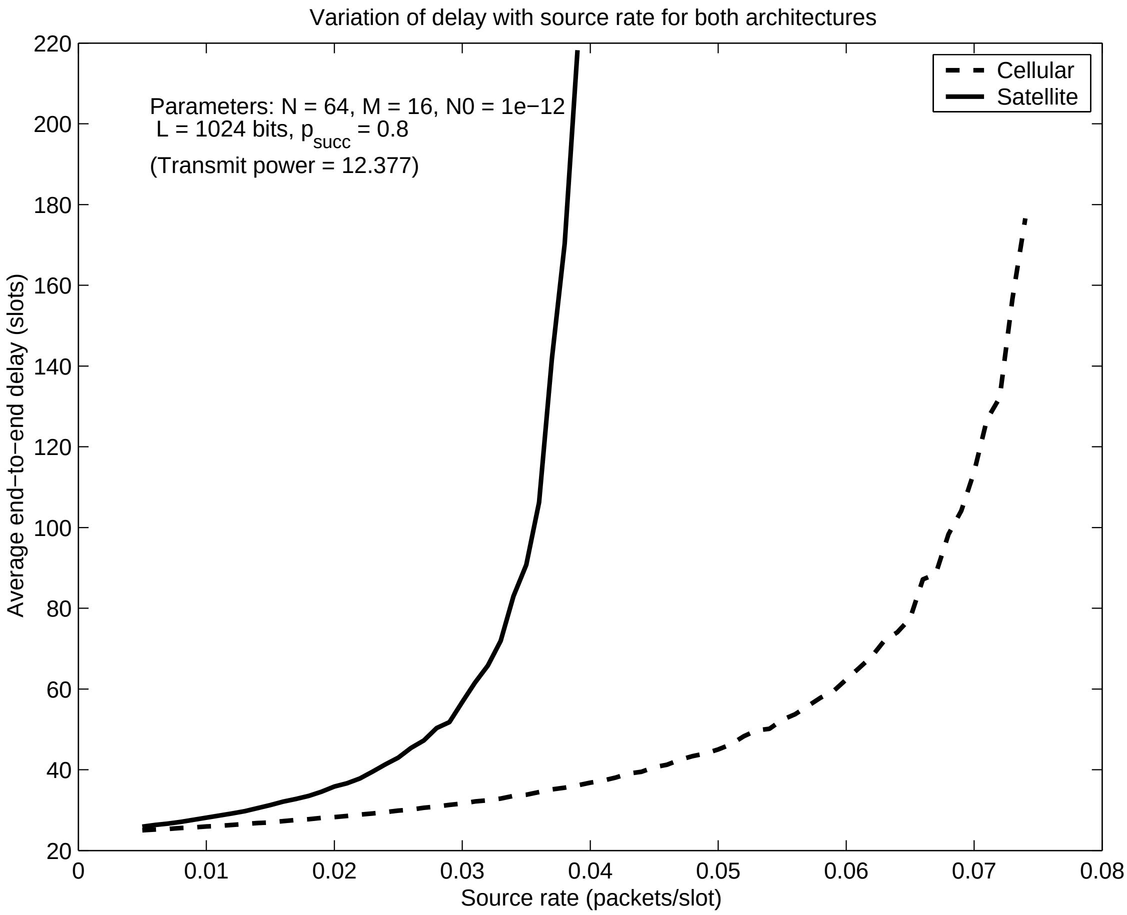

Figure 3.5: Delay vs. Source rate, both architectures, N = 64

The following conclusions can be reached from these figures.

1. As we can see from the figures, the cellular case always has a better delay performance than the satellite case, for all source rates and transmit powers.

2. We notice that as the transmit power increases, the delays become close to

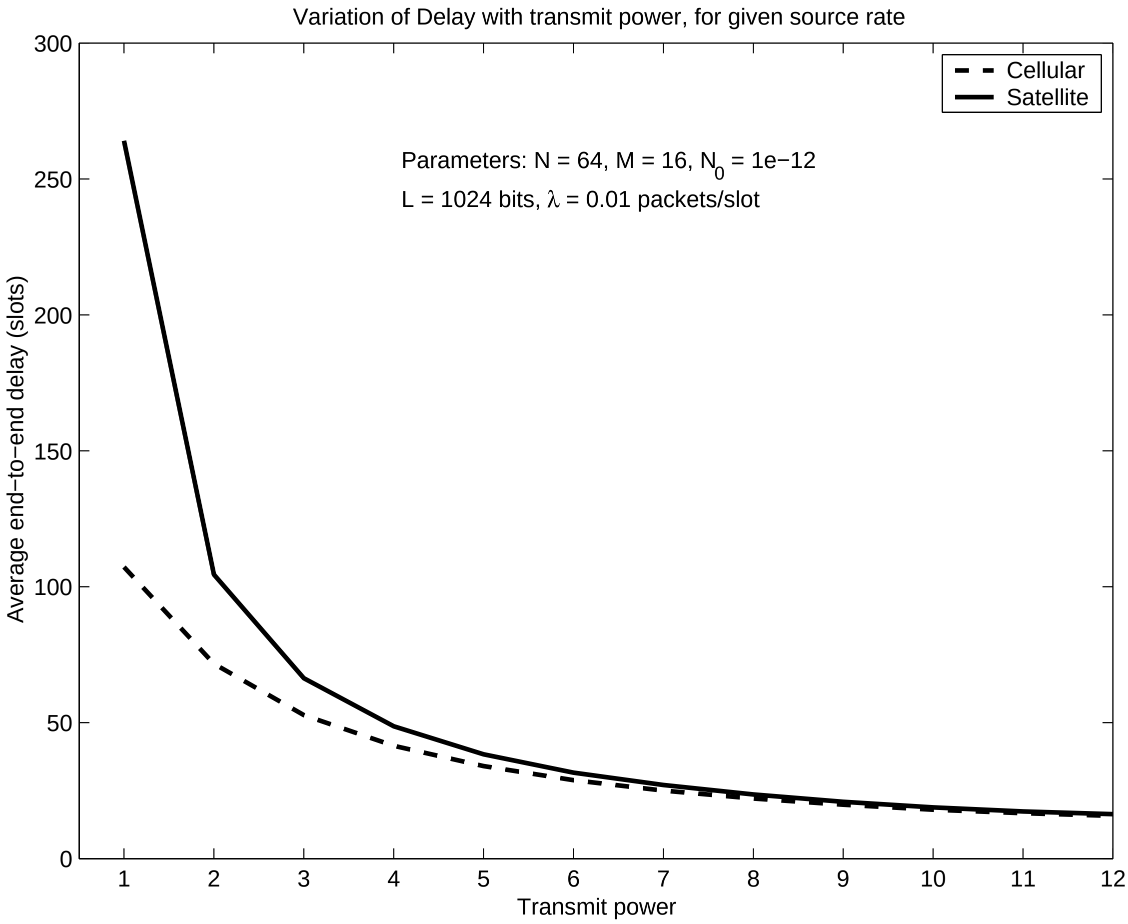

Figure 3.6: Delay vs. Transmit power for given source rate

each other, tending to a single transmission per packet.

3. For low transmission powers, the number of transmissions required per packet is high, hence the service times of the packets in the queues is high. The systems are therefore close to capacity (for the chosen source rate) when the transmit power is low. Under this case, we can see greater difference between the delays of the two architectures, as predicted by the sample-path comparison.

4. The effect of higher maximum throughput in the cellular case is also visible in figure 3.5. The satellite quickly reaches capacity (at around 0.034 packets per slot), whereas the delay for the cellular case at that load is quite low. The cellular system reaches capacity much later.

3.12.4 Variation with N

As the number of people using wireless services increases rapidly, the average number of wireless devices per unit area also increases. This prompts us to explore how the two architectures in consideration behave as we increase N, keeping all other parameters the same.

In order to gain some insight into how the architectures perform, let us first take a look at the average number of transmissions a packet in a single-queue system undergoes, when it has to reach N users. This is shown in figure 3.7, and can be used to come up with some interesting conclusions.

Because of the concave nature of this curve, we make the observation that

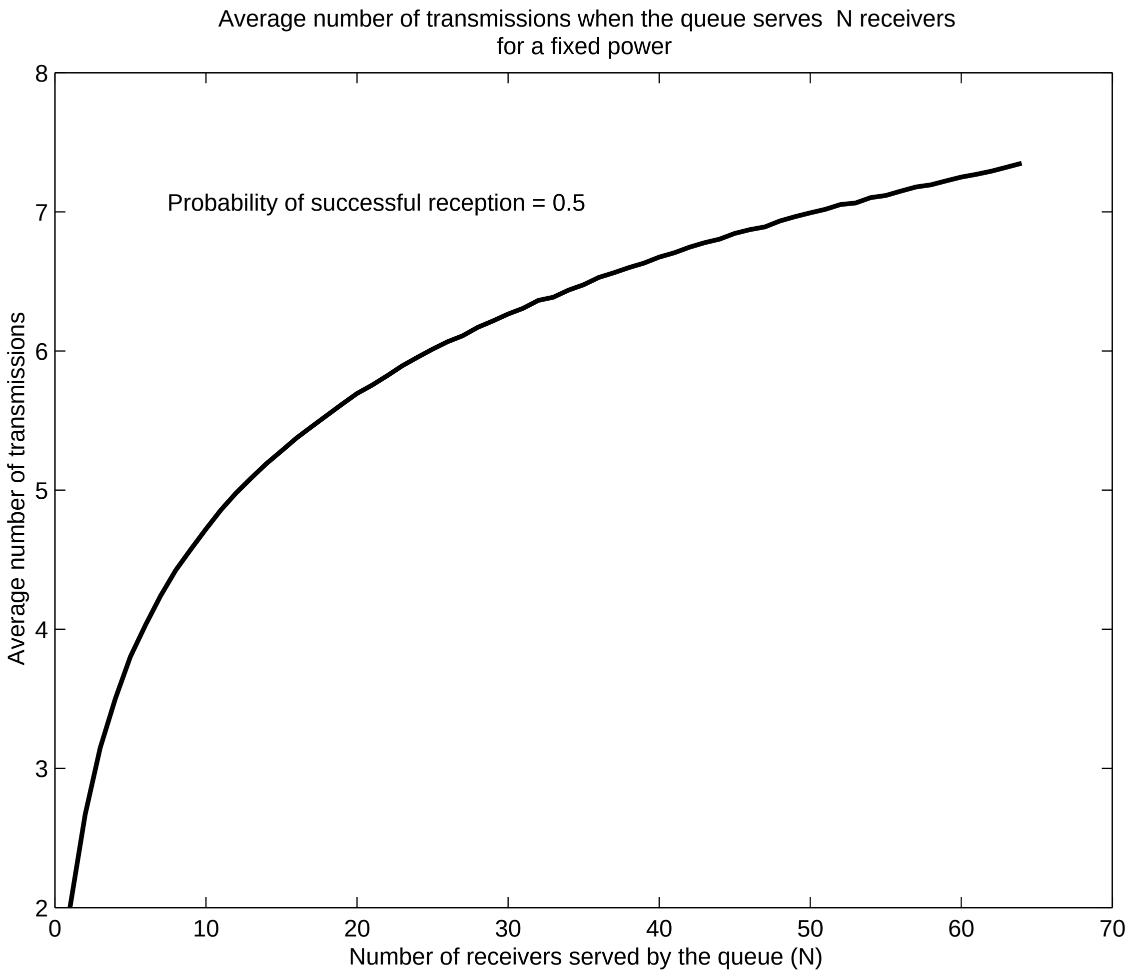

Figure 3.7: Variation with N of avg no of transmissions of a packet in a single-queue system

the average number of transmissions in a single-queue architecture (when there are N receiver nodes) is lower than the total number of transmissions when operating under the multiple-queue architecture (when each queue serves N/M receivers). In other words, $N_{tr}(N) \leq MN_{tr}(N/M)$, where N_{tr} is the average number of transmissions shown in figure 3.7. Also, we expect this difference to increase as N increases, keeping the number of base stations fixed. Therefore, we expect the energy consumption to also follow the same trend, i.e., as the number of nodes increases, the energy difference between the satellite and the cellular architectures increases. This can be verified through simulation, and shown in figure 3.8.

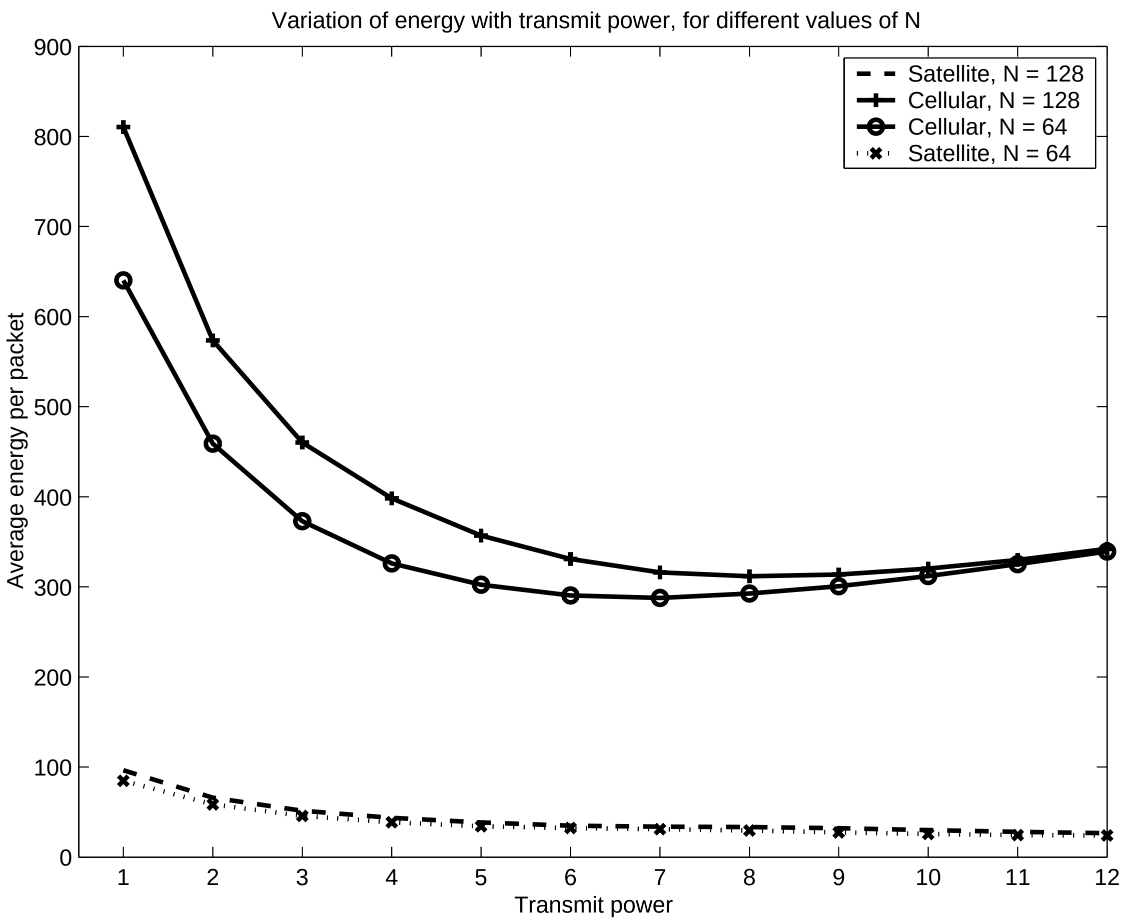

Figure 3.8: Energy vs. Transmit power for different N values

As we can see from the figure, the energy consumed per packet for the satellite architecture increases far lesser as compared to the cellular architecture when N is increased from 64 to 128. Owing to the scale of the numbers involved, the increase in energy in the satellite case is not very visible in figure 3.8. Therefore, we show the energy increase in the satellite architecture in figure 3.9.

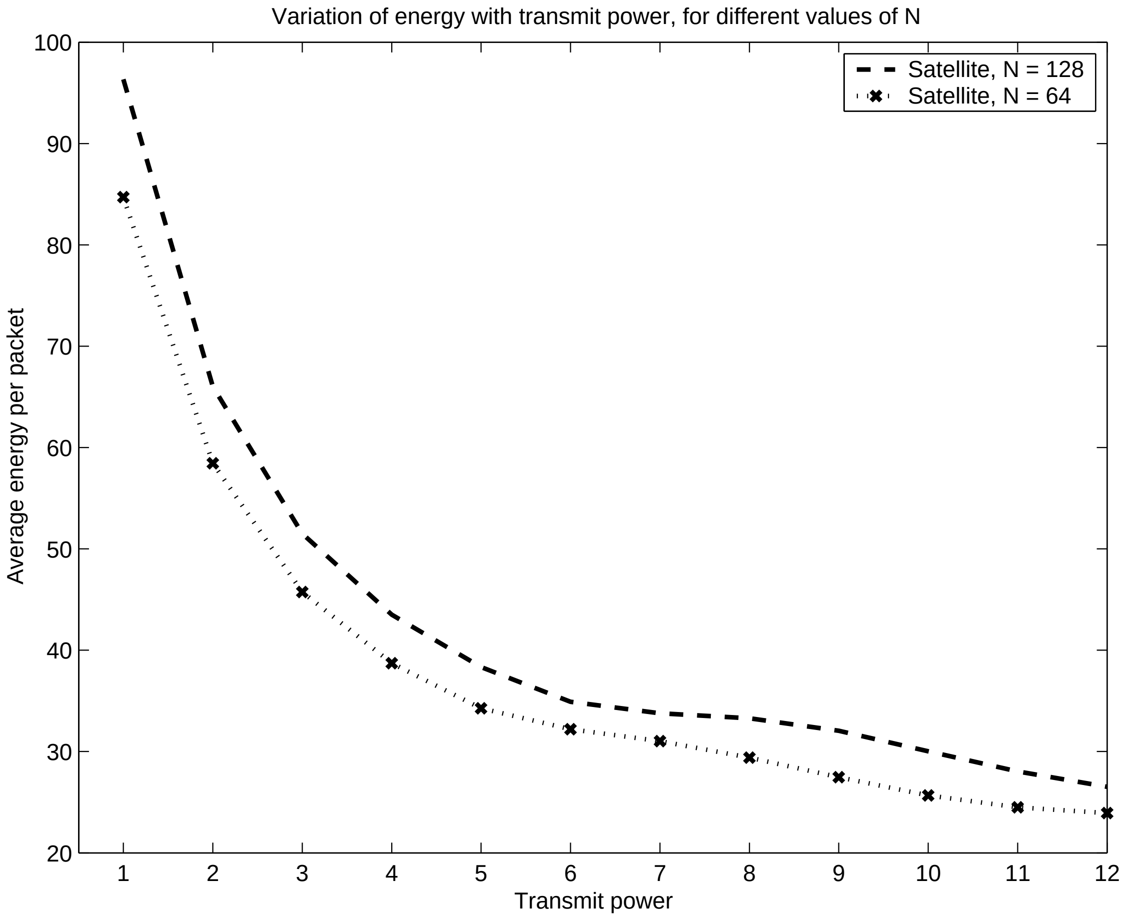

Figure 3.9: Energy increase in satellite architecture when N increases

The variation of delay with N is shown in figure 3.10. As we can see, the delay increases for both architectures as N increases. However, the increase is dependent on the source rate. For lower powers, the satellite architecture with $N = 128$ is much closer to the stability threshold than the satellite case with $N = 64$. Therefore, the delay is much higher in this case than in the other cases. The cellular case is quite far from being unstable, so the delays are not as affected by the increase in N.

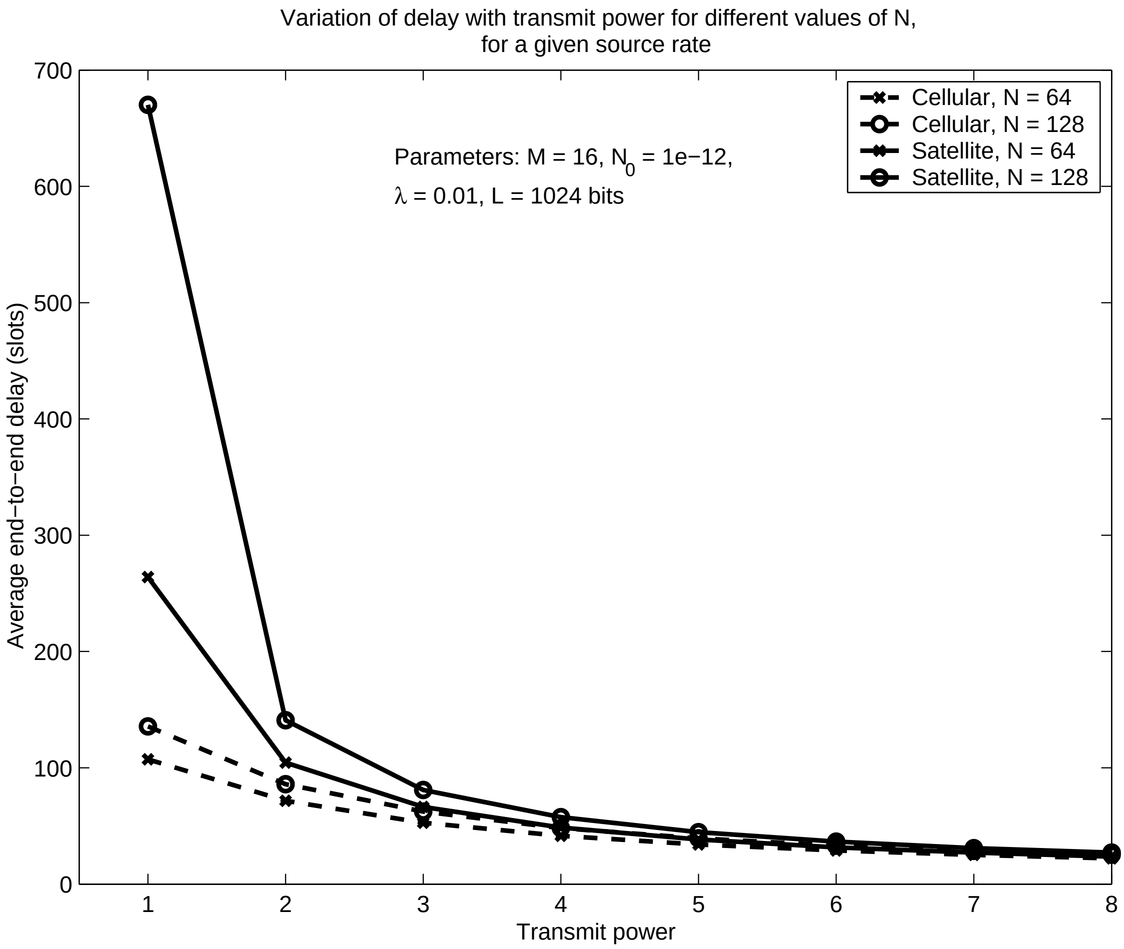

Figure 3.10: Delay vs. Power for different N values

3.13 Conclusions

In this chapter, we modeled the satellite and cellular systems using $Geo/G/1$ queues, and developed a formulation for comparing them on the basis of energy, throughput and delay. Using first queuing theory formulation, and then sample-path arguments, we see that the satellite is more energy-efficient than the cellular system. However, the cellular architecture provides a lower delay and a higher maximum stable throughput (for the same transmit power) than the satellite does. Thus, a trade-off between energy and delay is demonstrated in this context. Also, the systems were studied when the density of nodes (number of nodes per unit area) increases, and trends were established in this case as well.

Chapter 4

Hybrid Satellite-Cellular Networks

4.1 Introduction

In the previous chapter, we have seen the trade-offs associated with satellite and cellular networks, and how one performs better in terms of energy while the other performs better in terms of delay. It seems reasonable to then extend this idea into *hybrid* networks, which consist of both satellite and cellular components. This is an idea which has been proposed recently in the literature. Since most of the work has been done for voice applications or point-to-point links, we try to extend our results to hybrid networks and establish trade-offs in those as well. The key result that we obtain through simulation is that using hybrid networks, we can get higher throughputs than by using only one of the components. However, this again comes at the expense of energy consumption, which we shall demonstrate later.

4.2 Extension of Satellite and Cellular networks

We extend the existing models of satellite and cellular networks to a hybrid network model as follows. We assume that the nodes are now capable of receiving transmissions from both the satellite and the base station of the cell that they are a part of. In order to keep the comparison fair, we keep the total bandwidth of the system (W) constant. However, an additional parameter that arises in the context of hybrid networks is the portion of bandwidth allocated to the two components. Let us call this parameter α, where $0 \leq \alpha \leq 1$. $\alpha = 1$ indicates that all the bandwidth is alloted to the satellite, and $\alpha = 0$ indicates that all bandwidth is allocated to the terrestrial section. We still have the idea of frequency re-use in the base stations, which implies that each base station gets a bandwidth equal to $(1-\alpha)W/7$. The system is modeled as a set of $M+1$ parallel discrete-time $Geo/G/1$ queues (M base stations and 1 satellite queue). We assume the input processes to the satellite and cellular components to be independent Bernoulli processes with rate λ_s and λ_t respectively.

4.2.1 Discrete-time model

With the bandwidths of the two sections of the hybrid network being different, defining a slot time as the time taken for a packet transmission no longer works, since packet transmissions in the satellite and cellular sections can take different times that are not integral multiples of each other. Therefore, in order to retain the models that we have, we choose a slot length such that the packet transmission time in both sections are integral multiples of the slot time. This can be easily done for any α rational, as we see below.

Consider the total bandwidth W and the packet length L. For a given bandwidth split α, the packet takes time equal to $L/(\alpha W)$ in the satellite, and $7L/(1-\alpha)W$ in the cellular section. If α is rational, it can be cast in the form of $a/(a+b)$, where a and b are integers. Therefore, the packet transmission times can now be expressed in the following form.

$$P_{sat} = \frac{L(a+b)}{aW} \tag{4.1}$$

$$P_{cell} = \frac{7L(a+b)}{bW} \tag{4.2}$$

Therefore, if we now choose the slot length to be equal to $\Delta = L(a+b)/abW$, the packet transmission times are now integral multiples of Δ (i.e. $b\Delta$ at the satellite and $7a\Delta$ at the base stations).

4.2.2 Arrival rates

Since the slot lengths differ with α, we can no longer define the input rate in terms of packets per slot. Instead, we specify all input rates in packets per second, which then means that the input rate in terms of packets per slot varies with the fraction of bandwidth allocated to the two components.

4.2.3 Parameters

The hybrid network now has double the parameters as that of the single-component networks. We now have N nodes, $M+1$ queues, two arrival rates (terrestrial and satellite), two transmit powers, and a parameter indicating the fraction of bandwidth allocated to each component. The quantities of interest, i.e., energy, delay and throughput depend on all these parameters, which makes any meaningful comparison very difficult. Therefore, we study a simpler problem in the context of hybrid networks.

4.3 Problem Statement

For a given set of transmit powers, we would like to determine if the hybrid network architecture gives a higher maximum stable throughput than using only one of the components and allocating the bandwidth completely to it. We would also like to study the average energy consumed per packet, and show that in the case of the hybrid network, the average energy consumed per packet is greater than using only one of its components.

4.4 Energy consumption

From the model that we have been considering, the probability of successful reception of a packet does not depend on the bandwidth (and hence the transmit time). Therefore, the number of transmissions that a packet undergoes does not depend on the bandwidth split, but only on the transmit power chosen. However, the energy consumed per packet is equal to the $P_{tr}N_{trans}(P_{tr})\Delta$, which depends on the bandwidth split only through Δ. For the satellite (or cellular) architecture, denote the average energy consumed per packet (for a given transmit power) when the entire bandwidth is allocated to the satellite (or cellular) component by $E_{sat,f}(P)$ (or $E_{cell,f}(P)$) where the f indicates that bandwidth is fully allocated to it, and P is the transmit power. Therefore, the energy consumption for the two architectures when the bandwidth is split between them can be given by $E_{sat,f}(P)/\alpha$ and $E_{cell,f}(P)/(1-\alpha)$ for the same transmit power.

If the source generates packets at rates λ_s and λ_t packets per second, and the satellite and base stations use transmit powers P_s and P_t respectively, the average energy consumed per packet is given by the following expression.

$$E_{avg} = \left(\frac{\lambda_s}{\lambda_s + \lambda_t}\right) \frac{E_{sat,f}(P_s)}{\alpha} + \left(\frac{\lambda_t}{\lambda_s + \lambda_t}\right) \frac{E_{cell,f}(P_t)}{1-\alpha} \tag{4.3}$$

While we know that $E_{sat,f}(P) < E_{cell,f}(P)$ for given P, we cannot say the same about $E_{sat,f}(P_s)/\alpha < E_{cell,f}(P_t)/(1-\alpha)$, since they now depend on the value of α, P_s and P_t. However, it is clear that the energy expended is a linear combination of the two energy terms.

Say, for our choice of P_s, P_t, $E_{sat,f}(P_s)/\alpha < E_{cell,f}(P_t)/(1-\alpha)$. In this case, we achieve minimum energy by sending all packets to the satellite (i.e., $\lambda_t = 0$). Also, if we now allocate the entire bandwidth to the satellite (since the terrestrial component receives no packets), the energy required per packet reduces further to be equal to $E_{sat,f}(P_s)$. Similarly, if the inequality above were reversed, allocating the entire bandwidth and all packets to the cellular architecture would give us energy equal to $E_{cell,f}(P_t)$, which is still lower than what we would get if we used the hybrid architecture ($\lambda_s, \lambda_t > 0, 0 < \alpha < 1$). Therefore, we see that using the hybrid architecture leads to higher energy expenditure per packet than using only one of the components.

4.5 Maximum stable throughput

We have seen that using the hybrid architecture leads to higher energy consumption per packet. However, we would like to see if there are any benefits in doing so. We will see, through simulation, that for some choices of P_t and P_s, using the hybrid architecture can give us a higher maximum stable throughput than allocating the entire bandwidth to one of the components, but using the same transmit power.

From the model, we know that the average service time (and hence the maximum stable throughput) for the satellite (or terrestrial) queues depend on the transmit power P_s (or P_t) and the bandwidth split α. We determine whether a hybrid network can help increase throughput in the following fashion.

For any given pair of transmit powers $\{P_s, P_t\}$, we compute the average service time (and hence the stability threshold) for the satellite and terrestrial queues for different values of α between 0 and 1. Let us denote these by $\lambda_{max,t}(\alpha)$ and $\lambda_{max,s}(\alpha)$. The maximum stable throughput for the set (P_s, P_t, α) is then given by $\lambda_{max,t}(\alpha) + \lambda_{max,s}(\alpha)$. For any given pair $\{P_s, P_t\}$, we compute the maximum stable throughput for different values of α, and compute the α for which the maximum occurs. If this value of α is neither 0 or 1, this indicates that the hybrid architecture can provide a higher throughput than using only one of the components.

We show the result of the simulation in figure 4.1. The shading in the figure indicates the bandwidth split that gives the maximum throughput. White in this case refers to the entire bandwidth being allocated to the base stations, and black indicates that the entire bandwidth is allocated to the satellite. The various shades of grey indicate that hybrid architectures provide the maximum stable throughput. As we can see from the figure, for $P_s > P_t$ we find that the hybrid architecture can provide higher throughput. Also, as P_s increases, with respect to P_t, the bandwidth split that gives maximum throughput shifts more toward the satellite.

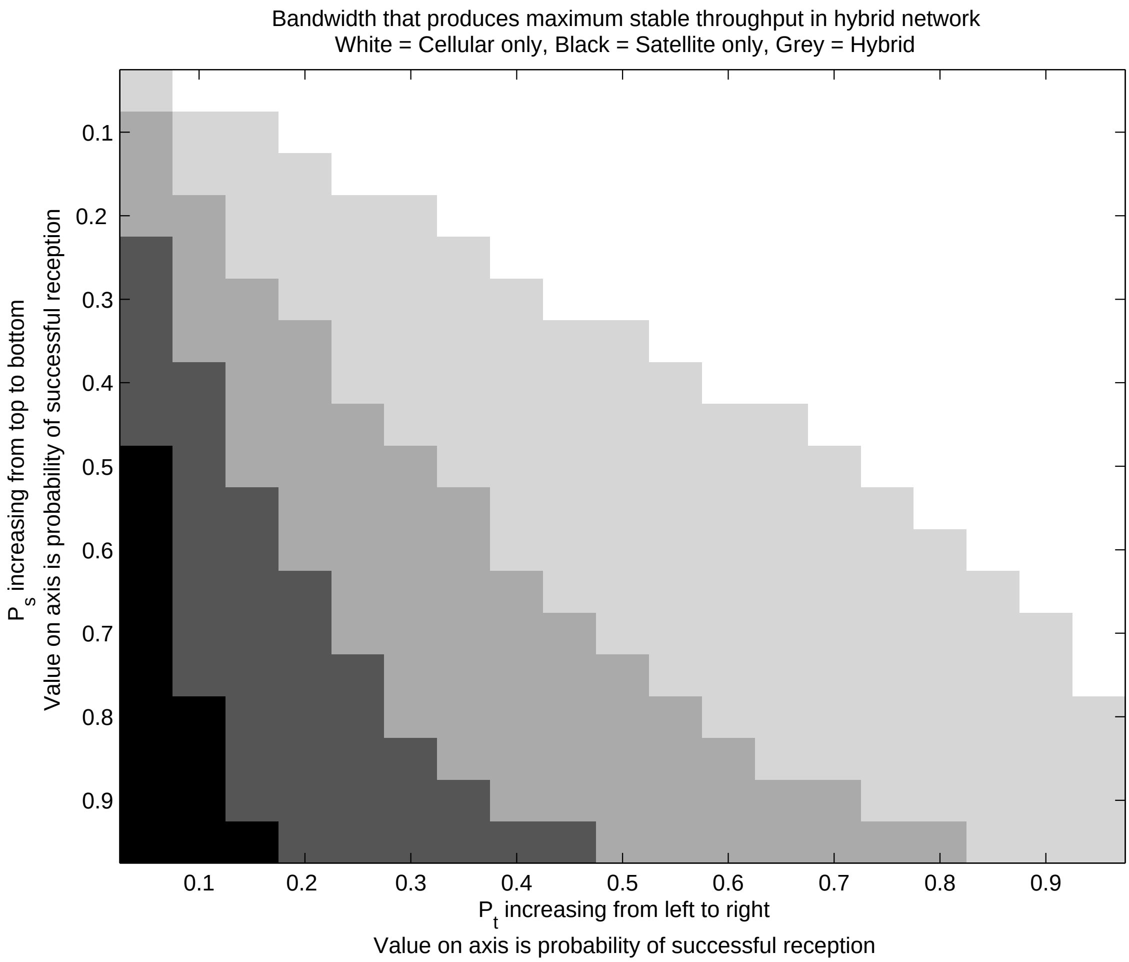

Figure 4.1: Hybrid networks provide higher throughput

Chapter 5

Conclusions and Future Work

In this thesis, we investigated the effect of power control and network architecture on the downlink performance of wireless networks. In the first part, we formulated the control problem of optimal-energy transmission, when a subset of the total number of nodes have to be reached. The problem was formulated as a Markov decision process, and the optimal policy was found to be of the threshold type. We studied the variation of the threshold with the choice of transmit powers. The performance of the optimal-energy policy was studied as the goal was varied, and an interesting relation between the shape of the energy-goal curve and the optimal policy was discovered. The results of this were extended to the multiple-power case, and the optimal policy was found to be of the separation type. This was then extended to approximate the continuous-power case, and the energy and delay performance with variable goal was studied. The structure of the optimal policy enables the power control mechanism to be simpler to implement, which could also lead to additional energy savings and lower cost.

In the second part, we compared the downlink performance of satellite and cellular networks in terms of energy and end-to-end delay. The two architectures were modeled as systems of queues, and existing queuing theory results were applied to this case. The lack of closed-form expressions for the delay led us to compare the architectures based on sample paths. We found that the satellite architecture is more energy efficient, however it experiences greater delays and lower throughput than the cellular architecture. This was further confirmed by simulations. The effect of increasing the node density was also explored.

The idea of hybrid network architecture proposed in the literature was also extended to the above model. It was found that hybrid networks could give higher throughputs than the individual components, at the expense of energy.

5.1 Future work

An important aspect that is missing in the power-control formulation is a formal proof of the optimal policy being of the threshold type. This is something that needs to be explored. Also, extensions of the channel model to more realistic cases, such as fading channels or correlated channels could be looked at. Modifications of the problem, where we also take the service time into the optimization criterion could also be useful.

A few enhancements are possible in the second part as well. For instance, we could modify the problem to one where there are finite buffers at the satellite and at the base stations, and compare blocking probabilities. Again, more realistic channel models could be used. Also, we have used Stop-and-wait ARQ, which is known to be inefficient for satellite channels. Therefore, other forms of ARQ could be studied, and performance comparisons could be found. Another important enhancement that can be considered is incorporating node locations with respect to the base stations in the cellular architecture and computing the actual energy, delay and throughput values as opposed to considering the bounds as we have done.

An important problem that can be looked at is the joint optimization of uplink and downlink in hybrid networks. When nodes have the option of transmitting to either the satellite or to their base stations, it would be instructive to consider the problem of how to choose between the two. This then leads to the joint consideration of uplink ARQ and MAC protocols with those on the downlink. Also, the enhancements that have been listed for the above section, such as finite buffers, improved ARQ schemes, improved channel models and incorporating node locations can all be extended to the hybrid network.

BIBLIOGRAPHY

[1] D. Ayyagari and A. Ephremides. Blocking analysis and simulation studies in satellite-augmented cellular networks. In *Seventh IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, volume 2, October 1996.

[2] Dimitri Bertsekas. *Dynamic Programming and Optimal Control: 2nd Edition.* Athena Scientific, 2000.

[3] Dimitri Bertsekas and Robert Gallager. *Data networks (2nd ed.).* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1992.

[4] M. Cagalj, J. Hubaux, and C. Enz. Minimum-energy broadcast in all-wireless networks: NP-completeness and distribution issues. In *MOBICOM 2002*, September 2002.

[5] K. Choi and S. Kim. Adaptive power/rate allocation for minimum mean transmission delay in CDMA networks. In *The 57th IEEE Semiannual Vehicular Technology Conference, 2003*, April 2003.

[6] A. T. Chronopoulos, P. Cotae, and S. Ponipireddy. Efficient power control for broadcast in wireless communication systems. In *WCNC*, 2004.

[7] Cyrus Derman. *Finite-State Markovian Decision Processes.* Academic Press, 1970.

[8] F. Filali, W. Dabbous, and F. Kamoun. Efficient planning of satellite-terrestrial hybrid networks for multicast applications. In *IEEE International Conference on Communications*, June 2001.

[9] D. Friedman. *Error Control for Multicasting in Satellite and Hybrid Communication Networks.* PhD thesis, University of Maryland, College Park, 2001.

[10] I.S. Gopal and J.M. Jaffe. Point-to-multipoint communication over broadcast links. *IEEE Transactions on Communications*, 32, September 1984.

[11] M. Grossglauser and D. Tse. Mobility increases the capacity of ad-hoc wireless networks. *IEEE Transactions on Networking*, 10(4), August 2002.

[12] J.B. Hamilton, C. Pike, R.C. Reinhart, and J.M. Savoie. A hybrid satellite/terrestrial network linking Europe and North America. *International Journal of Satellite Communications*, 17, 1999.

[13] Y.C. Hubbel. A comparison of the IRIDIUM and AMPS systems. *IEEE Network*, 11(2), March 1997.

[14] Transmitting Power Control in CDMA Cellular Systems. X. xia. In *AFRICON, 1999*, October 1999.

[15] J.E.Wieselthier, G.D.Nguyen, and A.Ephremides. On the construction of energy-efficient broadcast and multicast trees in wireless networks. In *INFOCOM 2000*, volume 2, March 2000.

[16] K.S. Khan. A hybrid system for global communications. In *Tenth International Conference on Digital Satellite Communications*, May 1995.

[17] M. Krunz and A. Muqattash. A power control scheme for MANETS with improved throughput and energy consumption. In *The 5th International Symposium on Wireless Personal Multimedia Communications*, October 2002.

[18] X. Li. Blocking and handoff performance analysis of a LEO satellite/terrestrial hybrid system. In *IEEE 2002 International Conference on Communications, Circuits and Systems*, 2002.

[19] M.Zorzi and B.R.R.Rao. Energy-constrained error control for wireless channels. *IEEE Personal Communications*, December 1997.

[20] N.Bambos and J.M.Rutnick. Mobile power management for maximum battery life in wireless communication networks. In *IEEE INFOCOM*, March 1996.

[21] P. Nuggehalli, V. Srinivasan, and R.R. Rao. Delay constrained energy efficient transmission strategies for wireless devices. In *IEEE INFOCOM*, 2002.

[22] A. Pantelidou. Scheduling transmissions in wireless ad-hoc networks with time-varying topologies. Master's thesis, University of Maryland, College Park, 2004.

[23] R.A. Raines and N.J. Davis. Personal communications via low earth orbit satellite communication networks. In *Military Communications Conference*, 1995.

[24] Theodore Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, Inc., 2002.

[25] E. Del Re, R. Fantacci, and G. Giambene. Performance analysis of dynamic channel allocation technique for terrestrial and satellite mobile cellular networks. In *Global Telecommunications Conference*, 1993.

[26] J.R. Simić and D.C. Simić. On the capacity of a cellular CDMA system employing power control on fading channels. In *4th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services*, October 1999.

[27] S. Soliman, C. Wheatley, and R. Padovani. CDMA reverse link open loop power control. In *Global Telecommunications Conference*, December 1992.

[28] Hideaki Takagi. *Queuing Analysis: A foundation of Performance Evaluation. Vol 3: Discrete-time systems*. North Holland, 1993.

[29] T.Girici and A.Ephremides. Optimal power control for wireless queuing networks. In *Conference on Information Systems and Sciences*, March 2004.

[30] E. Uysal-Biyikoglu, B. Prabhakar, and A. El-Gamal. Energy-efficient packet transmission over a wireless link. *IEEE/ACM Transactions on Networking*, 10(4), August 2002.

[31] Sergio Verdu. *Multi-user detection*. Cambridge University Press, 1998.

[32] P. Wan, G. Calinescu, and C. Yi. Minimum-power multicast routing in static ad hoc wireless networks. *IEEE/ACM Transactions on Networking*, 12(3), June 2004.

[33] J.E. Wieselthier and A. Ephremides. A new class of protocols for multiple access in satellite networks. *IEEE Transactions on Automatic Control*, 25(5), October 1980.

[34] C.G. Wildey. Satellite and cellular integration: A terminal manufacturer's perspective. In *Third Annual International Conference on Universal Personal Communications*, September 1994.

[35] W.W.S. Wong and E.S. Sousa. Single frequency broadcast system performance in a power-controlled CDMA cellular network. In *Wireless Communications and Networking Conference*, September 2000.

[36] W. Yang and E. Geraniotis. Performance analysis of networks of LEO satellites with integrated voice/data traffic. In *Military Communications Conference*, October 1993.

[37] Feng Zhu. http://www.ccs.neu.edu/home/zhufeng/security_manet.html.