Etiquette Modification for Unlicensed Spectrum: Approach and Impact[1]

Durga P. Satapathy[2] and Jon M. Peha[3]
Department of Electrical and Computer Engineering, Carnegie Mellon University

ABSTRACT

In *unlicensed spectrum*, any device is free to transmit without a government license that implies exclusive access. Such spectrum has significant benefits, but serious challenges must first be overcome. Foremost is the risk of drastic performance degradation and inefficient spectrum utilization, due to a lack of incentive to conserve shared spectrum. Previous work [8] has shown this problem to be a real possibility. This paper demonstrates that the solution lies in proper regulation of access to unlicensed spectrum and its usage. We present a choice of potential solutions that vary in the degree to which they solve the problem, and in their impact on performance.

I. INTRODUCTION

In *unlicensed spectrum*, any device is free to transmit without a government license that implies exclusive access. The Industry, Science and Medicine (ISM) bands have long been unlicensed, although most spectrum has traditionally been licensed [1]. The Federal Communications Commission (FCC) has greatly increased unlicensed allocations, which include the 30 MHz Unlicensed Personal Communication Services (UPCS) band [2], the 350 MHz National Information Infrastructure (NII) band [3], and the 59-64 GHz Millimeter Wave band [4,5]. The UPCS band is governed by a *spectrum etiquette* (known as the UPCS etiquette) [2,6], which is a set of rules regulating access to spectrum and its usage. Other bands have power and emission limits, but no spectrum etiquette. Unlicensed spectrum offers many benefits. It facilitates mobility of wireless applications. For example, for a wireless Local Area Network used by a sales team for demonstration, unlicensed operation is more practical than getting licenses for each client's location. Also, unlicensed bands promote spectrum sharing (as any device can transmit while others are idle,) which reduces trunking inefficiencies [7]. Unlicensed spectrum also facilitates experimentation and innovation, as it is readily accessible. Three challenges must be overcome to realize these benefits. First, there may be mutual interference, as devices can transmit at will. Second, applications using unlicensed bands may vary greatly, making it difficult to enforce efficient utilization for all applications. Third, and most difficult, there is little inherent incentive for devices to conserve shared spectrum. Thus, a device may overuse shared spectrum to improve its own performance, even if performance degrades for other devices. If this is common, the shared resource will be of little use. This phenomenon has been referred to as a *Tragedy of the Commons.* Previous work [8] has shown this problem can occur in unlicensed spectrum, even with the UPCS etiquette.

This paper demonstrates that an appropriate etiquette can avoid a tragedy of the commons, at the cost of reduced performance (even for isolated devices that do not face the problem.) We present a choice of solutions that trade off the risk of a tragedy of the commons versus reduced performance. Section 2 describes the scenario we use to compare the above tradeoffs. Section 3 demonstrates the risk of a tragedy of the commons in the UPCS bands. Sections 4 and 5 each present a potential solution, and illustrate its ability in avoiding a tragedy of the commons. Section 6 compares the resulting performance reduction. Section 7 presents our conclusions.

II. THE SCENARIO

In this section we describe the scenario we use to evaluate the potential of etiquette modifications to avoid a tragedy of the commons. How can this problem be avoided? All system designs involve tradeoffs between competing goals and interests. While conserving licensed spectrum is an important design goal, there is little incentive to conserve shared spectrum. Thus, in unlicensed spectrum, it is more likely that designers will adopt a *greedy* approach, where the more a device wastes shared spectrum in favor of its own goals, the more it is greedy. The amount of resources consumed with a transmission depend on the transmission duration, bandwidth, and coverage area (which is a function of transmission power.) Thus, greedy devices may have greater transmission duration, power, or bandwidth than necessary. None of the current approaches in unlicensed bands can deter greed. All unlicensed bands limit transmission power, which helps to reduce mutual interference, but not greed. The ISM bands additionally require spread-spectrum transmissions, which cannot deter greed either. The UPCS bands impose a "Listen Before Talk" (LBT) rule which requires devices to monitor the channel and transmit only if the signal energy is below a threshold throughout a specified monitoring time. While this helps to reduce interference, it does not avoid greed. The UPCS bands also limit transmission time and bandwidth, but this approach of setting upper limits on spectrum resources is ineffective, as devices may still be greedy up to the limit. Finding a limit optimal for all applications is also difficult. Furthermore, upper limits may result in spectral inefficiency, as devices are

---

[1] This material is based upon work supported in part by Bellcore, and by the Interstate Electronic Corporation.

[2] Doctoral Candidate; Carnegie Mellon University; Email: d.p.satapathy@ieee.org; www.contrib.andrew.cmu.edu/~dsaq/home.html

[3] Associate Professor, Carnegie Mellon University; Email: peha@ece.cmu.edu; http://www.ece.cmu.edu/afs/ece/usr/peha/peha.html

constrained even without contention for spectrum. To deter greed, we must penalize greedy devices. One option is to adjust parameters that strongly affect access to spectrum. For example, altering either the monitoring time or the LBT power threshold would affect a device's priority in accessing spectrum, and therefore, its performance. To be effective, the penalty must be proportional to the spectrum resources consumed. Thus, a greedy device would have a higher penalty than a nongreedy device. In this paper, we focus on avoiding greed in the transmission duration only, and develop *penalty functions* based on the monitoring time parameter to determine the penalty on devices. Specifically, we propose the following etiquette modification: We require a device to find the channel idle for a total period of at least its *penalty time* before it may contend for transmission, where this penalty time is an increasing function of the duration the channel was last held by the device. The penalty time is not required to be contiguous; it may be split into nonconsecutive intervals.

We now describe the specifics of our scenario. We consider n devices sharing a wireless channel, sufficiently close together to receive each other's transmissions. All devices require the same bandwidth. Thus, transmission power and bandwidth have no impact on device performance. We assume that bursts awaiting transmission are queued in an infinite buffer. Nongreedy devices release the channel immediately after all queued bursts are sent. Greedy devices continue to hold the channel after all queued bursts are sent, just to avoid the access delay for the subsequent bursts. Device $i : i \in \{1,2,3..n\}$ has load ρ_i. We assume $\rho_i > 0 \ \forall i$, and $\sum_{i=1}^{n} \rho_i < 1$, where total capacity is defined to be 1. Device i holds the channel for period H_i. It has greed T_i, which is its minimum holding time, so $H_i \geq T_i$. After holding for period T_i, Device i releases the channel immediately if its queue is empty; otherwise, it continues to transmit until its queue becomes empty. For period $X_i : X_i \leq H_i$, Device i has data queued and transmits 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. Let Device i have penalty P_i (which is a function of H_i.) After Device i pays its penalty, it must find the channel idle throughout the monitoring period M before it may transmit. We assume devices monitor with persistence, i.e., devices continuously sense the channel [9]. To simplify analysis, we do not consider the back-off provision in the UPCS etiquette, which requires devices to defer from accessing the channel for a random duration whenever the channel is found busy.

To make analysis tractable, we use a fluid flow model [10,11]. In this model, the amount of data received by Device i in any period τ is exactly $\rho_i \tau$, where ρ_i is the load of Device i. In practice, arrival rates may fluctuate somewhat. Whenever the delay caused by waiting for the other device to release the channel is greater than the delay caused by arrival rate fluctuations, the error percentage would be small, making this analysis even more accurate. To look at scenarios not addressed by a fluid flow model, we also consider traffic with Poisson arrival of bursts. As burst size becomes small, this approaches fluid flow. We to address the impact of varying burst sizes via simulation. The isochronous and asynchronous UPCS bands have different values of parameters such as the monitoring time. In our simulations, the parameters are typically consistent with the isochronous UPCS band.

III. THE TRAGEDY OF THE COMMONS

In this section we summarize earlier work [8,12] which demonstrated the risk of a tragedy of the commons in the UPCS bands. Consider the scenario described in Section 2. Greed always benefits an isolated device. When devices share spectrum, there is a cost involved. While a greedy device hoards the channel, the queue of bursts awaiting transmission at other devices grows. When the greedy device releases the channel, it may take much longer to reclaim it. It has been shown [8], for both fluid flow and bursty traffic, that while greed may benefit a device even with contention for spectrum, it always degrades performance of other devices, forcing them to also resort to greed, in order to regain their performance. Greed may reduce delay even if other devices are nongreedy. 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 T_j, Device j 's greed. With a fluid flow model for two devices contending for spectrum, this reaction function is $r_2(T_1) = \max\{T_1, 2M\}(1-\rho_1)/(\rho_1) - 2M$ [12]. Also, whenever $\rho_1 + \rho_2 + \min\{\rho_1, \rho_2\} < 1$, escalation of greed is inevitable, until both devices hold the channel as long as possible, and neither has adequate performance. We have observed this phenomenon for bursty traffic also [8].

IV. THE SQUARE ROOT PENALTY

In this section we develop a penalty function to avoid greed by imposing a penalty time requirement on devices. To prevent a tragedy of the commons, it is a necessary (but not sufficient) condition that a penalty function must deter greed when there is no contention for spectrum. Also, as this would prevent designers from adopting greedy strategies if contention is not anticipated, the first few generations of unlicensed devices are likely to be designed nongreedy, as long as the utilization of unlicensed bands remains low. Indeed, for low power devices meant to operate in isolation (e.g. indoor applications,) greed may never emerge with such a penalty function. By definition, any penalty function will always result in some performance loss for a solitary device, whether it avoids greed or not.

The optimal penalty function is one that prevents greed while imposing minimum penalty. The optimal penalty function for an isolated device is given by the following theorem for a fluid-flow model (See Appendix for proof.)

Theorem 1 : Of all penalty functions $P(H)$ that would discourage greed for a device in isolation, the smallest is $P(H)=\sqrt{M^2/4+MH}-M/2$ for all values of H (where $P(H)$ is the penalty time, H is the duration for which the channel was held by the device, and M is the monitoring period.)

We will henceforth refer to this as the *square root* penalty function. To see if it deters greed when two devices share spectrum, we employ simulation for the scenario described in Section 2. For traffic with mean burst transmission time 40 ms and devices at 0.1 load each, the square root penalty discourages greed for a device if the other is nongreedy, but not if the other device is greedy. The situation is worse for less bursty traffic. When devices are at 0.1 load with 0.5 ms mean burst transmission time, Device 1 becomes greedy even when Device 2 is nongreedy. The resulting behavior is shown by reaction functions in Figure 1. The solid line gives the reaction of Device 1 on the y-axis in response to Device 2's greed on the x-axis. The dashed line gives the reaction of Device 2 on the x-axis in response to Device 1's greed on the y-axis. If devices start with greed at point A, they will progress to points B, C, D, etc., greed increasing with each reaction. Clearly, the square root penalty cannot avoid a tragedy of the commons in this scenario.

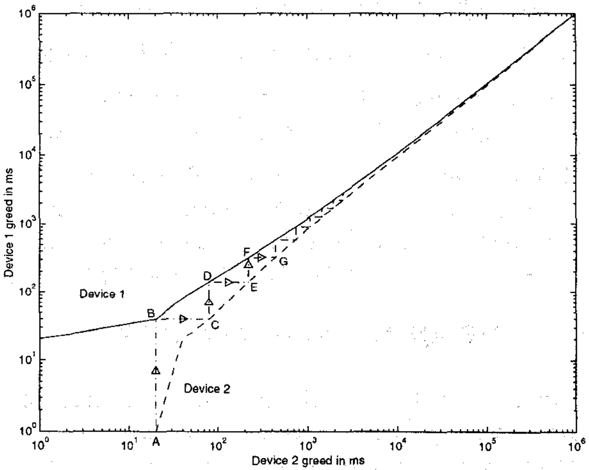

Fig. 1 Reaction Functions with square root penalty. Each device has 0.5 mean burst transmission time and 0.1 load.

V. THE LINEAR PENALTY

To avoid a tragedy of the commons in all circumstances, we clearly need a stronger penalty than the square root penalty. In this section, we show that one solution is $P(H)=H$, which we refer to as the *linear penalty*. We will evaluate its effectiveness in avoiding greed, first for a fluid flow model, and then for bursty traffic. We begin by considering a device in isolation. Theorem 1 shows, for a fluid flow model, that the square root penalty $P(H)=\sqrt{M^2/4+MH}-M/2$ deters greed for an isolated device. For $H=0$, $P(H)=0$ for both linear and square root penalty functions. As $dP(H)/dH$ for the linear penalty is greater than $dP(H)/dH$ for the square root penalty for all $H>0$, the linear penalty also avoids greed for isolated devices. We now address the case where there is contention for spectrum. Consider a fluid flow model for the scenario described in Section 2, where n devices compete for access to a wireless channel in unlicensed spectrum. Recall that for period X_i, Device i has data queued and transmits at the maximum rate possible, and then holds the channel further for duration I_i with an empty queue, transmitting data as it arrives. We define a cycle to begin whenever Device 1 completes transmission of its queued bursts. Cycle j begins with Device 1 holding the channel for period $I_1^{(j)}$, after which other devices may transmit several times while Device 1 pays its penalty $P_1^{(j)}=X_1^{(j-1)}+I_1^{(j)}$. Let the sum of the intervals for which the other devices hold the channel in cycle j be denoted by $K^{(j)}$. After Device 1 pays its penalty, it contends for the channel. As other devices may already have paid their penalty, Device 1 may lose contention many times before it can transmit. Let Device 1 find the channel idle for a total period $L_1^{(j)}$ after it pays its penalty and before it can transmit. As $L_1^{(j)}$ is Device 1's effective monitoring period, $L_1^{(j)}\geq M$. Device 1 transmits its total unfinished work $\rho_1(P_1^{(j)}+K^{(j)}+X_1^{(j)}+L_1^{(j)})$ in period $X_1^{(j)}$, so $X_1^{(j)}=\rho_1(P_1^{(j)}+K^{(j)}+X_1^{(j)}+L_1^{(j)})=(P_1^{(j)}+K^{(j)}+L_1^{(j)})\rho_1/(1-\rho_1)$. Device 1's average delay in the j^{th} cycle is its average unfinished work divided by its load ρ_1, and is given by $D_1^{(j)}=05(P_1^{(j)}+K^{(j)}+L_1^{(j)})^2/(1-\rho_1)(I_1^{(j)}+P_1^{(j)}+K^{(j)}+X_1^{(j)}+L_1^{(j)})$. $dD_1^{(j)}/dI_1^{(j)}\geq 0$ if and only if: $dP_1^{(j)}/dI_1^{(j)}+dK^{(j)}/dI_1^{(j)}+dL_1^{(j)}/dI_1^{(j)} \geq(P_1^{(j)}+K^{(j)}+L_1^{(j)})(1-\rho_1)/(P_1^{(j)}+K^{(j)}+L_1^{(j)}+2(1-\rho_1)I_1^{(j)})$. As $P_1^{(j)}=X_1^{(j-1)}+I_1^{(j)}$, and $X_1^{(j-1)}$ is independent of $I_1^{(j)}$, $dP_1^{(j)}/dI_1^{(j)}=1$. So $dD_1^{(j)}/dI_1^{(j)}\geq 0$ if and only if $dK^{(j)}/dI_1^{(j)}+dL_1^{(j)}/dI_1^{(j)}\geq-\rho_1(X_1^{(j)}+L_1^{(j)}+2I_1^{(j)}/(X_1^{(j)}+L_1^{(j)}+2\rho_1 I_1^{(j)})$. Thus, there is no incentive for greed if $dK^{(j)}/dI_1^{(j)}+dL_1^{(j)}/dI_1^{(j)}\geq 0$.

It is likely that $dK^{(j)}/dI_1^{(j)}\geq 0$, i.e., Device 1's increasing its greed would not result in other devices holding the channel for a lesser period. This is true when all other devices are nongreedy, for the following reasons. First, the number of bursts queued for transmission at the other devices will not decrease when they gain access. Second, other devices will have more chances to transmit because Device 1's penalty time will be longer. If other devices are greedy, then Device 1's holding longer may cause them to expect longer holding times from Device 1 in the future, thereby increasing their incentive to be greedy. As a result, the other greedy devices may or may not increase their greed, but there is no apparent reason for them to be less greedy.

In general, it is likely that $dL_1^{(j)}/dI_1^{(j)}=0$. If no other device has zero penalty when Device 1's penalty falls to zero, $L_1=M$ and $dL_1^{(j)}/dI_1^{(j)}=0$. Otherwise, Device 1 loses contention after monitoring for a period x, where $x<M$. If Device 1's increase in greed is much larger than M, the penalty increase would hardly alter L_1, i.e., $dL_1^{(j)}/dI_1^{(j)}=0$. For these cases, it is likely that $dK^{(j)}/dI_1^{(j)}+dL_1^{(j)}/dI_1^{(j)}\geq 0$, and the linear penalty would discourage greed. However, if Device 1 increases its greed by an amount $y<x$, it pays an extra penalty y and loses contention after a period $x-y$. Thus, $dL_1^{(j)}/dI_1^{(j)}=-1$ is possible for an increase in greed of order M (as $y<M$.) To discourage greed in this case, we need $dK^{(j)}/dI_1^{(j)}\geq 1-\rho_1(X_1^{(j)}+L_1^{(j)}+2I_1^{(j)})/(X_1^{(j)}+L_1^{(j)}+2\rho_1 I_1^{(j)})>0$, which is not necessarily true. Thus, the linear penalty cannot discourage all greed, but it can prevent greed from escalating much beyond the monitoring period.

Indeed, our simulations confirm these results. For bursty traffic with a mean burst transmission time of 40 ms, the linear penalty deters greed for a device when the other device is nongreedy. For less bursty traffic with 0.5 ms mean burst transmission time, greed cannot be avoided but it reaches equilibrium at 10 ms. Reaction functions in Fig. 2 show that the response to greed much higher than 10 ms is greed less by orders of magnitude. Thus, the linear penalty is sufficient to avoid a tragedy of the commons.

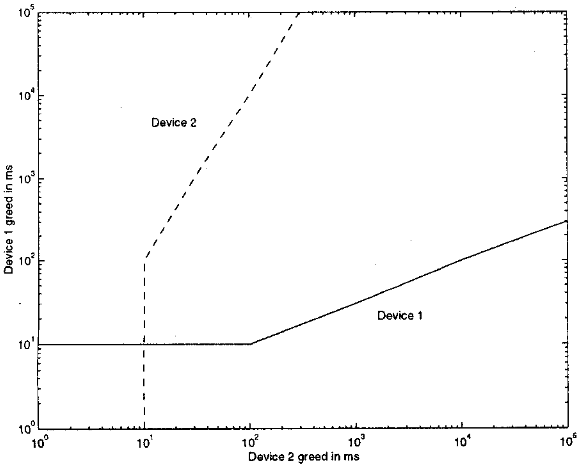

Fig. 2 Reaction Functions with linear penalty. Each device has 0.5 ms mean burst transmission time and 0.1 load.

VI. PERFORMANCE COMPARISON

In this section, we address performance implications of implementing penalty functions for an isolated device. The least delay occurs when there is no etiquette, as there would be no LBT rule. With an etiquette, the least delay occurs when there is no penalty. Delays are higher with the square root penalty, and even more so with the linear penalty, as Figures 3(a) and 3(b) show, for traffic with mean burst transmission times 0.5 ms and 40 ms respectively. We see that the delay with the linear penalty is not much higher than with no penalty, for low loads (less than 0.35.) The performance degradation with square root penalty versus no penalty is negligible for bursty traffic, but not otherwise. We also see that the linear penalty restricts the throughput to 0.5, whereas it can approach unity with the square root penalty, as with no penalty and with no etiquette.

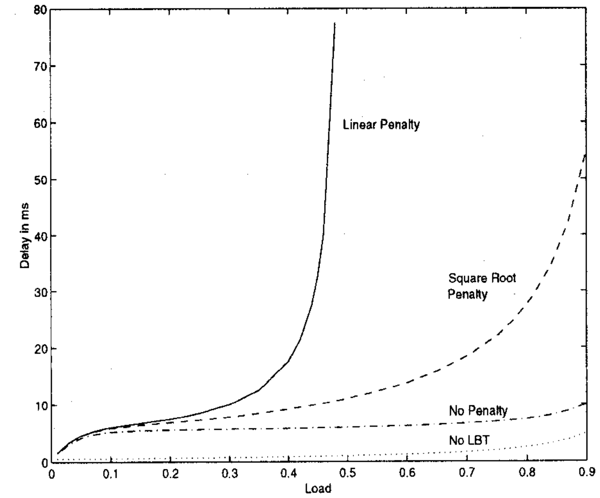

Fig. 3(a) Delay with 0.5 ms mean burst transmission time.

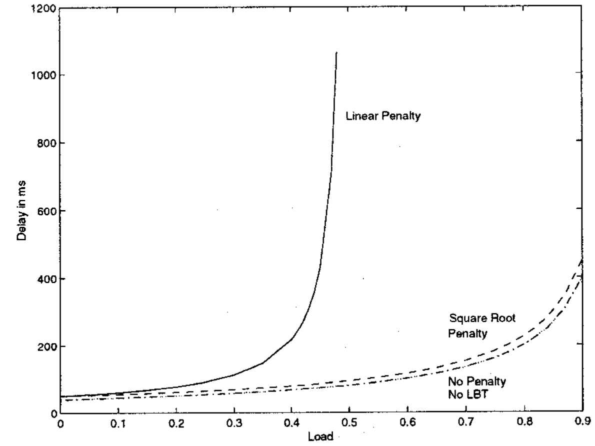

Fig. 3(b) Delay with 40 ms mean burst transmission time.

VII. CONCLUSION

Unlicensed spectrum has many benefits. It supports mobility of wireless applications, allows spectrum sharing, and facilitates experimentation and innovation. However, a device may overuse shared spectrum to improve its performance. Such greed is always beneficial for isolated devices. When devices contend for spectrum, greed can also lead to inadequate performance and inefficient spectrum utilization [8,12]. This paper shows that the problem can be solved by modifying the etiquette to include penalty functions that penalize greedy devices. Different penalty functions avoid greed to different degrees. At one extreme, a linear penalty function that equals the duration the channel was last held prevents escalation of greed in all scenarios, but limits throughput to 0.5. On the other, the UPCS etiquette (with no penalty function) fails to avoid escalation of greed, but allows relatively better throughput. A compromise is offered by a penalty function of order square root of the period the channel was last held. It deters greed for isolated devices with minimum penalty, and allows high throughput. However, it cannot always deter greed when spectrum is shared.

In general, the optimal penalty function is one that limits both the risk of a tragedy of the commons and the performance loss caused by the penalty to acceptable levels, where these levels vary with the applications and the frequency band. An etiquette with no penalty function is best if isolated operation is guaranteed. The square root penalty is best if contention will always be sufficiently rare such that devices will be designed for solitary use, e.g. for indoor applications, and for frequency bands that severely limit propagation distance. Many proponents of the NII bands and the Millimeter Wave bands described the intended use of the bands as meeting these criteria. However, this description was typically part of an argument why no etiquette is needed whatsoever, which would lead to a tragedy of the commons. The linear penalty is best if contention is a reasonable possibility, e.g. if coverage areas are not severely limited. Thus, when designing an effective etiquette, there may be tradeoffs between imposing penalties and imposing power limits.

Meanwhile, the FCC has increased allocations for unlicensed spectrum. If etiquette modifications are not effected soon, devices meant to operate in isolation may be designed greedy, and the risk of a tragedy of the commons will increase once these products are marketed. It would be prudent for wireless companies to demand etiquette modifications before their competitors market greedy devices, even if such modifications involve additional complexity or performance limitations in the short term. Consequently, before industry invests significantly in unlicensed bands, we suggest that industry actively participate in bringing about etiquette modifications.

APPENDIX

Proof: Consider a fluid flow model, with device load ρ. After holding the channel for period H, the device pays its penalty $P(H)$, and gets the channel back after duration M. It transmits its total unfinished work $\rho(P(H)+M+X)$ in period X. Thus, $X=\rho(P(H)+M+X)$. The average delay is

$$D=05(P(H)+M)\frac{P(H)+M+X}{P(H)+M+H}=\frac{05}{1-\rho}\frac{(P(H)+M)^2}{P(H)+M+H}.$$

$dD/dH\geq 0$ if and only if $dP(H)/dH\geq(P(H)+M)/(P(H)+M+2H)$. $P(H)$ is minimized if $P(0)$ is minimized, and $dP(H)/dH$ is minimized for all $H\geq 0$. By definition of penalty, $P(H)\geq 0$. So we set $P(0)=0$ and $dP(H)/dH=(P(H)+M)/(P(H)+M+2H)$. The solution is $P(H)=\sqrt{M^2/4+MH}-M/2$.

REFERENCES

[1] J. M. Peha, "Spectrum Management Policy Options," accepted for publication in *IEEE Communication Surveys*, March 1998, http://www.comsoc.org/pubs/surveys/.

[2] FCC, "Amendment of the Commission's Rules to Establish New Personal Communications Services," Memorandum Opinion and Order, FCC 94-144, Gen. Docket No. 90-314, June 9, 1994.

[3] FCC, "Amendment of the Commission's Rules to Provide for Operation of Unlicensed NII Devices in the 5 GHz Frequency Range," Report & Order, FCC 97-005, Gen. Docket No. 96-102, Jan. 9, 1997.

[4] FCC, "Amendment of Parts 2, 15, and 97 of the Commission's Rules to Permit Use of Radio Frequencies Above 40 GHz for New Radio Applications," First report and Order and Second Notice of Proposed Rule Making, FCC 95-499, Gen. Docket No. 94-124, Dec. 15, 1995.

[5] M. J. Marcus, "Recent Progress in U.S. Millimeter Wave Spectrum Management Policy," *IEEE MTT-S International Microwave Symposium Digest*, vol. 2, pp. 505-7, June 1996.

[6] D. G. Steer, "Wireless Operation in the Unlicensed Band," *Proc. Intl. Conf. on Universal Personal Communications*, Sept. 1994, pp. 586-90.

[7] H. Salgado-Galicia, M. Sirbu, and J. M. Peha, "A Narrow Band Approach to Efficient PCS Spectrum Sharing Through Decentralized DCA Access Policies," *IEEE Personal Communications*, vol. 4, no. 1, Feb. 1997, pp. 24-34.

[8] D. P. Satapathy and J. M. Peha, "Spectrum Sharing Without Licenses: Opportunities and Dangers," in "Interconnection and the Internet," book of Selected Papers From The 1996 Telecommunications Policy Research Conference, G. Rosston and D. Waterman (Eds.), Mahwah, NJ: Lawrence Erlbaum Associates,1997, pp. 49-75.

[9] I. Vukovic, and J. McKown, "Spectrum Sharing Under The Asynchronous UPCS Etiquette: The Performance of Collocated Systems Under Heavy Load," *Proceedings of ACM Mobicom*, Nov.1996, pp. 67-72.

[10] J. M. Peha, "Evaluating Scheduling in Integrated-Services Networks using a Semi-Fluid Flow Model," accepted to appear in *Telecommunication Systems: Modeling, Analysis,Design, and Management.*

[11] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic Theory of a Data-Handling System with Multiple Sources," *Bell Sys. Tech. Journal*, vol. 61, no. 8, Oct. 1982, pp. 1871-94.

[12] D. P. Satapathy and J. M. Peha, "Performance of Unlicensed Devices With a Spectrum Etiquette," *Proceedings of IEEE Globecom*, Nov. 1997, pp. 414-418.