Performance Of Unlicensed Devices With A Spectrum Etiquette
Durga P. Satapathy and Jon M. Peha
Department of Electrical and Computer Engineering, Carnegie Mellon University

Abstract
There are growing opportunities for the wireless industry to deploy products in unlicensed spectrum, where no licensing is required and devices share spectrum. Unlicensed spectrum offers several benefits, but there are serious challenges to overcome before these benefits can be fully realized. Foremost among these is the risk of inefficient utilization, resulting from a lack of incentive to conserve the shared spectrum. This phenomenon will be demonstrated in one scenario, and some techniques to deal with this problem will be presented.

1 Introduction
The wireless industry has actively and successfully petitioned the Federal Communications Commission (FCC) to allocate more unlicensed spectrum [1-4], which is spectrum in which users do not require a license to transmit. Such spectrum includes the unlicensed PCS band at 2 GHz [1], the unlicensed National Information Infrastructure (NII) band at 5 GHz [2], and the Millimeter Wave band at 59-64 GHz [3,4]. The unlicensed PCS band is governed by a spectrum etiquette [1,5], which is a set of rules dictating when, where and how may devices transmit. The other bands have power and emission limits, but there is currently no spectrum etiquette. By eliminating the need for licenses and associated delay, unlicensed spectrum offers several benefits. It facilitates the mobility of wireless applications, such as a wireless Private Branch Exchange (PBX) that is moved from one construction site to another. Operating in unlicensed spectrum is easier and more efficient than acquiring licenses for each PBX location. Another benefit is that unlicensed spectrum promotes spectrum sharing, as any device can transmit while other devices are idle. Consequently, it is especially useful for applications that require sporadic access to spectrum, and can tolerate widely varying access delays, like a wireless electronic mail service. As spectrum sharing eliminates trunking inefficiencies [6], it offers greater spectrum utilization. Finally, unlicensed spectrum facilitates experimentation and rapid deployment of new products, as it is a readily accessible test-bed for novel and rapidly evolving technologies.

However, to realize these potential benefits, there are three challenges to be overcome in unlicensed spectrum: First, as devices do not have exclusive access, they may cause mutual interference. Second, as devices may differ greatly in average data rate, transmission duration, or even the technology used, it becomes difficult to enforce efficient utilization for all devices. Third, as spectrum is shared, there is no inherent incentive to design devices to use the spectrum efficiently. As a result, individual devices may waste spectrum to improve their own performance to the extent that performance degrades significantly for other devices. If this is common, the shared spectrum will be of little use. This phenomenon has been referred to as a Tragedy of the Commons.

Given the significant advantages and potential problems of unlicensed spectrum, is it a viable alternative to licensed spectrum? If the utilization of unlicensed spectrum is likely to be highly inefficient, it would not be prudent to invest in the manufacture and deployment of unlicensed devices, nor would the recent unlicensed spectrum allocations be justified. Given that performance in unlicensed spectrum is uncertain, what strategies should designers of unlicensed equipment choose? If these strategies lead to inefficient spectrum utilization, can etiquette modifications provide incentives for efficient spectrum sharing? We address these issues in this paper.

Section 2 discusses how the spectrum etiquette approaches the challenges of unlicensed spectrum. Section 3 discusses whether designers may choose approaches that waste shared spectrum. Section 4 demonstrates that there is potential for significant performance gain with such strategies, but there is also a risk that performance may degrade considerably. Section 5 addresses techniques to provide incentives to conserve spectrum. Section 6 presents an example of a modification that alleviates the risk of poor utilization. In Section 7 we present our conclusions.

2 The Spectrum Etiquette
There are currently two approaches to alleviate mutual interference in unlicensed bands. Devices deployed in the Industry, Science and Medicine(ISM) bands are restricted to low power spread spectrum transmissions. This limits the coverage area of devices, which helps to reduce interference. However, as devices may transmit at will, they are still at risk of mutual interference. There is of course no scope for diversity. Furthermore, the power constraint results in inefficient spectrum utilization, since devices cannot transmit at greater power even when there is no contention for spectrum. The other approach is the spectrum etiquette for unlicensed PCS, which enforces a "Listen Before Talk" (LBT) rule, requiring devices to monitor the channel and transmit only if the signal energy detected is sensed to be below a threshold throughout the specified monitoring period. The etiquette has separate bands and rules for isochronous and asynchronous devices, with parameters like monitoring time and LBT threshold value varying from band to band. There are features designed to conserve spectrum, such as upper limits on transmission duration, power and bandwidth. Another provision is that devices may increase the LBT threshold by one dB for each dB reduction in their transmission power. This provides better access to devices with low power transmissions, thus promoting frequency reuse. It remains to be seen whether these provisions are sufficient and necessary.

1 This material is based upon work supported in part by Bellcore, and by the Interstate Electronic Corporation.
Section 3: To Be Or Not To Be Greedy?
All wireless system designs involve a trade-off between competing goals and interests, such as reducing equipment costs while optimizing reception quality. In licensed spectrum, where the spectrum consumed is typically the exclusive domain of the end users, conserving spectrum is an important design goal. Although conserving unlicensed spectrum is no less important from a system perspective, there is considerably less incentive to design individual devices to do so. Thus, in unlicensed spectrum, it is more likely that the best decision from the designer's selfish perspective is also a greedy approach, where the more a device wastes shared spectrum in favor of its own goals, the more we consider it to be greedy. The amount of resources a device consumes with a transmission depends on three factors: the transmission duration, bandwidth, and coverage area (which is a function of transmission power.) Thus, the transmissions of a greedy device have greater duration, bandwidth, or power, than is necessary. We will refer to these factors as the three dimensions along which devices may manifest greed.

To what extent would devices be greedy, and how does greed affect system performance? We address this issue for the duration dimension in the following scenario: Two devices share an unlicensed PCS channel, sufficiently close together to receive each other's transmissions. Device transmissions require the same bandwidth. Thus, transmission power and bandwidth have no impact on device performance. We assume that data awaiting transmission is queued in a buffer of infinite length. Devices are required to sense the channel to be idle throughout the monitoring period before transmitting. A non-greedy device releases the channel immediately after all queued data is sent. A greedy device holds the channel longer once it gets access, just to avoid the access delay whenever it has a data to transmit again. However, while the greedy device hoards the channel, the queued data awaiting transmission at the other devices grows. When the greedy device finally releases the channel, it may take much longer to reclaim it. Thus, greed may or may not be beneficial. Likewise, greed in power or bandwidth dimensions may not always improve performance [7].

If greed proves beneficial, then equipment designers will base their greedy strategy on their projections of the extent to which they will share spectrum with competing devices, and the greed of those devices. Since the strategies employed in one device can influence the optimal design for another, a useful measure of the resulting behavior is the reaction function $r_i(T_j)$, which is the optimal greed for device $i$ in response to that of device $j$. If Device 1 has greed $T_1$, then Device 2 minimizes its delay at greed $r_2(T_1)$. Such dynamic reactions to another device's greed may occur in one of two ways. It is possible that a system administrator can adjust greed on some devices, as one might change the maximum packet length in a LAN. In other systems, the extent of greed will be fixed when the device is manufactured. However, greed can change over time when equipment is replaced. For example, if most CB radios in use have a power of $P$, then someone buying a new radio will buy one with power $r(P)$, and the power levels will change generation after generation.

4 The Risk of Inefficient Utilization
We will use analysis to demonstrate the impact of greed on device performance, for the scenario described in Section 3. We will show that whenever a device becomes greedy, it degrades performance of the other device, causing the other device to become greedy as well. This results in the first device increasing its greed, thus starting a chain reaction, whereby each device always responds with more greed.

To make analysis tractable, we consider data to arrive at a constant rate, according to a fluid-flow model [8,9,10]. More precisely, the amount of data received by Device $i$ in any period $\tau$ is exactly $\rho_i \tau$, where $\rho_i$ is the load of Device $i$. We assume $\rho_1 > 0$, $\rho_2 > 0$, and $\rho_1 + \rho_2 < 1$, where total capacity is defined to be 1. Device $i : i \in \{1,2\}$ holds the channel for period $H_i$ every time it gets access. Device $i$ has greed $T_i$, which is its minimum holding time, so $H_i \geq T_i$. For period $H_i$, Device $i$ has data queued and is transmitting at the maximum rate possible. Let $I_i$ denote the period for which Device $i$ holds the channel with an empty queue, transmitting data as it arrives. Thus, $H_i = X_i + I_i$. $I_i = 0$ for a nongreedy device, and $I_i > 0$ for a greedy device. Devices start transmission only if the channel is sensed idle for a duration $M$. We assume devices monitor with persistence, i.e., devices continuously sense the channel [11]. To simplify analysis, we do not require devices to back-off, i.e., defer from attempting access for a random period when the channel is detected busy; introducing back-off will only give more incentive for greed.

Figure 1 shows Device 1's unfinished work as it varies over time, where unfinished work is the amount of time it would take to transmit all currently queued data. Let Device 2 start monitoring the channel at time 0. At that time, Device 1 has just relinquished the channel, and presumably has transmitted its queued data. Device 2 finds the channel free and begins transmission at time $M$. It then releases the channel at time $H_1 + M$. Device 1's queue increases from time 0 until $H_2 + 2M$, when it begins transmission. Its unfinished work built up at this time is given by $\rho_1(2M + H_2)$. Device 1’s queue length then starts decreasing and reaches zero at time $2M + H_2 + X_1$. The device continues to hold the channel, transmitting data as quickly as it arrives until time $2M + H_2 + H_1$. The cycle is then repeated.
The following theorems characterize the potential greedy behavior in this system. (See Appendix for proofs.) Theorem 1 shows the holding times $H_i$ and $H_j$ for devices 1 and 2 respectively, when neither device is greedy. Note that Device $i$ is never greedy if $\pi_i < H_i$, because the device always holds the channel for at least $H_i$ anyway. Therefore, delays are identical for any $T_i < H_i$. Theorem 2 shows the general reaction functions. Theorem 3 shows that even if device $i$ is nongreedy, if $H_i < 2 M$, then the other device is better off being greedy. To determine the ultimate results of these reaction functions, let Device 2 select an initial greed $T_2^{(0)}$. Devices 1 and 2 will then take turns responding to each other’s greed, i.e., for $t > 0$, $T_i^{(j+1)} = r_i(T_i^{(j)} - 1)$ and $T_j^{(j+1)} = r_j(T_j^{(j)} - 1)$. Theorem 4 shows that if one device is greedy, it will cause both devices to escalate their greed to infinity. Theorem 5 shows that whenever a device’s greed is increased, delay increases for the other device.

**Theorem 1:** If $T_i = T_j = 0$, then $H_i = \frac{2 M \rho_i}{1 - \rho_i - \rho_j}$ and $H_j = \frac{2 M \rho_j}{1 - \rho_i - \rho_j}$.

**Theorem 2:** $r_i(T_i) = \max\{T_i, 2 M\} (1 - \rho_i) / \rho_i - 2 M$ and $r_j(T_j) = \max\{T_j, 2 M\} (1 - \rho_j) / \rho_j - 2 M$.

**Theorem 3:** If $T_i < H_i$ and $H_j < 2 M$ then $r_i(T_i) > H_i$.

If $T_i < H_i$ and $H_j < 2 M$ then $r_j(T_j) > H_j$.

**Theorem 4:** If $T_i^{(j+1)} > H_i$, then $\lim_{j \to \infty} T_i^{(j)} = \infty$ and $\lim_{j \to \infty} T_j^{(j)} = \infty$.

**Theorem 5:** $\frac{\partial D_j}{\partial H_j} = 0$ and $\frac{\partial D_i}{\partial H_i} = 0$.

All together, these theorems show the potential for a tragedy of the commons. If $H_i < 2 M$ or $H_j < 2 M$, which occurs when $\rho_i + \rho_j + \min(\rho_i, \rho_j) < 1$, then it will inevitably lead to an escalation of greed until both devices hold the channel as long as possible, and neither has adequate performance.

## 5 Disincentives for Greed

In this section we will discuss techniques to discourage greedy design strategies. One way to control greed is to set upper limits on transmission duration, power, or bandwidth. However, an upper limit may result in spectrum inefficiency unless chosen appropriately. A device may use a resource as much as the upper limit, even if performance is adequate with less. Hence a higher than optimal limit would be inefficient. A lower than optimal limit may unnecessarily make spectrum unusable for some applications. Moreover, an optimal value in some scenarios would be highly ineffective in others. Another option is a flexible upper limit. For example, the maximum transmitted power allowed by the FCC etiquette is a function of bandwidth. Narrow bandwidth applications are permitted higher power spectral density, which provides an incentive for devices to not use excessive bandwidth. Although there is some flexibility in the dimension of resource consumption, this is still an upper limit, and there is no incentive to do better than the limit requires. Upper limits are indeed a blunt instrument for such a delicate problem.

To produce further incentives, devices that conserve spectrum can be given priority in accessing spectrum. To be effective, incentives must be based on parameters that strongly affect device performance. For example, altering either the monitoring time or the LBT power threshold in the current etiquette would affect a device’s chances of accessing spectrum. In the current FCC etiquette, monitoring time is fixed, but threshold is a function of transmit power. Other examples are the inter-burst gap, which is the minimum amount of time a device must wait to transmit after completing a transmission, and the back-off period, which is the minimum time a device has to wait before attempting to access a channel again once the channel is detected busy. We next explore the potential of such parameters to induce efficient spectrum sharing.

### Section 6: An Example of Etiquette Modification

In this section we analyze an example of etiquette modification using the monitoring time parameter. When a device ends a transmission, it must monitor the channel and find it idle for a total period of at least its penalty before it can contend for transmission, where this penalty time is an increasing function of the resources consumed. The penalty time is not required to be contiguous; it may be split into several nonconsecutive time intervals. Reconsider the scenario analyzed in Section 4, where Device 1 and Device 2 compete for access to a wireless channel. This time, however, for each device we add a penalty time $P_i$ equal to the period $H_i$ that Device $i$ holds the channel, and we let monitoring time be zero. Recall that $H_i = X_i + I_i$, where $X_i$ is the duration when Device $i$ transmits while it has data queued, and $I_i$ is the period for which it further holds the channel.

We define a cycle to begin whenever Device 1 completes transmission of its queued data. Cycle $j$ begins with Device 1 holding the channel for period $I_1^{(j)}$, after which Device 2 may transmit several times while Device 1 pays its penalty $P_1^{(j)} = X_1^{(j)} + I_1^{(j)}$. Let the total duration that Device 2 holds the channel in cycle $j$ be denoted by $K_2^{(j)}$, which need not be contiguous. Device 1 then transmits its total unfinished work $\rho_1(P_1^{(j)} + K_2^{(j)} + X_1^{(j)})$ in period $X_1^{(j)}$, so $X_1^{(j)} = \rho_1(P_1^{(j)} + K_2^{(j)} + X_1^{(j)}) = (P_1^{(j)} + K_2^{(j)}) / (1 - \rho_1)$. The average delay $D_1^{(j)}$ for Device 1 in the $j^{th}$ cycle is its average unfinished work divided by $\rho_1$. So

$$D_1^{(j)} = \frac{5}{2}(P_1^{(j)} + K_2^{(j)}) - \frac{P_1^{(j)} + K_2^{(j)} + X_1^{(j)}}{I_1^{(j)} + P_1^{(j)} + K_2^{(j)} + X_1^{(j)}}$$

$$dD_1^{(j)} = \frac{[P_1^{(j)} + K_2^{(j)} + I_1^{(j)}(1 - \rho_1)](P_1^{(j)} + K_2^{(j)})}{[P_1^{(j)} + K_2^{(j)} + I_1^{(j)}(1 - \rho_1)]^2} \frac{dP_1^{(j)} + K_2^{(j)}}{dI_1^{(j)}} - \frac{2[P_1^{(j)} + K_2^{(j)} + I_1^{(j)}(1 - \rho_1)]^3}{2[P_1^{(j)} + K_2^{(j)} + I_1^{(j)}(1 - \rho_1)]^2} \frac{d(P_1^{(j)} + K_2^{(j)})}{dI_1^{(j)}} + (1 - \rho_1)$$
\[ \frac{dD_{i}^{(j)}}{dt} \geq 0 \text{ if and only if } \frac{dP_{i}^{(j)}}{dt} + \frac{dK_{i}^{(j)}}{dt} \geq -\frac{(1-\rho_{i})(P_{i}^{(j)}+K_{i}^{(j)})}{(1-\rho_{i})^{2}+2(1-\rho_{i})I_{i}^{(j)}} \]

Since \( P_{i}^{(j)} = X_{i}^{(j-1)} + I_{i}^{(j)} \), \( \frac{dP_{i}^{(j)}}{dt} \) is independent of \( I_{i}^{(j)} \). Thus, \( \frac{dD_{i}^{(j)}}{dt} \geq 0 \) if and only if \( \frac{dK_{i}^{(j)}}{dt} \geq -\frac{X_{i}^{(j-1)}+2I_{i}^{(j)}}{1-\rho_{i}} \).

Thus, there is always a disincentive for greed if \( \frac{dK_{i}^{(j)}}{dt} > 0 \), i.e., if Device 1’s holding the channel longer causes Device 2 to also hold longer. This is likely to be the case for the following reasons. First, Device 2 will presumably have more queued data to transmit when it gains access. Second, Device 2 will have more opportunities to transmit if it wishes because Device 1’s penalty time will be longer. For both these cases, this condition is likely to hold if Device 2 is nongreedy. If Device 2 may be greedy, then Device 1’s holding longer may cause Device 2 to expect longer holding times from Device 1 in the future, thereby increasing Device 2’s incentive to be greedy. However, this injects many complexities that will be addressed further in forthcoming work.

7 Conclusion

Unlicensed spectrum has several advantages. It supports mobility of wireless applications, allows spectrum sharing, and facilitates experimentation and innovation. However, as individual devices have little inherent incentive to conserve spectrum, they may hoard shared resources to improve their performance. In this paper, we have demonstrated, in a simple scenario, that greedy behavior is sometimes rewarded, and it can also lead to grossly inefficient spectrum utilization and inadequate performance. The most practical way to deal with this problem is to modify the etiquette. We have shown that etiquette modifications can effectively solve the problems of greed in the above scenario, although we do not claim that this is the most efficient solution. Future work will seek etiquette modifications that discourage greed in a variety of scenarios while minimizing the penalty on devices that have exclusive access to spectrum.

Meanwhile, demand for more unlicensed spectrum is high, and the FCC has increased the allocations for unlicensed spectrum. At this point, there is still little evidence that all provisions in the current FCC etiquette are either necessary or sufficient. It is possible that many existing restrictions in the etiquette could be relaxed without penalty, thereby simplifying designs. It is also possible that spectrum would be utilized inefficiently, unless etiquette modifications are effected. The recent decision by the FCC to allow unlicensed operation in the 5 GHz range with requiring devices to follow a spectrum etiquette may serve to establish the need for an etiquette, if device performance degrades significantly without an etiquette. However, such experimentation runs the risk that an etiquette may be imposed after industry has already invested in the manufacture of unlicensed devices. Consequently, before industry invests significantly in deploying products and services in the PCS, NII or millimeter wave unlicensed bands, caution is advised until more is known.

APPENDIX

Theorem 1: If \( T_{1} = T_{2} = 0 \), then \( H_{1}^{\ast} = \frac{2M\rho_{1}}{1-\rho_{1}+\rho_{2}} \) and \( H_{2}^{\ast} = \frac{2M\rho_{2}}{1-\rho_{1}+\rho_{2}} \)

Proof: When \( T_{1} = T_{2} = 0 \), \( H_{1}^{\ast} = X_{1}^{\ast} \) and \( H_{2}^{\ast} = X_{2}^{\ast} \). Device 1 transmits unfininshed work \( \rho_{1}(2M+H_{1}^{\ast}+H_{2}^{\ast}) \) in time \( H_{1}^{\ast} \),

Thus, \( H_{1}^{\ast} = \rho_{2}(2M+H_{1}^{\ast}+H_{2}^{\ast}) \) and \( H_{2}^{\ast} = \rho_{1}(2M+H_{1}^{\ast}+H_{2}^{\ast}) \) from symmetry. Thus, \( H_{1}^{\ast} = \frac{2M\rho_{1}}{1-\rho_{1}+\rho_{2}} \) and \( H_{2}^{\ast} = \frac{2M\rho_{2}}{1-\rho_{1}+\rho_{2}} \) \( (1) \)

Theorem 2: \( r_{1}(T_{1}) = \max(T_{1},2M)(1-\rho_{1})/\rho_{1} \) if \( r_{2}(T_{2}) = \max(T_{2},2M)(1-\rho_{2})/\rho_{2} \)

Proof: Device 1 transmits its total unfinished work \( \rho_{1}(2M+H_{2}^{\ast}+X_{1}) \) in period \( X_{1}^{\ast} \), so \( X_{1}^{\ast} = \frac{2M+H_{1}^{\ast}+X_{1}}{1-\rho_{1}} \). Similarly, \( X_{2}^{\ast} = \frac{2M+H_{1}^{\ast}+X_{1}}{1-\rho_{2}} \) \( (2) \)

Device 1 has average unfinished work \( 0.5\rho_{1}(2M+H_{2}^{\ast}) \) for fraction of time \( 2M+H_{2}^{\ast}+X_{1}/2M+H_{1}^{\ast} \), so its average unfinished work divided by \( \rho_{1} \) gives its delay \( D_{1} = 0.5(2M+H_{1}^{\ast})+X_{1}/2M+H_{1}^{\ast} \) \( (3) \)

Case 1: \( T_{1} > T_{2} \): As \( H_{2} = T_{2} \), \( \frac{\partial D_{1}}{\partial H_{1}} = 0.5(2M+H_{1}^{\ast})(2M+T_{1}+H_{1}^{\ast})^{2} < 0 \)

Thus, while \( T_{1} > T_{2} \), Device 1 would choose to increase \( H_{1} \).

Case 2: \( T_{2} \leq T_{1} \): As \( H_{1} = T_{1} \), \( D_{1} = 0.5(2M+X_{1})+\frac{2M+X_{1}^{*}}{2M+X_{1}} \)

\( \frac{\partial D_{1}}{\partial H_{1}} = \frac{0.5(2M+H_{1}^{\ast})(\rho_{1}(H_{1}^{\ast}+2M)-2M(1-\rho_{1}))}{(1-\rho_{1})(1-\rho_{2})(2M+H_{1}^{\ast})^{2}} > 0 \) if and only if \( \rho_{2}(H_{1}^{\ast}+2M) > 2M(1-\rho_{2}) \), i.e., \( X_{1}^{\ast} = \frac{2M+H_{1}^{\ast}+X_{1}}{1-\rho_{2}} \)

Thus, Device 1 would increase \( H_{1} \) and Device 2 would choose to increase \( H_{2} \).

Considering both cases, Device 1 would choose \( T_{1} = H_{1} \) at which \( X_{1}^{\ast} = \max(T_{1},2M) \) if possible. (\( H_{1} \) may be larger even if Device 1 is nongreedy.) From Equation 2, we get \( T_{1} = \max(T_{1},2M)(1-\rho_{2})/\rho_{2} \) \( (4) \)

By symmetry, \( T_{2} = \max(T_{2},2M)(1-\rho_{2})/\rho_{2} \) \( (5) \)

Theorem 3: If \( T_{2} < H_{2}^{\ast} \) and \( H_{2}^{\ast} < 2M \) then \( r_{1}(T_{1}) > H_{1}^{\ast} \)

If \( T_{2} < H_{2}^{\ast} \) and \( H_{2}^{\ast} < 2M \) then \( r_{1}(T_{1}) > H_{1}^{\ast} \)

Proof: \( r_{1}(T_{1}) = \max(T_{1},2M)(1-\rho_{1})/\rho_{1} - 2M \). As \( H_{1}^{\ast} = \frac{2M\rho_{1}}{1-\rho_{1}+\rho_{2}} \), \( 2M\rho_{1}/(1-\rho_{1}+\rho_{2}) < 2M \), subtracting \( 2M \), we get \( 2M\rho_{1}/(1-\rho_{1}+\rho_{2}) < 2M(1-2\rho_{2})/\rho_{1} \), i.e., \( r_{2}(T_{2}) > H_{2}^{\ast} \) \( (6) \)

By symmetry, \( r_{1}(T_{1}) > H_{1}^{\ast} \) if \( T_{2} < H_{2}^{\ast} \) and \( H_{2}^{\ast} < 2M \) \( (7) \)

Theorem 4: If \( T_{2} > H_{2}^{\ast} \) then \( \lim_{i \to \infty} t_{i}^{(1)} = \infty \) and \( \lim_{i \to \infty} t_{i}^{(2)} = \infty \)
Proof: Let $S_2^{(0)}=T_2^{(0)}$. We define $f_2(S_2)=S_2(1-\rho_2)/\rho_2-2M$ and $f_1(S_2)=S_2(1-\rho_1)/\rho_1-2M$. $f_2$ and $f_1$ are monotonically increasing. Let $S_2^{(i)}=f_2(S_2^{(i-1)})$ and $S_2^{(i)}=f_2(S_2^{(i-1)})$. Let $T_2^{(i)}=\eta(T_2^{(i-1)})$ and $T_2^{(i)}=T_2^{(i-1)}$. By Theorem 2, $\eta(T_2)=\max(f_1(T_2),2M(1-\rho_1)/\rho_2-2M)\geq f_2(T_1)$ (6) and $\eta(T_1)=\max(2f_2(T_1),2M(1-\rho_1)/\rho_2-2M)\geq f_2(T_1)$ (7). We will now prove three propositions, to prove Theorem 4.

Proposition 1: $T_2^{(i)} \geq S_2^{(i)} \forall i \geq 0$

Proof by induction Basis: For $i=0$, $T_2^{(0)} \geq S_2^{(0)}$ holds as $T_2^{(0)}=S_2^{(0)}$.

Inductive step: If $T_2^{(i)} \geq S_2^{(i)}$, we show $T_2^{(i+1)} \geq S_2^{(i+1)}$. As $T_2^{(i+1)}=\eta(T_2^{(i)}) \geq f_2(\eta(T_2^{(i)}))$, and $\eta(T_2^{(i)}) \geq f_2(\eta(T_2^{(i)}))$, we get $f_1(\eta(T_2^{(i)})) \geq f_2(\eta(T_2^{(i)}))$. Thus $T_2^{(i+1)} \geq S_2^{(i+1)}$. As $T_2^{(i+1)} \geq S_2^{(i+1)} \geq f_2(S_2^{(i+1)})$. Thus we get $T_2^{(i)} \geq f_1(S_2^{(i)}) \geq f_2(S_2^{(i)})$. As $T_2^{(i)} \geq S_2^{(i+1)}$, we have $T_2^{(i)} \geq S_2^{(i+1)}$ and $T_2^{(i+1)} \geq S_2^{(i+1)}$.

Proposition 2: $S_2^{(i)}=(1+(1-\rho_1-\rho_2)/\rho_1\rho_2)(S_2^{(0)}-H_2^*)+H_2^* \forall i \geq 0$

Proof by induction Basis: For $i=0$, proposition is true as $S_2^{(0)}=S_2^{(0)}$. Inductive step: If proposition holds for $i$, we show it holds for $i+1$. If $S_2^{(i)}=(1+(1-\rho_1-\rho_2)/\rho_1\rho_2)(S_2^{(0)}-H_2^*)+H_2^*$ for some $i$, then

$S_2^{(i+1)}=f_1(S_2^{(i)})=(1-\rho_1)(1-\rho_2)S_2^{(i)}/\rho_1-2M/\rho_2$

$S_1^{(i+1)}=(1+(1-\rho_1-\rho_2)/\rho_1\rho_2)(S_2^{(0)}-H_2^*)+H_2^*$

Proposition 3: If $S_2^{(0)}-H_2^*$ then $\lim_{i \to \infty} S_2^{(i)}=\infty$ and $\lim_{i \to \infty} S_1^{(i)}=\infty$

Proof: $S_2^{(i)}=(1-\rho_2-\rho_2)/\rho_2-2M$, so if $\lim_{i \to \infty} S_2^{(i)}=\infty$, then $S_2^{(i)}=\infty$ as $\rho_1+\rho_2<1$.

We now prove Theorem 4. From Propositions 1, we have $S_2^{(i)}=\infty$ if $S_2^{(0)}=H_2^*$. From Proposition 3, we have $T_2^{(i)} \geq S_2^{(i)} \forall i \geq 0$. Thus, $\lim_{i \to \infty} T_2^{(i)}=\infty$ if $S_2^{(0)}=H_2^*$. As $T_1^{(i)}=\max(2M(1-\rho_2)/\rho_2-2M, \lim_{i \to \infty} T_1^{(i)})=\infty$ if $\lim_{i \to \infty} T_1^{(i)}=\infty$.

As $T_2^{(0)} \geq S_2^{(0)}$, then $\lim_{i \to \infty} T_2^{(i)}=\infty$ and $\lim_{i \to \infty} T_1^{(i)}=\infty$.

Theorem 5: \[
\frac{\partial D_1}{\partial H_2}(0, 0) = 0
\]

Proof: $D_1=0.5(2M+H_2)/(2M+H_2+X_1)/(2M+H_2+H_2)$ from Equation 3, where $X_1=\rho_1(2M+H_2)/(1-\rho_1)$. If $T_1 \leq X_1$ then

$D_1=0.5(2M+H_2)$ and $\frac{\partial D_1}{\partial H_2}=0.5$. If $T_1 > X_1$ then

$D_1=0.5(2M+H_2)^2/(1-\rho_1)\cdot (2M+H_2+X_1)/(2M+2T_1) \cdot (1-\rho_1)(2M+H_2+T_1) > 0$

By symmetry, the same is true for $\frac{\partial D_2}{\partial H_2}$.