Chapter 4

Cooperative MAC for Cognitive Radio Network with Energy Harvesting and Randomized Service Policy

Providing wireless communication services is becoming more challenging due to spectrum scarcity problem; one technique to approach this problem is the cognitive radio technology in which the unlicensed users (or SUs or cognitive users) are allowed to exploit unused spectrum by the licensed users (or PUs) so that the spectrum utilization is improved and consequently the spectral efficiency increases [1], [4]. The primary user can use the channel at any time as long as it has a packet to transmit, while the coexistence of the secondary user with primary user is allowed provided that the secondary user does not violate some Quality of Service (QoS) requirements of the PU.

Cooperative scenarios have been lately introduced in which a cooperating terminal relays packets for other terminals over the so called relay channel in order to increase the channel availability for its own packets [34]. Similar scenarios are proposed in which the PU has the authority to access the channel whenever it has a packet to transmit. A primary packet unsuccessfully transmitted by the PU and successfully transmitted to the SU is stored in a relay queue at the SU. On the other hand, the SU waits for the opportunity of an idle instant to transmit either the relayed packets or its own packets and in most studies priority is given to the relayed traffic in a way that guarantees QoS requirements of the PU. Randomized cooperative policies for cognitive radio system are also introduced in [35]. These scenarios have proved to enhance cognitive node performance.

Energy harvesting and finiteness of energy have also gained a lot of interest recently. Several works have considered the losses in connectivity periods due to the limited available energy. Despite the advancement in energy harvesting and rechargeable batteries, the study of networks with energy harvesting nodes is still in its infancy. The common objectives were usually to maximize the lifetime of the network whose nodes are powered by rechargeable batteries, while maintaining a certain degree of connectivity [39].

In this chapter, we study the effect of finite energy sources and energy harvesting on the stability and the average packet delay of cooperative multiple access for cognitive radio system. We focus on the class of randomized cooperative policies, whereby the SU admits the PU's packet and serves it with certain probabilities. Besides, we study the effect of energy limitation on the way that tuning the admission and service probabilities affects the stable throughput of the PU and SU. Note that, the analysis involves an interaction between packet queues and the battery. We solved this difficulty by using *stochastic dominance technique*.

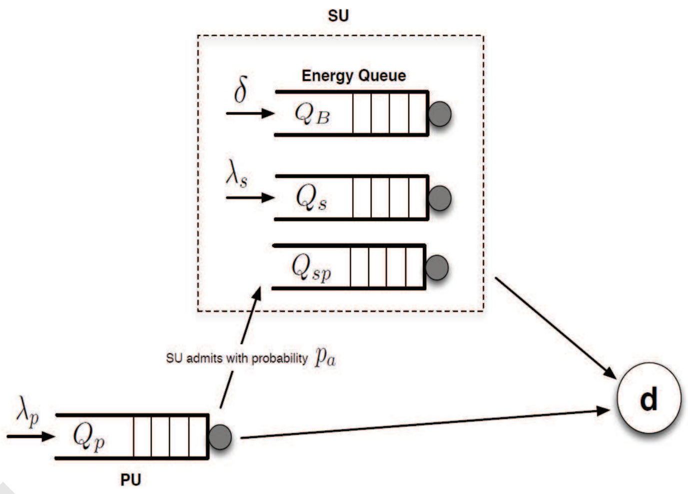

FIGURE 4.1: System model

4.1 System Model

Fig. 4.1 depicts the model of the system under consideration. The system is comprised of a PU and a SU equipped with infinite capacity buffers, transmitting their packets to a common destination d. Each user has an infinite queue, Q_p and Q_s, to store fixed length packets. Also, the SU has another relay queue, Q_{sp}, to store the packets overheard from the PU. The arrival processes at the two queues, Q_p and Q_s, are modeled as Bernoulli arrival processes with means λ_p and λ_s, respectively [44]. Under our system model, the average arrival rates are λ_p and λ_s packets per time slot, and lie in the interval [0,1][1]. The arrival processes at both users are independent of each other, and are independent and identically distributed (i.i.d) across time slots.

To store energy, the SU has a battery modeled as an energy queue, Q_B. Energy is assumed to be harvested in a certain unit and one unit of energy is consumed by each transmission attempt. The energy harvesting process is modeled as Bernoulli arrival processes with mean δ [44]. Under our system model assumptions, the average energy arrival rate is δ energy units per time slot, and is bounded as $0 \leq \delta \leq 1$ [44].

The channel is slotted in time and a slot duration equals one packet transmission time. A successful transmission requires receiving the entire packet without error, otherwise, the packet is discarded. Moreover, we assume that the SU performs perfect sensing. Thus, the system is collision-free, since at most one user is allowed to transmit in a given slot. For

[1]The maximum service rate in our model is 1 packet/slot, since the slot duration equals one packet transmission time, then the arrival rates must be less than 1 otherwise the system will be unstable.

a transmission to be successful, the channel must not be in outage, i.e. the received SNR should not be smaller than a pre-specified threshold. This threshold is the minimum value of the SNR required by the receiver to perform an error-free decoding. Let f_{pd}, f_{sd}, and f_{ps} denote the probability of success between the PU and destination, the SU and destination, and the PU and SU, respectively. It is assumed throughout this chapter that $f_{pd} < f_{sd}$. We assume that a perfect (error-free) feedback channel exists via which the destination sends a feedback to acknowledge the reception of packets. Thus, an ACK is sent if a packet is correctly received. The SU overhears and exploits this feedback.

Next, we describe our PU and SU channel access policy. We assume that the PU has the priority to transmit a packet whenever Q_p is non-empty. An ACK is heard by both users in the network if the packet is successfully decoded by the destination. Thus, the packet exits the system. If the packet is not successfully received by the destination but successfully decoded by the SU, Q_{sp} either buffers the packet with probability p_a or discards it with probability $(1 - p_a)$. This constitutes the probabilistic relaying admission policy. If the packet is buffered in Q_{sp}, the SU sends back an ACK to announce successful reception of the PU's packet. Therefore, the packet is dropped from Q_p and becomes the responsibility of the SU to deliver it to the destination. If the packet is neither successfully received by the destination nor decoded by the SU or decoded but not admitted to Q_{sp}, it is kept at Q_p for retransmission in the next time slot. When the PU is idle, the SU has the opportunity to transmit a packet depending on the battery and data queues status. If the battery queue is empty, then the SU will not be able to transmit a packet. In contrary, if the battery queue is not empty, then the SU transmits a packet from either Q_s or Q_{sp} with probabilities p_q and $(1 - p_q)$, respectively. If the packet is successfully decoded by the destination, it sends back an ACK and the packet exits the system. Otherwise, it is kept at its queue for later retransmission.

4.2 Stable Throughput Region

In this section, we characterize the stability region of the system in Fig. 4.1 under the proposed randomized service policy with probabilistic relaying and energy constraint at the SU queues. In particular, we characterize the shrinkage in the stability region due to the limited energy, which constitutes one of the major contributions of this work. Moreover, we study the effect of tuning system parameters, (p_a, p_q), on the stability region of the system and how it may help increasing the throughput of a certain user by tuning these parameters within different cases, depending on the PU and SU performance requirements and QoS constraints.

Stability can be loosely defined as having a certain quantity of interest bounded. In the queuing theory context, we are interested in the queue size being bounded. For a rigorous definition of stability under more general scenarios, see [48] and [49].

Lemma 4.1. *For our system with energy limitations, and for a fixed value of* (p_q, p_a)*, the system is stable if the arrival rates of* Q_p *and* Q_s *satisfy the following conditions:*

$$\lambda_p < \frac{\delta(1-p_q)f_{sd}(f_{pd}+p_a f_{ps}(1-f_{pd}))}{p_a f_{ps}(1-f_{pd})}, \quad \lambda_s < p_q f_{sd}\delta. \tag{4.1}$$

Proof. If the arrival and service processes of a queueing system are strictly stationary, then one can apply Loynes' theorem to check for stability conditions [52]. This theorem states that if the arrival process and the service process of a queueing system are strictly stationary, and the average arrival rate is less than the average service rate, then the queue is stable, otherwise it is unstable.

For Q_p stability, the condition $\lambda_p < \mu_p$ must be satisfied, where μ_p denotes the service rate of Q_p. A packet departs Q_p if it is successfully received by the destination or is decoded by the SU and is admitted to its relay queue. Thus, μ_p is given by

$$\mu_p = f_{pd} + p_a f_{ps}(1-f_{pd}). \tag{4.2}$$

It is worth noting that, the service rate of packets in both queues, Q_s and Q_{sp}, depends on the state of the battery queue, Q_B at the secondary node and vise versa. This results in an interacting system of queues, and complicates the stability region characterization. We bypass this hurdle by using the *Dominant System* concept originally proposed by Rao and Ephremides in [48] in which we assume that Q_s and Q_{sp} continue to transmit dummy packets even when they are empty. This system "stochastically dominates" our system, that is the SU queues lengths in the dominant system are always larger than the SU queues lengths in our system if both, the dominant system and our system, start from the same initial state and have the same arrivals and encounter the same packet losses.

By this dominant system, the battery queue, Q_B, is decoupled from Q_s and Q_{sp} and forms a discrete-time M/M/1 system with arrival rate δ and service rate $(1-\lambda_p/\mu_p)$. The energy is consumed if and only if the PU's queue is empty which occurs with probability $(1 - \lambda_p/\mu_p)$. Therefore, we have two different cases depending on comparing δ to $(1 - \lambda_p/\mu_p)$. If $\delta > (1 - \lambda_p/\mu_p)$, the role of Q_B is ruled out as the energy arrival rate is greater than the energy consumption rate and the energy queue will saturate (no energy limitation). In this case, the stability conditions are derived from the stability of the data queues only as studied in [35] which are given by

$$\begin{aligned} \lambda_p &< \frac{f_{sd}(1-p_q)\left[f_{pd} + p_a f_{ps}(1-f_{pd})\right]}{f_{sd}(1-p_q) + p_a f_{ps}(1-f_{pd})}, \\ \lambda_s &< p_q f_{sd}\left[1 - \frac{\lambda_p}{f_{pd} + p_a f_{ps}(1-f_{pd})}\right]. \end{aligned} \tag{4.3}$$

On the other hand, if $\delta \leq (1 - \lambda_p/\mu_p)$, the effect of Q_B prevails as the system becomes energy-limited. We expect the stability region to shrink, compared to the no energy limitation case, which is shown later. It follows from Little's theorem that Q_B is non-empty for a fraction of time $\frac{\delta}{(1-\lambda_p/\mu_p)}$. We will consider this case in our analysis. For Q_{sp} stability, the following condition must be satisfied

$$p_a f_{ps}(1-f_{pd})\frac{\lambda_p}{\mu_p} < (1-\lambda_p/\mu_p)(1-p_q) f_{sd}\frac{\delta}{(1-\lambda_p/\mu_p)}. \tag{4.4}$$

A PU's packet is buffered at Q_{sp} if the link between the PU and the destination is in outage which happens with probability $(1 - f_{pd})$, whereas the link between the PU and the SU is not in outage which happens with probability f_{ps}, and the packet is admitted to Q_{sp} which occurs with probability p_a, while Q_p is not empty which has a probability of λ_p/μ_p. This explains the left hand side of (4.4) which is the rate of packet arrivals to the SU relay queue. The right hand side represents the service rate seen by the packets of Q_{sp}. A packet departs the relay queue if Q_p is empty, Q_{sp} is selected to transmit a packet which occurs with probability $(1-p_q)$, the link between the SU and the destination is not in outage and the battery queue is non-empty which occurs with probability $\frac{\delta}{(1-\lambda_p/\mu_p)}$. Rearranging the terms of the above inequality yields the following condition on the maximum achievable arrival rate at the PU

$$\lambda_p < \left[\frac{\delta(1-p_q)f_{sd}}{p_a f_{ps}(1-f_{pd})}\right]\mu_p. \tag{4.5}$$

substituting from (4.2) in (4.5) we get

$$\lambda_p < \frac{\delta(1-p_q)f_{sd}(f_{pd}+p_a f_{ps}(1-f_{pd})}{p_a f_{ps}(1-f_{pd})}. \tag{4.6}$$

From the condition $\delta \leq (1-\lambda_p/\mu_p)$, we conclude that λ_p cannot exceed the value $(1-\delta)\mu_p$. Therefore, (4.6) provides a tighter bound on λ_p than the condition $\lambda_p < \mu_p$.

For Q_s stability, the following condition must be satisfied

$$\lambda_s < p_q f_{sd}(1-\lambda_p/\mu_p)\frac{\delta}{(1-\lambda_p/\mu_p)}, \tag{4.7}$$

which leads to

$$\lambda_s < p_q f_{sd}\delta. \tag{4.8}$$

Using the same argument, a packet departs Q_s if Q_p is empty, Q_s is selected to transmit a packet, the link between the SU and the destination is not in outage, and the battery queue is non-empty. This explains the service rate seen by the packets of Q_s given in the right hand side of (4.7) which is independent of primary service and arrival rates and its queue state. As a result, it does not depend neither on the state of the Q_p nor on p_a. The reason for this behaviour will be explained later. By (4.6) and (4.8), we establish the result in (4.1). □

Next, we study the effect of tuning p_q and p_a on the stability region of the system. At first we begin by varying p_q while keeping p_a constant, followed by varying p_a while keeping p_q fixed.

Lemma 4.2. *In case of $\delta \leq (1-\lambda_p/\mu_p)$, the maximum achievable arrival rate at the PU, λ_p, is monotonic decreasing in both p_q and p_a. Furthermore, for a fixed λ_p, the maximum achievable arrival rate at the SU, λ_s, is monotonic increasing in p_q and does not depend on p_a.*

On the other hand, for the case of $\delta > (1-\lambda_p/\mu_p)$, the maximum achievable arrival rate at the PU, λ_p, is monotonic decreasing in p_q. It is monotonic increasing in p_a if p_q lies in the interval $\left(0, 1-\frac{f_{pd}}{f_{sd}}\right)$, and is monotonic decreasing in p_a if p_q lies in the interval $\left(1-\frac{f_{pd}}{f_{sd}}, 1\right)$. Furthermore, for a fixed λ_p, the maximum achievable arrival rate at the SU, λ_s, is monotonic increasing in both p_q and p_a.

Proof. For $\delta \leq (1 - \lambda_p/\mu_p)$, taking the derivative of the maximum achievable arrival rate at the PU , λ_p given by (4.6), with respect to p_q yields

$$\frac{\partial \lambda_p}{\partial p_q} = \frac{-\delta f_{sd}(f_{pd} + p_a f_{ps}(1 - f_{pd}))}{p_a f_{ps}(1 - f_{pd})}. \tag{4.9}$$

Since p_a, f_{sd}, f_{ps}, f_{pd}, and δ are all positive numbers less than one, then $\frac{\partial \lambda_p}{\partial p_q}$ is negative definite irrespective of the choice of $p_a > 0$. Therefore, the maximum achievable λ_p is monotonic decreasing in p_q when $\delta \leq (1 - \lambda_p/\mu_p)$.

By taking the derivative of (4.6) with respect to p_a yields

$$\frac{\partial \lambda_p}{\partial p_a} = \frac{-\delta f_{ps}(1 - f_{pd}) f_{sd} f_{pd}}{(p_a f_{ps}(1 - f_{pd}))^2}. \tag{4.10}$$

Since p_a, f_{sd}, f_{ps}, f_{pd}, and δ are all positive numbers less than one, then $\frac{\partial \lambda_p}{\partial p_q}$ is negative definite independent of the choice of $p_q > 0$. Therefore, the maximum achievable λ_p is monotonic decreasing in p_a when $\delta \leq (1 - \lambda_p/\mu_p)$.

At the SU side, taking the derivative of (4.8), with respect to p_q yields

$$\frac{\partial \lambda_s}{\partial p_q} = \delta f_{sd}. \tag{4.11}$$

Since f_{sd} and δ are positive numbers less than one, then $\frac{\partial \lambda_s}{\partial p_q}$ is positive definite irrespective of the choice of p_a. Therefore, the maximum achievable λ_s in case of $\delta \leq (1 - \lambda_p/\mu_p)$ is monotonic increasing in p_q.

Also, the maximum achievable arrival rate at the SU, λ_s, does not depend on p_a. This behavior can be explained as follows: As the number of admitted packets from PU to SU's relay queue increases (which depends on the probability p_a), the amount of energy, which is consumed to deliver these packets, increases. This means that, the effect of p_a vanishes by the additional consumed energy. Also, SU will not be able to utilize the free time slot unless the battery has energy. Therefore, λ_s does not depend on p_a, while it depends on δ.

The case of $\delta > (1 - \lambda_p/\mu_p)$ has been proven in [35]. It is worth noting that, p_q does not affect the relation between λ_p and p_a in case of $\delta \leq (1 - \lambda_p/\mu_p)$, while it affects the relation between them in case of $\delta > (1 - \lambda_p/\mu_p)$. Perhaps an intuitive explanation for this behavior is the following: in case of $\delta \leq (1 - \lambda_p/\mu_p)$, the energy limitation problem

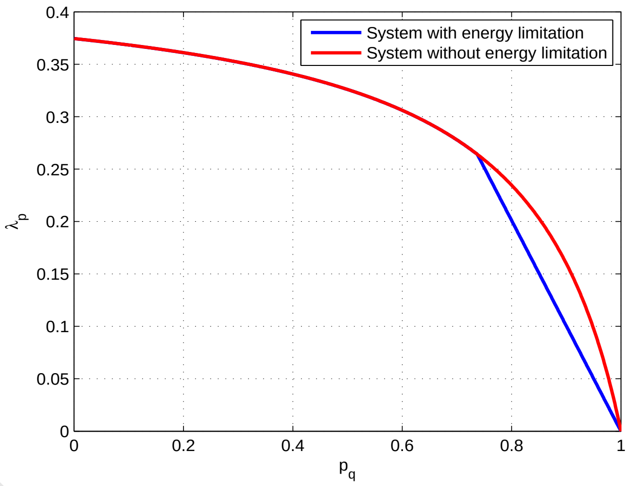

FIGURE 4.2: Maximum achievable λ_p versus p_a for $p_a = 0.5$ and $\delta = 0.42$.

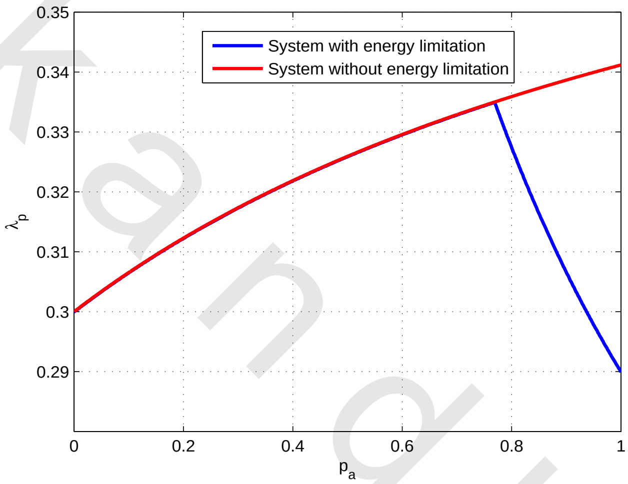

FIGURE 4.3: Maximum achievable λ_p versus p_a for $p_q = 0.5$ and $\delta = 0.42$.

influences the situation. This means that, using the SU as a relay in this case will decrease the PU throughput whatever the channel quality between the SU and the destination. This occurs due to the energy limitation at SU node which prevents the SU from transmitting if the battery queue is empty. □

In Fig. 4.2 and Fig. 4.3, we depict the effect of tuning (p_q, p_a) on the maximum achievable λ_p. The PU throughput of both systems, with and without energy limitation, is plotted against p_q and p_a for $\delta = 0.42$. The system parameters are chosen as follows: $f_{pd} = 0.31$, $f_{ps} = 0.42$, and $f_{sd} = 0.8$. It is worth noting that, the split point of the two curves in both figures is the point after which the condition $\delta \leq (1 - \lambda_p/\mu_p)$ holds and the system becomes energy-limited. According to these figures, the maximum achievable arrival rate at the PU decreases

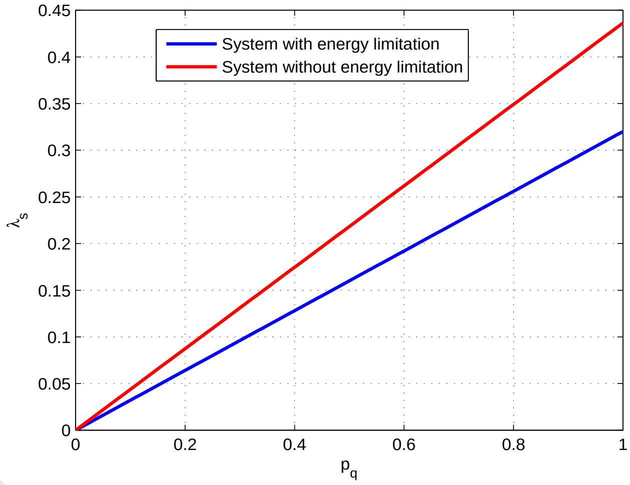

FIGURE 4.4: Maximum achievable λ_s versus p_a for $p_a = 0.5$ and $\delta = 0.42$.

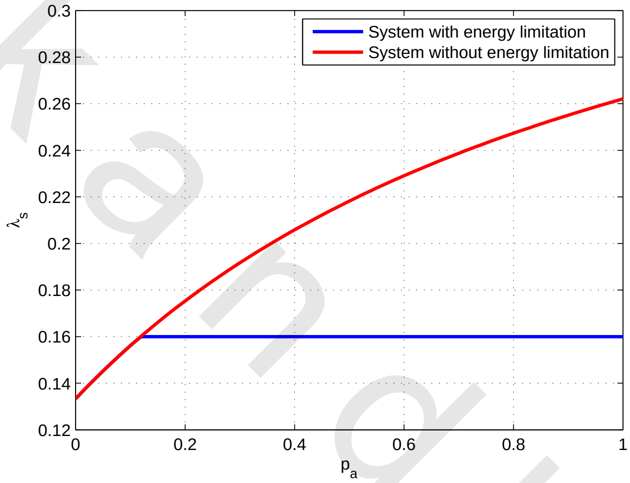

FIGURE 4.5: Maximum achievable λ_s versus p_a for $p_q = 0.5$ and $\delta = 0.42$.

with the increase of p_q and p_a for the case $\delta \leq (1 - \lambda_p/\mu_p)$. The relation between λ_p and p_q is intuitive, since increasing the value of p_q gives more chance for transmitting the SU own packets as opposed to the PU's relayed packets. This, in turn, reduces the degree of cooperation the PU experiences from the SU and, hence, the maximum achievable λ_p decreases. For the energy-limited scenario (beyond the curves split point), the shortage in energy affects the transmission operation of the SU queues. Therefore the SU cannot serve the the relayed packet from PU if the battery queue is empty. As a result, the maximum achievable λ_p decreases with the increase of p_a. Moreover, the energy limitation causes a loss in the PU throughput as shown in the figures.

In Fig. 4.4 and Fig. 4.5, we depict the effect of tuning (p_q, p_a) on the maximum achievable

λ_s. The SU throughput is plotted against p_q and p_a for $\delta = 0.42$, respectively. We also plotted the SU throughput without energy limitation. The channel success probabilities are the same as above. Once more, the curves split point in both figures is the point after which the condition $\delta \leq (1 - \lambda_p/\mu_p)$ holds and the system becomes energy-limited. According to these figures, the maximum achievable arrival rate at the SU increases with the increase of p_q while it remains constant with the change of p_a for the case $\delta \leq (1 - \lambda_p/\mu_p)$. The relation between λ_s and p_q is intuitive, since increasing p_q leads to an increase in the number of SU own packets to be served. Thus, we conclude that increasing p_q is always in favor of the SU. On the other hand, the maximum achievable λ_s is constant whatever the value of p_a due to energy restriction. Also, the energy limitation causes a loss in the SU throughput as shown in the figures.

We present next a complete characterization of the boundary of the stability region for the whole system. In case of $\delta \leq (1 - \lambda_p/\mu_p)$, we find the union over $0 \leq p_a \leq 1$ and $0 \leq p_q \leq 1$ of the stability regions given in (4.1). For the region defined, we should either maximize λ_s for a given λ_p or maximize λ_p for a given λ_s. We consider maximizing $\lambda_p = \dfrac{\delta(1-p_q)f_{sd}(f_{pd}+p_a f_{ps}(1-f_{pd})}{p_a f_{ps}(1-f_{pd})}$ under the condition $\lambda_s = p_q f_{sd} \delta$. It has been proven that λ_p is monotonically decreasing in p_a. We can get the minimum value of p_a from the condition $\delta < (1 - \lambda_p/\mu_p)$ which is $p_a = \frac{\lambda_p - (1-\delta) f_{pd}}{(1-\delta) f_{ps}(1-f_{pd})}$. For p_q, we can get its value from the given condition in our optimization problem which is given by $p_q = \frac{\lambda_s}{f_{sd}\delta}$. Now, substitute by p_a and p_q in λ_p to get a relation between λ_p and λ_s which can be characterized as

$$\lambda_s = \delta f_{sd} + (1-\delta) f_{pd} - \lambda_p. \tag{4.12}$$

This equation is valid in the region $0 \leq p_a \leq 1$ which implies that $(1-\delta)f_{pd} \leq \lambda_p \leq (1-\delta)(f_{ps}(1-f_{pd}) + f_{pd})$. By parallel arguments, we can characterize the boundary of the stability region in case of $\delta > (1 - \lambda_p/\mu_p)$ which is given by

$$\lambda_s = f_{sd} - \left[\frac{f_{sd} + f_{ps}(1-f_{pd})}{f_{pd} + f_{ps}(1-f_{pd})} \right] \lambda_p. \tag{4.13}$$

The stability region boundary depends on the values of δ, f_{ps}, f_{pd} and f_{sd}. If $\delta \leq \frac{f_{ps} - f_{pd} f_{ps}}{f_{ps} + f_{sd} - f_{pd} f_{ps}}$, the boundary of the stability region for the whole system is given by (4.14) on the top of the next page and if $\delta > \frac{f_{ps} - f_{pd} f_{ps}}{f_{ps} + f_{sd} - f_{pd} f_{ps}}$, the boundary of the stability region of the whole system is given by (4.15) on the top of the next page.

$$\lambda_s = \begin{cases} \delta f_{sd} & 0 \leq \lambda_p \leq (1-\delta) f_{pd} \\ \delta f_{sd} + (1-\delta) f_{pd} - \lambda_p & (1-\delta) f_{pd} \leq \lambda_p \leq (1-\delta)(f_{ps}(1-f_{pd}) + f_{pd}) \\ 0 & (1-\delta)(f_{ps}(1-f_{pd}) + f_{pd}) \leq \lambda_p \leq 1 \end{cases} \quad (4.14)$$

$$\lambda_s = \begin{cases} \delta f_{sd} & 0 \leq \lambda_p \leq (1-\delta) f_{pd} \\ \delta f_{sd} + (1-\delta) f_{pd} - \lambda_p & (1-\delta) f_{pd} \leq \lambda_p \leq (1-\delta)(f_{ps}(1-f_{pd}) + f_{pd}) \\ f_{sd} - \left[\frac{f_{sd}+f_{ps}(1-f_{pd})}{f_{pd}+f_{ps}(1-f_{pd})}\right] \lambda_p & (1-\delta)(f_{ps}(1-f_{pd}) + f_{pd}) \leq \lambda_p \leq \dfrac{f_{sd}(f_{pd}+f_{ps}(1-f_{pd}))}{f_{sd}+f_{ps}(1-f_{pd})} \\ 0 & \dfrac{f_{sd}(f_{pd}+f_{ps}(1-f_{pd}))}{f_{sd}+f_{ps}(1-f_{pd})} \leq \lambda_p \leq 1 \end{cases} \quad (4.15)$$

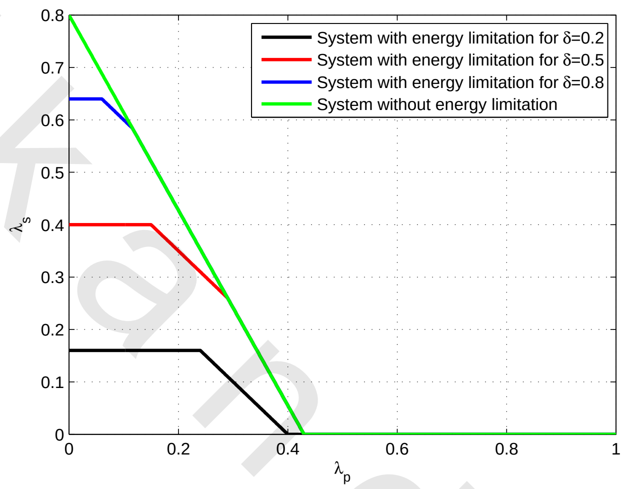

FIGURE 4.6: The stability boundaries for our system with $\delta = 0.2$, 0.5 and 0.8 and the system with no energy limitation.

In Fig. 4.6, the stability boundaries of the proposed system with energy limitation are plotted for $\delta = 0.2$ where $\delta \leq \frac{f_{ps}-f_{pd}f_{ps}}{f_{ps}+f_{sd}-f_{pd}f_{ps}}$ and $\delta = 0.5$ and 0.8 where $\delta > \frac{f_{ps}-f_{pd}f_{ps}}{f_{ps}+f_{sd}-f_{pd}f_{ps}}$, respectively. We also plotted the stability boundary of the system without energy constraint. From the figure, we can conclude that due to the energy constraint, there is a loss in the stability region recognized by gap between the stability boundaries of the systems. This loss occurred as Q_s cannot be served at a rate greater than the energy harvesting rate multiplied by the channel outage rate. This explains why the stability boundaries of the system with energy constraint cut the λ_s axis at $\lambda_s = \delta f_{sd}$.

4.3 Delay analysis

In this section, we perform the delay analysis of the energy-limited system using the moment generating function. As mentioned before, we have an interacting system of queues. Such interacting system complicates the calculation as we have to solve a moment generating function with 3 variables (the states of Q_s, Q_{sp} and Q_B) which is very complex to analyze. To bypass this difficulty, we assume that the battery Q_B loses energy whenever the PU queue, Q_p, is empty irrespective of SU queues state. This means that, if the PU queue is empty, an energy unit will be lost even if the SU queue, which is selected to be served, is empty. As a result, we characterize the upper bound of the average packet delay.

For the case of $\delta \leq (1-\lambda_p/\mu_p)$, to characterize the average delay encountered by the packets of the PU as well as the SU, we have to calculate the average length for each queue. It is worth noting that, service processes at both Q_s and Q_{sp} depend on the state of Q_p. However, Q_s and Q_{sp} are independent, i.e., having independent arrivals and departures. So that, we can use the moment generating function and follow the same footsteps in [35] to calculate the average length of Q_s and Q_{sp}. The moment generating function of the joint lengths of Q_p and Q_s is defined as

$$\begin{aligned} G(x,y) &= \lim_{t\to\infty} E\left[x^{Q_p^t} y^{Q_s^t}\right] \\ &= \lim_{t\to\infty} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} x^i y^j P\left[Q_p^t = i, Q_s^t = j\right], \end{aligned} \tag{4.16}$$

where E and P denote the statistical expectation and the probability operators, respectively.

Thus, the sequence of characterizing N_s goes as follows. First, we derive $G(x,y)$, then take its derivative with respect to y and put $x = y = 1$. After following previous procedures and calculating N_s we can note that, a term $\beta = \frac{\delta}{1-\frac{\lambda_p}{\mu_p}}$ will appear multiplied by p_q. This term refers to the prolongation that happens to the average delay encountered by packets of the secondary queue. That is why N_s is given as following

$$N_s = \frac{\lambda_p\lambda_s A + (\lambda_s^2 - \lambda_s)B(B+\lambda_p)}{BC}, \tag{4.17}$$

where

$$
\begin{aligned}
A =& p_q f_{sd}\beta[f_{pd} + p_a f_{ps}(1 - f_{pd}) - 1] \\
B =& f_{pd} + p_a f_{ps}(1 - f_{pd}) - \lambda_p \\
C =& (\lambda_s - p_q f_{sd}\beta)[f_{pd} + p_a f_{ps}(1 - f_{pd})] + p_q f_{sd}\beta\lambda_p.
\end{aligned}
\tag{4.18}
$$

Also, by following same procedure to calculate the average length of Q_{sp}, N_{sp}, we will recognize that, the term β will appear multiplied by $(1 - p_q)$. This term refers to the prolongation that happens to the average delay encountered by packets of the secondary-relay queue. Therefore, N_{sp} can be characterized as

$$
N_{sp} = \frac{m\lambda_p^2 + n\lambda_p}{\alpha\lambda_p^2 + \beta'\lambda_p + \gamma},
\tag{4.19}
$$

where

$$
\begin{aligned}
m =& p_a f_{ps}(1 - f_{pd})[\frac{(1 - p_q) f_{sd}\beta - f_{pd}}{f_{pd} + p_a f_{ps}(1 - f_{pd})} \\
& - (1 - p_q) f_{sd}\beta - p_a f_{ps}(1 - f_{pd})] \\
n =& p_a f_{ps}(1 - f_{pd})[f_{pd} + p_a f_{ps}(1 - f_{pd}) \\
\alpha =& (1 - p_q) f_{sd}\beta + p_a f_{ps}(1 - f_{pd}) \\
\beta' =& [f_{pd} + p_a f_{ps}(1 - f_{pd})][-2(1 - p_q) f_{sd}\beta - p_a f_{ps}(1 - f_{pd})] \\
\gamma =& (1 - p_q) f_{sd}\beta[f_{pd} + p_a f_{ps}(1 - f_{pd})]^2.
\end{aligned}
\tag{4.20}
$$

For the PU, we can easily calculate N_p by observing that Q_p is a discrete-time M/M/1 queue with arrival rate λ_p and service rate μ_p. Thus, applying the Pollaczek-Khinchine formula [53], N_p is directly given as

$$
N_p = \frac{-\lambda_p^2 + \lambda_p}{f_{pd} + p_a f_{ps}(1 - f_{pd}) - \lambda_p}.
\tag{4.21}
$$

The average packet delay for each queue can be calculated as follows.

$$
D_p = \frac{N_p + N_{sp}}{\lambda_p}, \quad D_s = \frac{N_s}{\lambda_s},
\tag{4.22}
$$

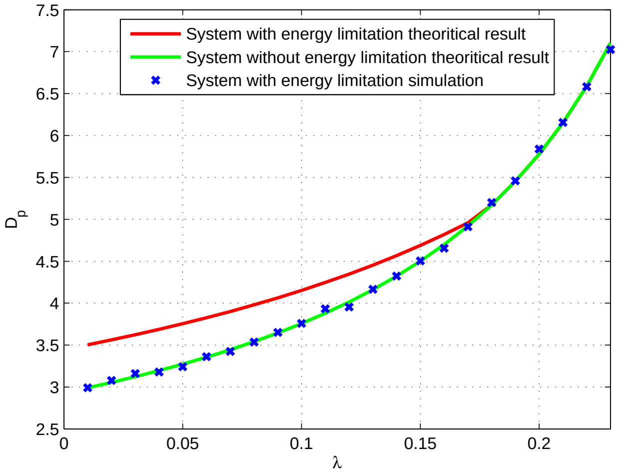

FIGURE 4.7: Comparison between simulation results and theoretical results of the PU delay at $p_a = 1$, $p_q = 0.5$ and $\delta = 0.7$.

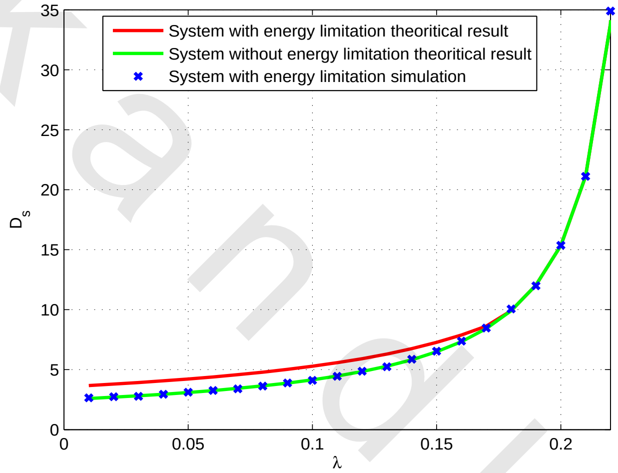

FIGURE 4.8: Comparison between simulation results and theoretical results of the SU delay at $p_a = 1$, $p_q = 0.5$ and $\delta = 0.7$.

where D_p and D_s are the average packet delay for the primary and secondary users, respectively.

In Fig. 4.7 and Fig. 4.8, the average delay encountered by the packets of the PU and SU, respectively, are plotted against λ where we choose $\lambda_p = \lambda_s = \lambda$ for simplicity. These figures compare the average delay of the system without energy limitations to the upper bound derived above for the delay of the energy-limited system. Moreover, simulation results for the energy-limited system's delay are presented. In these figures, p_a, p_q and δ are fixed at 1, 0.5 and 0.7, respectively.

It is noted from Fig. 4.7 and Fig. 4.8 that the simulated average delay of the energy-limited system coincides with the delay of the system without energy limitations for all values of λ. For values of λ satisfying the condition $\delta \leq (1 - \lambda_p/\mu_p)$, the upper bound on the delay is within 17% of the simulation results at the point with maximum difference. However, for values of λ satisfying the condition $\delta > (1 - \lambda_p/\mu_p)$ (when the battery queue is always saturated), the upper bound coincides with the simulation results and the delay of the system without energy limitations. This coincidence is expected, since when the battery queue becomes saturated the SU will always have energy to transmit, and the system falls back to the system with no energy limitations.

On the other hand, the coincidence between the two systems (with and without energy limitations) for values of λ satisfying the condition $\delta \leq (1 - \lambda_p/\mu_p)$ seems counter intuitive. We can interpret this result using the following conjecture: "Inside the stable throughput region of the energy-limited system the battery queue is always saturated". If this is the case, then the finiteness of energy will only lead to a reduction of the stable throughput region as proved above. However, inside the stability region the average delay is not affected.