NTIA Special Publication SP-04-409

Proceedings of the International Symposium on Advanced Radio Technologies March 2–4, 2004

J. Wayde Allen, General Chair
Timothy X Brown, Program Co-Chair
Douglas C. Sicker, Program Co-Chair
Jeanne Ratzloff, Publications

U.S. DEPARTMENT OF COMMERCE
Donald L. Evans, Secretary

Michael D. Gallagher, Acting Assistant Secretary
for Communications and Information

March 2004

DISCLAIMER

Certain commercial equipment, components, and software are identified to adequately present the underlying premises herein. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the equipment, components, or software identified is the best available for the particular applications or uses.

INTERNATIONAL SYMPOSIUM
ON
ADVANCED RADIO TECHNOLOGIES

March 2 - 4, 2004 • Boulder, Colorado

Sponsored by:

Institute for Telecommunication Sciences (ITS),
National Telecommunications and Information Adm. (NTIA)

National Institute of Standards and Technology (NIST)

University of Colorado, Department of Interdisciplinary Telecommunications

U.S. Department of Commerce, Boulder Laboratories

2004 ISART Technical Committee:

Timothy X Brown, Program Co-Chair, University of Colorado

Douglas C. Sicker, Program Co-Chair, University of Colorado

Wayde Allen, General Chair, NTIA/ITS

Eldon Haakinson, NTIA/ITS

Dale Hatfield, University of Colorado

Nobuo Ikeda, Research Institute of Economy, Trade and Industry, Japan

Stanley Jedrus, University of Pittsburgh

William Lehr, MIT

Roger Marks, NIST

Preston Marshall, DARPA

Robert Matheson, NTIA/ITS

Durga P. Satapathy, Global Markets Group Strategy, Sprint

PREFACE

This marks the second year that we have published an ISART Proceedings. I am particularly proud of this accomplishment since I believe that the Proceedings provides the radio community with a valuable resource. Not only is this a peer reviewed publication including ideas that span a number of related disciplines, it also gives the speakers a place to provide more detail about their theoretical concepts, data, and/or methodology. Additionally, the call-for-papers process used to develop the Proceedings helps the Technical Program Committee by giving presenters a way of showing through their submissions what they think is important. Without this mechanism, the
technical program would consist exclusively of invited papers. The call-for-papers process allows for the development of technical sessions that might not otherwise have been considered. This is important for a conference like ISART which is intended to help forecast the development of the radio art.

There are a few changes this year. The conference has been scaled back to three days rather than four. We also decided to open the conference with a two hour tutorial session to provide some detailed background in these areas:

- Mobile Geolocation
- Modern Spectrum Management Alternatives,
- SAFECOM: The State of Public Safety Communications.

This should provide a solid foundation of knowledge upon which the rest of the speakers can build.

Finally, while this year's underlying theme is Geolocation and Location-based Services, we have developed the technical program around what we believe to be a holistic view of the radio art. ISART strives to bring together a diverse collection of people from academia, business, and government to discuss issues related to radio technology in a common forum. The goal is to reach beyond plain technical know-how to the technical, business, and regulatory forces that shape the radio landscape today to help us understand where the technology will be tomorrow.

J. Wayde Allen, NTIA/ITS
ISART General Chair

Table of Contents

Page

Accuracy Enhancements for TDOA Estimation on Highly Resource Constrained Mobile Platforms

Kumar Gaurav Chhokra[1], Theodore Bapty[1], Jason Scott[1], Mitch Wilkes[2]
[1]Institute for Software Integrated Systems, Vanderbilt University
{kumar, bapty, jscott}@isis-server.isis.vanderbilt.edu
Tel: (615) 343 7567, Fax: (615) 343 7440

[2]Electrical Engineering and Computer Science, Vanderbilt University
{mitch.wilkes@vanderbilt.edu}

Abstract.

Over the past few years, there has been an immense thrust in geolocation, surveillance, tracking and location-aware systems and services. The advent of compact, low power, high processing-power DSPs have made possible several tasks which were infeasible just a few years ago. TDOA estimates on such energy and resource constrained platforms suffer from the lack of a coherent sampling clock. We present two related techniques of Doppler-frequency and time shift correction for such platforms. The techniques are formally developed, analyzed, and then compared from an implementation and performance perspective.

1. Introduction

Over the past few years, there has been an immense thrust in the fields of geolocation, surveillance, tracking and location-aware systems and services. The advent of compact, low power, high processing-power digital signal processors have made possible several tasks which were considered infeasible just a few years ago. The applications abound in utilitarian (E911, product-customized services, in-building tracking, etc.) and military sectors (surveillance, un-manned reconnaissance, target location etc.) [1-4].

Traditional TDOA / TOA systems used for geolocation depend heavily on fixed array based processing and typically employ high-precision, high-cost components. Such systems also have long stand off ranges, making it difficult to closely track and follow the movements of targets emitting low power signals in congested areas [5]. Proximity to target and mobility of the tracking devices gather greater relevance in urban environments.

Unmanned aerial vehicles (UAVs) and organic aerial vehicles (OAVs) equipped with geolocation electronics can counter the problems with long-standoff systems. They can afford safety, security and stealth for the reconnaissance mission. However, they pose interesting technical constraints. Robustness and reliability under military use conditions demand that the devices be compact, rugged and mechanically stable. Stealth in reconnaissance implies a small, silent design (i.e. limited communication bandwidth) while, endurance and usability demand an extensive operating life (i.e. low power). These UAVs are imagined as a fleet of dispensable mobile devices, implying a relatively strong constraint on the manufacturing and operating cost.

These real-world constraints manifest themselves in the algorithmic and signal processing domains as well. Physically distinct, passive RF and mobile units imply the lack of a common clock, thus obviating array based processing. Stealth prevents the UAVs from synchronizing on a regular basis to maintain a common clock, complicating accurate TOA estimates. Also, a reduced communication bandwidth between the devices negates the usual techniques of correlating received signals for obtaining TDOA estimates.

These constraints suggest a technique where much of the detection, identification and TOA are done locally and the actual target locus computation is relegated to a remote base station with bandwidth-constrained inputs from the individual UAV sensors. Specifically, each UAV must compute absolute TOA using *a-priori* information such as a signal template (known signal features, synchronization sequences, etc). TOA processing must then compensate for the relatively low precision GPS clocks and unique (possibly varying) sample rates at the sensors.

This paper[‡] describes two techniques for correcting the errors in TOA/TDOA estimates due to incongruent sampling frequencies of the received and template signals. Section 2 describes and mathematically models the effects of disparate sampling frequencies as a relative time companding (RTC) problem. Section 3 introduces the Doppler shift correction technique for narrow band signals. Section 4 explains the motivation for a time domain correction. Section 5 develops the time-shift correction technique. Section 6 compares and contrasts the two techniques with respect to

[‡] This work has been supported by DARPA under the WASP contract F30602-02-2-0206

performance characteristics and computational efficiency with field data.

2. Mathematical preliminaries

A signal emanating from a remote source and monitored in the presence of noise at two spatially separate sensors may be mathematically modeled as:

$$x_1(t) = f(t) + n_1(t) \quad (1a)$$

$$x_2(t) = mf(t + D) + n_2(t) \quad (1b)$$

where $f(t)$, $n_1(t)$ and $n_2(t)$ are real, jointly stationary random processes. The signal $f(t)$ is assumed to be uncorrelated with noise $n_1(t)$ and $n_2(t)$. D is the delay between the two received signals.

The traditional technique of detecting a signal of interest (template) buried in a data stream corrupted by additive Gaussian random noise is to use a matched filter or a generalized cross correlator (GCC) [6]. Since the GCC approach may be viewed as pre-filtering the two signals with whitening filters before a usual cross-correlation, we focus on the correlation technique for simplicity. The correlation function may be written as [7]:

$$Rx_1x_2(\tau) = \int_{-\infty}^{\infty} x_1(t)x_2(t+\tau)dt \quad (2)$$

The value of τ that maximizes (2) provides an estimate of the delay D.

2.1. Scaling due to disparate sampling frequencies

Since the signal processing components on the UAVs are subjected to a wide gamut of operating conditions, the operating clock and sampling frequencies are seldom the same as those in the test environments. In particular, the frequency at which the template was constructed is generally different from the frequency at which the incoming signal is sampled. [8] and [9] provide a good discussion of the effects of RTC on the cross-correlation operation. [10] extends the argument to quadratic delays.

Fig. 1 illustrates the effects of different sampling frequencies. The signal with a nominal Fourier transform $F(\omega)$ is sampled at two sampling frequencies, f_{s1} and f_{s2}, with $f_{s1} < fs2$. The sampled signal has different normalized bandwidths given by (3) where $\omega_{si} = 2\pi fs_i$.

$$\hat{\omega}_{b_1} = \frac{\omega_b}{\omega_{s_1}} \text{ and } \hat{\omega}_{b_2} = \frac{\omega_b}{\omega_{s_2}} \quad (3)$$

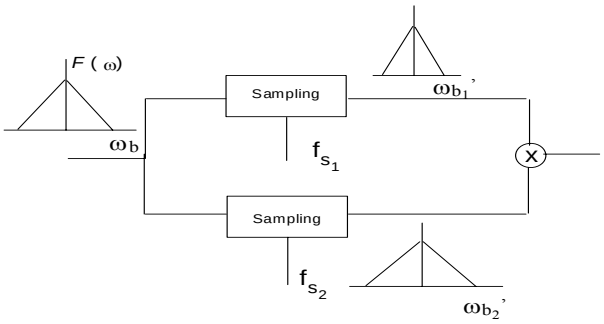


Figure 1 different sampling frequencies lead to frequency companding

Clearly, the higher sampling frequency produces a smaller normalized bandwidth, i.e., $\omega_{b2} < \omega_{b1}$. When viewed in the time domain, the differing relative bandwidths manifest themselves as time scaling artifacts.

3. Time scaling and Doppler shift

The disparate sampling frequencies of the template and the UAVs produce correlation artifacts, leading to erroneous time delay and TOA estimates. For relatively narrow-band signals and fairly similar sampling frequencies, it is now shown that the RTC between the template and received signal may be approximated by Doppler shifts. Weiss [11] gives a good explanation of the narrow band criterion. The signals under consideration fall well within the domain of narrow band representation.

Let the template be represented by $f(t)$ and the received signal by $g(t)$. Both signals are considered to be real and continuous as this eases analysis.

Let the received signal suffer both a scaling, given by the scale parameter s, and a time delay, τ:

$$g(t) = f(s(t-\tau)) \quad (4)$$

Taking the Fourier transform on both sides, we have,

$$G(\omega) = s^{-1}F(\omega/s)e^{-j\omega s\tau} \quad (5)$$

Let $s = 1 - a$, then

$$G(\omega) = 1/(1-a)F(\omega/(1-a))e^{-j\omega(1-a)\tau} \quad (6)$$

if $a << 1$, we have $\frac{1}{1-a} \approx 1 + a$ (7)

$$G(\omega) = (1+a)F(\omega(1+a))e^{-j\omega(1-a)\tau} \quad (8)$$

For narrow band signals,

$$F(\omega_0 + \delta\omega) = 0, |\delta\omega| > \frac{\omega_b}{2} \quad (9)$$

where ω_0 is the center radian frequency and ω_b is the bandwidth of the signal. Hence the above expression for $G(\omega)$ may be approximated by:

$$G(\omega) \approx (1+a)F(\omega + a\omega_0))e^{-j\omega\tau}e^{j\omega a\tau} \quad (10)$$

Noting that both a and τ are small, so that the second exponential maybe approximated by unity, we get,

$$G(\omega) \approx (1+a)F(\omega + a\omega_0))e^{-j\omega\tau} \quad (11)$$

Taking the inverse Fourier transform, we obtain,

$$g(t) \approx f(t-\tau)e^{-j\omega_d(t-\tau)} \quad (12)$$

where $\omega_d = a\omega_o = (1\text{-}s)\ \omega_o$ (13)

The above expression indicates that for small scale factors, a Doppler shift can approximate the effect of scaling.

3.1 Compensating for RTC with Doppler shifts

Observe that the form of the companded narrow band signal closely resembles the kernel of the cross-ambiguity function. Hence, the cross-ambiguity function is well suited to correct for the "shift" introduced due to relative companding. [8], [12] present several techniques of using cross-ambiguity functions (CAFs) for determining the relevant delay and scale parameters.

For our case, the ratio of the sampling frequencies of the two signals, *s,* is always known: the sampling frequency of the template is known *a priori*, and can in fact be accurately controlled through correct construction. The instantaneous sampling frequency for the data obtained at a sensor may be estimated using several introspection algorithms. Also, the center frequency of the transmitted signal may be known a priori by construction. For arbitrary signals, it may also be estimated as the mean frequency as defined in [12], [13] by

$$\omega_0 = \frac{\int_{-\infty}^{\infty} \omega F(\omega)\, d\omega}{\int_{-\infty}^{\infty} F(\omega)\, d\omega} \quad (14)$$

To estimate the TOA, a generalized cross correlator (GCC) may be used on the RTC compensated signals. As before, if $f(t)$ represents the template and $g(t)$ the received signal, then we create a new signal, $f_1(t)$ which is a frequency shifted version of the template to compensate for the current operating frequency of the sensor.

$$f_1(t) = f(t)e^{-j\omega_d t} \quad (15)$$

Signals $f_1(t)$ and $g(t)$ are then fed through matched filter to obtain an estimate for the time delay, τ.

$$\tau = \arg\max \left| \int_{-\infty}^{\infty} f_1(u) g^*(u+t)\, du \right| \quad (16)$$

$$\tau = \arg\max \left| \int_{-\infty}^{\infty} f(u) g^*(u+t) e^{-j\omega_d t}\, du \right| \quad (17)$$

The above expression may be viewed as a cross-ambiguity function evaluated at a given shift ω_d. This circumvents the need to compute several CAFs for differing values of scale and delay.

4. Implementation issues in Doppler RTC correction

The Doppler shift technique compensates for RTC by modulating the template signal with a complex exponential as shown in (15). Usually, the complex exponential is computed using Euler's expansion:

$$\exp(j\omega_d t) = \cos(\omega_d t) + j\sin(\omega_d t) \quad (18)$$

For a discrete time system, the above evaluation must be performed for each sampling instant. Given a template of length N, this produces an N x 1 vector of samples of the modulating complex exponential

$$\mathbf{e} = \cos(\omega_d \mathbf{t}) + j\sin(\omega_d \mathbf{t}) \quad (19)$$

where, $\mathbf{t}[k] = k/\omega_s$ is the vector of sampling instants, with $0 \le k \le N, k \in Z$ and sampling frequency ω_s. Computing the elements of this complex exponential in terms of trigonometric functions is both computationally and temporally intensive. Furthermore, once the vector has been computed, its application to the real template vector requires N complex multiplications. Examining the structure and nature of this calculation and applying a few engineering assumptions, we can significantly reduce this burden.

For values of s fairly close to 1, or equivalently, when a is fairly small and when the period over which the Doppler correction must be employed is limited (If it is large, then the time-bandwidth product condition for narrow band signals is violated, making the Doppler shift assumption invalid), the product $\omega_d t$ is fairly small. Thus, we can make the following approximations

$$\cos(\theta) \approx 1 \text{ and } \sin(\theta) \approx \theta \quad (20)$$

$$\exp(j\omega_d t) = 1 + j(\omega_d t) \quad (21)$$

Expressing this as a vector of discrete samples, we have

$$\mathbf{e}[k] = 1 + j\omega_d \mathbf{t}[k] \quad (22)$$

This computation represents enormous savings in creating the correction vector, **e**. The structure of each e[k] also enables us to apply the correction efficiently. The Doppler shifting of the template may be accomplished in a single pass, leading to an O(N) computation routine.

5. Alternative time shift based correction

The Doppler correction method compensates for relative companding by approximating the frequency scaling by a frequency shift: the modulation of the template sequence by a complex exponential shifts the spectrum of the template is to approximately line up with that of the scaled signal.

The Doppler correction has the advantage of offering a very intuitive and simple technique of compensating for the different sampling frequency between the nodes and the templates. While it also obviates the need to evaluate the passive ambiguity function for a large number of scales and shifts, it still suffers from a computational perspective.

As explained earlier, the sensor nodes are constrained in memory and computational power. In particular, the TI C67x DSPs used to provide the signal processing capabilities have a limited internal cache. When implementing computationally and temporally intensive operations such as a GCC, it is imperative that the memory access requirements be controlled to meet hard real-time constraints. Ensuring that the data and result vectors fit in the fast internal memory is a proven and routinely employed technique [14].

Both the Doppler correction and approximation techniques suffer in this regard because they convert normally real data vectors into complex vectors, doubling the initial memory requirement and causing performance to degrade. For e.g., an 8K point real-FFT takes a significantly lesser time than a complex-FFT of the same length.

The performance loss accrued over several iterations can mean the failure of a real-time constraint. Such a situation can occur in the computation of the cross spectral density (CSD) [15] of the received signal with the template.

Let us examine the relative companding problem again. As before, let

$$g(t) = f(s(t-\tau)) \quad (23)$$

where, $g(t)$, $f(t)$ are, respectively, the received and template signals, s is the scaling factor and τ is the time delay introduced.

The equations below recall the matched filtering operation.

$$R_{gf}(u) = \int_{-\infty}^{\infty} g(t) f(t+u) dt \quad (24)$$

$$\tau = \arg\max \left| \int_{-\infty}^{\infty} g(t) f(t+u) dt \right| \quad (25)$$

when $s = 1$, a value of $t = \tau$ maximizes R_{gf}. For the present case ($s \sim 1$, and narrow band signals), the value of t over a period (T_1, T_2) that maximizes R is given by,

$$\tau_m(T_1, T_2, t) = \arg\max \left| \int_{T_1}^{T_2} f(st - s\tau) f^*(t+u) dt \right| \quad (26a)$$

$$\tau_m(T_1, T_2, t) = u, \ni s(t-\tau) = (u+t) \quad (26b)$$

$$\therefore \tau_m(T_1, T_2, t) = (s-1)t - s\tau \quad (26c)$$

Notice that this delay $\tau_m(T_1, T_2, t)$ is not fixed as in the previous case, but is "smeared" in time.

We wish to approximate this time varying delay $\tau_m(T_1, T_2, t)$ by a constant $\tau_d(T_1, T_2)$ chosen according to a least squared error criterion. Let the sum of the squared errors, $\in(\tau)$, be given by

$$\in(\tau_d) = \int_{T_1}^{T_2} e^2(t) dt = \int_{T_1}^{T_2} \left(\tau_m(t) - \tau_d\right)^2 dt \quad (27)$$

where (T_1, T_2) is the interval over which the two sequences are compared. Using the Leibniz integral rule [16], we have

$$\frac{\partial}{\partial \tau_d} \in(\tau_d) = \int_{T_1}^{T_2} \frac{\partial}{\partial \tau_d} \left(\tau_m(t) - \tau_d\right)^2 dt + \left(\tau_m(T_1) - \tau_d\right)^2 \frac{\partial}{\partial \tau_d}(T_1) - \left(\tau_m(T_2) - \tau_d\right)^2 \frac{\partial}{\partial \tau_d}(T_2) \quad (28)$$

For τ_d^o, the desired optimal value, the LHS of (28) vanishes.

$$\therefore \tau_d{}^o = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \tau_m(t) dt = (s-1)\frac{T_2 - T_1}{2} - s\tau \quad (29)$$

Over a given interval [T_1, T_2], the effects of relative companding between $f(t)$ and $g(t)$ may be minimized in the least squares sense by constructing a time-shifted version of $f(t)$, denoted $\tilde{f}(t)$, as

$$\tilde{f}(t) = f(t + \tau_d{}^o) = f(t + (s-1)\frac{T_2 - T_1}{2} - s\tau) \quad (30)$$

In the discrete domain, the above equation may be expressed as

$$\tilde{f}[n] = \tilde{f}(\frac{2\pi n}{\omega_s}) = f(\frac{2\pi n}{\omega_s} + (s-1)\frac{T}{2} - s\tau), \quad (31)$$

where ω_s is the angular sampling frequency of f.

The analog signal $f(t)$ is usually not available to permit arbitrary time-shifting of the template signal. $\tilde{f}$ would then need to be evaluated by interpolating at "arbitrary" time instants. While this is achievable, the interpolation in the time domain usually comes with a computationally expensive price tag.

An easier alternative may be found by migrating to the frequency domain. Taking the Fourier transform of the preceding equation, we obtain

$$\tilde{F}(\omega) = F(\omega)e^{j\tau_d\omega} \quad (32)$$

where $F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt$ is the Fourier transform of $f(t)$. Also note that we have dropped the superscript of τ_d in favor of brevity. As with the Doppler correction, it can be argued that the nature of τ_d and ω allow the correction factor $e^{j\tau_d\omega}$ to be computed as an approximation

$$e^{j\tau_d\omega} = 1 + j\tau_d\omega \quad (33)$$

The time-shift based approach has the following notable advantages

1. Smaller memory foot print for $F(\omega)$ computation and reduced memory latency: For real signals, the Fourier transform exhibits Hermitian symmetry. This may be exploited to reduce the number of computations required to compute the entire FFT. It also means that the input vector to the FFT routine is a real-vector, requiring half as much memory as its complex counterpart, reducing memory access times.
2. Ability to precompute $F(\omega)$: since the template is known by design, $F(\omega)$ may be computed offline. This obviates its computation at run-time entirely. This introduces huge savings in computation time.
3. Facility to efficiently implement correction in the frequency domain: As with the Doppler shift correction, we can utilize the nature of the correction to devise an efficient method of applying the correction in the frequency domain. This has the added advantage of allowing an *in situ* computation, bringing further reduction in memory access latencies and computation time.

6. Experimental Results

The figure 2 shows the schematic setup on which these algorithms were implemented and tested. The sensor nodes consisted of an FRS receiver connected to the relevant electronics. A "global" clock was derived using GPS signals one pulse per second (PPS) signals. The GPS signals were filtered and processed to reduce the effects of any timing jitters.

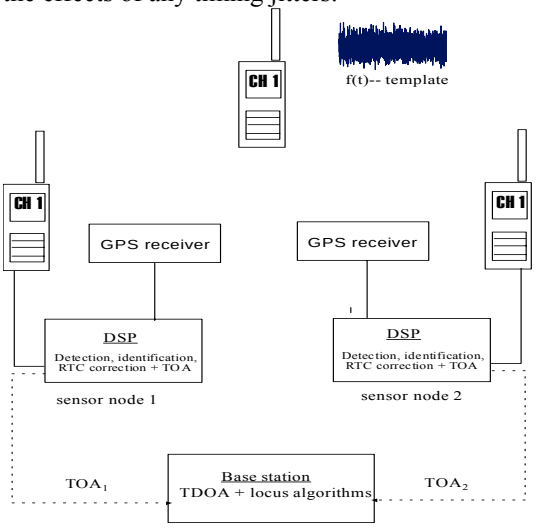


Figure 2. Schematic test setup. A FRS radio transmits a pseudo random sequence containing the template. The two nodes find the template and send TOA estimates to the remote base station

A 1-sec, 5 kHz band-limited pseudo-random sequence served as the template. A signal source placed at a controlled distance from both receivers transmitted a sequence containing the template. The sensor nodes had on-board a pre-manufactured version of the template. The detection and TOA operations were performed locally on the nodes and the TOA estimates were transferred to a PC (base-station) containing the TDOA location determining algorithms. The nominal sampling frequency at all the nodes was 480 kHz.

We present the results of the various experiments in Figure 3-5 and Table 1.

7. Conclusion

The availability of powerful low cost embedded DSPs with an impetus in wireless communications has led to a very large interest in developing self-contained distributed geolocation systems. On such systems, location estimates using TDOA techniques suffer due to inconsistencies in manufacturing and operating conditions. We have presented for low bandwidth, highly resource constrained real-time embedded systems, the related techniques of Doppler frequency and time shifting to compensate for the adverse effects of different sampling frequencies on TOA/TDOA estimates. While the Doppler shift correction is found

to be more accurate, the time-shift technique is more lucrative from a performance perspective.

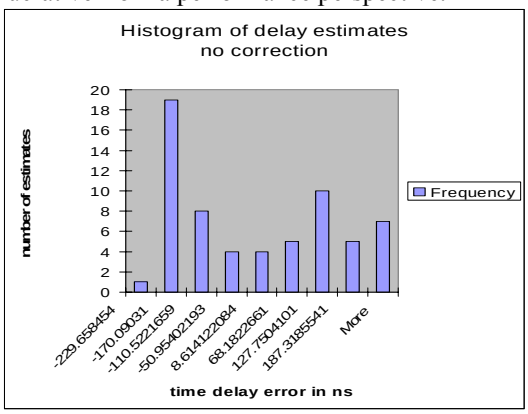


Figure 3 histogram of error in TDOA estimates without any correction applied

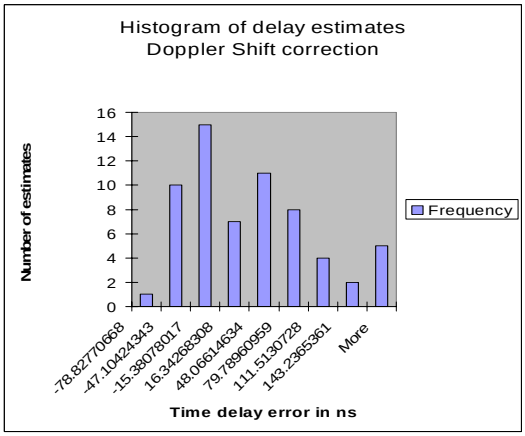


Figure 4 Histogram of errors in TDOA estimates with Doppler shift correction

Table 1 Mean and standard deviation of errors in TDOA estimates

Correction type	Mean error (n sec)	Std dev. (n sec)
None	-30.57	153.85
Doppler shift	17.18	69.46
Time Shift	29.48	60.05

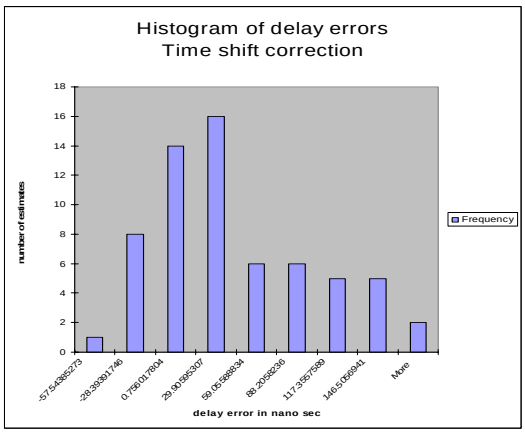


Figure 5 Histogram of errors in TDOA estimate when time-shift correction is used

References

1. Warrior, J et al; "They know where you are"; IEE. Spectrum, Vol.40, no.7, Jul 2003; pp. 20- 25
2. Kumar, S., Shepherd, D.; "SensIT: Sensor information technology for the war fighter"; Proc. Inter. Conf. on Information Fusion, Vol. 1, 2001
3. Bahl, P et. al.; "RADAR: an in-building RF-based user location and tracking system", INFOCOM 2000, 19th Ann. Jt. Conf. IEEE Comp. and Com. Soc. Proc. IEEE, Vol.2, 2000, pp. 775-784
4. Drane, C. et. al. "Positioning GSM telephones" IEE. Comm. Mag. Vol.36, Iss.4 Apr 1998 pp. 46-54, 59
5. Stotts, L.B.: Unattended ground sensor related technologies; an army perspective, Proc. SPIE. Vol. 4040. (2000) 2–10
6. Knapp, C., Carter, G.; "The generalized correlation method for estimation of time delay"; IEE. Trans. Acous., Speech. Sig. Proc. Vol.24, Aug 1976 pp. 320- 327
7. Proakis J; Manolakis, D; "Digital Signal Processing, Principles Algorithms, and Applications", Prentice Hall, Third Ed., 1996.
8. Betz, J.; "Effects of uncompensated relative time companding on a broad-band cross correlator" IEEE Trans. Acous., Speech, Sig. Proc, Vol.33, Jun 1985
9. Remley, W.; "Correlation of signals having linear delay", Acoust. SOC. Amer., vol. 35, pp. 65-69, Jan. 1963.
10. W. B. Adams et al.; "Correlator compensation requirements for passive time delay estimation with moving source or receivers"; IEE. Trans. Acoust., Speech, Signal Processing, vol. 28, pp.158-168, Apr. 1980.
11. Weiss, L.G.; "Wavelets and wideband correlation processing", IEE. Sig Proc Mag Vol.11, Jan 1994 pp13-32
12. Stein, S; "Algorithms for Ambiguity function processing"; IEE. Trans Acou., Spch. & Sig. Proc. Vol.29, Iss.3, Jun 1981 pp 588- 599
13. Ulman, R.; Geraniotis, E; "Wideband TDOA FDOA processing using summation of short-time CAF's", IEE. Trans. Sig. Proc. Vol.47, no 12, Dec 99
14. Dharia A. et al; "Signal Processing Examples Using the TMS320C67x Dig. Sig. Proc. Library (DSPLIB)", Texas Instruments, App. Rpt SPRA947
15. Welch, P.; "A method based on time averaging over short, modified periodograms"; IEEE Trans. on Audio and Electro., Vol.15, Iss.2, Jun '67 pp. 70- 73
16. Leibniz Integral Rule, mathworld.wolfram.com

Enhanced Location Estimation via Pattern Matching and Motion Modelling

Harald Kunczier, Hermann Anegg

ftw. Forschungszentrum Telekommunikation Wien, Donau-City-Str. 1, A-1220 Vienna

Email: {kunczier, anegg}@ftw.at,

Phone: +43-1-5052830-0, Fax: +43-1-5052830-99

ABSTRACT

Mobile location based applications are in the first place designed for urban areas where providers find a high density of customers. Exactly in these environments conventional trilateration location techniques often lack performance due to multipath propagation. Attention has thus be drawn to fingerprint localization methods allowing to localize mobile users in such areas. These have already proved that accuracies below 100 meters are possible in heavy urban areas, but can further be improved significantly when using several fingerprints than relying only on a single one. In this paper we will present a method to optimally combine consecutive position estimates utilizing a motion model for the targeted user. We show trial results from the city of Vienna where we have successfully applied the method and compare it with the single fingerprint case. The achieved accuracy in 90% of all cases has improved from above 100 meters with the single fingerprint method to below 70 meters using the proposed method. This is adequate for most location applications.

I. INTRODUCTION

Fingerprint methods [1], [2], [3] estimate the position of a target (e.g. mobile user) by comparing location dependent parameters (e.g. received power levels) with beforehand measured samples. Accuracy for a single snapshot ranges from about 300 to below 100 meters in urban and heavy urban areas. Most of the location based applications and E911 in the US however do require a significantly higher accuracy.

One strategy of improving the accuracy is to increase the number of pre-measured samples in the database. This is undesired since a larger number of samples requires a higher effort to deploy and maintain the fingerprint based location system. A more promising approach is therefore to rely on a sequence of position estimates and compute the most probable one. This does not effect the size of the database, but does unfortunately increase the required localization time. We will see however that already three consecutive snapshots significantly increase the accuracy even for slow moving pedestrians.

In this paper we propose an algorithm similar to Kalman filtering to utilize several consecutive snapshots instead of relying on a single one. The algorithm combines uncertain position information from several snapshots with a mobility model for the targeted user to enhance the final position estimate. We avoid using a mobility model which assumes a deterministic realization of the velocity and direction [4], but instead combine deterministic behavior with randomness to mimic actual human behavior.

The rest of the paper is organized as follows. In section II we shortly review a pattern matching based localization method which will serve as position estimator for a single measured fingerprint. In section III we show the mobility model used to simulate the motion of the user and apply it to a single position estimate. In section IV we finally update the propagated position estimate with a new position estimate to improve the overall accuracy. We further present results where we have applied the method in the city of Vienna. Finally we conclude in section V.

II. SINGLE POSITION ESTIMATES

Before we start to improve the localization accuracy by combining several estimates we first define an estimator for the probability of being at a position over all considered positions. We therefore briefly review the method proposed in [3] which will serve as a simple sin-

gle fingerprint estimator. Bayesian networks[1] are used there to represent a position by describing dependencies between the different measured Cell IDs (serving cells and ordered neighboring cells according to the received power levels) at a position. The Bayesian networks are then trained with pre-measured data. In our test area in the city of Vienna we used equally spaced measurements every 5 meters. The final search of the mobile's position is then a comparison of the mobile's current fingerprint $\boldsymbol{f}$ containing the received serving and neighboring cells and all models in the expected target area (e.g. the area of the serving cell). For the comparison we use the marginal likelihood as a scoring method to identify the optimal model according to

$$\mathcal{L}(\lambda_{\boldsymbol{i}}) := p(\boldsymbol{f}|\lambda_{\boldsymbol{i}}) = \int p(\boldsymbol{f}|\lambda_{\boldsymbol{i}}, \boldsymbol{\theta_i}) p(\boldsymbol{\theta_i}) d\boldsymbol{\theta_i} \quad (1)$$

with $\lambda_{\boldsymbol{i}}$ being the Bayesian network at position $\boldsymbol{i} = (x, y)$, $\boldsymbol{f}$ the fingerprint of the mobile we want to localize and $\boldsymbol{\theta_i}$ the parameters of the Bayesian network which are updated during the training with the measured samples.

Maximization over all Bayesian networks within the area of the serving cells results in the best matching Bayesian network and thus in the best estimate for a single position. The resulting accuracy within our target area is shown in Fig. 5 (dashed line).

We should note at this point that the considerations concerning the fingerprint method address GSM in this paper. We would like to stress however that this the all the methods introduced here can be applied to any location dependent parameter of the mobile system in general.

III. User Mobility model

The mobility model we propose here attempts to mimic human movement behavior to predict the new position of a mobile user. This is important, since we avoid the approach of estimating the position of a user simply as the mean position computed from several single localization estimates. The result would suffer from a systematically increasing error for increasing velocity of the target, caused by the larger spatial separation of the different position estimates.

Instead we initially rely on the probability density distribution of the positions given the measured location dependent parameter of the mobile resulting from (1):

$$p(\boldsymbol{i}|\boldsymbol{f}) := p(\lambda_{\boldsymbol{i}}|\boldsymbol{f}) = \frac{p(\lambda_{\boldsymbol{i}})}{p(\boldsymbol{f})} \tilde{\mathcal{L}}(\lambda_{\boldsymbol{i}}) = \gamma \mathcal{L}(\lambda_{\boldsymbol{i}}), \quad (2)$$

[1]For an introduction see e.g. [5], [6]

with γ being a constant, if we assume no prior knowledge about the occurrence of either a certain position or a certain fingerprint.

In order to combine this information with information from the next fingerprint at time $t+T$ we use a mobility model which will change our believes about the initial position taken at time t. The variance of the first estimate will thus increase since the user might move ahead during the time T. We are not so sure anymore where the user actually is located.

For the mobility model we make three assumptions:

1) The user will normally move with constant velocity $\boldsymbol{u}$ for the time under consideration. (This is about a few seconds).
2) Physical obstacles, other persons, etc. are viewed as perturbations $\boldsymbol{v}$ upon the constant velocity trajectory from assumption one.
3) The user will try to reestablish it's constant velocity (equal to $\boldsymbol{v} = 0$), once he was perturbed.

In general these assumptions result from the tendency of a person to maximize his personal utility, which includes to avoid deceleration and acceleration processes [7].

For a single physical dimension we therefore model the targeted user's motion as a dynamic linear system and write:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t) + \boldsymbol{G}\boldsymbol{w}(t) \quad (3)$$

with

$$\begin{aligned} \boldsymbol{x} &= \begin{bmatrix} x \\ v \end{bmatrix} \\ &= \begin{bmatrix} \text{user's position at time t} \\ \text{user's velocity variation around its constant speed at time t} \end{bmatrix}. \end{aligned} \quad (4)$$

The vector $\boldsymbol{u} = [u, 0]$ is a deterministic vector and addresses assumption one by denoting the constant velocity. The random vector $\boldsymbol{w} = [0, w]$ represent white Gaussian noise and models our second assumption where the user is perturbed by obstacles and suddenly has to change his velocity. The resulting speed difference between his current speed and his desired velocity is denoted by the variable v. In such a case the user will try to reach its personal optimal speed u again and thus will change his speed until the term v becomes zero. This indicates the speed difference v to be correlated in time; if the user does not move with his desired speed u at time t, it is likely that he still moves with different speed than u at time $t + \tau$ for sufficient small τ.

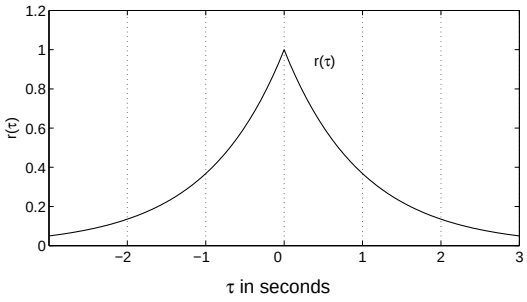


Fig. 1. Correlation function $r_w(\tau)$ of user's perturbation caused velocity v.

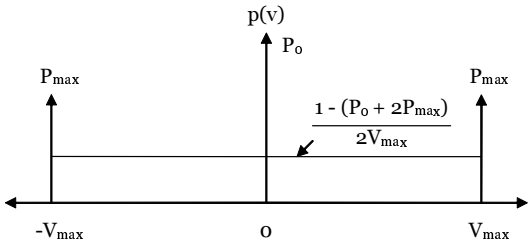


Fig. 2. Probability density model of the user's velocity

Singer used in his paper about tracking [8] a similar model but incorporates acceleration also. We have omitted to model acceleration here, since the acceleration period of the users under consideration (mainly pedestrians) is assumed to be small compared to the systems time constants. For the time correlation of v we assume (refer also to Fig. 1):

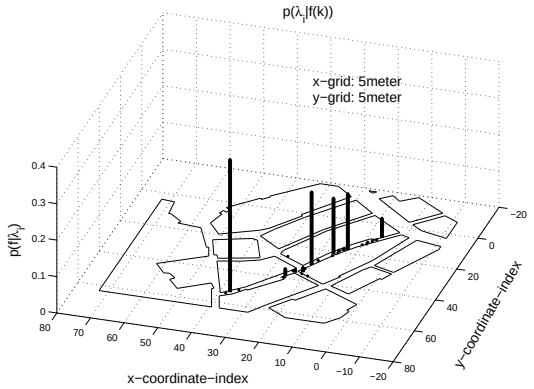
$$r_v(\tau) = E\{v(t)v(t+\tau)\} = \sigma_m^2 e^{-\alpha|\tau|} \qquad (5)$$

where σ_m^2 is the variance of the difference speed v and α is the reciprocal of the random difference speed time constant. We assume $\alpha = \frac{1}{0.2}$ and for σ_m^2 we use the same approach as in [8]: We construct the variance assuming that the user may increase or decrease his speed due to perturbation by a maximum value V_{max} $(-V_{max})$. He will do so with a probability P_{max}. The user will not change his velocity with probability P_0 and will speed up or down between the limits according to a uniform distribution (Fig. 2). We can then write for the variance

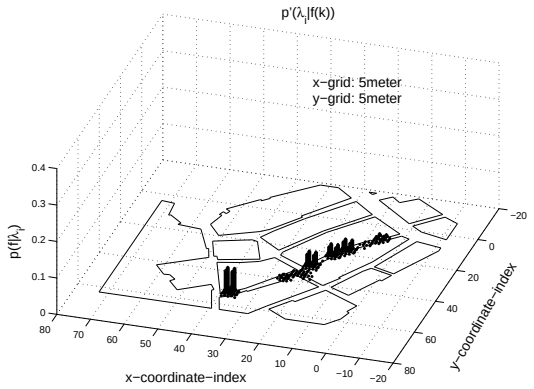
$$\sigma_m^2 = \frac{V_{max}^2}{3}(1 - 4P_{max} - P_0). \qquad (6)$$

Deriving further the power density spectrum from (5) and interpreting the result as being produced by a shaping filter driven by white Gaussian noise w we get for the corresponding differential equation:

$$\dot{v}(t) = -\alpha v + w(t) \quad \text{with} \quad \sigma_w^2(\tau) = 2\alpha\sigma_m^2\delta(\tau) \quad (7)$$



(a) probability density of the position estimate given the fingerprint $\boldsymbol{f}$ at time k



(b) probability density of the position estimate given the fingerprint $\boldsymbol{f}$ at time k+1.

Fig. 3. Impact of the mobility model on a position estimate. The probability of a single position is not so sure anymore. The possible movement of a user broadens the variance of $p(\lambda_{\boldsymbol{i}}|\boldsymbol{f}(k))$

The remaining desired velocity u we model as a random variable with its density constructed by the superposition of three Gaussian shaped curves:

$$\begin{aligned} f(u) &= \frac{(1-\mathrm{w})}{2}\mathcal{N}(-u_m, \sigma_{u_m}^2) + \mathrm{w}\mathcal{N}(0, \sigma_{u_0}^2) + \\ &+ \frac{(1-\mathrm{w})}{2}\mathcal{N}(u_m, \sigma_{u_m}^2) \end{aligned} \qquad (8)$$

The two Gaussian shaped curves with mean $-u_m$ and u_m represent the users moving forward or backwards. The curve in the middle denotes a motionless (or almost motionless) user. The weighting factor $\mathrm{w} \in [0, 1]$ allows to control the percentage of motionless users.

The dynamic linear system equation (3) is now specified completely by

$$\boldsymbol{F} = \begin{bmatrix} 0 & 1 \\ 0 & -\alpha \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

and represents the motion of a user in a single physical dimension.The extension into a second dimension is straight forward if we assume independence between the cartesian coordinates. For convenience we keep the same names for the variables, but introduce the indices x and y to describe the physical dimension.

Assuming a new position estimate every T seconds and applying the state-space-method to (8) we write for the discrete mobility equation

$$\boldsymbol{X}(k+1) = \boldsymbol{\Theta}(T,\alpha)\boldsymbol{X}(k) + \boldsymbol{B_d}(k)\boldsymbol{U}(k) + \boldsymbol{W}(k) \quad (9)$$

where

$$\begin{aligned} \boldsymbol{X} &= [x, v_x, y, v_y]^T \\ \boldsymbol{U} &= [u_x, 0, u_y, 0]^T \\ \boldsymbol{B_d}(T) &= \int_t^{t+T} \boldsymbol{\Theta}(t-\tau,\alpha)\boldsymbol{B}d\tau \end{aligned}$$

$\boldsymbol{X}$ is the dynamic state vector containing the position and the velocity for both cartesian dimensions. $\boldsymbol{U}$ is the desired deterministic speed of the user and $\boldsymbol{W}(k)$ is a discrete-time zero-mean white Gaussian noise with statistics according to

$$E\{\boldsymbol{W}(k)\} = \boldsymbol{0}$$

$$E\{\boldsymbol{W}(k)\boldsymbol{W}^T(j)\} = \left\{ \begin{array}{cc} \boldsymbol{Q}(k) & j = k \\ \boldsymbol{0} & j \neq k \end{array} \right\}.$$

and

$$\boldsymbol{Q}(k) = \boldsymbol{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 2\alpha{\sigma_m}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\alpha{\sigma_m}^2 \end{pmatrix}$$

The matrices $\boldsymbol{\Theta}$ and $\boldsymbol{B_d}$ are the state transition matrices to link the system at the times k and $k+1$.

Since $\boldsymbol{F}$ is time invariant the state transition matrix $\boldsymbol{\Theta}(T,\alpha)$ can easily be obtained by [9]

$$\boldsymbol{\Theta}(T,\alpha) = \mathcal{L}^{-1}\{(s\boldsymbol{I} - \boldsymbol{F})^{-1}\} \quad (10)$$

where $\mathcal{L}$ denotes the Laplace transformation. This results in

$$\boldsymbol{\Theta}(T,\alpha) = \begin{pmatrix} 1 & \frac{1}{\alpha}(1-e^{-\alpha T}) & 0 & 0 \\ 0 & e^{-\alpha T} & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{\alpha}(1-e^{-\alpha T}) \\ 0 & 0 & 0 & e^{-\alpha T} \end{pmatrix} \quad (11)$$

We see from (9) that $\boldsymbol{X}(k+1)$ is Gaussian if $\boldsymbol{X}(k)$ is either Gaussian or deterministic and since we assume a known initial position at time $k=0$ the density function $p_{\boldsymbol{X}(k+1)}(.)$ is completely determined by the mean and covariance given by [9]:

$$\boldsymbol{m_X}(k+1) = \boldsymbol{\Theta}(T,\alpha)E\{\boldsymbol{X}(k)\} + \boldsymbol{B_d}(T)\boldsymbol{U}(k) \quad (12)$$

$$\begin{aligned} \boldsymbol{P_{XX}}(k+1) &= \boldsymbol{\Theta}(T,\alpha)E\{\boldsymbol{X}(k)\boldsymbol{X}^T(k)\}\boldsymbol{\Theta}(T,\alpha)^T + \\ &+ \int_0^T \boldsymbol{\Theta}(T-\tau,\alpha)\boldsymbol{Q}\boldsymbol{\Theta}(T-\tau,\alpha)^T d\tau \end{aligned} \quad (13)$$

Letting the mobile user start at the initial position $\boldsymbol{X}(k=0)$ and with velocity $\boldsymbol{U}(k=0), v_x = 0, v_y = 0$ the mean results according to (12) in

$$\boldsymbol{m_X}(k+1) = \boldsymbol{X}(k=0) + T\boldsymbol{U}(k=0) \quad (14)$$

The covariance computes to

$$\boldsymbol{P_{XX}}(k+1) = 2\alpha{\sigma_m}^2 \begin{pmatrix} p11 & p12 & 0 & 0 \\ p12 & p22 & 0 & 0 \\ 0 & 0 & p11 & p12 \\ 0 & 0 & p12 & p22 \end{pmatrix} \quad (15)$$

where

$$p11 = \frac{1}{2\alpha^3}\left(-e^{-2\alpha T} + 4e^{-\alpha T} - 3 + 2\alpha T\right)$$

$$p12 = \frac{1}{2\alpha^2}\left(e^{-2\alpha T} - 2e^{-\alpha T} + 1\right)$$

$$p22 = \frac{1}{2\alpha}\left(1 - e^{-2\alpha T}\right).$$

IV. Combined Position Estimate and User Mobility Model

We are now able to propagate the optimal position estimate $\hat{\boldsymbol{i}}(k)$ at time k into the estimate $\hat{\boldsymbol{i}}'(k+1)$ at time $k+1$.

We therefore treat the position $\hat{\boldsymbol{i}}(k)$ of the user as random variable and use (2) to describe its probability density. By adding the distance $\boldsymbol{X}(k+1)$ which the user has moved during time period T we receive the new position to be:

$$\hat{\boldsymbol{i}}'(k+1) = \hat{\boldsymbol{i}}(k) + \boldsymbol{A}\boldsymbol{X}(k+1). \quad (16)$$

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Thus we are able to compute $\hat{\boldsymbol{i}}'$ from the position estimate $\hat{\boldsymbol{i}}(k)$ at time k and the mobility model's contribution $\boldsymbol{X}(k+1)$.

Figure 3 illustrates the impact of the motion model. It shows a section of the target area in the inner city of Vienna. The x- and y-coordinates are given as indices of a 5 times 5 meter measurement grid. The z-axis shows the probability $p(\lambda_{\boldsymbol{i}}|\boldsymbol{f})$ of being at a certain position $\boldsymbol{i}$. In Fig. 3(a) a first estimate for the fingerprint $\boldsymbol{f}$ at time k is shown. A simple maximum likelihood estimator would localize the user at the position with the indices $(38, 52)$. Fig. 3(b) shows the situation at time $k+1$ after $T = 3$ seconds. The reliability of the first estimate is reduced by the possible movement of the user. The probability of being at the position $(38, 52)$ is reduced by about 80% compared to time t.

Let us now consider the incorporation of the position estimate $\hat{\boldsymbol{i}}(k+1)$ which becomes available at time $k+1$. We now combine this estimate with the one $\hat{\boldsymbol{i}}'(k+1)$ propagated over time.

Since we still do not know whether to trust the propagated position estimate or the newly available estimate more, we combine them according to

$$\hat{\boldsymbol{i}}^*(k+1) = (\boldsymbol{I} - \boldsymbol{K})\hat{\boldsymbol{i}}'(k+1) + \boldsymbol{K}\hat{\boldsymbol{i}}(k+1). \quad (17)$$

whit $\boldsymbol{K}$ denoting a blending factor and $\boldsymbol{I}$ being the identity matrix.

To find the blending factor $\boldsymbol{K}$ we chose to minimize the estimator's variance and limit $\boldsymbol{K}$ to be between $\boldsymbol{0}$ and $\boldsymbol{I}$. This is equal to a minimization of the major diagonal of the covariance matrix of the estimator $\hat{\boldsymbol{i}}^*(k+1)$ and we write:

$$\frac{d(tr(\boldsymbol{P}_{KK}))}{d\boldsymbol{K}} = 0 \quad (18)$$

with

$$\begin{aligned} \boldsymbol{P}_{KK} &= E\{\hat{\boldsymbol{i}}^*(k+1)\hat{\boldsymbol{i}}^*(k+1)^T\} - \\ &- E\{\hat{\boldsymbol{i}}^*(k+1)\}E\{\hat{\boldsymbol{i}}^*(k+1)\}^T \quad (19) \\ & \quad (20) \end{aligned}$$

Assuming $\hat{\boldsymbol{i}}'$ and $\hat{\boldsymbol{i}}$ uncorrelated and applying a straightforward differential calculus approach utilizing

$$\frac{d\,(tr(\boldsymbol{AB}))}{d\boldsymbol{A}} = \boldsymbol{B}^{\boldsymbol{T}} \quad \boldsymbol{A}, \boldsymbol{B} \text{ square}$$

$$\frac{d\left(tr(\boldsymbol{ACA}^{\boldsymbol{T}})\right)}{d\boldsymbol{A}} = 2\boldsymbol{AC} \quad \boldsymbol{C} \text{ symmetric}$$

we find for the blending factor

$$\boldsymbol{K} = \frac{2(\boldsymbol{Q}^{\boldsymbol{T}} - \bar{\boldsymbol{i}}'\bar{\boldsymbol{i}}'^{\boldsymbol{T}}) + \bar{\boldsymbol{i}}'\bar{\boldsymbol{i}}^{\boldsymbol{T}} + \bar{\boldsymbol{i}}\bar{\boldsymbol{i}}'^{\boldsymbol{T}}}{2(\boldsymbol{Q} + \boldsymbol{R} - (\bar{\boldsymbol{i}}' - \bar{\boldsymbol{i}})(\bar{\boldsymbol{i}}' - \bar{\boldsymbol{i}})^{\boldsymbol{T}})} \quad (21)$$

TABLE I

TEST CAMPAIGN'S PARAMETER SETTING IN THE CITY OF VIENNA

Parameter	Description	Value
α	reciprocal difference speed time constant	5
V_{max}	maximal speed increase due to perturbation	1.5 meter/s
$-V_{max}$	maximal speed decrease due to perturbation	1.5 meter/s
P_{max}	probability of maximal speed increase	0.1
P_0	probability of no perturbation	0.6
$\|um\|$	mean speed of moving user	1.5 meter/s
$\sigma^2_{\frac{u}{y}m}$	variance of forward/backward moving user	0.25
σ^2_{u0}	variance of motionless user	0.0025
w	proportion of motionless user	0.1
T	time period between measured fingerprints	3s

Fig. 4. Map of test area in the city of Vienna. Source of the map: [10]

with

$$\bar{\boldsymbol{i}} = E\{\boldsymbol{i}\}, \quad \bar{\boldsymbol{i}'} = E\{\boldsymbol{i}'\}$$

$$\text{and} \quad \boldsymbol{Q} = E\{\boldsymbol{i}'\boldsymbol{i}'^T\}, \quad \boldsymbol{R} = E\{\boldsymbol{i}\boldsymbol{i}^T\}.$$

The time propagated measurement $\hat{\boldsymbol{i}}'(k+1)$ can now be updated according to (17) and we receive a final estimator for the position at time $k+1$. The same procedure can easily be applied for following time periods. It has to be noted however that we compute for every time step sums of random variables which involves a convolution. A simple tracking will thus be inefficient in terms of computational effort. For the improvement of position estimates however, where only a few time steps are considered the method is suitable.

To test our method we use a heavy urban area in the downtown area of Vienna. A map is shown in Fig. 4. For the initial training of the Bayesian networks to perform the single position estimate we use 10 samples per position. For the localization we choose the time period T between two consecutive measurements to be 3 seconds to allow a pedestrian to move at least several meters. For a sum of all parameters chosen refer to Tab. I. The resulting accuracy is shown in Fig. 5 (solid line). The error $e = \|\hat{\boldsymbol{i}}^* - \boldsymbol{i}\|$ is defined as the difference between the true and the estimated position. We can see, that the 90% margin is below an error of 70

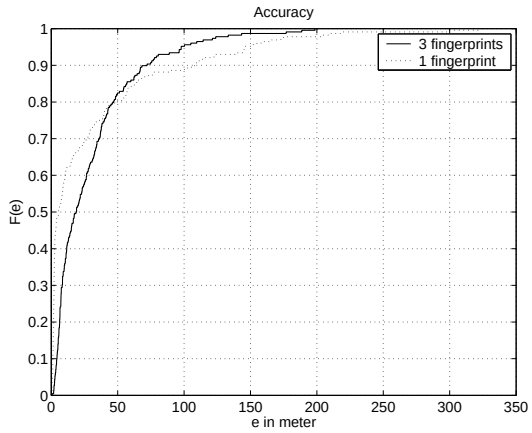


Fig. 5. Accuracy of the method utilizing one single fingerprint (dashed) and three consecutive fingerprints (solid). Total sample size: 280

meters compared to more than 100 meters for the single estimation case (dashed line). On the other side, due to the combination of several position estimates, positioning errors up to about 50 meters are more likely to occur. The main achievement however is the reduction of outliers which classifies the method to be suitable for most location based services, especially if a deployment in densely populated heavy urban areas is intended.

V. Conclusions

In this paper we have presented a method to improve the accuracy of a simple pattern matching based position estimate by applying a motion model and combining several consecutive fingerprints. We have then applied the method to trail measurements taken in the city of Vienna and have achieved an accuracy of about 70 meters in 90% of all cases and less than 40 meters in 67% of all cases. Limitations to the method apply if the target user is very slowly moving and the underlying localization method show the same probability densities of the positions for all three consecutive fingerprints. In this case no accuracy improvement can be expected.

Acknowledgment

This work is supported within the Austrian competence center program ***K**plus* and by the member companies of ftw. (Mobilkom Austria AG and Dipl.-Ing. Dr. Hermann Bühler GmbH).

References

[1] S. Ahonen and H. Laitinen, "Database correlation method for UMTS location," *IEEE Vehicular Technology Confererence*, 2003.

[2] H. Laitinen, J. Lähteenmäki, and T. Nordström, "Database correlation method for GSM location," *IEEE VTC 2001 Spring Conf.*, May 2001.

[3] H. Kunczier and H. Anegg, "Enhanced cell id based terminal location for urban area location based applications," *to be presented at IEEE Consumer Communications and Networking Conference*, January 2004.

[4] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonpriorirized handoff procedures," *IEEE Transactions On Vihicular Technology*, vol. 35, no. 3, August 1986.

[5] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning bayesian networks: The combination of knowledge and statistical data," Microsoft Research, Advanced Technology Division, Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, Technical Report MSR-TR-94-09, 1995.

[6] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*, ser. Statistics for Engineering and Information Science. 175 Figth Avenue, New York, NY 10010, USA: Springer Verlag New York, Inc., 1999.

[7] D. Helbing, "A mathematical model for the behavior of pedestrians," *Behavioral Science*, vol. 36, pp. 298–310, 1991.

[8] R. A. Singer, "Estimating optimal tracking filter performance for manned maneuvering targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-6, no. 4, July 1970.

[9] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, ser. Mathematics in Science and Engineering. Academic Press New York, 1979, vol. 1.

[10] Wien-Grafik Redaktion, "Stadtplan mit Adressensuche," www.wien.at ©1995-2001 wien.at: Magistrat der Stadt Wien, Rathaus, A-1082 Wien, 2003.

Mobile Transmitters Tracking Using Geodetic Models with Multiple Receivers

Ming-Wang Tu and François Patenaude
Communications Research Centre Canada
3701 Carling Avenue, Box 11490, Station H
Ottawa, ON, Canada, K2H 8S2
Tel: 613-998-9262, 613-990-5878, Fax: 613-990-8842
ming-wang.tu@crc.ca, francois.patenaude@crc.ca

Abstract

The paper presents two geodetic models using data from two or three receivers to track the locations of mobile transmitters.

To apply the techniques in this study, one first has to obtain the estimated angles of arrival (AOAs) and their standard deviations (SDs) from two or three receivers to each transmitter. The histogram-based algorithms in [1] can be used to calculate the estimated AOAs and AOA SDs from a receiver to various transmitters and can be extended for mobile transmitters and fixed receivers with a two/three-receiver (2R/3R) fixing. Once the estimated AOAs and AOA SDs from each receiver to the mobile transmitters are obtained, both of the Spherical and WGS84 geodetic models [2], [3] are used to process the estimated data for each receiver to track the locations of the mobile transmitters with a 2R or 3R fixing. To consider the statistical variations during the location tracking, the concept of confidence ellipse (CE) [4] is applied.

Three sets of simulated data were processed and their results showed significant accuracy for various scenarios. The results of this study demonstrated the effectiveness of using both of the geodetic models to track mobile transmitters with a 2R or 3R fixing.

In general, both of the geodetic models provide a simple and efficient way to track mobile transmitters. The approach is particularly applicable for receivers using fast wideband scanning devices and where items such as AOAs and their instantaneous SDs from several channels are reported per second.

I. Introduction

In mobile communication, the problem of position determination of a mobile transmitter has been studied extensively, particularly in the context of military operations and governmental spectrum licensing. Position estimation can be enhanced by combining geodetic modeling with direction finding (DF) techniques and receivers' global positioning system (GPS) [2] data. In this study, multiple (two and three) receivers were used to perform the DF fixing. Two geodetic models (Spherical and WGS84) were used.

The CRC's Spectrum Explorer [5] can be used as the receiver to scan, collect and pre-process wireless signals. Each scenario may include multiple channels (frequency bands), in or out of regulation. Each channel may have multiple users from various AOAs with different signal-to-noise ratios (SNRs). Each user may use various modulations such as amplitude modulation (AM), frequency modulation (FM), etc. The estimated AOAs and AOA SDs of users for each receiver can be used for the 2R or 3R fixing. With the known GPS coordinates of the receivers and the geodetic models, one can track the mobile transmitters accurately.

Three simulated data sets will be processed using both of the geodetic models for the 2R and 3R fixings. The detailed results will be shown later in this study.

II. Geodetic Models

As described in [2], a point Q on an ellipsoid is determined by (ϕ, λ) = (latitude, longitude). The latitude is the angle between the normal at Q and the plane of the Equator. For an ellipsoid, the normal at Q does not go through the center point. The longitude is the angle between the plane of the meridian of Q and the plane of a reference meridian through Greenwich.

In this study, the GPS data include each receiver's coordinates (ϕ, λ). The AOA is equivalent to the azimuth, defined as the angle with respect to the North Pole in a clockwise direction. Since the earth can be modeled as a sphere or an ellipsoid, the corresponding two models, Spherical and WGS84, were used.

1. Spherical Model

To derive formulas for the transformation of a sphere, two basic laws of spherical trigonometry [3] are used. Referring to the spherical triangle in Fig. 1, with three points having angles ($\underline{A}$, $\underline{B}$, $\underline{C}$) on the sphere, and three great circle angles ($\underline{a}$, $\underline{b}$, $\underline{c}$) connecting them, the Laws of Sines/Cosines declare that

$$\sin \underline{A} / \sin \underline{a} = \sin \underline{B} / \sin \underline{b} = \sin \underline{C} / \sin \underline{c}\,, \quad (1)$$

$$\cos \underline{c} = \cos \underline{b} \cos \underline{a} + \sin \underline{b} \sin \underline{a} \cos \underline{C}\,, \quad (2)$$

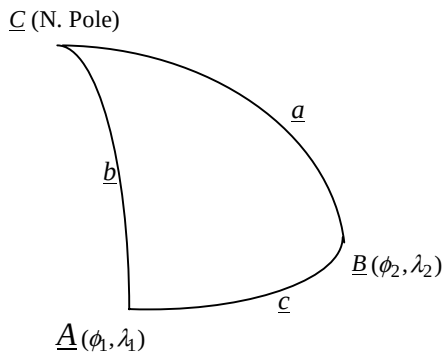


Fig. 1. Spherical triangle

$$\cos\underline{C} = -\cos\underline{B}\cos\underline{A} + \sin\underline{B}\sin\underline{A}\cos\underline{c}. \quad (3)$$

If $\underline{C}$ is placed at the North Pole, it becomes the angle between the two meridians extending to $\underline{A}$ and $\underline{B}$. If $\underline{A}$ is the starting point on the sphere and $\underline{B}$ the second, $\underline{c}$ is the great circle angle between them, and angle $\underline{A}$ is the azimuth *Az* east of north which point $\underline{B}$ bears to point $\underline{A}$. If (ϕ_1, λ_1) and (ϕ_2, λ_2) are for $\underline{A}$ and $\underline{B}$, respectively, the great arc distance between $\underline{A}$ and $\underline{B}$ is *S*, and the radius of the earth is *R* (=6367.445 km), then

$\underline{A} = Az, \sin\underline{a} = \cos\phi_2, \sin\underline{b} = \cos\phi_1,$

$\sin\underline{C} = \sin(\lambda_2 - \lambda_1), \underline{c} = S/R.$

Then (1) and (2) become

$$\sin Az/\cos\phi_2 = \sin(\lambda_2 - \lambda_1)/\sin\underline{c}, \quad (4)$$

$$\cos\underline{c} = \sin\phi_1\sin\phi_2 + \cos\phi_1\cos\phi_2\cos(\lambda_2 - \lambda_1). \quad (5)$$

From (4), one can get

$$\cos Az = [\cos\phi_1\sin\phi_2 - \sin\phi_1\cos\phi_2\cos(\lambda_2 - \lambda_1)]/\sin\underline{c}. \quad (6)$$

From (5) and (6), one can get

$$\phi_2 = \arcsin(\sin\phi_1\cos\underline{c} + \cos\phi_1\sin\underline{c}\cos Az). \quad (7)$$

Given (ϕ_1, λ_1) and (ϕ_2, λ_2) for receivers 1 and 2 (R1 and R2), (5) can be used to find the *S* (D12) between R1 and R2. Then the corresponding *Az* (A12, A21) between R1 and R2 can be found by using (4) and (6). For transmitter 1 (T1) with AOA1 from R1 and AOA2 from R2, by properly subtracting (A12, A21) from (AOA1, AOA2), one can find the two internal angles at the R1 and R2 corners of the R1-R2-T1 triangle. Then the internal angle at the T1 corner can be calculated by using (3). The D13 between R1 and T1 can then be calculated by using (1). Then the problem with given D13 and AOA1 from R1 is next. (5) and (7) can be used to find the (ϕ_3, λ_3) which is the location of T1. The same procedure can be repeated for multiple AOA1 and AOA2 sets for various mobile transmitters. The above 2R fixing procedure can be extended to a 3R fixing by processing two receivers at a time.

2. WGS84 Model

The following mid-latitude formulas [2] (8), (9) and (10) can be used to find the arc distance *S* and the two azimuths between known points at (ϕ_1, λ_1) and (ϕ_2, λ_2):

$$S\sin Az = Nl\cos\phi[1 - (l\sin\phi)^2/24 + (1 + \eta^2 - 9\eta^2t^2)b^2/24V^4], \quad (8)$$

$$S\cos Az = Nb'\cos(l/2)[1 + (1 - 2\eta^2)(l\cos\phi)^2/24 + \eta^2(1 - t^2)b'^2/8V^4], \quad (9)$$

$$\Delta Az = l\sin\phi[1 + (1 + \eta^2)(l\cos\phi)^2/12 + (3 + 8\eta^2)b^2/24V^4], \quad (10)$$

where

$\phi = (\phi_1 + \phi_2)/2, l = \lambda_2 - \lambda_1, b = \phi_2 - \phi_1, t = \tan\phi, \eta = e'\cos\phi,$

$e'^2 = (a^2 - b^2)/b^2, V^2 = 1 + \eta^2, f = (a - b)/a, c = a^2/b,$

$N = a/\sqrt{1 - f(2 - f)\sin^2\phi}, V = c/N, b' = b/V^2,$

and *a* (=6378.137 km) is the semimajor axis (equatorial radius) of earth; *b* is the semiminor axis (polar radius) of earth; *f* (=1/298.257223563) is the flattening; e' is the second eccentricity and *c* is the radius of curvature.

(10), (11) and (12) can be used iteratively to find the (ϕ_2, λ_2) at a given *S* and azimuth *Az* east of north from (ϕ_1, λ_1):

$$l = S\sin Az[1 + (l\sin\phi)^2/24 - (1 + \eta^2 - 9\eta^2t^2)b^2/24V^4]/N\cos\phi, \quad (11)$$

$$b' = S\cos Az[1 - (1 - 2\eta^2)(l\cos\phi)^2/24 - \eta^2(1 - t^2)b'^2/8V^4]/N\cos(l/2). \quad (12)$$

Note that the *S* between the estimated location of a transmitter and a known receiver in WGS84 model can be estimated by using (4) and (5) in the Spherical model (note the $\underline{c}$=*S*/*R*) combined with the averaged spherical-to-ellipsoidal correction factor. From [3], the ratio for the length of a radian of latitude along a meridian on the sphere to that on the ellipsoid is

$$C_m(\phi) = R(1 - e^2\sin^2\phi)^{3/2}/[a(1 - e^2)], \quad (13)$$

and the ratio for the length of a radian of longitude along a parallel on the sphere to that on the ellipsoid is

$$C_p(\phi) = R(1 - e^2\sin^2\phi)^{1/2}/a, \quad (14)$$

where $e = (1 - b^2/a^2)^{1/2}$ is the eccentricity of the ellipsoid. Given (ϕ_1, λ_1) and (ϕ_2, λ_2), the averaged spherical-to-ellipsoidal correction factor is

$$C_{sp\_ellips} = \left[\sqrt{C_m^2(\phi_1) + C_p^2(\phi_1)} + \sqrt{C_m^2(\phi_2) + C_p^2(\phi_2)}\right]/2. \quad (15)$$

The relationship between the *S* in WGS84 model (S_{WGS84}) and the *S* in Spherical model (S_{sp}) is

$$S_{WGS84} = S_{sp} / C_{sp\_ellips} . \quad (16)$$

Given (ϕ_1, λ_1) and (ϕ_2, λ_2) for R1 and R2, (8) and (9) can be used to find the *S* (D12_WGS84) between R1 and R2. Then the corresponding *Az* (A12, A21) between R1 and R2 can be found by using (8), (9) and (10). Again, for T1 with AOA1 from R1 and AOA2 from R2, by properly subtracting (A12, A21) from (AOA1, AOA2), one can find the two internal angles at the R1 and R2 corners of the R1-R2-T1 triangle. Then the internal angle at the T1 corner can be calculated by using (3). Note that since the Spherical model equation (3) is used in the WGS84 model, (15) has to be used. That is, D12_WGS84 should be multiplied by (15) to get the D12_Spherical, which can be then used to find the internal angle at the T1 corner. The D13_Spherical between R1 and T1 can then be calculated by using (1). Then the D13_Spherical should be divided by (15) to get the D13_WGS84. Then the problem with given D13_WGS84 and AOA1 from R1 is next. (10), (11) and (12) can be used to find the (ϕ_3, λ_3) which is the location of T1. Again, the same procedure can be repeated for multiple AOA1 and AOA2 sets for various mobile transmitters. The above 2R fixing procedure can be extended to a 3R fixing by processing two receivers at a time.

III. 2R/3R Fixing and Confidence Ellipse (CE)

Both of the 2R and 3R fixings were investigated in this study. The concept of CE was also applied.

1. 2R Fixing

The detailed calculation procedure for the 2R fixing has been described in both the Spherical and WGS84 model sections. One can repeat the procedure for multiple AOA1 and AOA2 sets for various mobile transmitters to track those mobile transmitters.

2. 3R Fixing

In Fig. 2, *K*, *L* and *M* represent the receivers and *A*, *B*, *C* the corresponding corners of a triangle. From [4], the best estimated location of the transmitter is at *V*, the meeting-point of lines *AT* and *BU*, where

$$\frac{BT}{TC} = \frac{D_M^2 \sigma_{\psi M}^2 \sin^2 BCA}{D_L^2 \sigma_{\psi L}^2 \sin^2 ABC}, \quad (17)$$

$$\frac{CU}{UA} = \frac{D_K^2 \sigma_{\psi K}^2 \sin^2 CAB}{D_M^2 \sigma_{\psi M}^2 \sin^2 BCA}, \quad (18)$$

where $D_K = KC, D_L = LA, D_M = MB$ and $\sigma_{\psi J}$ is the AOA SD from receiver *J* (=*K*,*L*,*M*). By Menelaus' Theorem [6],

$CA \times VU \times BT = UA \times BV \times TC$, thus

$$\frac{BV}{VU} = \frac{BT}{TC} \times \frac{CA}{UA} = \frac{BT}{TC} \times (1 + \frac{CU}{UA}) . \quad (19)$$

The detailed calculation procedure for the 2R fixing has been described in both the Spherical and WGS84 model sections. For various mobile transmitters, one can repeat the 2R fixing procedure for multiple AOA1, AOA2 and AOA3 sets for the 3R fixing to find the sets of (*A*, *B*, *C*). Once the sets of (*A*, *B*, *C*) are found, one can use (17), (18) and (19) to find the *V*s to track those mobile transmitters.

3. Confidence Ellipse

The CE with probability *P* is the probability that the *V* of a mobile transmitter will lie within the area bounded by an elliptical contour with semimajor axis *r* and semiminor axis *s*. The related equations [4] are defined as follows:

$$\frac{X^2}{r^2} + \frac{Y^2}{s^2} = -2\log_{\varepsilon}(1-P) , \quad (20)$$

where

$$\frac{1}{r^2} = 2\kappa - \nu \tan\varphi , \ \frac{1}{s^2} = 2\mu + \nu \tan\varphi , \ \kappa = \sum_J \frac{\sin^2 \theta_J}{\sigma_{\psi J}^2 D_J^2} ,$$

$$\mu = \sum_J \frac{\cos^2 \theta_J}{\sigma_{\psi J}^2 D_J^2} , \ \nu = \sum_J \frac{\sin\theta_J \cos\theta_J}{\sigma_{\psi J}^2 D_J^2} , \ \tan 2\varphi = -\frac{2\nu}{\kappa - \mu},$$

and θ_J is the AOA from receiver *J*. Note that *X* and *Y* rotate through an angle φ relative to the coordinates *x* and *y*. Also the $(r, s, \kappa, \mu, \nu, \varphi)$ and the CE vary as the AOA varies. The CE can be applied to both of the 2R and 3R fixings.

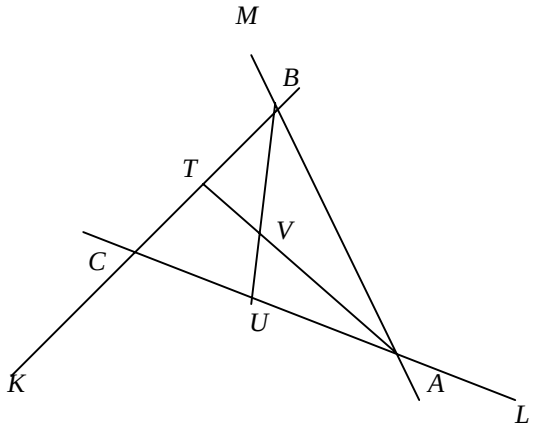


Fig. 2. 3R Fixing

IV. Simulated Results

Three receivers located at Mont Royal (R1) and St-Remi (R2) in Quebec, Canada and a dummy location (R3) were simulated with (ϕ, λ) equal to $(45.50^\circ, -72.39^\circ)$, $(45.28^\circ, -72.33^\circ)$ and $(45.43^\circ, -72.68^\circ)$, respectively. For the Spherical model, the arc distance R1_R2 (D12) = 25.207 km, R2_R3 (D23) = 31.979 km and R3_R1 (D31) = 23.893 km. For the WGS84 model, D12 = 25.211 km, D23 = 32.058 km and D31 = 23.965 km. One and two transmitters with estimated 25-dB SNR were also simulated. In this study, AOAi_Tj represents the true azimuth from receiver i to transmitter j. For the case of one motionless transmitter, the AOAi_T1 set was used, while T1&T2 sets with two indices (AOAi_T1&2(1,2)) were used for the case of two mobile transmitters.

From [1], the AOA SDs among various single-receiver scenarios with 25-dB SNR are around 2°. Thus, to simulate multi-receiver scenarios with a 25-dB SNR, normal distributions with 1° and 2° SD were used with various seeds for each snapshot as the signal model of the AOA SDs of the transmitters from each receiver. A snapshot refers to multiple wideband scans of data that are averaged to produce estimates of the AOAs and AOA SDs. Note that the estimated AOAs and AOA SDs in [1] for a single-receiver scenario can be used as the inputs from each receiver in this study. Certainly, the estimated AOAs and AOA SDs from each receiver to the transmitters can be obtained by methods other than [1].

For the 2R fixing, (R1, R2) and (AOA1, AOA2) were used. For the 3R fixing, (R1, R2, R3) and (AOA1, AOA2, AOA3) were used. In this study, 20 snapshots of simulated data were generated for each case using the sets of the AOAi_Tj with an AOA SD for transmitter j from receiver i. Among the 20 snapshots, (*r_max*, *s_max*) were calculated by finding the maximum of (*r*s, *s*s), and the corresponding CE (represented by an ellipse) was calculated for each case. In related figures, the (*r_max*, *s_max*)_Tj of the CE was displayed in km. *Vavg*_Tj, the center of a CE in (latitude_degree, longitude_degree), was calculated by averaging *V*_Tjs among 20 snapshots. The true location of Tj (True_Tj in (latitude_degree, longitude_degree)) was calculated by using zero AOA SDs and was represented by a star.

For each scenario, both of the Spherical and WGS84 models were used and the recalculated AOAi_Tj (Rec_AOAi_Tj, azimuth from Ri to *Vavg*_Tj) was calculated. The AOAi_Tj error (Er_AOAi_Tj = Rec_AOAi_Tj – AOAi_Tj) was calculated in degree to check the accuracy. The root-mean-square distance (RMSD_Tj in km) between *V*_Tjs and the True_Tj among 20 snapshots was calculated. The corresponding area of each CE (AREACE_Tj in km^2) was also calculated. For the cases of mobile transmitters, Er_AOAi_Tj(1,2), True_Tj(1,2), *Vavg*_Tj(1,2), (*r_max*, *s_max*)_Tj(1,2), RMSD_Tj(1,2) and AREACE_Tj(1,2) were calculated.

Simulated Test1 Scenario: Motionless T1 for a 2R fixing, (AOA1, AOA2) = $(225^\circ, 315^\circ)$ with (I): 1° AOA SD and *P* = 50% CE; (II): 1° AOA SD and *P* = 99% CE; (III): 2° AOA SD and *P* = 99% CE. For the Spherical and WGS84 models, among 20 snapshots, the results of (I) are shown in Fig. 3 and Fig. 4, respectively. The results of (II) are shown in Fig. 5 and Fig. 6, respectively. The results of (III) are shown in Fig. 7 and Fig. 8, respectively. The related results are shown in TABLE I.

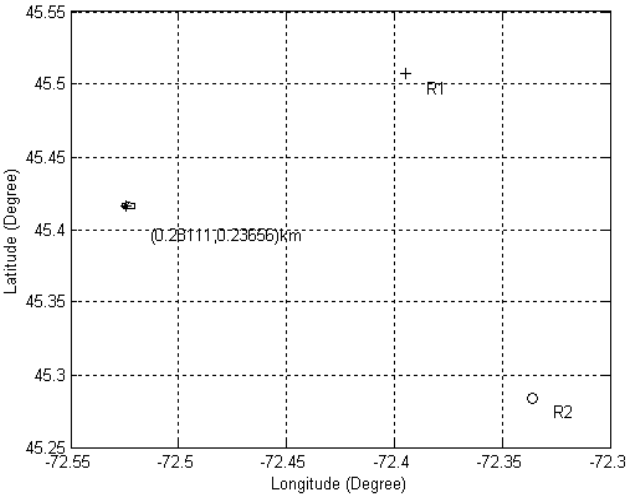


Fig. 3. Test1, Spherical, 1° SD, 50% CE, Avg_20snaps

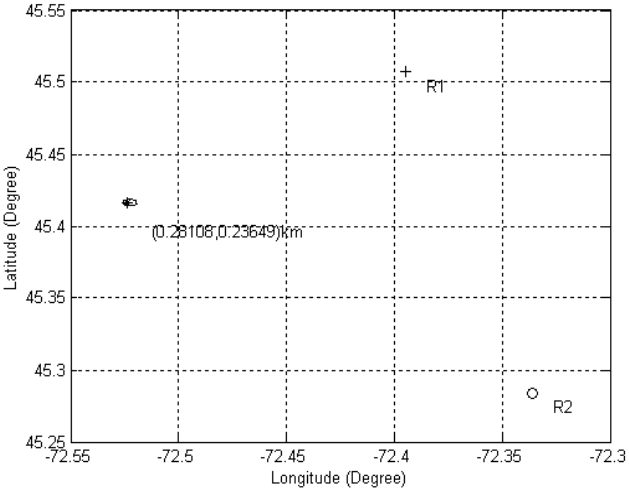


Fig. 4. Test1, WGS84, 1° SD, 50% CE, Avg_20snaps

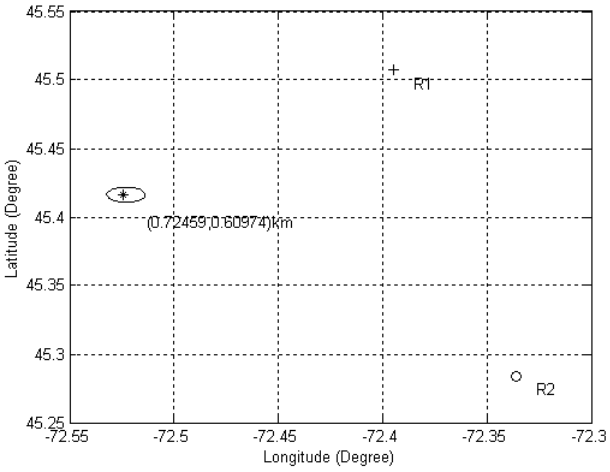


Fig. 5. Test1, Spherical, 1° SD, 99% CE, Avg_20snaps

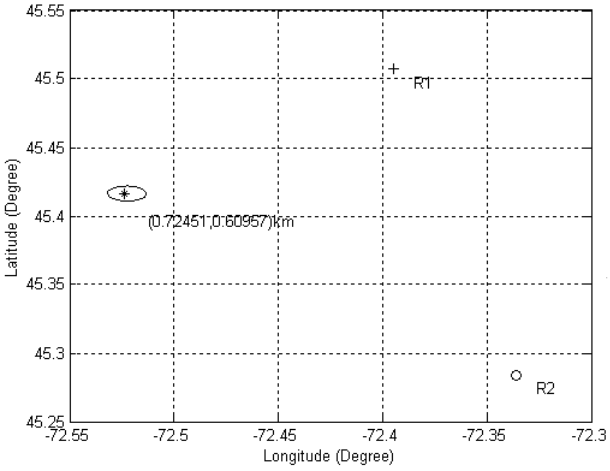


Fig. 6. Test1, WGS84, 1° SD, 99% CE, Avg_20snaps

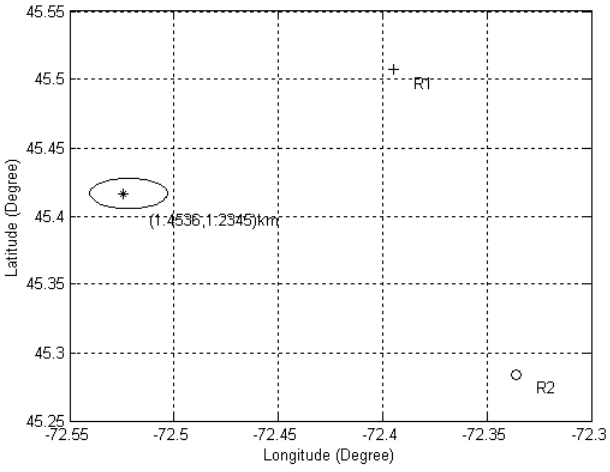


Fig. 7. Test1, Spherical, 2° SD, 99% CE, Avg_20snaps

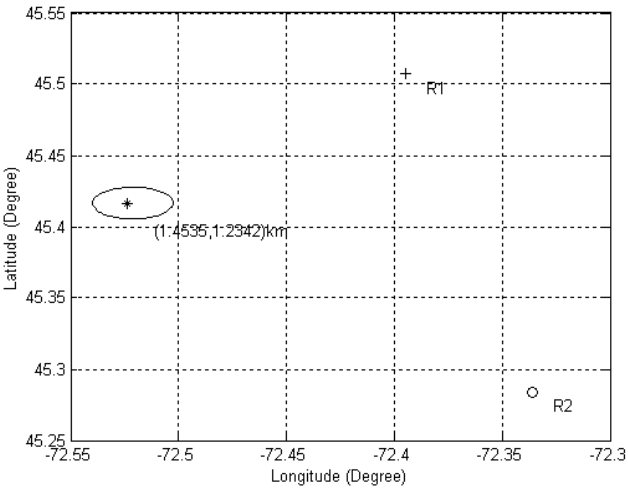


Fig. 8. Test1, WGS84, 2° SD, 99% CE, Avg_20snaps

Simulated Test2 Scenario: Mobile T1&T2 for a 2R fixing with 2° AOA SD and P = 99% CE. AOA1,2_T1(1,2) = $(225^\circ, 228^\circ), (315^\circ, 318^\circ)$, AOA1,2_T2(1,2) = $(235^\circ, 238^\circ), (325^\circ, 328^\circ)$. Among 20 snapshots, the results for the Spherical and WGS84 models are shown in Fig. 9 and Fig. 10, respectively. The related results are shown in TABLE I.

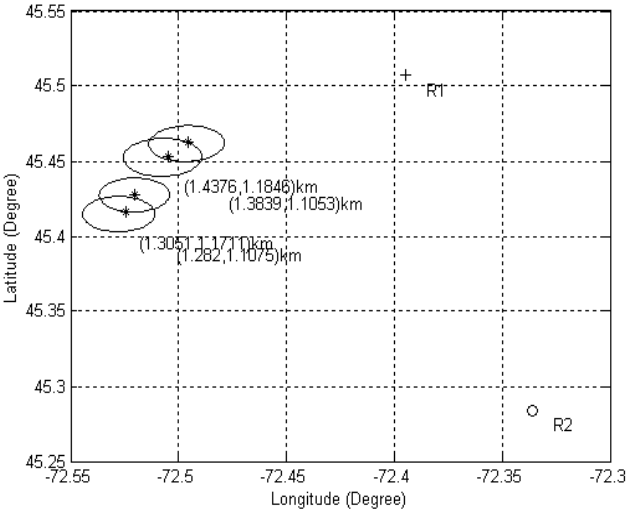


Fig. 9. Test2, Spherical, 2° SD, 99% CE, Avg_20snaps

Simulated Test3 Scenario: Mobile T1&T2 for a 3R fixing with 2° AOA SD and P = 99% CE. AOA1,2,3_T1(1,2) = $(225^\circ,228^\circ),(315^\circ,318^\circ),(90^\circ,93^\circ)$, AOA1,2,3_T2(1,2) = $(235^\circ,238^\circ),(325^\circ,328^\circ),(80^\circ,83^\circ)$.

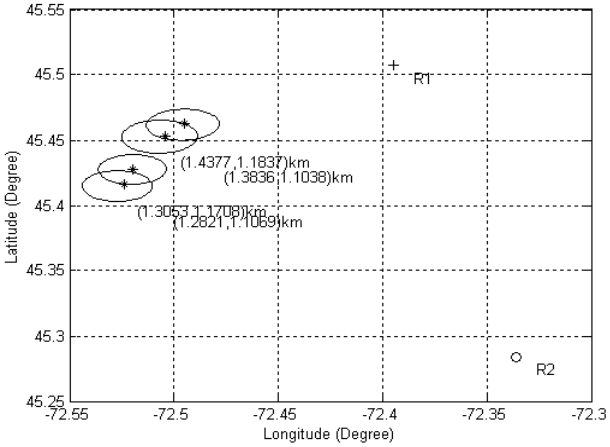


Fig. 10. Test2, WGS84, 2° SD, 99% CE, Avg_20snaps

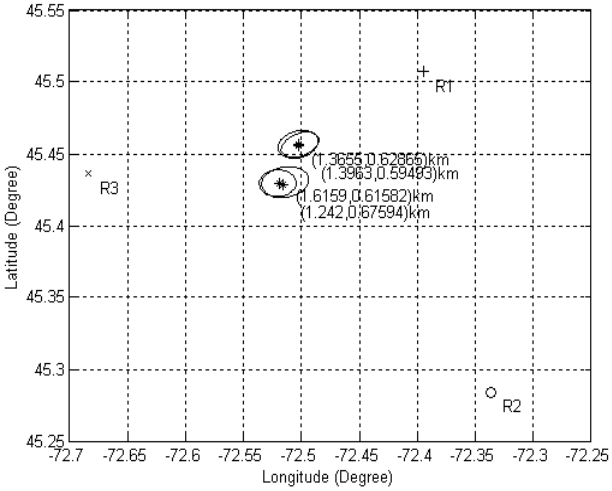


Fig. 11. Test3, Spherical, 2° SD, 99% CE, Avg_20snaps

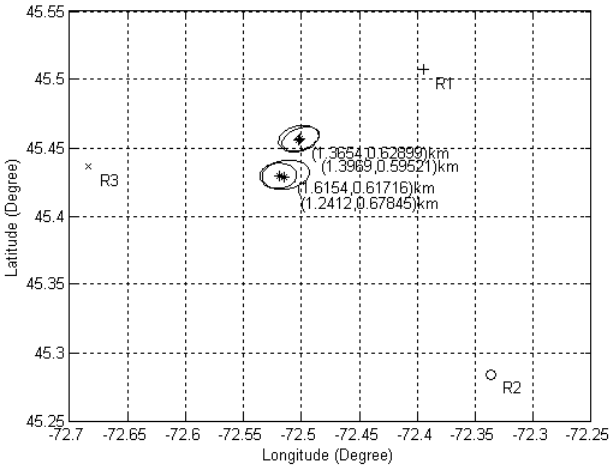


Fig. 12. Test3, WGS84, 2° SD, 99% CE, Avg_20snaps

Among 20 snapshots, the results for the Spherical and WGS84 models are shown in Fig. 11 and Fig. 12, respectively. The related results are shown in TABLE I.

From TABLE I Test1 part, one can see that the (Er_AOAi_Tj, *Vavg*_Tj, RMSD_Tj) are independent of CEs (that is, 50% and 99% CEs have the same (Er_AOAi_Tj, *Vavg*_Tj, RMSD_Tj)).

However, the higher percentage the CE has, the larger (*r_max*, *s_max*) and AREACE_Tj are. Note that the capability to locate a transmitter properly relies on the accuracy of the area enclosed by the CE, which in turn, depends on the accuracy of *Vavg* and (*r_max*, *s_max*). Among the 2R and 3R fixings, the AOA SDs are involved in calculation of both *Vavg* and (*r_max*, *s_max*) for the 3R fixing, while for the 2R fixing, the AOA SDs are only involved in calculation of (*r_max*, *s_max*).

From TABLE I Test2&3 parts, the (Er_AOA1_Tj, Er_AOA2_Tj) in the 2R fixing are smaller than their counterparts in the 3R fixing. This may come from the involvement of the AOA SDs in (17), (18) and (19) to calculate *Varg* for the 3R fixing and the effect of the AOA SDs not being used to calculate *Varg* for the 2R fixing. Nevertheless, one can see that the Er_AOAi_Tjs are small (less than 2.25 SD of AOA SDs) in both the 2R and 3R fixings, which shows the accuracy of the two geodetic models in this study is significant. However, since the AOA SDs are not used to calculate *Vavg* for the 2R fixing, with the extra information from the third receiver plus the AOA SDs, the accuracy of *Vavg* should be higher for the 3R fixing. Also, the accuracy of (*r_max*, *s_max*) should be higher for the 3R fixing with the extra information from the third receiver, this can be seen from TABLE I that AREACE_Tjs of the CEs in the 2R cases are generally larger than their counterparts in the 3R cases.

The size of a CE at a certain percentage provides direct information of how confident a transmitter is located within that CE. That is, the larger the CE is, the higher the confidence is. At a certain confidence percentage, a smaller CE indicates a more accurate estimate of a transmitter's location. Thus, with higher accuracy of *Vavg*, (*r_max*, *s_max*), and smaller AREACE, the 3R fixing can locate the transmitters better than the 2R fixing with the tradeoff that one more receiver is needed. Due to different calculations for *V*, the True_Tjs in the 2R cases have different values from their counterparts in the 3R cases. From the above discussion, one can see that the reason why the RMSD_Tjs in the 2R cases are not always larger than their counterparts in the 3R cases may be caused by the inaccuracy of True_Tjs in the 2R cases.

As per the Spherical and WGS84 models, their estimated results are close to each other in this study.

TABLE I. RESULTS FOR 20 SNAPSHOTS

	Spherical	WGS84
Simulated Test1 (I): AOA1=225, AOA2=315 dgs, 1-dg SD, 50% conf.		
Er_AOA1/2_T1(dg)	-0.21141/0.24453	-0.21142/0.24459
True_T1(la_dg,lo_dg)	(45.4162, -72.5244)	(45.4162, -72.5238)
*Vavg*_T1(la_dg,lo_dg)	(45.4164, -72.5231)	(45.4165, -72.5226)
(*r_max*, *s_max*)_T1(km)	(0.28111, 0.23656)	(0.28108, 0.23649)
RMSD_T1(km)	0.38072	0.38077
AREACE_T1 (km^2)	0.20891	0.20883
Simulated Test1 (II): AOA1=225, AOA2=315 dgs, 1-dg SD, 99% conf.		
Er_AOA1/2_T1(dg)	-0.21141/ 0.24453	-0.21142/0.24459
True_T1(la_dg,lo_dg)	(45.4162, -72.5244)	(45.4162, -72.5238)
*Vavg*_T1(la_dg,lo_dg)	(45.4164, -72.5231)	(45.4165, -72.5226)
(*r_max*, *s_max*)_T1(km)	(0.72459, 0.60974)	(0.72451, 0.60957)
RMSD_T1(km)	0.38072	0.38077
AREACE_T1 (km^2)	1.388	1.3874
Simulated Test1 (III): AOA1=225, AOA2=315 dgs, 2-dg SD, 99% conf.		
Er_AOA1/2_T1(dg)	-0.43712/0.49647	-0.43715/0.49652
True_T1(la_dg,lo_dg)	(45.4162, -72.5244)	(45.4162, -72.5238)
*Vavg*_T1(la_dg,lo_dg)	(45.4166, -72.5218)	(45.4167, -72.5212)
(*r_max*, *s_max*)_T1(km)	(1.4536, 1.2345)	(1.4535, 1.2342)
RMSD_T1(km)	0.76158	0.76167
AREACE_T1 (km^2)	5.6374	5.6359
Simulated Test2: AOA1_T1(1,2)=225,228, AOA2_T1(1,2)=315,318, AOA1_T2(1,2)=235,238, AOA2_T2(1,2)=325,328 dgs, 2-dg SD, 99% conf.		
Er_AOA1_T1(1,2)(dg)	0.32178, -0.047916	0.32177, -0.047892
Er_AOA2_T1(1,2)(dg)	-0.80228, -0.040003	-0.80221, -0.039927
True_T1(1)(la_dg,lo_dg)	(45.4162, -72.5244)	(45.4162, -72.5238)
True_T1(2)(la_dg,lo_dg)	(45.4277, -72.5205)	(45.4278, -72.5199)
*Vavg*_T1(1)(la_dg,lo_dg)	(45.4148, -72.5277)	(45.4149, -72.5272)
*Vavg*_T1(2)(la_dg,lo_dg)	(45.4275, -72.5205)	(45.4276, -72.52)
(*r_max*, *s_max*)_T1(1)(km)	(1.3051, 1.1711)	(1.3053, 1.1708)
(*r_max*, *s_max*)_T1(2)(km)	(1.282, 1.1075)	(1.2821, 1.1069)
RMSD_T1(1,2)(km)	0.72998/ 0.94992	0.73003/ 0.72995
AREACE_T1(1,2) (km^2)	4.8017/ 4.4606	4.801/ 4.4584
Er_AOA1_T2(1,2)(dg)	0.41135, -0.47495	0.41132, -0.47493
Er_AOA2_T2(1,2)(dg)	-0.48043, -0.29088	-0.48038, -0.29083
True_T2(1)(la_dg,lo_dg)	(45.4531, -72.5048)	(45.4531, -72.5043)
True_T2(2)(la_dg,lo_dg)	(45.463, -72.4955)	(45.4631, -72.4951)
*Vavg*_T2(1)(la_dg,lo_dg)	(45.4526, -72.5074)	(45.4527, -72.5069)
*Vavg*_T2(2)(la_dg,lo_dg)	(45.4618, -72.4963)	(45.4619, -72.4958)
(*r_max*, *s_max*)_T2(1)(km)	(1.4376, 1.1846)	(1.4377, 1.1837)
(*r_max*, *s_max*)_T2(2)(km)	(1.3839, 1.1053)	(1.3836, 1.1038)
RMSD_T2(1,2)(km)	0.87361/ 0.78357	0.87379/ 0.87371
AREACE_T2(1,2) (km^2)	5.3501/ 4.8054	5.3463/ 4.798
Simulated Test3: AOA1_T1(1,2)=225,228, AOA2_T1(1,2)=315,318, AOA3_T1(1,2)=90,93, AOA1_T2(1,2)=235,238, AOA2_T2(1,2)=325,328, AOA3_T2(1,2)=80,83 dgs, 2-dg SD, 99% conf.		
Er_AOA1_T1(1,2)(dg)	2.4732, 0.68219	2.459, 0.6615
Er_AOA2_T1(1,2)(dg)	4.4999, 0.34127	4.4743, 0.3225
Er_AOA3_T1(1,2)(dg)	2.6763, 0.17662	2.6639, 0.15849
True_T1(1)(la_dg,lo_dg)	(45.4289, -72.5158)	(45.4288, -72.5153)
True_T1(2)(la_dg,lo_dg)	(45.4291, -72.5194)	(45.4291, -72.5189)
*Vavg*_T1(1)(la_dg,lo_dg)	(45.4303, -72.5141)	(45.4303, -72.5137)
*Vavg*_T1(2)(la_dg,lo_dg)	(45.4295, -72.5206)	(45.4295, -72.5201)
(*r_max*, *s_max*)_T1(1)(km)	(1.6159, 0.61582)	(1.6154, 0.61716)
(*r_max*, *s_max*)_T1(2)(km)	(1.242, 0.67594)	(1.2412, 0.67845)
RMSD_T1(1,2)(km)	1.4094/ 0.68348	0.87823/ 0.87822
AREACE_T1(1,2) (km^2)	3.1262/ 2.6375	3.132/ 2.6455
Er_AOA1_T2(1,2)(dg)	0.81012, -1.657	0.81611, -1.6443
Er_AOA2_T2(1,2)(dg)	1.0897, -2.1499	1.1006, -2.1406
Er_AOA3_T2(1,2)(dg)	0.89494, -2.2815	0.90839, -2.2694
True_T2(1)(la_dg,lo_dg)	(45.4556, -72.5026)	(45.4557, -72.5021)
True_T2(2)(la_dg,lo_dg)	(45.4569, -72.5022)	(45.457, -72.5017)
*Vavg*_T2(1)(la_dg,lo_dg)	(45.4564, -72.5013)	(45.4565, -72.5008)
*Vavg*_T2(2)(la_dg,lo_dg)	(45.4566, -72.503)	(45.4567, -72.5025)
(*r_max*, *s_max*)_T2(1)(km)	(1.3655, 0.62865)	(1.3654, 0.62899)
(*r_max*, *s_max*)_T2(2)(km)	(1.3963, 0.59493)	(1.3969, 0.59521)
RMSD_T2(1,2)(km)	0.81975/ 0.95306	0.90192/ 0.90193
AREACE_T2(1,2) (km^2)	2.6968/ 2.6097	2.6981/ 2.612

V. Conclusions

The paper presents two geodetic models using data from two or three receivers to track the locations of mobile transmitters. To apply the techniques, one first has to obtain the estimated AOAs and AOA SDs from two or three receivers to each transmitter using his own method. The histogram-based algorithms in [1] can be used to calculate the estimated AOAs and AOA SDs from a receiver to various transmitters. Once the estimated AOAs and AOA SDs from each receiver to the mobile transmitters are obtained, both the Spherical and WGS84 geodetic models can be used to process the estimated data for each receiver to track the mobile transmitters with a 2R or 3R fixing.

The results showed the effectiveness of using both of the geodetic models to track mobile transmitters with a 2R or 3R fixing. One can see that the accuracy of the algorithms is significant. The 3R fixing can locate the transmitters better than the 2R fixing, with the tradeoff being that one more receiver is needed. Results using the Spherical and WGS84 models are very similar.

In general, both of the geodetic models provide a simple and efficient way to track mobile transmitters. The approach is particularly applicable for receivers using fast scanning devices and where items such as AOAs and their instantaneous SDs from several channels are reported per second. In the near future, measured data for a 2R or 3R fixing with mobile transmitters will be used to test the capability of the algorithms in this study.

VI. References

[1] M.-W. Tu and F. Patenaude, “Channel Usage Classification Using Histogram-Based Algorithms for Fast Wideband Scanners,” Proceedings of the 6th Annual International Symposium on Advanced Radio Technologies, Boulder, Colorado, March 2004.

[2] G. Strang and K. Borre, “Linear Algebra, Geodesy, and GPS,” 1997 Wellesley-Cambridge Press.

[3] J.P. Snyder, “Map Projections – A Working Manual,” U.S. Geological Survey Professional Paper 1395, United States Government Printing Office, Washington: 1987.

[4] R.G. Stansfield, “Statistical Theory of D.F. Fixing,” J. Inst. Elec. Engrs., Vol. 94, part IIIA, 15, pp. 762-770, 1947.

[5] P. Chahine, M. Dufour, E. Matt, J. Lodge, D. Paskovich and F. Patenaude, "Monitoring of the Radio-Frequency Spectrum with a Digital Analysis System: An Update," Proceedings of the 16th International Wroclaw Symposium on EMC, Poland, June 2002.

[6] B. Grünbaum and G.C. Shepard, “Ceva, Menelaus, and the Area Principle,” Math. Mag. 68, 254-268, 1995.

IP Wireless Networks for Digital Video and Data along Highway Right Of Way

Ashwin Amanna
Center for Technology Deployment -Virginia Tech Transportation Institute
Phone: (540) 231-6349
FAX: (540) 231-1555
ashwin@vtti.vt.edu

Dr. Aaron Schroeder
Center for Technology Deployment - Virginia Tech Transportation Institute

ABSTRACT

This paper presents the use of Commercial off-the-shelf (COTS) wireless Internet technology to meet the security, mobility and safety needs of departments of transportation (DOTs). COTS wireless is an economical, scalable alternative to traditional fiber optics and telephony communications solutions. A virtual Ethernet network is created along a highway right-of-way (ROW) by installing wireless point-to-point links in a serial fashion that can extend upwards of 30 miles per section from a base node. This local area network (LAN) becomes a seamless extension of the DOT's communications for field devices such as cameras, RWIS, traffic sensors, and field personnel. This paper discusses the design and architecture issues of serial wireless LANs used in a transportation setting based on real world deployments and outdoor testing on Virginia's Smart Road transportation test bed. Digital video applications along wireless networks are specifically addressed.

1.0 The needs of the DOT

Departments of Transportation (DOTs) are under increasing pressure to maintain control over their widespread infrastructure. The desire to provide secure and accurate information to travelers is pushing the existing DOT communications infrastructure to the limit. In an ideal world, fiber optics would be available along every interstate right-of-way (ROW) and along every major arterial. Dedicated home-run fibers would be available for traffic monitoring cameras and the myriad of other DOT field devices, such as weather sensors (RWIS), acoustic sensors, variable message signs (VMS), license plate readers (LPR), and HAR. The quantity of field devices that DOTs desire can number into the thousands along ever major stretch of interstate and major arterial.

The reality, however, is far from ideal. In the early days of ITS, hundreds of millions of dollars were spent on dedicated camera fiber optic systems; however, those days are long gone. The cost for fiber optic installation is more than DOTs are willing to spend, and the long time frame for bringing a system online is longer than they want to wait. Recently, several DOTs have attempted to develop public/private partnerships to deploy large-scale fiber optic networks using DOT ROW. These attempts have often failed due to lack of interest from the financially ailing telecom industry. Traditional telephony solutions, such as DSL, ISDN, phone modems, and T1s, are viable alternatives; however, the bandwidth can be limiting, each individual installation incurs a monthly bill, and these options may not be available in highly rural areas.

2.0 The wireless alternative

Recent advances in wireless technologies have made this communications medium a viable, economical, and scalable alternative for DOTs. The infrastructure requirements are a fraction of fiber optic installation, with minimal disruption to existing infrastructure. With appropriate infrastructure in place, a wireless network can go online within hours, as opposed to the months of construction required for a fiber optic network. Wireless links can be used as temporary installations until a fiber optic network becomes available or can be made permanent for long-term use. From a scalability issue, adding another wireless link over a small distance is much more reasonable than extending a fiber optic network. Furthermore, the use of open standard wireless IP devices ensures that the owners do not cubbyhole themselves into one type of proprietary technology and costly services contracts from one vendor.

To compare the costs of wireless to fiber optics and traditional telephone solutions, a mock scenario was developed. The scenario involved placing two cameras, two VMS, and one traffic speed/count sensor along an interstate ROW and collecting the data back at a DOT District Headquarters. The district office was located several miles from the interstate, and the overall length

covered between devices was over 8 miles. Three options were analyzed: 1) constructing a new fiber optic network from the DOT District HQ to each field device; 2) installing individual telephony subscription services to each device; 3) installing a wireless network extending from the DOT District HQ to each device.

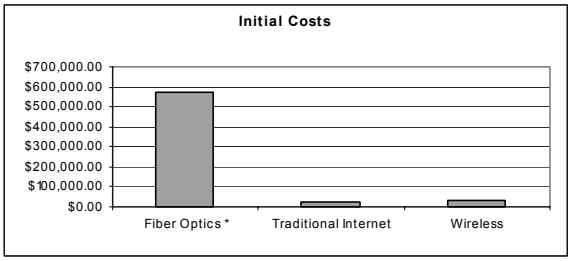


Figure 1 Initial cost comparison for mock scenario

The traditional wire line solution and wireless solutions had similar up front costs. However the recurring monthly charges applied to each device soon made wireless the more cost effective approach.

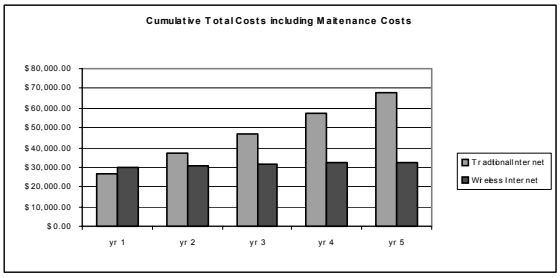


Figure 2 Five year cost comparison between wireless and traditional telephone data solution

The wireless network becomes the property of the DOT; therefore, the recurring monthly costs associated with individual ISDN and phone lines are minimized. The bandwidth capabilities of wireless can support streaming video, which is the most bandwidth- intensive application. If the network can handle video, then it can certainly handle the low data rates associated with other field devices, such as RWIS and traffic counters. Overall, wireless systems provide a good value based on the cost per installation versus the available bandwidth they provide.

A limitation on long distances with many products is the requirement of clear line-of-sight (LOS). One advantage for the DOTs is that outside of owning a mountain top or having access to cell towers, the best line-of-sight through a region is the existing interstate infrastructure. Since the DOT owns the ROW along an interstate, it can use it to create wireless LANs easily. Non Line of Sight (NLOS) or near line of sight systems are available, however their cost can be much higher than other LOS systems. If an NLOS system can replace the installation of several nodes of a LOS system, than the increased cost can be balanced by the savings of reducing the number of nodes.

2.1 COTS versus proprietary solutions

Traditionally, the transportation industry has shied away from off-the-shelf products in favor of environmentally hardened proprietary design-build solutions. While this approach has its merits, it often results in a DOT being "on the hook" to one vendor and one specific technology. This can lead to expensive service contracts and a lack of desire to change or upgrade technologies even though a better, cheaper solution usually becomes available over time. COTS products are driven by a market much larger and more dynamic than the transportation arena. This marketplace breeds a phenomenon that is to the buyer's advantage: capabilities increase while, at the same time, costs come down. In addition, COTS products tend to follow a standard, making interoperability and use of different vendors easy. In many situations, improved products roll out on three to five-year timelines. Radios that we purchased for $2,000 three years ago have been replaced with products that are three times as fast for half the cost.

Placing emphasis on the infrastructure for these wireless nodes is more prudent than focusing on the absolute state of the art military grade system. Proper grounding and clean power are a must for a sound installation. Wireless and video server technologies are improving on a 2-3 year time frame with new, better and often times cheaper products becoming available. By specifying a good solid CCTV camera combined with a COTS video server, the DOT can cheaply upgrade the system to better video server technology as it becomes available. And as long as there is a good place to hang an antenna, and power a radio, the wireless system can be upgraded relatively easy as higher throughput systems become available or as the demands on the system increase.

2.2 Suggested Architectures

The design of a WLAN will depend on several factors, including desired capabilities, terrain, and available infrastructure. In general, they can be constructed in the following architectures: single point-to-point, point-to-multipoint, serial point-to-point, and client-Access Point (point-to-multipoint).

Point-to-multipoint architectures are the most robust because each link is independent of the other link. In serial point-to-point, each previous link is dependent on subsequent links. However, for most linear highway environments, a serial application is the only option. The number of "hops" that a serial network can go depends on the technology used and the requirements placed on the network. One drawback of a serial daisy- chained wireless network is that the available bandwidth begins to degrade over successive hops. This will be discussed in greater detail later. The serial LANs can be installed as a completely self-contained network terminating at a DOT office, or they can be set up to interface with the Internet through a T1 or better connection. With this type of design, the remote network and associated field devices can be accessed from anywhere on the Internet.

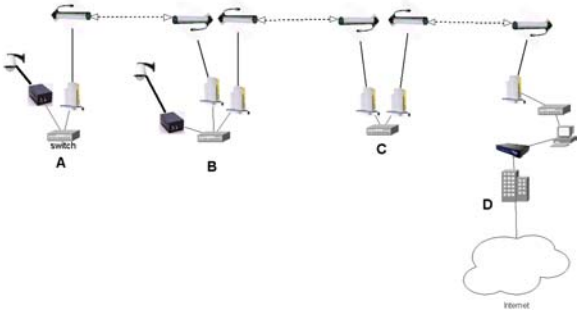


Figure 3 Conceptual diagram of a serial wireless LAN

A typical scenario on a highway ROW would involve telephone pole height towers placed approximately 1-3 miles apart, depending on the terrain and highway topography. With 802.11b technology, an 8-hop system still has over 500kb of bandwidth available at the furthest node, which is adequate to support a jpeg still or MPEG4 streaming video. A base node or Internet connection should be placed in the middle, with wireless nodes extending in two directions away from it. With an 802.11b system, this segment could span 16 to 24 miles of highway ROW.

Newer technologies, such as 802.11a and g point-to-point systems start out with a higher aggregate throughput than 802.11b and, therefore, could extend well past 8 hops and still have enough bandwidth to support MPEG 4 streaming video. The video or traffic data could then be disseminated to the agencies and public that needs it.

Figure 4 Typical installation of a repeater node and IP camera

2.3 Interference and Security Issues

With any use of unlicensed spectrum there is always the potential for interference. With linear serial wireless LAN's along DOT right of way, there are some ways in which interference can be mitigated. Narrow beam directional antennas are used with these point to point links. This limits the amount of interference from outside sources. Additionally, the distances between links are relatively small due to the distance constraint placed on the smaller heights of towers/telephone poles that will most likely be used. The smaller distance combined with using a higher gain antenna designed for long distance links also serves to limit potential interference. Finally, antenna polarities can be changed, and cycling through available channels for a quieter frequency can help mitigate around interference.

Securing these linear wireless networks follows the same strategy for securing any intranet system. There are no absolutes in security, only discrete levels of security where each stair step provides additional levels of security at the cost of time and money. The owner of the network needs to determine what they want to protect, who they want to protect it from and how much they are willing to pay in time and money. Securing a camera image from the roadside that may eventually be served out to the public may not pose a very high security priority. However, protecting a roadside variable message sign from unauthorized access certainly is a high priority.

General security recommendations include: only using backbone wireless links and not Access Point – Client links, turning on the vendor specific wireless encryption between individual links, using a router with a VPN system at the interface with the wireless LAN and the

Internet, and authorizing networking for only known MAC addresses of roadside radios and devices.

3.0 Real world deployments and current research

VTTI has worked with VDOT to deploy the serial wireless architecture in Virginia. In addition to real world deployments, VTTI also has a wireless test bed on a controlled research highway.

3.1 Route 460 WLAN

The Virginia Tech Transportation Institute (VTTI) installed its first serial WLAN with VDOT over two years ago along route 460 in Christiansburg and Blacksburg, Virginia. The system was designed to provide a communications infrastructure for digital IP cameras for traffic monitoring purposes. The wireless network extends in three directions from the VTTI Smart Road Control room. The entire WLAN is networked as a stand-alone private system. It interfaces with the Virginia Tech Internet network via a router. This particular system has a maximum of five wireless hops away from the base node. At the three endpoints of the system, a wireless access point is available for client access into the system. These APs are disabled unless required for use by field personnel.

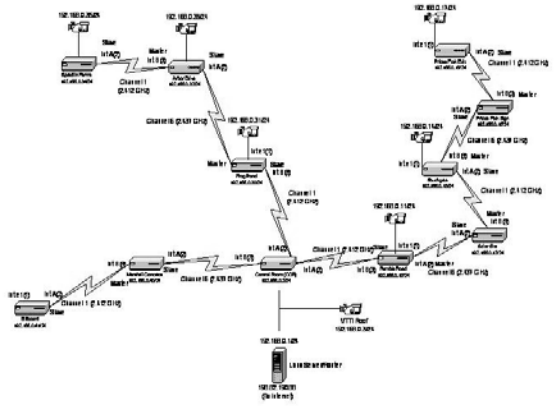


Figure 5 Network diagram of Rt. 460 Wireless Camera System

802.11b COTS products by Orinoco were used along with JVC network PTZ cameras. Most of the cameras utilize MJPEG compression. With a serial wireless network, all devices on the network share the available bandwidth. Streaming video from all the cameras at the same time places a significant draw upon the WLAN. VTTI recommends grabbing JPEG stills on a timed interval, displaying them in a matrix, and then streaming from one or two cameras at a time as needed. For a traffic monitoring concept of operations, this architecture is appropriate. This architecture is in contrast to the desired method of installing home-run fiber optic cables to each individual camera.

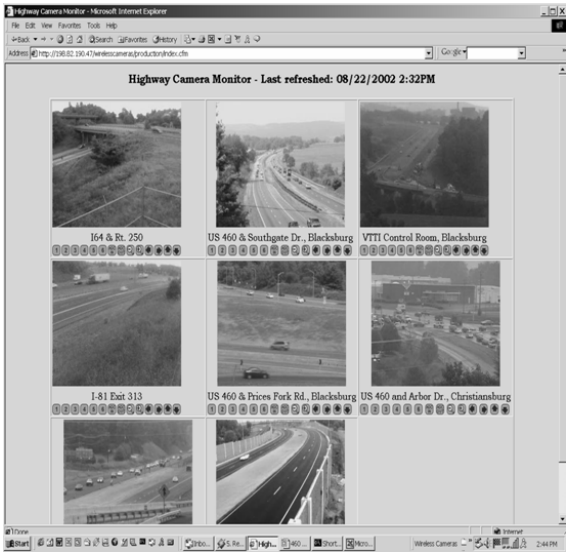


Figure 6 Matrix of JPEG still images

VTTI is currently tracking all operations and maintenance of the Route 460 WLAN to track the long term costs of the system. As part of this analysis, Knowledge Skills Assessments (KSAs) are being developed for the design, deployment and maintenance of the system to help VDOT determine what skills they have in house and what skills they will need to contract or hire to make use of WLANs in their operations.

VTTI is currently under contract with the Salem District of VDOT to design and install two 5-mile sections of WLAN along Interstate 81. The system consists of 21 nodes, 12 cameras, 7 acoustic sensors and 2 Internet backdrops. The system will utilize the newest 802.11a or 802.11g point-to-point COTS solutions paired with current COTS MPEG4 video servers used in conjunction with environmentally rated CCTV dome cameras.

3.2 Smart Road 2 mile wireless backbone with seamless AP coverage

Virginia's Smart Road is located at VTTI in Blacksburg. This highway is a closed test track used for various types of controlled transportation research. VTTI deployed a backbone serial wireless LAN down the highway and added access point coverage to create seamless coverage across 2 miles of 2 lanes and shoulders of the Smart Road. 802.11b technologies are not designed for mobile

applications, and the intent in developing this system was to analyze the ability of the 802.11b standard to operate in a mobile environment.

The wireless backbone was created using Orinoco ROR-1000 outdoor routers, as used on the Route 460 WLAN. Directional Yagi antennas were mounted on the top of existing light poles to transmit the backbone signal up and down the road. Two 120º sector antennas were mounted lower on the light poles to provide AP coverage up and down the road.

Figure 7 Yagi backbone antennas and sector AP antennas

While the technology was not designed with mobile applications in mind, the system works admirably at speeds ranging from 5mph up to highway speeds of 60mph. A client laptop inside the vehicle connected to the first AP upon entering the roadway. As the vehicle continued down the roadway, the client computer would associate with a new AP further down the road as the signal strength from the first AP grew weaker and reached a threshold level where the client looks for stronger signal. This re-association to new APs continued through the length of the roadway. During mobile tests, VTTI used network analyzing software to measure throughput from the client computer to a stationary computer back at the command center. As mentioned earlier, the available bandwidth degrades as the number of hops away from the base node increases.

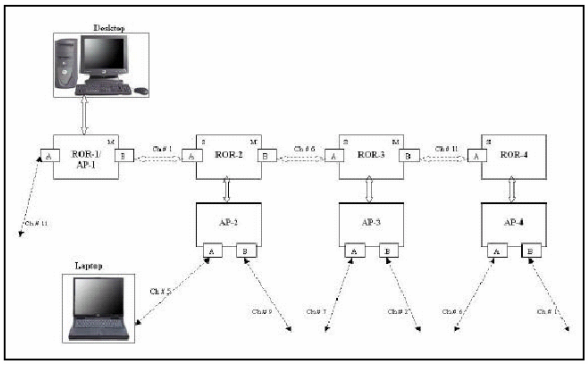


Figure 8 Diagram of Smart Road Wireless Backbone and AP system[1]

Table 1 Throughput from mobile client on Smart Road wireless network[1]

Connected to:	Static (Mbps)	20mph (Mbps)	40mph (Mbps)	60mph (Mbps)
AP-1	4.337	4.4023	4.1322	4.2823
AP-2	3.383	3.3654	3.1561	3.1568
AP-3	2.233	2.1893	2.1543	2.2058
AP-4	1.049	1.1986	1.1940	0.9824

3.3 Smart Road Reconfigurable Wireless Test bed

Currently, VTTI has developed a reconfigurable wireless test bed on the Smart Road. Using temporary antenna poles that are easy to move, networks of over 8 "hops" can be created. In addition, an AP can be added at each end of the system to connect to a client to simulate an additional two hops. At each node, custom-designed Single Board Computers (SBCs) have been installed. These mini computers are used with the top-of- the-line network simulation software to allow benchmark readings of the wireless network performance to be taken.

Figure 9 Aerial view of the Smart Road - Blacksburg, VA

The first system installed on the test bed was an 8-hop backbone Orinoco 802.11b system that could be expanded to 10 wireless hops with the addition of APs on either end. The Orinoco system is an older technology that is currently being phased out for newer products. In addition to 802.11b, we are testing 802.11a and 802.11g, point to point systems.

The test bed will be used to install varied devices in the field and then to benchmark their network performance. Criteria that will be measured include TCP and UDP throughput as well as ping delay times and signal-to-noise ratios in varied weather conditions. The UDP measurement is the most applicable for streaming digital video as it is a "connectionless" transfer. The research is not limited to just wireless devices: VTTI is also testing multiple digital video servers that assess their capabilities when used on serial wireless networks.

As discussed earlier, one of the major issues to consider when dealing with serial wireless LANs is the bandwidth degradation that occurs at each node. When dealing with devices on a network, especially digital video, the main design criteria is the bandwidth draw of the device in relation to the available bandwidth of the system.

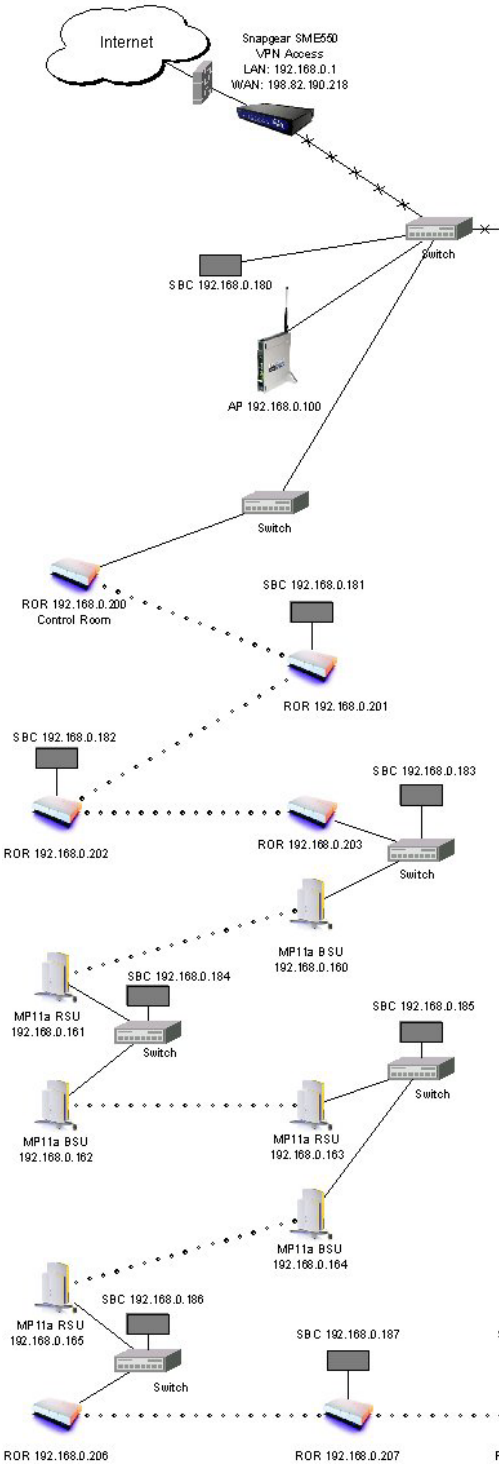


Figure 10 Network diagram of Smart Road reconfigurable wireless test bed

3.4 Analysis of Orinoco ROR-1000 802.11b Eight Hop Serial Wireless Network

The network performance of an eight hop wireless LAN using Orinoco ROR-1000 802.11b radios was characterized using NET IQ Chariot, NET IQ Qcheck, and simple FTP transfers between laptop computers. Over 3000 records per hop were taken with Chariot for UDP characterization. Chariot would not work with Orinoco ROR product line for testing TCP throughput, so FTP transfers were used instead. Using regression analysis it was determined that the UDP throughput decreases by 4% per hop. There was a linear relationship between UDP throughput and number of hops for the Orinoco System that can be described by the equation:

Throughput = 4.524 – 0.159(# of hops).*

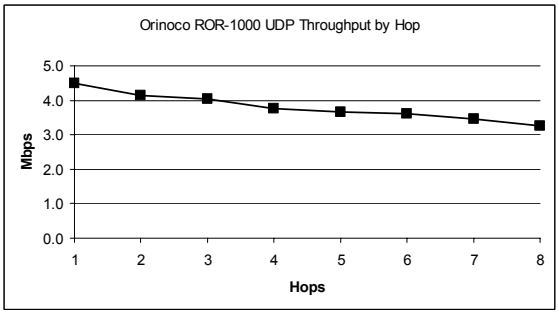


Figure 11 Average UDP throughput at each hop illustrating throughput degradation

TCP data was gathered by placing laptops at each node and performing FTP transfers between laptops. As expected, UDP throughput is higher than TCP throughput due to the connectionless nature of UDP.

3.5 Analysis of other systems in 3 hop configurations

We have procured several other radios and set them up in 3 hop configurations in order to determine if they are suitable for a serial type of architectures. Some radios, such as the Tsunami Quick Bridge products would not operate in a serial configuration over multiple hops and are more suited for individual point to point or single link point to multipoint architectures. Currently we are testing the Proxim MP.11a 802.11a radio, the Proxim MP.11 802.11b radio that has replaced the Orinoco product line, and the Buffalo Tech 802.11g wireless bridge/AP radio.

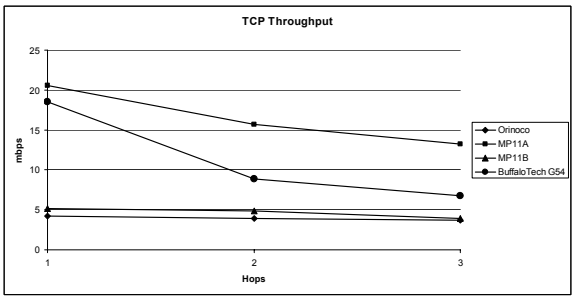


Figure 12 TCP throughput comparison over 3 hops

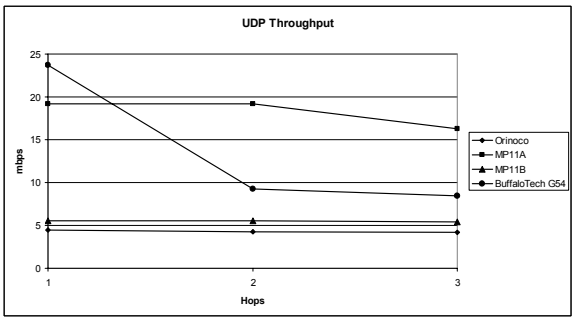


Figure 13 UDP throughput comparison over 3 hops

4.0 Design methodology for DOTs

The first steps required for designing a wireless LAN are to determine the quantity and type of devices that will be placed on it and that will be sharing the bandwidth. Cameras are by far, the most bandwidth-intensive devices that will most likely be used in the field. Therefore, they tend to be the driving force in defining the requirements of a WLAN.

Once the number of cameras is determined, the next step is to define what type of image is required and where it will be viewed. In other words, what kind of clarity, picture size, and streaming quality is required? Is a delay between when a PTZ command is issued and when the command is realized on screen acceptable? For example, in a security application where one must have the ability to pan and follow a specific vehicle or individual, a delay on the front-end compression or delay in transmission of a pan/tilt command might be unacceptable. However, this delay is perfectly acceptable in a strictly traffic monitoring application, where the defining questions are: Is traffic moving? If not, why?

Common digital video compression algorithms include MJPEG, MPEG1, MPEG2, and MPEG4. Each compression method will have a range of bit rates that the camera uses when streaming video. The chosen bit rate will affect the clarity of the picture, and depending on the

specific compression method and manufacturer it may affect the frame rate. Larger pictures sizes will naturally require more bandwidth because they are sending more information.

All of these factors need to be taken into account to determine the optimal design of the WLAN. One suggestion is to make use of JPEG stills as opposed to streaming video for general monitoring applications where several cameras are sharing wireless bandwidth. For example, if a network has 9 cameras on it, it is more bandwidth efficient to grab 9 JPEG stills every 30 seconds than it is to stream video from all 9 cameras at the same time. In addition, an operator scanning multiple cameras can focus easier on a still image than multiple streaming small scale images. Cameras can be streamed continuously as needed for more detailed monitoring.

If the terrain allows it, creating point to point links between each camera and a central location is the equivalent of home run fibers to each camera.

5.0 Next steps and conclusion

VTTI is continuing its wireless research and plans to publish a full report on the results of the network benchmarking later this year. The goal is to begin testing several more products within the coming year. VTTI's experiences with the development of the Salem I-81 WLAN will help further the current knowledge base in WLANs and their use with digital video. We are available to help any state DOT with wireless and digital video training, specification development and wireless design.

There seems to be no slowdown for the wireless industry in the near future, and new products are being developed yearly. On the national-standards level, a new standard is currently in development called 802.16. This standard is specifically for backbone point- to-point applications. It will have its own dedicated spectrum and will be designed with higher throughputs and with the demands of long-distance point-to-point communications in mind. In addition, non-line-of-sight and near-line-of-sight systems will certainly come down in price, making them more available for large-scale deployments. In addition a DSRC (Dedicated Short Range Communications) standard is in development specifically for vehicle-to-vehicle and vehicle-to-roadside communications.

Security, safety and mobility are the driving forces in ITS today. Infrastructure assets of the DOT are now considered targets and require monitoring that was never dreamt of before. The driving public craves more information, especially video to make its traveling decisions. DOTs need to place devices in the field to gather this information, which requires two main items: power and communications.

The time for accepting wireless as a viable alternative is here. The costs are well within the means to deploy systems on a permanent or temporary basis. The speed in which they can be installed means that the field device can be placed within months instead of years. While it is by no means and end-all solution, wireless is definitely a viable option for DOTs to extend their communications network.

1. F. Aziz, "Implementation and Analysis of Wireless Local Area Networks for High-Mobility Telematics," Masters Thesis submitted to Virginia Tech University, p. 124, May 2003.

Local Spectrum Sovereignty: An Inflection Point in Allocation

Mike Chartier

Director of Regulatory Policy, Corporate Technology Group, Intel

5000 W. Chandler Blvd Chandler, AZ 85226

(646) 267-4071

mike.s.chartier@intel.com

Abstract

Orders of magnitude changes in technology have dramatically altered the way radio spectrum is used since it was codified a "public resource" in the US in 1927.

Although some proponents of spectrum policy reform believe comprehending this technological change calls for a complete over-hall of US spectrum regulations, a wholesale re-write is neither justified nor possible.

Use of radio technology spans a huge range of human activity from the use of power drills and digital circuitry, to RFIDs on products, to satellite communications, broadcast, and radio astronomy.

Moreover domestic mission critical applications such as defense and public safety, as well as international obligations, preclude immediate changes to the rules governing some spectrum.

However this paper demonstrates that a true "disruptive event" HAS occurred in radio technology, and that this disruptive event enables and calls for an inflection point in spectrum regulation.

A fortuitous accident of propagation characteristics, pre-defined operating parameters, and the resulting products and usage models that evolved, effectively created a sphere of local spectrum sovereignty, empowering the local property owner with de facto spectrum rights.

Contrary to claims that the success of the FCC's unlicensed regime in general and WLANs in particular, are the result and proof of a successful "commons", it is local property owners exercising their de factor rights that have prevented a "Tragedy of the Commons".

By recognizing and codifying this de facto right, the Commission can propagate the value of this phenomenon beyond the restrictions imposed by the aforementioned accidents, for instance enabling longer range applications such as WISPs

Accordingly this paper advocates the establishment of local spectrum sovereignty, where the right to use some frequencies, and freedom from interference in using them, is attached to the property where they are used.

The critical issue, low transaction cost dispute resolution of interference claims, can be facilitated by the FCC with an ex ante definition of a per se nuisance and minimal equipment regulations.[1]

A current FCC proceeding provides a low-risk opportunity for implementing local spectrum sovereignty today in unlicensed bands, this paper propose specific regulations, which if adopted, would establish such rights.

The views expressed in this paper are those of the author, and do not necessarily represent the views of Intel.

[1] Thanks to Kevin Werbach for his insight in the role of product liability in dispute resolution in SUPERCOMMONS: Toward a Unified Theory of Wireless Communication, forthcoming in March 2004 issue of the Texas Law Review http://werbach.com/research/supercommons.pdf, and special thanks to Ellen Goodman whose forthcoming *Spectrum Rights in the Telecosm to Come*, __ SAN DIEGO L.REV. ___ (forthcoming February 2004), will be a seminal piece in spectrum management.

Introduction

Orders of magnitude changes in technology, enabling productive and novel usage and business models have dramatically altered the way radio spectrum is used since radio spectrum was codified a "public resource" in the US in 1927.

Although some proponents of spectrum policy reform believe comprehending this technological change calls for a complete over-hall of US spectrum regulations, a wholesale re-write is neither justified nor possible.

Use of radio technology spans a huge range of human activity from the use of a power drill and digital circuitry, to RFIDs on products, to satellite communications, broadcast, and radio astronomy.

Moreover domestic mission critical applications such as defense and public safety, as well as international obligations, preclude immediate changes to the rules governing some spectrum.

Accordingly reform must occur within an existing framework of commitments, and established ecosystems.

The trend to more market oriented solutions involving auctions and exclusive, flexible usage rights, should, and will likely continue.

Likewise some scenarios, such as spurious and unintentional emissions will continue to be most efficiently handled with a Pigouvian approach to pollution mitigation.

However this paper demonstrates that a true "disruptive event" has occurred in radio technology, and that this technological disruptive event enables an inflection point in spectrum regulation.

A fortuitous accident of propagation characteristics, the operating parameters defined for some frequency bands, and the resulting products and usage models that evolved, effectively limit their use to the immediate locale, empowering the local property owner with de facto spectrum rights.

Contrary to claims that the success of the FCC's unlicensed regime in general and WLANs in particular, are the result and proof of a successful "commons", it is local property owners exercising their de factor rights that have precluded a "Tragedy of the Commons".

By recognizing and codifying this de facto right, the Commission can propagate the value of this phenomenon beyond the restrictions imposed by the aforementioned accidents, for instance enabling longer range applications such as WISPs.

This paper proposes exactly that, by advocating local spectrum sovereignty- where the right to use some frequencies, and freedom from interference in using them, is attached to the property where they are used.

The salient issue, dispute resolution of interference claims, can be accomplished by an ex ante definition of a per se nuisance by the FCC. Moreover that the applicability of this standard (the transaction cost in detecting and identifying an interferer) could also be enabled in current a FCC proceeding.

This paper provides the concrete steps for implementing local spectrum sovereignty today in unlicensed or license exempt bands via this proceeding.

In part 1 I show how technology has created a sphere of spectrum activity over which federal jurisdiction is no longer justified nor required.

In part 2 I describe why there is no "commons" in this sphere.

In part 3 I respond to the FCC's current NPRM on Cognitive Radios to show how a local sovereignty solves the Commission's objective of enabling longer-range uses for rural applications, while retaining the tremendous innovation fostering characteristics of the current unlicensed regime. I address the major concerns of dispute resolution, impact on innovation, QoS, market incentives, and possible dislocations.

1. Technology mitigation of the justification for Federal regulation

1.1. Federal Justification

1.1.1. Interstate

Prior to initial US regulation and for decades after, the perceived essence of radio was its ability to conquer distances, bridge oceans. This lack of spatial constraint, or borderless characteristic was a key element responsible for the belief in spectrum as "public property".

In its early years "how far away" quickly became the dominant metric for users. "Advertisements for receiving sets reflected the obsession with distant radio stations"…[the] "lure of distant stations grips the radio fan"[2]

Compounding this early end-user "pull" for distant access (which was supplanted as the technology matured for a focus on content) was a "push" from broadcasters to reach more and more listeners. Although "localism" quickly emerged as one of the pillars of "public interest", "distance" provided the key

[2] Smulyan, S. (1994), Selling Radio. Smithsonian Institute Press, Washington. Pg 15

driver for initial user demand and advances in receiver technology; as well as supply-side economies of scale and increases in transmitter power and range.

It was this "Interstate" feature that provided fundamental justification for federal regulation in the first place.

As Bensman puts it "here was the unique approach to the right of federal control of the air-waves, by affirming the right of authority via the commerce clause of the Constitution, which to this day underpins government control."[3]

1.1.2. POLLUTION

The second classic justification for federal regulation is that in some cases because a large number of entities could be affected by emissions, it was more efficient for the government to regulate rather than allowing the parties to negotiate. A Coase himself states:

> In the standard case of a smoke nuisance, which may affect a vast number of people engaged in a wide variety of activities, the administrative costs might well be so high as to make any attempt to deal with the problem within the confines of a single firm impossible. An alternative solution is direct government regulation. Instead of instituting a legal system of rights, which can be modified by transactions on the market, the government may impose regulations which state what people must or must not do and which have to be obeyed. Thus, the government (by statute or perhaps more likely through an administrative agency) may, to deal with the problem of smoke nuisance, decree that certain methods of production should or should not be used (e.g. that smoke preventing devices should be installed or that coal or oil should not be burned) or may confine certain types of business to certain districts (zoning regulations).[4]

1.1.3. FREE SPEECH

The last justification for federal intervention was the realization that broadcasting provided a powerful medium for disseminating information and shaping public opinion.

The 1920's saw explosive growth in broadcasting, irrational exuberance applied to the stock prices of the new pioneering companies, and no known way to extract any profit. As Hoover stated in 1924 "The largest unsolved question is the entire problem of remunerations for the broadcasting stations."[5]

Although the business model of advertisement supported entertainment, was yet to be decided, it became obvious in the early twenties that the more households you could reach with one broadcast, the more you could spread (once you figured out how) the cost of high priced entertainment.

And so the correct architecture of the system was known prior to figuring out the revenue stream. The task was to simultaneously provide the same content to geographically dispersed stations to reach a larger audience.

David Sarnoff argued that "as long as 559 broadcasting stations in this country are maintained, the situation is hopeless," and found the solution in a few super-power stations which will reach every home in the country"[6]

The vision that Sarnoff evoked, that of a giant broadcaster, blanketing the country with a single signal, contributed to the true motivating factor for government control.

Berle and Means in their contemporary study of American business document the prevailing concern, if not apprehension, of mega corporations and their management: "the corporation has, in fact become both a method of property tenure a means of organizing economic life." "whereby the wealth of innumerable individuals has been concentrated into huge aggregates…The power attendant upon such concentration has brought forth princes of industry, whose position in the community is yet to be defined"[7]

It was more the threat of monopoly control of voices, rather than lack of competition in the economic sphere, or a technical interference over-exploitation problem, that drove regulation of the airways.

As Hazlett pointed out in 1990 "In the event any misunderstanding had arisen that placed interference control as the primary aim of federal legislation, Dill was pointedly direct "there is much agitation and much

---

[3] Bensman, M. (2000) The Beginning of Broadcast Regulation In The Twentieth Century. McFarland & Company, Inc., North Carolina. Pg 100

[4] The Problem of Social Cost; RONALD COASE; originally published in *The Journal of Law and Economics* (October 1960).

[5] Bensman, M. (2000) The Beginning of Broadcast Regulation In The Twentieth Century. McFarland & Company, Inc., North Carolina. 126

[6] ibid46

[7] Berle & Means (1933), The Modern Corporation and Private Property. The Macmillan Company, New York. Pg. 1

resentment to day over the chaos of the air, but that does not concern me so seriously as the problems of the future. Chaos in the air will be righted as a matter of business" 'Dills concerns were devoted to monopoly and political fairness over the airwaves, both derived from his belief that radio broadcasting would become an important, powerful means of expression"[8]

Leaving a critical means of communication to the mercy of the market was unacceptable for Congress.

1.2. Mitigation, what has changed.

1.2.1. Interstate to local

In the 21st century conquering distance is as attractive as it was in 1927.

The use of radio communications via satellite, images broadcasted from mars, and radio astronomy have pushed literally to the far corners of the universe.

However technology and business models have also driven the Ether to be used for decades more and more as a short haul carrier from a few miles as in cellular, to a few yards as in remote controls and garage door openers, down to inches or feet as in Bluetooth or RFIDs.

In particular the use of radio as a means to network computing devices, Wireless Local Area Networks (WLAN), has created tens of billions of dollars of economic value.

Radio spectrum usage that is very localized to a specific property, whether a home, office building, Starbucks, Airport, or Washington Square Park appears to be entrenched on a global basis.

1.2.2. Speech

Contrasted with 1927, the public has a myriad of available ways to electronically access information and entertainment. Indeed from cable and satellite networks to the Internet., "information overload" is often cited as a problem with our vast choices.

Moreover, government regulation by restricting private use, might actually hinder free expression, rather than protect a plurality of voices as it was initially intended.

1.2.3. Pollution

As mentioned above, a Pigouvian approach is only warranted where the number of parties involved would make negotiations very costly.

In the area under discussion in this paper, where emissions are restricted to the immediate vicinity, only a small number of parties are involved, and so transaction cost would not be high enough to justify the regulatory burden.

2. Commons Myth, De Facto Local Control

The dramatic success of the FCC's unlicensed regime, and Wi-Fi in particular, has been claimed by some as proof of a viable "commons".

However this is a misinterpretation of the situation.

While the arbitration mechanisms of the 802.X standards allow for coexistence of a finite amount of similar devices, too many devices trying to operate simultaneously will degrade the system just like any other network.

Moreover wi-fi devices also must share the spectrum with a panoply of device that have no means of coordination such as cordless phones, baby monitors and micro-wave ovens.

The fundamental reason that a so-called "Tragedy of Commons" has been avoided for the bulk of Wi-Fi deployments, is that the corporate or campus IT department or homeowner controls the deployment of devices in their domain.

The combination of low power limits and propagation characteristics in the unlicensed bands, limit the effective range of these devices to the immediate vicinity.

The property owner, by regulating the operation of devices in the area of their control maintains a working environment for all.

This is a highly efficient mechanism. Similar to a firm internalizing transaction costs, the business, homeowner, or campus administrator trades offs which devices to allow based on their utility and impact on others.

2.2. Business

In addition QoS issues, security concerns have driven corporate I.T. departments to regulate the deployment and use of WLAN equipment such as Intel's policy on non-I.T. department deployed, or "experimental" WLANS:

> Failure to fulfill the above terms and conditions [*for non- IT WLANs*] will result in I.T.'s

[8] Hazlett, T. (1990); The Rationality of U.S. Regulation of the Broadcast Spectrum. Journal of Law & Economics, Volume XXXIII, No. 1, April1990

disconnecting and or taking possession of the Experimental W-LAN Access Points.

2.3. Campus

Campus administrators regulate the deployment and use of Wi-Fi competing devices, as demonstrated by Carnegie Mellon's policy:

> While we will not actively monitor use of the airspace for potential interfering devices, we will seek out the user of a specific device if we find that it is actually causing interference and disrupting the campus network. In these cases, Computing Services reserves the right to restrict the use of all 2.4 GHz radio devices in university-owned buildings and all outdoor spaces on the Carnegie Mellon Campus. [9]

2.4. Home

I had installed wired Ethernet (CAT 5) in my home and so deployed a wireless LAN only recently when my wife got a lap-top.

She discovered while using her lap-top in a room far away from the access point, that simultaneous use of our (expensive) 2.4 GHz phone would cause her internet connection to stop working. Accordingly we replaced the expensive 2.4 GHz phones with (cheaper) 900 Mhz ones, problem solved.

However later wanting the caller ID feature on the 2.4GHz phone she reconnected it in a different location, trading off a smaller amount of interference for the added feature.

This behavior is the epitome of an efficient Coasian firm- internalizing transactions costs and optimizing resources in a way neither regulation nor market transactions could achieve.

2.5. Common Mistake

Commoners erroneously believe that the way to propagate the success of this regime is for the Commission to mandate specific service requirements (broadband packet based digital transmission) for bands, and specific arbitration or sharing "etiquette" rules for equipment.

Attempting to substitute the highly efficient and successful market mechanism with an ex ante definition of "fair" spectrum use is problematic at best and probably impossible. This is because the "Digital Migration" has de-coupled service from transport, there is no longer a fixed "service" (such as voice call) or use that can be "achieved" with some minimum spectrum use.

Devices operating in unlicensed spectrum exploit many different technical parameters in their use of spectrum, such as power, bandwidth, time, etc.

Attempting to define ex ante transmit power etiquettes are particularly problematic. Modern air interfaces maximize bandwidth as a function of S/N, which of course varies with transmit power. "Range" is no longer a simple fixed parameter: It's a given bandwidth at a certain distance, that's dependent on transmit power.

3. Practical Steps

As mentioned earlier de facto land-owner spectrum sovereignty is an accident enabled by a fortuitous coincident of the FCC power limits, propagation characteristics of the particular frequencies, and resultant physical nexus of control.

However this breaks down when different physical areas, or changes in power or frequency ranges are considered, and there is strong economic incentives to propagate the success achieved beyond these physical constraints.

In particular, the current limits severely curtail, or preclude many longer-range applications that would be very beneficial in rural environments.

The solution is to recognize and codify the de facto right into a de jure one.

The Commission is addressing this exact issue in a recently adopted Notice of Proposed Rulemaking on Cognitive Radio Technologies & Software Defined Radios[10]

In this proceeding the Commission recognizes that the current power limits for certain part 15 devices, unduly preclude their application in rural settings.

> The lower population density and the greater distances between people in rural areas can make it difficult for certain types of unlicensed operations at the current Part 15 limits to provide adequate signal coverage. Such operations include Wireless Internet Service Providers (WISPs) and wireless LANs operated between buildings or other locations with a large separation between transmitters. These operations could potentially

[9] Airspace Guideline for 2.4 GHz Radio Frequency at Carnegie Mellon University

[10] ET Docket 03-108; In the Matter of Facilitating Opportunities for Flexible, Efficient, and Reliable Spectrum Use Employing Cognitive Radio Technologies & Authorization and Use of Software Defined Radios

benefit from higher power limits in rural areas, which would result in greater transmission range.[11]

Accordingly this proceeding provides an excellent vehicle for the implementation of local spectrum sovereignty, where its application can achieve precisely the goal the Commission seeks.

Moreover, it would provide QoS and innovation benefits beyond what the Commission envisions.

3.1. Defining the Right

A fundamental finding of the Spectrum Policy Task Force was that spectrum policy models must be "based on clear definitions of the rights and responsibilities of both licensed and unlicensed spectrum users, particularly with respect to interference and interference protection.[12]"

Rights in the unlicensed space have heretofore been constructed as a right to act, to use certain equipment with certain operating parameters such as power, frequency, modulation etc. In fact users are specifically forbidden from claiming any interference.

In the current proceeding the Commission is again proposing to define the right as the ability to use a higher output power base on a sensing of the environment. However such a proposal may be problematic.

Fundamentally the issue is potential interference at the receiver, and so sensing the environment at the transmitter may be a poor substitute.

The Commission is attempting to guess at a universal transmit power to balance increased interference vs added utility, over a myriad of settings.

The market is the only method found successful for solving such poly-centrc problems of determining what users making what trade offs, in what settings, should be made.

Simply, the solution is to tell potential operators you MAY transmit at a higher power, UNLESS you cause interference to someone.

To enable such a paradigm the key right to define is one of freedom from interference; the metric of what constitutes establish a per se nuisance.

Once established it the gives parties the certainty needed to negotiate and arrive at optimum solutions.

This de-centralization of the dispute process also allows local authorities (whether the super of an apartment building, police, or even courts) to settle disputes.

Therefore rather than trying to establish a maximum output power for a transmitter, the commission should define an interference level, which when demonstrated to exist in a premises constitutes a per se nuisance, from which a user has the right to claim relief.

In order to minimize dislocation, a level should be set that closely approximates typical existing conditions.

For instance, the Commission should look at typical scenarios such as adjacent Wi-Fi users in an apartment building.

Using existing Part 15 the maximum power limits, and allowing for free space propagation loss and losses for intervening walls, a value of -50 dbm might be a viable threshold for a per se nuisance.

The FCC rulemaking process would vet all the issues with concerned parties to determine a good value

Accordingly rather than a Cognitive radio regulation, a Local Spectrum Use regulation should be codified into part 15 as followed:

§ 15.206 Local Spectrum Use

(a) Devices operating under the provisions of § 15.247 may operate with a power level greater than the maximum permitted in these sections under the conditions specified in paragraph (b) of this section.

(b) Owners of property may operate intentional radiators on their property at the higher power limits specified in paragraphs (a) subject to the following conditions:

i. Operators must register on the FCC website www.xxx.

ii. Operators of devices must cease operation if interference is demonstrated to be caused by them on property not their own. For the purpose of this paragraph such demonstration shall be:

1. A signal level in bands designated in 15.247 of –50dbm with a measurement bandwidth of 1.25 MHz, measured in accordance with procedure defined in xx; or

2. An indication from a device certified under this part that incorporates a mechanism for monitoring the band and detecting and displaying a signal level in access of –50dbm , and ID of the

[11] ibid @ 53
[12] SPTFR @ 3

interfering signal, and/or approximate location of the interfering transmitter which can be correlated to the FCC database of registered operators.

3.2. Dispute resolution

As articulated by Ellen Goodman in her forthcoming piece[13]:

> neither side has examined with any degree of specificity how its proposed model of spectrum management would actually function. Interference is the eight hundred pound gorilla in the spectrum policy debate. … despite the centrality of interference to the current administrative system, and to any legal regime in the future, surprisingly little thought has been given to the variety of interference scenarios and their relevance to the law.

Adoption of the above-proposed rule would establish a definite and verifiable metric, which would make ascertaining infringement simple and hence minimize transaction costs.

Equipment manufacturers would take advantage of the new regime to gain competitive advantage for their products.

Wi-Fi devices already have capabilities for monitoring signal level, devices that allowed users to protect their "air-space" by proving interference would have added value.

Current 802.x WLANs broadcast identifiers. A validated level about the nuisance level correlated with the ID would prove causation.

Alternately in the case where an ID was not embedded or readable in the signal, a first order approximation of the direction of the signal and its received strength, matched against the FCC registration database (all of which could be automated), should facilitate easy identification of interferers.

And so market forces would be enough to make sure devices that accurately detect and "defend" local spectrum get deployed because companies would advertise the feature as allowing consumers to protect rights.

Once an infringement was established, negotiations could then proceed at the pace dictated by the parties.

A WISP operating at a higher power, which was found to cause interference would have multiple means to settle with the claimant such as:

- They could offer to reconfigure the claimants home network to make it more immune to interference, for instance by adding access points; or
- They could offer to compensate the user with free Internet service; or
- The WISP could reconfigure its own network to eliminate or lesson the interference.

3.2.1. Market Incentives

In addition to the incentives for equipment mentioned above, a market would also develop for ancillary products that mitigate interference to allow for higher power such as directional antennas.

Likewise it might be expected, as is the case with other property, that owner ship would create incentive for investment to improve the "property", for instance people might take proactive measure to make their homes more immune from noise.

3.3. QoS

Another problem with the existing unlicensed regime is that commercial entities who wish to offer a commercial service have no way to guarantee a Quality of Service to their customers. By establishing definite rights from interference, WISPs would now have a mechanism to calculate costs involved in delivering a fixed Qos. Local owners would be free to sell or lease his rights to a larger aggregator, who would then be able to guarantee a level of service.

3.4. Impact on Innovation

Perhaps one of the greatest critiques of current spectrum regulation is that new or novel uses are ex ante prohibited until they can prove non-interference to existing users.

This represents a huge cost of entry and has a chilling effect on innovation.

Establishing local sovereignty will finally create an environment for low cost experimentation by permitting innovation until an ex post interference finding.

Also to be considered is the impact of a fixed standard for RF nuisance. Unlike the standard for an audio nuisance (where human hearing isn't likely to change and 45 db will always be annoying) -50 dbm might, as technology evolves, look more and more arbitrary.

[13] Supra 1

New uses and technology might require a greater immunity from interference, or higher power applications might generate a higher potential level of interference.

However the establishment of a level now would not preclude innovation in either case.

If a lower level of interference is of value to a property owner, a market could be expected to evolve for methods to make a property more "quiet" for instance UV coatings on windows also reduce emissions.

Likewise if a higher transmit power application appeared to have great potential, a market would evolve to contain higher emissions to the immediate vicinity, such as with directional antennas. Also the provider always has the opportunity to negotiate with claimants for the right to transmit.

3.5. Possible Dislocations and Disruptions

Giving users of spectrum in the unlicensed band a right to claim freedom from interference could invoke images of upsetting the existing equilibrium resulting in rampant interference claims overcrowding local courts. It is unlikely that the codification of the 45-decibel audio limit by New York City in 1972, created such rampant noise nuisance claims.

Regardless, in this instance the issue is moot because the new rules as I have proposed apply only to new, higher-power operation.

Operation of existing devices of lower power would constitute a safe harbor against interference claims.

Over time as the benefits accrued from innovation unleashed by the establishment of local spectrum sovereignty, market forces would develop and deploy products and architectures that would take advantage of the new regime.

In turn this ecosystem (of new products, architectures and usage models) based on local sovereignty would becomes the dominant force. This would allow the existing regime to be sunset with minimal dislocation.

4. Conclusion

Wi-Fi works because of de facto land-owner rights.

Recognizing and codifying these rights would propagate this success allowing its application beyond restrictions imposed by regulated physical limits, while preserving the great innovation fostering characteristics of the current unlicensed regime.

The current FCC NPRM provides an excellent, low-risk opportunity, to trial local spectrum sovereignty.

Channel Usage Classification Using Histogram-Based Algorithms for Fast Wideband Scanners

Ming-Wang Tu and François Patenaude
Communications Research Centre Canada
3701 Carling Avenue, Box 11490, Station H
Ottawa, ON, Canada, K2H 8S2
Tel: 613-998-9262, 613-990-5878, Fax: 613-990-8842
ming-wang.tu@crc.ca, francois.patenaude@crc.ca

Abstract

The paper presents a set of histogram-based algorithms to determine the wireless channel usage or station occupancy of fixed frequency channels.

The results of this study show the effectiveness of using the histogram-based algorithms to analyze and estimate the channel usage with a significantly small number (down to 25) of scanned samples. Both simulated and measured data sets were processed and their results show significant accuracy for various scenarios, including the line-of-sight (LOS) and multipath frequency modulation (FM) signal cases. However, the results of the LOS amplitude modulation (AM) signal cases show multiple peaks in the signal-to-noise ratio (SNR) histograms due to the amplitude variation of the AM signals. Thus, the histogram-based algorithms in this study solely cannot be used to classify the channel usage of the LOS AM cases. The AM type and its characteristics have to be identified first by other algorithms. Then the averaged AM results from the histogram-based algorithms can be used as estimates for classification.

In general, the histogram-based algorithms provide a simple and efficient way to classify the LOS channel usage. The approach is particularly applicable for fast wideband scanning devices where items such as power levels, SNRs, angles of arrival (AOAs) and AOA instantaneous standard deviations (ISDs, that is, SDs at each scan) from several channels are reported per second.

I. Introduction

For each transmitter-receiver scenario, a wideband scanning device such as the CRC's Spectrum Explorer (SE) [1] could be used to scan, collect and preprocess wireless signals to form a channel data set. The wideband nature of the scanner allows several channel measurements to be taken in a short period. Each channel may have multiple users from various AOAs with different SNRs. Users may use various modulations such as AM, FM, etc. Each channel data set resulting from the preprocessing by the scanning device may include information such as power levels, SNRs, AOAs and AOA ISDs. The preprocessed SNR, AOA and AOA ISD samples are used to generate related histograms for active channels. Statistically, each lobe of a variable histogram from various scans is linearly related to the distribution density of that variable, given that the number of scans is large enough [2]. Thus, those histograms provided valuable information of related signals. The purpose of this study is to develop a set of histogram-based algorithms to determine the wireless channel usage or the station occupancy of a fixed frequency channel.

II. Concept and Algorithms

The results presented in [3] make the assumption that the channel impulse response $h(t,\theta)$ as a function of time and azimuth angle is a separable function, or

$$h(t,\theta) = h(t)h(\theta), \quad (1)$$

where $h(t)$ is the time domain impulse response, and

$$h(\theta) = \sum_{l=0}^{\infty}\sum_{k=0}^{\infty}\beta_{kl}\delta(\theta - \Theta_l - \omega_{kl}), \quad (2)$$

is the angular domain impulse response, where β_{kl} is the received signal amplitude of the *k*th arrival in the *l*th cluster, Θ_l is the mean azimuth AOA of the *l*th cluster, and ω_{kl} is the azimuth AOA of the *k*th arrival in the *l*th cluster, relative to Θ_l. In this study, for non-amplitude-varying signals (like FM), the number of AOA clusters (histogram lobes) corresponds to the number of transmitters and the number of arrivals within an AOA histogram lobe corresponds to the potential multipath effect, given the assumption that the AOA histogram lobes are distinct. This study doesn't include the multi-user scenarios where both the AOA and SNR histogram lobes are not distinct.

Since there is a quasi-linear relationship between the power levels and SNRs, the SNR histogram is linearly related to the histogram of power levels which are proportional to energy levels, given the assumption of

additive white Gaussian noise. From [4] and [5], the cluster energy is linearly related to the relative delay of signals, assuming that all incident waveforms are identical. Thus, the relative delay of signals is linearly related to SNRs. That is, the lower the SNRs are estimated, the farther apart the transmitters (users) are located, assuming the same transmitted power level is used. For amplitude-varying signals (like AM), the effect of intra-signal amplitude variation also needs to be considered.

The SNR, AOA and AOA ISD histograms from the SE's preprocessed data were generated from 1000 scans for various scenarios. Note that the AOA and AOA ISD histograms were only generated for each valid SNR histogram lobe (lobe_peak/*HistSNR*_peak > 0.8). For a SNR histogram, its isolated lobes *CS(M,3)* were found where *M* is the final grouped lobe number; *CS(k,1 to 3)* are the leftmost, peak and rightmost sample indices of the *k*th pre-lobe. For each *k*, the pre-lobes are grouped to *M* lobes using the (Matlab) algorithm:

```
if (CS(k,3)-CS(k,2))<=2; if CS(k+1,2)-CS(k,2)<3
   if HistSNR(CS(k,2))>=HistSNR(CS(k+1,2));CS(k,3)=CS(k+1,3);
   else CS(k,2)=CS(k+1,2); CS(k,3)=CS(k+1,3); end; end; end.
```

The purpose of this grouping operation is to distinguish distinct clusters that are taken to represent separate transmitters.

The averaged ISD (AISD) was obtained by weight averaging the AOA ISD histogram lobe. The weight average of a histogram lobe is:

$$WA = \sum_{m=m_s}^{m_e}[m \times Hist(m)] \Big/ \sum_{m=m_s}^{m_e} Hist(m), \qquad (3)$$

where m_s, m_e are the start and end sample indices within a histogram lobe, respectively.

For an AOA histogram, its isolated lobes *CA(N,3)* were found where *N* is the final grouped lobe number; *CA(k,1 to 3)* are the leftmost, peak and rightmost sample indices of the *k*th pre-lobe. For each *k*, the pre-lobes are grouped to *N* lobes using the (Matlab) algorithm:

```
if (CA(k,3)-CA(k,2))<AISD
  if HistAOA(CA(k,2))>HistAOA(CA(k+1,2));CA_C(k,1)=CA(k,1);
    CA_C(k,2)=CA(k,2); CA_C(k,3)=CA(k+1,3);
  else CA_C(k,1)=CA(k,1); CA_C(k,2)=CA(k+1,2);
    CA_C(k,3)=CA(k+1,3); end; n_cc=2; n_rc=N'-n_cc;
  while (CA_C(k,3)-CA_C(k,2)<=2*AISD)&n_rc>0
        if HistAOA(CA(n_cc+1,2))<=HistAOA(CA_C(k,2))
          CA_C(k,3)=CA(n_cc+1,3);
        else CA_C(k,2)=CA(n_cc+1,2); CA_C(k,3)=CA(n_cc+1,3);
        end; n_cc=n_cc+1; n_rc=N'-n_cc; end;
elseif CA(k+1,2)-CA(k+1,1)<=AISD
       if HistAOA(CA(k+1,2))>=HistAOA(CA(k,2))
CA_C(k,1)=CA(k,1);CA_C(k,2)=CA(k+1,2);CA_C(k,3)=CA(k+1,3);
       elseif HistAOA(CA(k+1,2))<HistAOA(CA(k,2))
CA_C(k,1)=CA(k,1);CA_C(k,2)=CA(k,2);CA_C(k,3)=CA(k+1,3);
       end; end;
CA(k,1)=CA_C(k,1);CA(k,2)=CA_C(k,2);CA(k,3)=CA_C(k,3); end;
```

where *N'* is the grouped lobe number computed on the fly. Again, the goal is to distinguish separate clusters and therefore different transmitters. The standard deviations (SDs) of AOAs were then calculated from the isolated AOA histogram lobes.

Each valid SNR histogram lobe was processed independently. The weight average of each valid SNR histogram lobe is an estimate of the actual SNR of the detected signal. Within each valid SNR histogram lobe, the angularly weighted average of each valid AOA histogram lobe (lobe_peak/*HistAOA*_peak > 0.5) is an estimate of the actual AOA of the detected signal. The angularly weighted average (in degree) of a histogram lobe is:

$$AWA = \frac{180}{\pi} \times Ang\left[\sum_{m=m_s}^{m_e} e^{j\frac{\pi}{180}[m \times Hist(m)]}\right], \qquad (4)$$

where $Ang[\bullet]$ represents the phase angle of $\bullet$ in radians. Again, m_s, m_e are the start and end sample indices within a histogram lobe, respectively.

The estimated pair (SNR, AOA) represents an estimated signal from a transmitter (user) within an active channel. A user has to be detected in a channel most of the time (50% in this study) among scans for that channel to qualify as an active channel.

III. Simulated and CRC-measured Data

The simulated data of the SE's preprocessed results were generated using Matlab for 1000 scans. Normal distributions were used with various seeds for each trial as the signal models of the SNRs and AOAs to form the internal data sets.

The CRC-measured data were obtained using the SE to scan, collect and preprocess the FM and AM signals within various frequency bands in various scenarios. The FM voice signal was within 3 bands (30~90, 460~470 and 902~928 MHz) and had 15 kHz peak frequency deviations. The AM digital random sequence signal was within 902~928 MHz with 3.6 kHz bandwidth and used 60% modulation index. A single transmitter/receiver (TX/RX) pair was used for each scenario. Some scenarios were clear LOS, some were with the transmitter near a building and some with the receiver near trees. The multipath effect could be observed among scenarios. For each scenario, the height of the receiver was either low (about 8 feet above ground) or high (about 25 feet above ground) while the height of the transmitter was always low. For scenarios 1 to 6, the receiver was located at the origin (0°, 0 meter). For scenario 7, the transmitter was located at the origin. The location information of each

scenario is shown in TABLE I and Fig. 1. For each scenario, around 5000 data scans were collected by the SE with about 300 ms between consecutive scans. For each scan, only one 25 kHz channel was used. The channel and the CRC-transmitted signal used the same frequency center. The data collected within each channel were preprocessed using a 2.5 kHz FFT resolution (400 μs time span for each scan) to determine the SNRs, AOAs and AOA ISDs of every detected channel.

TABLE I. CRC-MEASURED SCENARIOS

Scenario	TX Location (degrees, meters)	Description
1	(0, 110)	Clear LOS
2	(108, 100)	Clear LOS
3	(340, 110)	Clear LOS
4	(16.5, 110)	Clear LOS
5	(33, 110)	Clear LOS
6	(122, 300)	Beside a building
	RX Location (degrees, meters)	
7	(10, 350)	Behind 20' high trees

For a channel to be deemed active, it had to be detected more than 500 times during the first 1000 scans. For each active channel, the corresponding SNRs, AOAs and AOA ISDs from the 1000 scans were gathered to form the internal data sets.

For both of the simulated and CRC-measured data, once the internal data sets were obtained, various histograms were generated from the samples of the SNRs, AOAs and AOA ISDs. The results included estimated SNR (E_S in dB), estimated AOA (E_A in degrees), AOA AISD (AI_A in degrees) and AOA SD (SD_A in degrees) vs. SNR in dB.

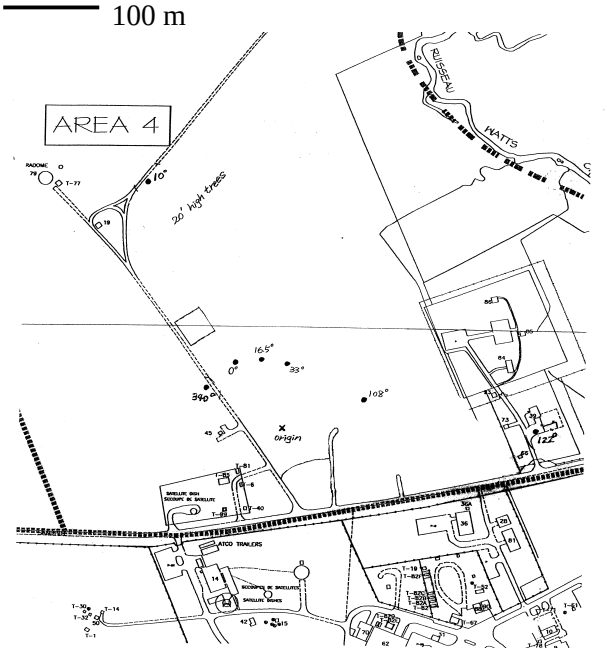


Fig. 1. Map for CRC-measured scenarios

IV. Selected Results

Simulated Test Scenario: Three signals, 20° AOA with 2° SD for signal one, 40° AOA with 3° SD for signal two, 60° AOA with 4° SD for signal three, 5 to 40-dB SNRs with 2-dB SD for each signal. The SNR and AOA histograms at 20-dB SNR are shown in Fig. 2 and Fig. 3, respectively. The results are shown in TABLE II. The corresponding accuracy plots are show in Fig. 4 and Fig. 5. The E_S and E_A results closely approximate the true value. The spread of signal three simulated the multipath effect. Signal two and signal three have some interference. One can see that the SD_A1s are close to their true values since the AOA histogram lobe of signal one is distinct. Note that in this simulation, the AOA SDs were set purposely to be independent of the SNRs. In practice, as the SNR

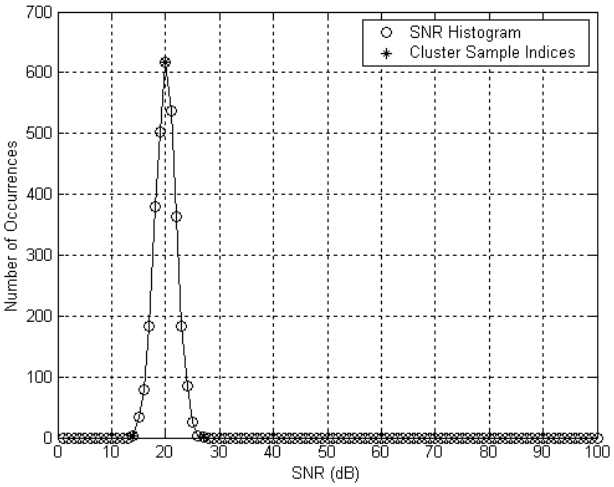


Fig. 2. Simulated Test SNR histogram at 20-dB SNR

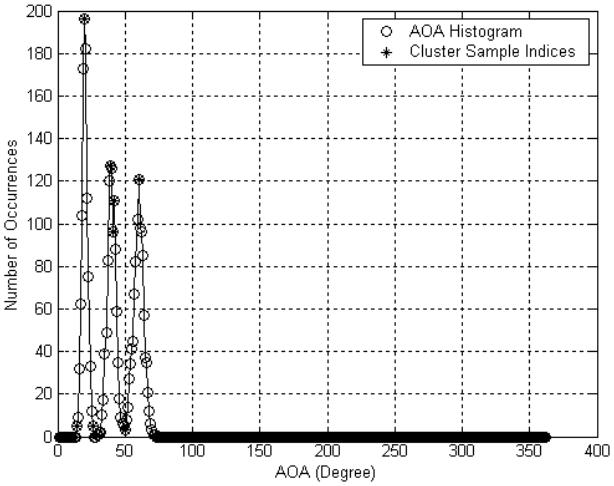


Fig. 3. Simulated Test AOA histogram at 20-dB SNR

increases, the AOA SD decreases, which will be shown in the results of the CRC-measured data later on. It is not a goal of this study to find a proper simulation model between the AOA SDs and the SNRs.

CRC Test18 Scenario: This corresponds to Location Scenario 6 (transmitter at (122°, 300 meters) beside a building) with a 905 MHz FM signal and a high receiver. The SNR and AOA histograms at 20-dB SNR

TABLE II. SIMULATED TEST RESULT

SNR	5	10	15	20	25	30	35	40
E_S	5.129	9.9943	14.96	19.957	24.9	29.9	35.0	39.814
E_A1	19.954	20.071	20.086	19.980	19.990	20.2	20.0	20.044
E_A2	39.777	39.790	40.337	39.957	40.205	40.4	39.9	40.028
E_A3	59.730	59.901	60.360	60.333	59.818	60.4	59.9	60.050
SD_A1	2.0001	2.0013	2.0001	2.0033	2.0032	2.00	2.00	2.0016
SD_A2	2.8839	2.8868	2.5853	2.5821	2.5823	3.16	2.58	2.8957
SD_A3	3.4577	2.9001	3.1633	3.1634	3.1938	3.16	3.16	2.8725

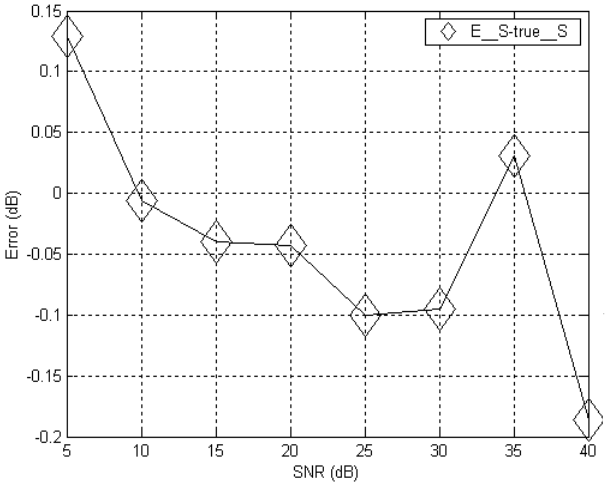


Fig. 4. Simulated Test SNR accuracy

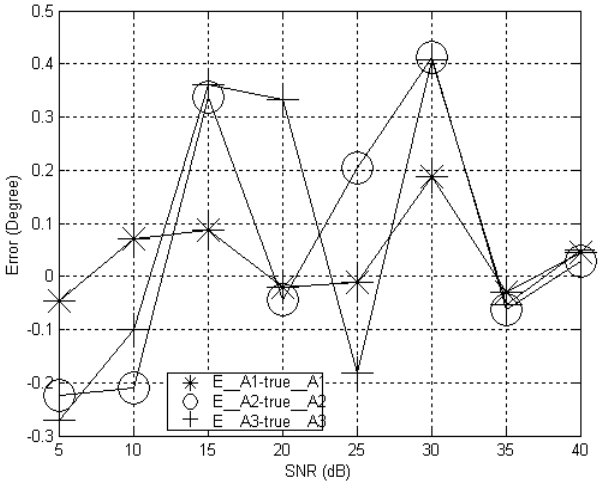


Fig. 5. Simulated Test AOA accuracy

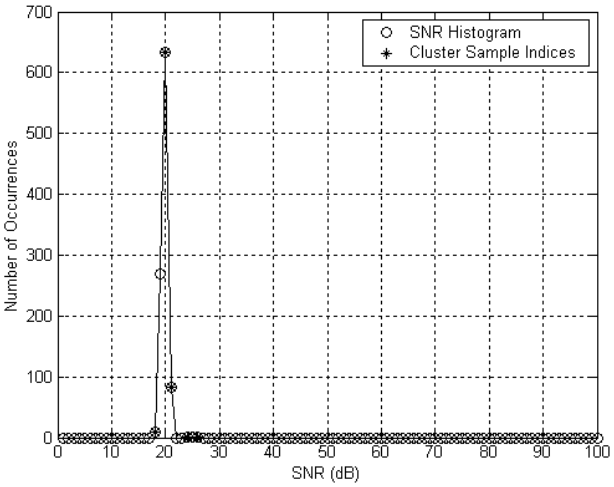


Fig. 6. CRC Test18 SNR histogram at 20-dB SNR

are shown in Fig. 6 and Fig. 7, respectively. The results are shown in TABLE III. The corresponding accuracy plots are shown in Fig. 8 and Fig. 9. Although a building is close to the transmitter in this case, the multipath effect is not very significant. One can see that there is only one AOA histogram lobe and AI_A ≈ SD_A for various SNRs, especially from 20 to 35 dB. Note that the AOA static error comes from inaccurate angle alignment between the transmitter and receiver.

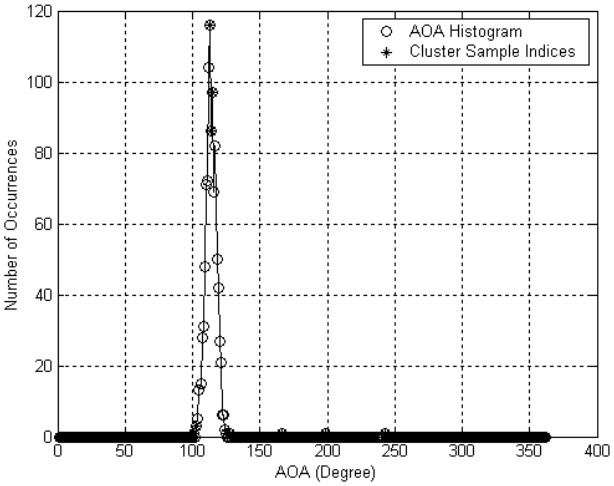


Fig. 7. CRC Test18 AOA histogram at 20-dB SNR

TABLE III. CRC TEST18 RESULT

SNR	5	10	15	20	25	30	35	40
E_S	5.7	10.5	15.0	19.8	27.4	32.0	36.7	40.8
E_A	113.58	114.73	115.59	113.57	117.89	117	116	109.87
AI_A	21.248	13.152	7.7716	4.5715	2.0420	1.21	0.74	0.1770
SD_A	26.664	11.326	5.9228	4.0318	2.0029	0.82	0.82	1.4204

CRC Test23 Scenario: This corresponds to Location Scenario 7 (receiver at (10°, 350 meters) behind 20' high trees) with a 39 MHz FM signal and a high receiver. The SNR and AOA histograms at 20-dB SNR are shown in Fig. 10 and Fig. 11, respectively. The results are shown in TABLE IV. The corresponding accuracy plots are shown in Fig. 12 and Fig. 13. In this case, some 20' trees block the receiver which is at high position (about 25 feet in height). With the low blocking trees in front of the high receiver, the multipath effect from the trees is not very significant, when the signal frequency is relatively low (39 MHz). The accuracy of SNR and AOA is better than the result in Test18. Again, AI_A ≈ SD_A for SNRs from 20 to 35 dB. The effect introduced by both buildings and trees is an appropriate topic for future study.

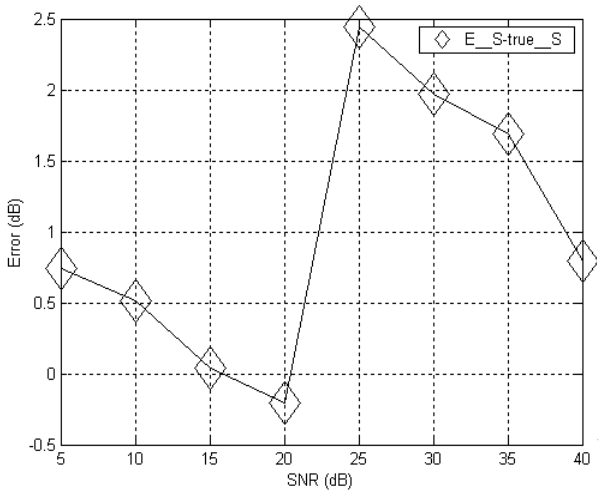


Fig. 8. CRC Test18 SNR accuracy

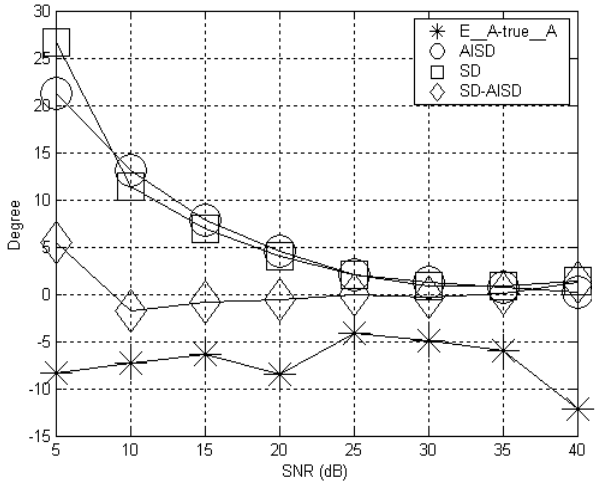


Fig. 9. CRC Test18 AOA accuracy

CRC Test17 Scenario: This corresponds to Location Scenario 6 (transmitter at (122°, 300 meters) beside a building) with a 905 MHz AM signal and a high receiver. The SNR and AOA histograms at 20-dB SNR are shown in Fig. 14 and Fig. 15, respectively. The results are shown in TABLE V. Although a building is close to the transmitter in this case, the multipath effect is not very significant. The corresponding accuracy plots are shown in Fig. 16, Fig. 17 and Fig. 18. There are two peaks in the SNR histogram due to the amplitude variation of the AM signal. Note that the AM type and its characteristics have to be identified by

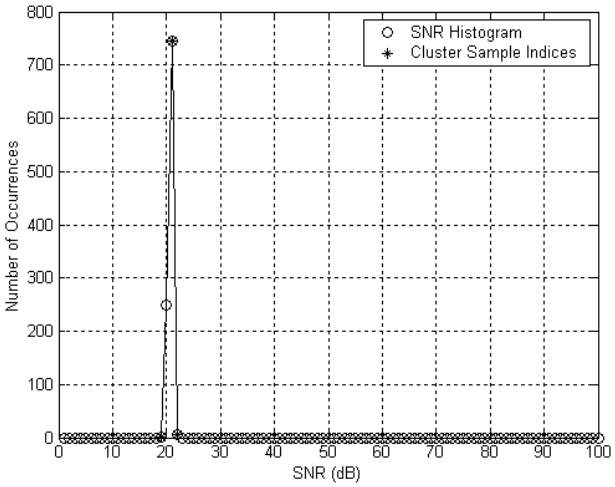


Fig. 10. CRC Test23 SNR histogram at 20-dB SNR

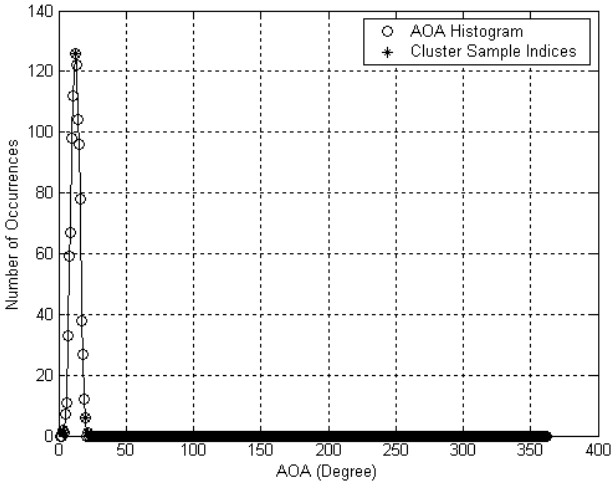


Fig. 11. CRC Test23 AOA histogram at 20-dB SNR

TABLE IV. CRC TEST23 RESULT

SNR	5	10	15	20	25	30	35	40
E_S	6.0043	11.095	15.504	20.756	25.936	30.8	35.9	41.237
E_A	4.5275	13.709	12.703	12.388	12.422	12.0	12.4	12.657
AI_A	20.759	11.531	6.9510	3.8350	2.1150	1.15	0.79	0.0750
SD_A	23.299	10.686	6.3475	3.4539	1.7096	0.82	0.52	0.5241

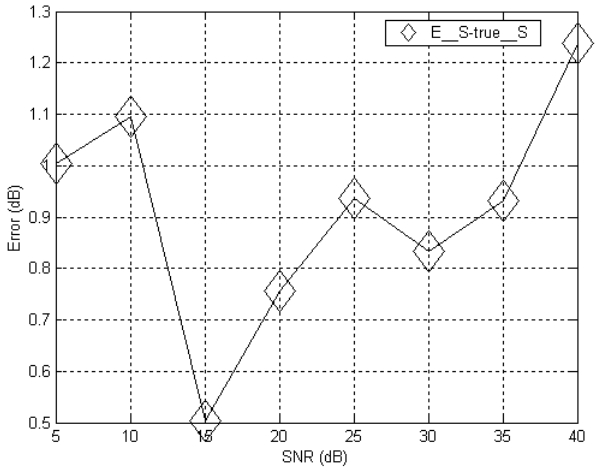


Fig. 12. CRC Test23 SNR accuracy

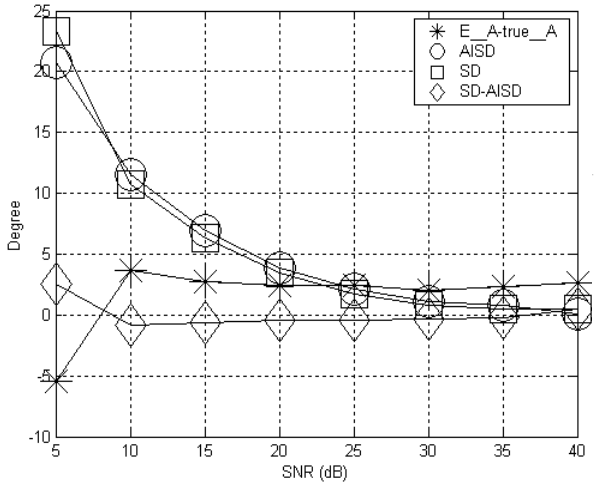


Fig. 13. CRC Test23 AOA accuracy

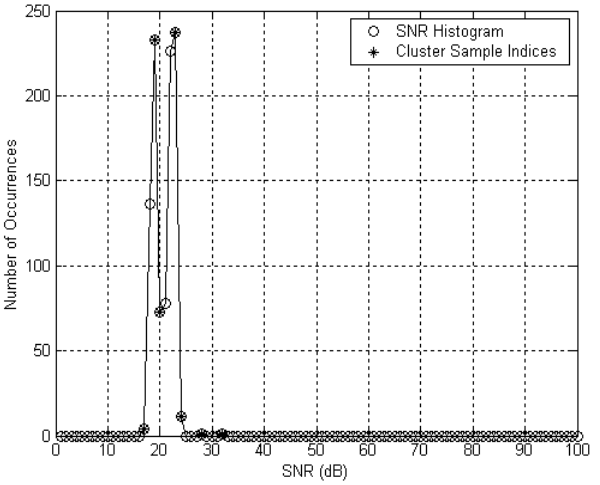


Fig. 14. CRC Test17 SNR histogram at 20-dB SNR

other algorithms ([1], [6] and [7]). Since 60% modulation index with 3.6 kHz bandwidth was used, and the time step between consecutive scans of the SE was not fixed (about 300 ms), the amplitudes of the scanned AM samples varied among scans. For the extreme case, the AM's maximum amplitude can be as large as four times (about 12 dB) its minimum. From Fig. 14, since the smallest valid SNR is 17 dB (the leftmost star in the main lobe) and the largest valid SNR is 23 dB (the rightmost star in the main lobe), the result is reasonable. The final results (E_Savg, E_Aavg) are the average of (E_S1, E_A1) and (E_S2, E_A2). The averaged SNR and AOA accuracy is close to the FM counterpart in Test18. Again, the AOA static error comes from inaccurate angle alignment between the transmitter and receiver. This AM effect is also a proper topic for future study.

Study of Limitation: Two signals (Signal_1 and Signal_2) were simulated with 1 to 5-dB SNR SDs and 1 to 5° AOA SDs. Signal_1 has 5 to 40-dB SNRs with 5-dB increment at 21° AOA. Signal_2 has 40-dB SNR at 20° AOA. Since the AOA SDs of both signals vary, it simulates the various AOA fluctuations between

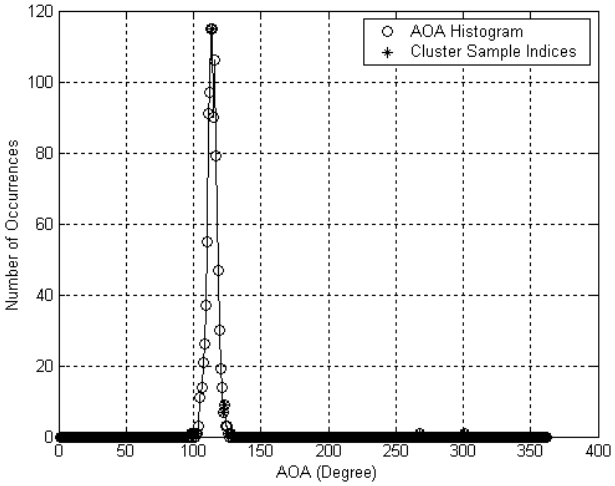


Fig. 15. CRC Test17 AOA histogram at 20-dB SNR

TABLE V. CRC TEST17 RESULT

SNR	5	10	15	20	25	30	35
E_S1	7.0068	9.4573	14.349	18.841	23.754	29.224	33.384
E_S2	-	12.476	17.585	22.056	27.042	32.513	36.690
E_A1	114.96	112.75	113.55	113.58	114.81	114.19	111.54
E_A2	-	114.78	112.95	113.81	115.10	114.32	111.42
AI_A1	19.368	14.653	8.4600	5.0516	2.6765	1.4971	0.9963
AI_A2	-	10.676	5.9791	3.3824	1.8153	1.0352	0.7481
SD_A1	21.162	15.924	7.8082	4.9334	2.0090	1.4271	1.1189
SD_A2	-	10.466	4.8993	3.1677	2.0024	1.1326	1.1211
E_Savg	-	10.967	15.967	20.449	25.398	30.869	35.037
E_Aavg	-	113.77	113.25	113.70	114.96	114.26	111.48

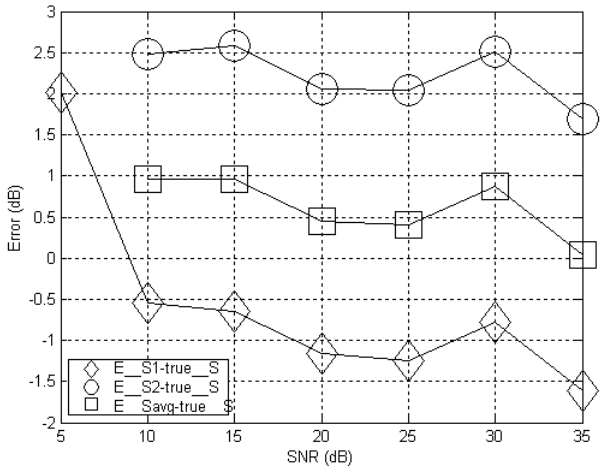


Fig. 16. CRC Test17 SNR accuracy

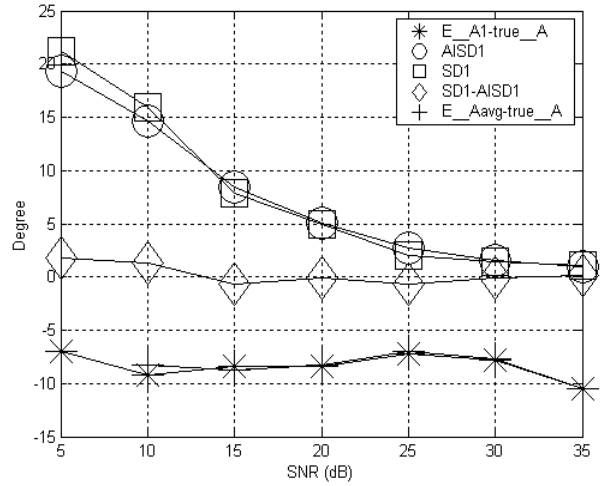


Fig. 17. CRC Test17 AOA accuracy - part 1

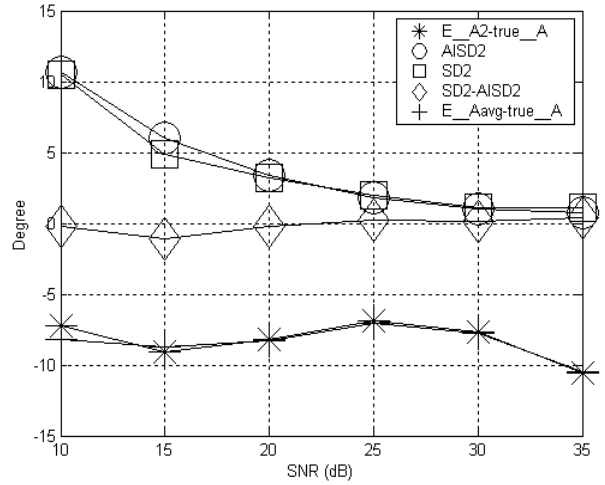


Fig. 18. CRC Test17 AOA accuracy - part 2

two close signals in the aspect domain. Also, although Signal_2 has fixed SNR, the SNR of Signal_1 (SNR1) and the SNR SDs of both signals vary, which simulates the various SNR fluctuations between two signals. Like the Simulated Test Scenario, the AOA SDs were set purposely to be independent of the SNRs. For scenarios with *I*-dB SNR SD and I° AOA SD, the corresponding E_S*I* and E_A*I* for both signals are obtained. The estimated results of Signal_1 and Signal_2 for 1000 scans are shown in TABLE VI and TABLE VII, respectively. From both tables, the smaller AOA and SNR SDs and the bigger SNR of Signal_1 are, the more accurate the E_S and E_A become. However, as the SNR1 increases to 40 (35 in one case) dB, both signals tend to be merged. Also, the estimations fail when AOA and SNR SDs are big (4, 5-deg/dB in this study) and SNR1 is low (5, 10-dB in this study). In general, when the AOA and SNR SDs are not significant, the histogram-based algorithms can estimate close signals with great accuracy. The estimations of Signal_1 and Signal_2 for 25 scans are shown in TABLE VIII and TABLE IX, respectively. From both tables, one can see that the estimations are not as accurate as the 1000-scan cases. However, again, when the AOA and SNR SDs are not significant, the histogram-based algorithms can estimate close signals with great accuracy, even when the number of scans is low.

TABLE VI. RESULT FOR SIGNAL_1 (1000 SCANS)

SNR1	5	10	15	20	25	30	35	40
E_S1	5.5007	9.958	14.958	19.958	24.958	30.0	35.1	39.990
E_A1	21.002	21.020	21.020	21.020	21.020	21.0	21.0	-
E_S2	4.9085	9.921	14.921	19.921	24.921	30.0	34.7	39.968
E_A2	21.076	21.071	21.071	21.071	21.071	21.1	21.0	-
E_S3	5.8986	9.9716	14.876	19.876	24.814	30.3	37.3	39.723
E_A3	21.005	21.117	21.117	21.117	21.118	21.1	-	-
E_S4	-	9.5632	14.750	19.720	25.041	29.7	34.3	39.237
E_A4	-	21.157	21.152	21.139	21.11	21.1	21.0	-
E_S5	-	-	17.137	22.397	25.539	30.1	35.2	38.677
E_A5	-	-	21.021	21.017	20.115	21.2	20.1	-

TABLE VII. RESULT FOR SIGNAL_2 (1000 SCANS)

SNR1	5	10	15	20	25	30	35	40
E_S1	40.021	40.021	40.021	40.021	40.021	40.0	40.0	39.990
E_A1	20.007	20.007	20.007	20.007	20.007	20.0	20.0	-
E_S2	40.031	40.031	40.031	40.031	40.031	40.0	39.8	39.968
E_A2	20.014	20.014	20.014	20.014	20.014	20.0	20.1	-
E_S3	38.644	38.644	38.644	38.644	38.619	40.1	37.3	39.723
E_A3	19.941	19.941	19.941	19.941	19.942	19.9	-	-
E_S4	-	38.302	38.302	38.554	38.548	38.6	40.6	39.237
E_A4	-	20.139	20.139	20.130	20.111	20.2	20.1	-
E_S5	-	37.878	37.878	37.853	37.580	40.9	40.2	38.677
E_A5	-	20.034	20.034	20.034	19.947	19.8	20.2	-

TABLE VIII. RESULT FOR SIGNAL_1 (25 SCANS)

SNR1	5	10	15	20	25	30	35	40
E_S1	5.08	10.08	15.08	20.08	25.08	30.1	35.1	40.04
E_A1	21	21	21	21	21	21	21	-
E_S2	5.2	10.2	15.2	20.2	25.2	30.2	34.9	40.16
E_A2	20.960	20.960	20.960	20.960	20.960	21.0	21.1	22.191
E_S3	5.6	9.5714	14.571	19.571	24.571	29.6	-	41.167
E_A3	21.268	-	-	-	-	-	-	23.385

TABLE IX. RESULT FOR SIGNAL_2 (25 SCANS)

SNR1	5	10	15	20	25	30	35	40
E_S1	40	40	40	40	40	40	40	40.04
E_A1	19.75	19.75	19.75	19.75	19.75	19.8	19.8	-
E_S2	40.12	40.12	40.12	40.12	40.12	40.1	39.7	40.16
E_A2	19.240	19.240	19.240	19.240	19.240	19.2	19.4	19.088
E_S3	41.533	41.533	41.533	41.533	41.533	41.5	41.5	41.167
E_A3	-	-	-	-	-	-	-	18.120

V. Conclusions

The purpose of this study is to determine the wireless channel usage or the station occupancy of fixed frequency channels using a wideband scanning device.

The results of this study show the effectiveness of using the histogram-based algorithms to analyze and estimate the channel usage using a significantly small number of scanned samples. Both simulated and measured data sets were processed and their results show significant accuracy for various scenarios, including the LOS and multipath FM signal cases. However, the results of the LOS AM signal cases show multiple peaks in the SNR histograms due to the amplitude variation of the AM signals. Thus, the histogram-based algorithms in this study solely cannot be used to classify the channel usage of the LOS AM cases. The AM type and its characteristics have to be identified first by other algorithms. Then the averaged AM results from the histogram-based algorithms can be used as estimates for classification. As per the multipath AM cases, the histogram-based algorithms may not be able to do the classification correctly.

In general, the histogram-based algorithms provide a simple and efficient way to classify the LOS channel usage correctly. Further study of the multipath effect and signal types other than AM and FM should be done in the future to test the robustness of the algorithms.

VI. References

[1] P. Chahine, M. Dufour, E. Matt, J. Lodge, D. Paskovich and F. Patenaude, "Monitoring of the Radio-Frequency Spectrum with a Digital Analysis System: An Update," Proceedings of the 16th International Wroclaw Symposium on EMC, Poland, June 2002.

[2] J.H. Jo, M.A. Ingram and N. Jayant, "Angle Clustering in Indoor Space-Time Channels Based on Ray Tracing," Proc. *IEEE Veh. Technol. Conf.*, pp. 2067-2071, 2001.

[3] Q. Spencer, M. Rice, B. Jeffs and M. Jensen, "A Statistical Model for the Angle-of-Arrival in Indoor Multipath Propagation," Proc. *IEEE Veh. Technol. Conf.*, pp. 1415-1419, 1997.

[4] R.J.-M. Cramer, R.A. Scholtz and M.Z. Win, "Evaluation of an Ultra-Wide-Band Propagation Channel," *IEEE Trans. Antennas Propagat.*, vol. 50, pp. 561-570, May 2002.

[5] Q. Spencer, B. Jeffs, M. Jensen and A. Swindlehurst, "Modeling the Statistical Time and Angle of Arrival Characteristics of an Indoor Multipath Channel," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 347-360, Mar. 2000.

[6] D. Boudreau, C. Dubuc, F. Patenaude, M. Dufour, J. Lodge, "A Fast Automatic Modulation Recognition Algorithm and its Implementation in a Spectrum Monitoring Application," Proceedings of the Military Communications Conference (MILCOM 2000), Los Angeles, California, United Sates, October 2000.

[7] E.E. Azzouz and A.K. Nandi, "Automatic Modulation Recognition of Communication Signals," 1996 Kluwer Academic Publishers.

Measurement and Analysis of Urban Spectrum Usage

Allen Petrin[(1)], Paul G. Steffes[(2)]

[(1)] Georgia Institute of Technology School of Electrical and Computer Engineering, 777 Atlantic Dr., Atlanta, GA, USA, 30332-0250; me@allenpetrin.com; 404-894-5280 (phone); 404-894-5935 (fax),

[(2)] Georgia Institute of Technology School of Electrical and Computer Engineering, 777 Atlantic Dr., Atlanta, GA, USA, 30332-0250; ps11@prism.gatech.edu; 404-894-3128 (phone); 404-894-5935 (fax)

To increase spectrum utilization, a thorough understanding is needed of its current usage profile. While some coarse information can be attained from spectrum licenses, essential details including the location of transmitters, transmitter output power and antenna type are often unknown. Additionally licenses do not specify how often the spectrum is being occupied if at all. Furthermore the local environment effects the propagation of radio waves; while this effect can be simulated, the results offer only moderate precision. Hence to categorize spectrum usage, measured data is vastly preferable to theoretical analysis.

This paper presents the results of measured spectra from 400 MHz to 6.4 GHz in urban Atlanta, GA USA. This study improved on past ones by resolving spectrum usage azimuthally, in polarization, and in time. The often-dynamic nature of spectrum usage necessitates the analysis of its usage over time. To provide accurate and substantive information on spectrum usage more then one billion data samples where taken. This data was analyzed to produce information on spectrum usage levels and characteristics. Additional analysis of the data was used to find low probability of intercept (LPI) signals in passive user bands. The information gathered from this study will be used to develop frequency agile radio protocols that maximize the amount of spectrum reused and lessen the possibility of inference.

Spectrum Study

To maximize the utility of the radio spectrum, knowledge of its current usage is beneficial. A spectrum study was initiated to provide multidimensional usage information and characteristics. This study improves upon past ones by resolving spectrum into nearly all its possible constituents. Figure 1 displays the dimensions that are assessed. For this paper only the urban location type (Atlanta) shown in figure 2 has been performed. Finally, to provide a statistically valid model of the spectral environment a large number of data samples where taken.

The implementation of this spectrum study is limited by the capabilities of measurement equipment. It is desirable to sense every emitted signal, but sensitivity is limited by receiver noise, intermodulation and also gain of the receiving antenna. Additionally the volume of data that the study produces it limited by the time it takes to perform the study. For a study to be statistically relevant enough must be collected. The volume of data measured is limited by the amount of time allocated for the study.

Spectrum Study Variables

- Frequency (400 MHz – 6.4 GHz)
- Time (short term usage)
- Time Period (6 discrete time periods per day)
- Polarization (Linear: Horizontal & Vertical)
- Azimuth (6 directions)
- Location type (Urban, Suburban, Rural)

Figure 1

Spectrum Measurement and Analysis System

This spectrum study required the design and construction of several hardware and software subsystems. Collection of the spectrum data required an antenna system (including an azimuthal positing system), an RF-subsystem, spectrum analyzer, and finally a data acquisition and control system, which is shown in figure 3. The mining of the data to produce

Atlanta Measurement Site

Figure 2

information is accomplished by several analysis programs.

Antenna System

A high gain antenna system was chosen to increase the system's sensitivity and to resolve spectrum the in azimuthal directions. Four antennas with 8 dBi to 9 dBi gain (depending on frequency) are able to cover from 100 MHz to 8 GHz with both linear polarizations (horizontal and vertical). The near constant gain over the frequency range of the antennas also provides close to constant beamwidth. These antennas are mounted on a rotating mast that offers good line-of-sight to the urban Atlanta area. Azimuthal positing is remotely controlled by the data acquisition and control system. Six azimuthal directions are used to azimuthally resolve spectrum and offer omni-directional sensitivity.

RF-Subsystem

Filtering and amplification is performed by the RF-subsystem. This system is connected to all the antennas and also serves as an antenna selector. A matrix of filters with an octave or less of bandwidth is used to reduce the creation of intermodulation in the subsequent stages of the system. After filtering, signals pass through a low noise amplifier (LNA) with a high (+27 dBm) third-order intercept point. The LNA is needed to lower the total systems noise temperature, since the spectrum analyzers has a very high noise figure (27 dB to 29 dB depending on frequency). Across the 400 MHz to 6.4 GHz frequency range, the total systems noise figure ranges from 6 dB to 7 dB. The filter and LNA combination results in an instantaneous spurious free dynamic range that is better than that for the spectrum analyzer. Hence, the spectrum analyzer limits the systems intermodulation performance and thus sensitivity.

As with all spectrum measurement system components the RF-subsystem is remotely controlled by the data acquisition and control system. All antenna and filter selection is verified by the used of position sensors and checked by software after any change in state. Additionally the RF-subsystem integrated a noise diode that is used to calibrate the complete RF system (except for the antennas and their short connecting cables). These features produce a highly reliable self calibrating system with built-in fault detection.

Data acquisition and control system

An automated system was developed to control the position of the antennas, choose the desired antenna and filter, perform calibration, and communicate with the spectrum analyzer. Another design requirement was very high reliability; this is needed for unattended multi-week data acquisitions. The system developed to meet these requirements incorporates fault detection and correction at all levels. Only a hardware failure can produce a fatal error, every other fault mode is accommodated. The complete spectrum measurement system has achieved better than 99.999% operational reliability. The only fatal error occurred as a result of an electromechanical microwave switch failure which was detected at the instant it failed. The system was able to identify the exact component which needed replacing. Additionally, the software keeps statistics on the health of each subsystem (to the component level for the RF-subsystem), recording the time and number of corrected fault events. This allows for the prediction and preventive replacement of components before they fail completely.

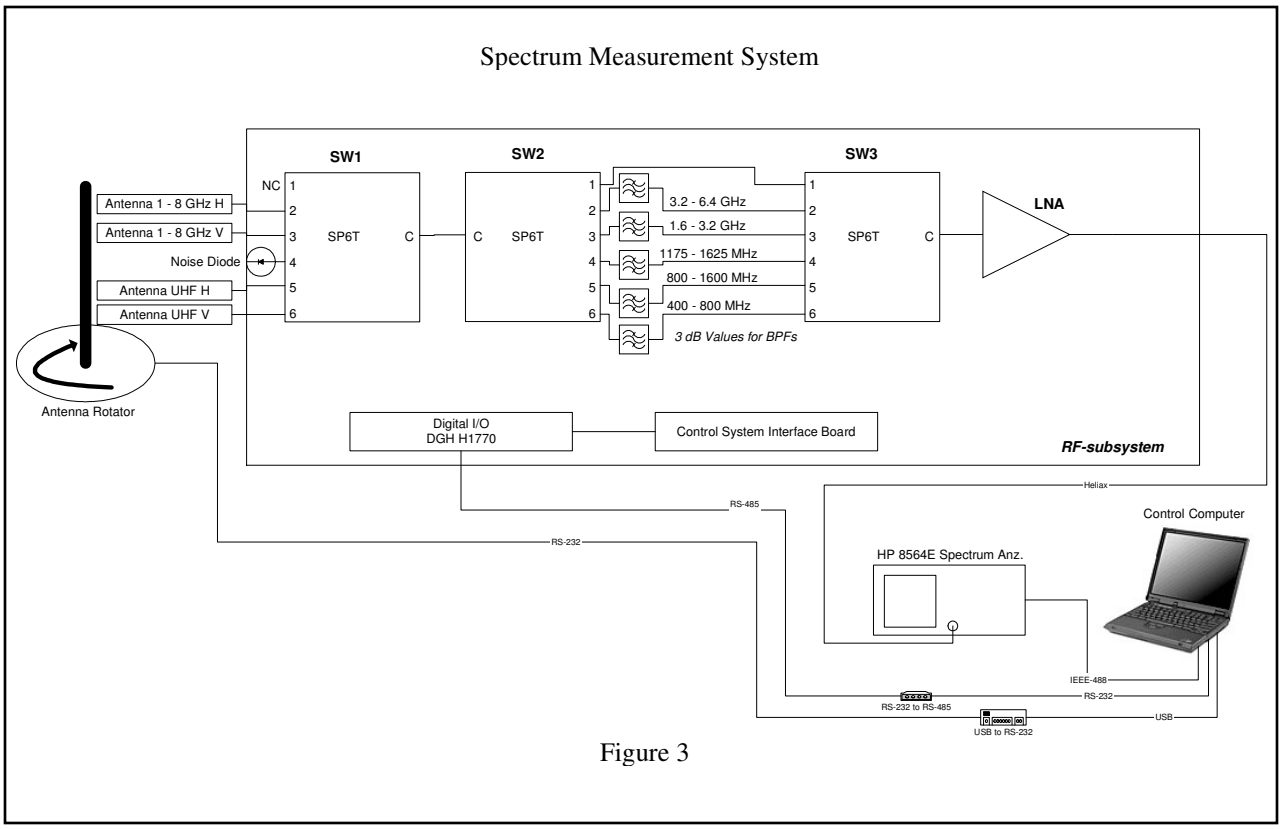


Figure 3

The data acquisition and control software integrates two operating modes: interactive and batch processing. The latter is used for multi-week unattended data acquisitions. A scheduling program was designed to minimize the time needed to perform multidimensional spectrum studies. All data is saved in a file as a collection of objects, with each object containing one sweep from the spectrum analyzer and 20 other essential parameters. This format retains all the data produced by the spectrum analyzer in its raw form, thus allowing for latter post-processing.

A modified scheduling program with more frequent measurements was employed to detect low probability of intercept (LPI) signals in passive user bands.

Analysis Software

The post-processing of raw collected data permits extensive data mining of the measured spectrum. First, the data is calibrated with the help of the noise diode and 6,000 data points for each 6 MHz frequency range (12,000 data points are used to calculate noise figure). Initial analysis software was developed to demonstrate the multidimensional usage of spectrum. Additional software is being developed to examine usage patterns in spectrum and determine other characteristics.

Results and Conclusion

More than 2 billion spectrum measurements were taken in urban Atlanta over several weeks. Figure 4 display initial analysis of the 1.2 GHz to 1.4 GHz range. These plots are just a few of the hundreds required to minimally describe Atlanta's spectral environment.

Because of the great wealth of information produced from the Atlanta spectrum study, only a minute fraction of it can be presented in print. Hence, the authors have decided to offer post-processed data in Excel format online: www.measuredspectrum.com. Additionally, interested parties can attain the complete Atlanta spectrum study in its raw data form. This data can be used to develop frequency agile protocols and assist in testing their ability to transparently use spectrum. Other interference analyses and spectrum utility maximizations are possible with the use of this rich data set.

The authors are in the process of performing a suburban spectrum study which will be followed by a rural study. These three studies will assist in classifying spectrum usage by location type.

Spectrum Measurement Dimensions for 1200 MHz to 1400 MHz

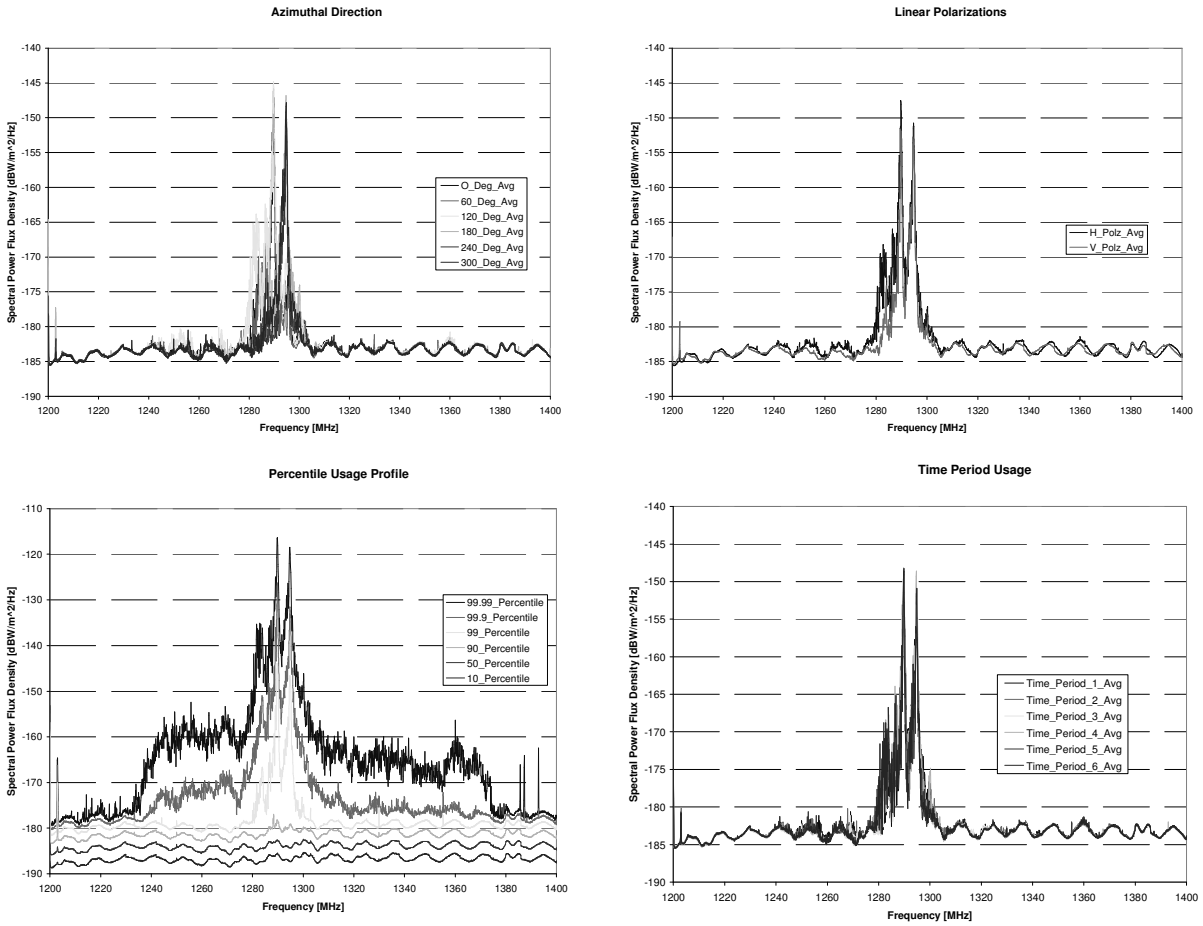


Figure 4

Analog Front-End Cost Reduction for Multi-Antenna Transmitter

Edmund Coersmeier, Ernst Zielinski, Klaus-Peter Wachsmann
Nokia, Meesmannstrasse 103, 44807 Bochum, Germany

edmund.coersmeier@nokia.com, ernst.zielinski@nokia.com , peter.wachsmann@nokia.com

Abstract *– Multi-antenna systems provide the option to enhance the data rates and to improve the overall system performance. Therefore it is important that the transmitter provides high signal accuracy for all different signal branches. On the one hand high signal accuracy can be reached by employing expensive, high-end analog front-ends. On the other hand a more cost efficient solution might be a low cost analog front-end in combination with digital, software-based pre-adjustment algorithms to guarantee high precision output signals. A direct-conversion transmitter analog front-end architecture for an OFDM multi-antenna system will be proposed and analog filter imperfections are compensated based on a digital filter pre-equalizer. Because of the direct-conversion technique it is important to combine the pre-equalizer with a digital IQ sample estimator, which derives the IQ values from the envelope signal without a down-modulation process. This paper provides a mathematical and architectural description and simulation results of the software-based IQ estimation and filter pre-equalizer.*

Index Terms *– Multi-antenna, MIMO, OFDM, direct conversion, filter pre-equalization, IQ estimation*

I. Introduction

Multi-antenna systems enable high data-rates and a good system-performance [1], [2] if they can provide good signal accuracy already at the transmitter output [3]. This is problematical if a low-cost direct conversion architecture has been chosen, which offers a cheap and low-power implementation with the drawback of imperfect I- and Q-signal accuracies. Reasons for the imperfections can be imperfect analog base band filters and unbalanced I- and Q-branch amplification.

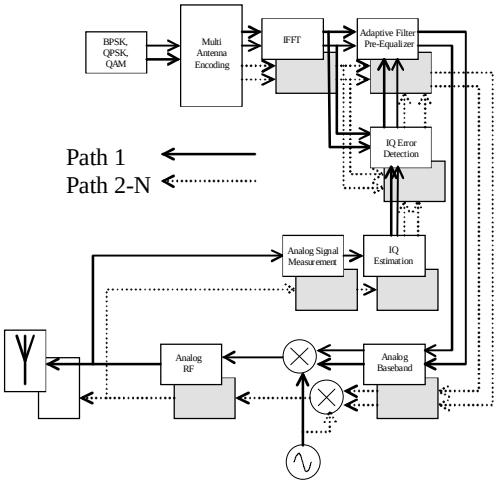


Figure I-1 Multi-antenna OFDM transmitter.

Hence analog signal imperfections have to be removed digitally, shown in Figure I-1. The transmitter generates first low rate QPSK or QAM constellation points, which will be mapped onto the different carriers [4]. The multi-antenna coding inserts e.g. diversity or spatial-multiplexing approaches [1], [2]. This is followed by the signal conversion from the frequency domain to the time domain. After the IFFT operation has been finished the pre-equalizer takes care about the analog filter imperfections [6], [7], [8]. This block is able to pre-modify the ideal digital data stream. The digital pre-modification effects will be compensated later by the introduction of the unwanted analog imperfections, such, that the final result at the antenna input port leads to an ideal signal. The digital pre-compensation block receives its correction coefficients from the error detection algorithm, which calculates the error based on the comparison between ideal and real I- and Q-samples. Hence an I- and Q-sample estimation from the RF-envelope signal is required. Typically this information is not available at this stage, because the RF signal does not provide an access to the base-band equivalent signals I and Q without a complete down-modulation process.

The purpose of this paper is to provide a simple IQ sample estimation combined with a digital filter pre-equalization algorithm. These algorithms will be

implemented in each branch of the multi-antenna transmitter. The digital components save costs, because cheap, non precise analog filters can be incorporated. Because of the multi-antenna approach it saves even more costs by implementing the algorithms via software to a Digital-Signal-Processor (DSP). The analog components will not change quickly their amount of imperfections during a certain period of time and hence an algorithm reuse of all compensation blocks via a DSP can have significant influence to the overall system complexity in terms of hardware gate count and power consumption.

II. Digital IQ Estimation

In direct conversion architectures the I- and Q-branches are fed from the digital base band via two independent DACs to the analog base band. After separate low-pass filtering and appropriate amplification of each branch the up-conversion to the RF range takes place. In case of a multi-antenna system with N – transmitter antennas this architecture has to be installed N times. At the N antenna outputs it is desired to have the best possible signal accuracy available. This can be reached by installing precise, but most probably expensive analog components for each I- and Q-branch and for each transmitter path. An advantageous alternative is the installation of low cost analog components with less precision and additionally digital compensation techniques to remove the analog imperfections via a cheap solution. Therefore I- and Q-signal extraction from the RF-envelope needs to be done. The extraction is required to estimate reliably the wanted IQ samples from the analog RF-envelope without a down-modulation process on the transmitter side.

In this paper the estimated IQ samples are used for the digital pre-equalization process. To pre-equalize the analog base-band filters there has to be employed filter imperfection estimation. Such an error detection could be done by subtracting the non-ideal IQ samples $\tilde{I}[n]$ and $\tilde{Q}[n]$ from the ideal IQ samples $I[n]$ and $Q[n]$.

$$\begin{aligned} e_I[n] &= I[n] - \tilde{I}[n] \\ e_Q[n] &= Q[n] - \tilde{Q}[n] \end{aligned} \qquad (1)$$

In case of ideal output samples at the antenna port the differences between the wanted and the transmitted signals equals zero and no pre-equalization need to be activated. Assuming that there are imperfections present then the difference is unequal to zero in both branches from equation (1). To enable the required measurement the non-ideal IQ samples have to be extracted from the analog envelope signal. This will be done digitally by comparing two consecutive analog and digital IQ pairs. The analog samples are measured at the antenna input port, the corresponding digital samples before the DAC operation. Two consecutive analog samples are described by equation (2). The amplitude $|A_a[n]|$ and the amplitude $|A_a[n-1]|$ are measured at the time instances n and $n-1$. Only the left sides of both branches in equation (2) can be measured physically at the antenna input port. The elements of the right side have only theoretical meaning.

$$\begin{aligned} |A_a[n]| &= \sqrt{\tilde{I}[n]^2 + \tilde{Q}[n]^2} \\ |A_a[n-1]| &= \sqrt{\tilde{I}[n-1]^2 + \tilde{Q}[n-1]^2} \end{aligned} \qquad (2)$$

At the same time it is necessary to measure the corresponding digital sample amplitudes. This is shown in equation (3).

$$\begin{aligned} |A_d[n]| &= \sqrt{I[n]^2 + Q[n]^2} \\ |A_d[n-1]| &= \sqrt{I[n-1]^2 + Q[n-1]^2} \end{aligned} \qquad (3)$$

The digital amplitudes at the time instances n and $n-1$ on the left side need to be calculated, because the I- and Q-branches provide $I[n], Q[n]$ and $I[n-1], Q[n-1]$ separately. For convenience further digital or analog component latency has been neglected.

The corresponding analog and digital amplitudes need to be compared. In ideal case the corresponding amplitudes should equal by omitting a certain constant amplification factor.

$$\begin{aligned} |A_d[n]| &\overset{!}{=} |A_a[n]| \\ |A_d[n-1]| &\overset{!}{=} |A_a[n-1]| \end{aligned} \qquad (4)$$

Then the digital pre-equalization error detector from equation (1) would have indicated no error. But in practice the analog base band filters might add signal imperfections. To calculate the missing signals the following relationship will be taken into account.

$$\begin{aligned} \frac{\tilde{I}^2[n]}{\tilde{I}^2[n-1]} &\overset{!}{=} \frac{I^2[n]}{I^2[n-1]} \\ \frac{\tilde{Q}^2[n]}{\tilde{Q}^2[n-1]} &\overset{!}{=} \frac{Q^2[n]}{Q^2[n-1]} \end{aligned} \qquad (5)$$

Equation (2), equation (3) and equation (5) can be used to re-formulate equation (4). This leads to

$$\tilde{I}[n] = s_I \cdot \sqrt{A_a[n]^2 - \tilde{Q}[n]^2} \qquad (6)$$

and

$$\tilde{Q}[n] = s_Q \sqrt{\frac{Q[n]^2}{Q[n-1]^2} \cdot \frac{A_a^2[n-1] \cdot \frac{I[n]^2}{I[n-1]^2} - A_a^2[n]}{\frac{I[n]^2}{I[n-1]^2} - \frac{Q[n]^2}{Q[n-1]^2}}} \qquad (7)$$

First equation (7) needs to be solved and after that equation (6) can be taken into account. The signals s_I and s_Q provide the digital sample signs. They have been stored in parallel and it is assumed that the analog imperfections will not disturb the sign of the analog samples. This will be almost true, because one takes care about imperfect analog filters and no random channels.

But generally it is no problem if every now and then some wrong IQ estimates will occur. Because the estimation procedure has been designed to operate in conjunction with an adjustment feedback loop. The feedback loop employs a low-pass filtering process, which automatically removes the influence of wrong IQ decisions from the IQ estimate algorithm in equation (6) and (7). Such wrong estimates might happen because of the missing down-modulation on the transmitter side there is no information about the signal phase available. This missing phase information is not critical as long as the overall system imperfections are generated simply by imperfect analog components and not by a random channel. Figure II-1 provides the differences between the ideal and estimated I-values during a pre-equalizer's adaptation process. After all filter imperfections have been compensated by the pre-equalizer there has been left no differences between the ideal and estimated I-values.

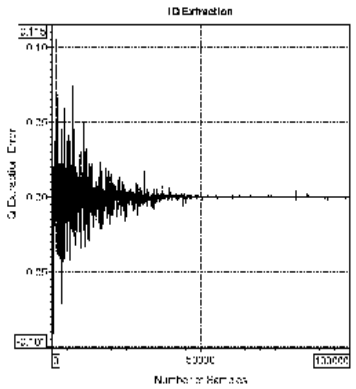


Figure II-1 Estimated I-branch error during a pre-equalizer's adaptation process.

This section has presented a mathematical description of an IQ sample extraction algorithm from the analog RF-envelope. Without the need for a down-modulation process on the transmitter side it is possible to estimate the imperfect I- and Q-samples, which are required to enable the compensation techniques for direct-conversion front-end architectures.

III. Filter Pre-Equalization

In this section there will be introduced an LMS based pre-equalizer [7], [8], which does not operate with complex coefficients, but with real ones. This is unusual but makes it possible to handle I-branch and Q-branch imperfections independently. The I-branch and Q-branch filter imperfections are generated by the analog base band filters, which are two real filters. The IQ amplitude error detection will be done via equation (1).

To update the pre-equalizer's filter coefficients successfully the gradient has to be calculated based on the approximated system identification [7]. The approximation of the analog filters will be simple tap-delay lines providing the same latency as the analog filters contain. Equation (8) provides the gradient of the LMS approach.

$$\hat{\underline{\nabla}}\left\{\hat{E}\left\langle e^2[n]\right\rangle\right\} = -2 \cdot e[n] \cdot \underline{D}[n] \cdot \underline{h}^{\#}[n] \qquad (8)$$

The approximation-based gradient is updated on a sample-by-sample base and depends on the measured error value e[n] and the delayed input signal. The mentioned delay corresponds to the approximated analog filter peak. Figure III-1 provides the difference between the proposed pre-equalizer with an approximation-based gradient, a gradient based on an ideal system identification and a deterministic gradient.

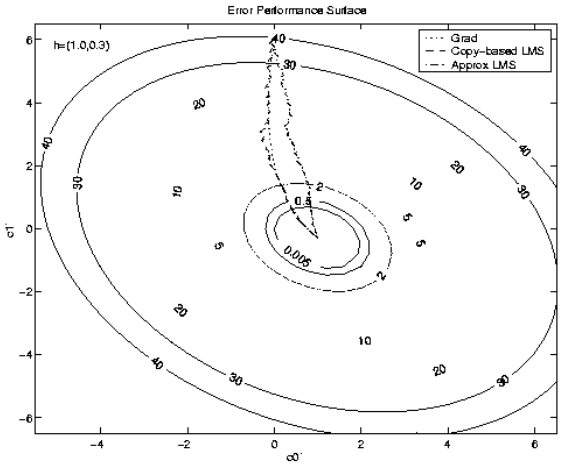


Figure III-1 Gradients on the level curve diagram.

The approximation-based gradient takes a different route but reaches the optimal filter vector as the other algorithms. Figure III-2 shows the three gradients from another camera position.

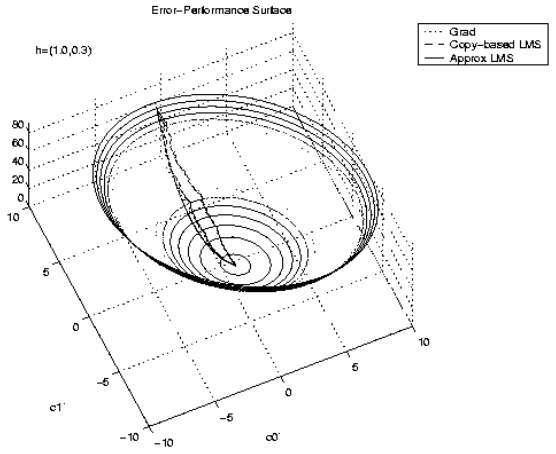


Figure III-2 Gradient behaviour from bird's eye view

Based on the gradient there can be calculated the pre-equalizer's coefficient update. There have to be calculated for both branches independent correction coefficients. This is described by equation (9). The new coefficients at the time n+1 will be calculated from the current coefficients at the time n and an additional addend.

$$\underline{c}_{I,Q}[n+1] = \underline{c}_{I,Q}[n] + \mu\, e_{I,Q}[n]\,\underline{D}_{I,Q}[n]\,\underline{h}^{\#}_{I,Q}[n] \qquad (9)$$

The addend consists out of four factors. First the constant μ describes the step width. The step width defines the loop accuracy, loop adaptation speed or loop bandwidth, respectively. Because the expected filter imperfections will not change over a very long period of time the loop bandwidth needs not to be large and hence the loop accuracy can be high. The second factor is the calculated error from equation (1). After that the product of the ideal input data matrix $\underline{D}$ and an approximation $\underline{h}^{\#}$ of the analog filters $\underline{h}_{I,Q}$ follows. The new coefficient vector leads to a better signal equalization and if the optimium adaptive filter vector has been reached the adaptation loop is in equilibrium.

Combined with the IQ estimation there can be build an adaptive filter pre-equalization system to enable low cost analog front-ends.

IV. Software Architecture for further cost reduction

From the architecture point of view it will be advantageous to implement the algorithms as software code via a Digital-Signal-Processor [9]. The mathematical operations from equation (1), (6), (7) and (9) are good candidates to be handled by the DSP. This is true because the analog filter imperfections do not change quickly. Hence the IQ sample estimation, error calculation and the coefficient update need not to be done as quickly as practical possible. Changing the block based hardware implementation from Figure I-1 it is possible to end up with a much more flexible and cost-reducing architecture by employing a DSP.

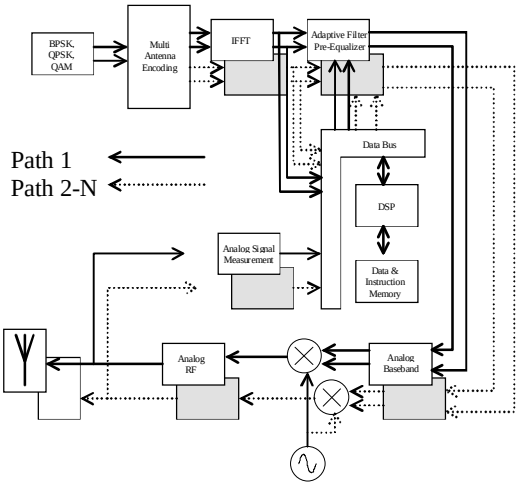


Figure IV-1 Software-based IQ estimation, pre-equalizer's error and coefficient update calculation.

Figure IV-1 shows a software-based transmitter part for the IQ sample estimation, the pre-equalizer's error detection as well as the coefficient update. The data bus establishes the connections between the DSP and the envelope input signal and the pre-equalizer's adaptive filters, respectively. The instructions for the different algorithms, which have been implemented via dedicated hardware in Figure I-1, are stored now in the instruction memory in Figure IV-1. Additional control SW, which is responsible to guarantee the correct order of the different algorithm operations, needs to be provided as well. Besides the instructions the DSP requires the data from the digital base band and the analog front-end. The information is stored in the data memory and used by the DSP instructions to calculate the new coefficient update.

Once the coefficients have been updated they can be provided via the bus to the pre-equalizer's adaptive filters. The filters are still implemented via dedicated HW because the signal pre-modification needs to operate on the base of the user data rate. From the instructions point of view the DSP could handle the adaptive filtering process as well. But in that case a significant higher processor clock rate needs to be

considered. Such a high clock rate might increase the power consumption of the DSP to an unwanted value.

V. System Performance

This chapter shows a new analysis about the performance decrease by introducing sub-optimal analog filters and the corresponding signal improvements through the digital pre-equalization setup. There has been investigated an IEEE802.16a based OFDM system including 16-QAM and 64-QAM.

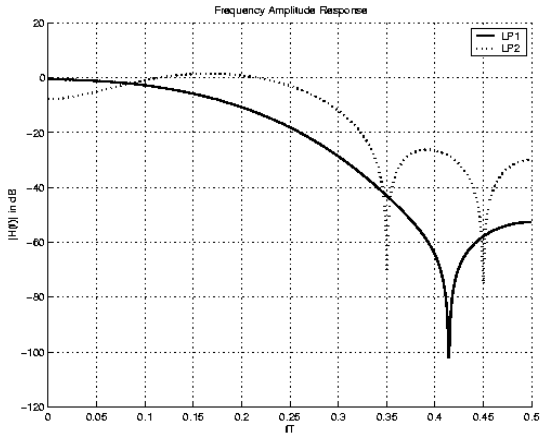


Figure V-1 Sub-optimal filter transfer functions.

Because of the cost reduction for the analog front-end one assumes imperfect filters. Based on the transfer functions from Figure V-1 there can be expected a significant decrease of the transmitted signal accuracy. Figure V-2 shows possible inaccuracies for a 16-QAM signal.

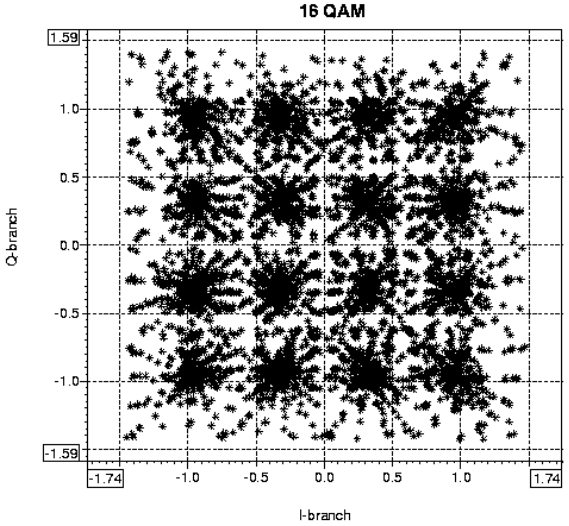


Figure V-2 Imperfect 16-QAM constellation diagram.

After the pre-equalization process has been enabled the imperfections are reduced significantly already by a 3-coefficient adaptive filter. Figure V-3 shows that the constellation points are much more precise but not perfect.

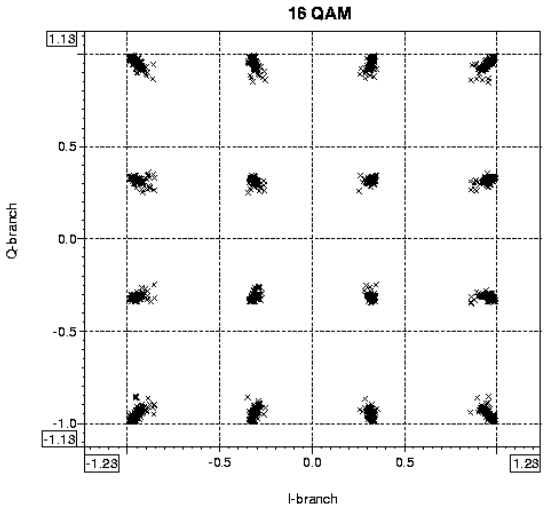


Figure V-3 3-coefficients pre-equalizer

By employing 19 coefficients perfect signal accuracy at the transmitter's output can be reached. This is shown in Figure V-4. Hence a digital adaptive filter can allow the use of low-cost, imperfect analog filters.

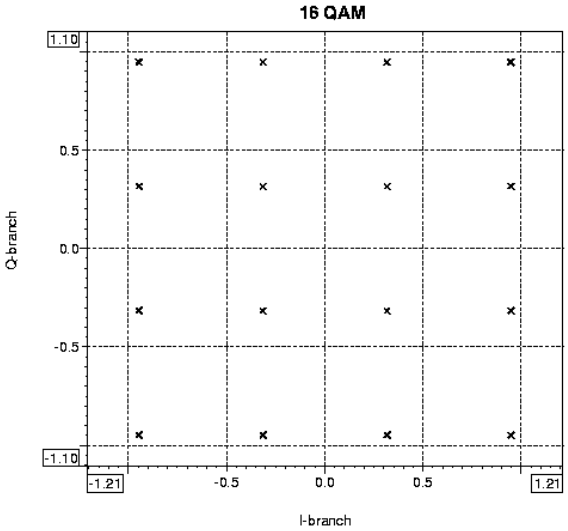


Figure V-4 Perfect pre-equalized 16-QAM signal.

Besides the signal accuracy it is possible to measure the imperfections via BER curves as well. Figure V-5 and Figure V-6 provide simulation results for 16-QAM and 64-QAM, respectively.

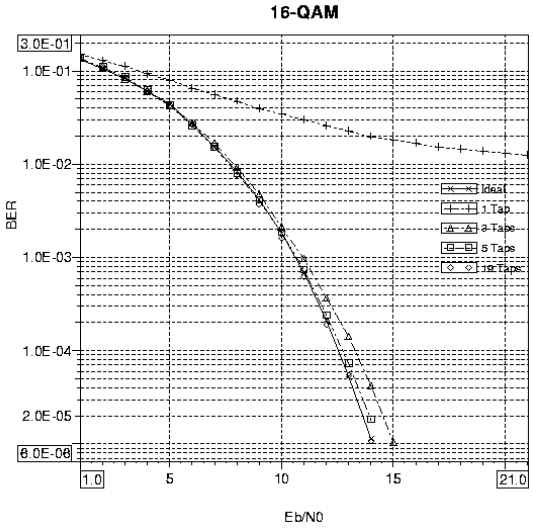


Figure V-5 16-QAM BER curves with different pre-equalizer coefficient numbers.

Non-frequency selective corrections employ only 1 coefficient and cannot remove the imperfect analog filter influences. They adjust just the signal's amplitude. A BER floor makes the overall transmitter performance pure. Increasing the number of pre-equalizer coefficients leads to better performances. In case of a 64-QAM there is a very high BER floor without corrections and also a 3-coefficients pre-equalizer suffers still significant losses. With 19 coefficients the desired performance is provided.

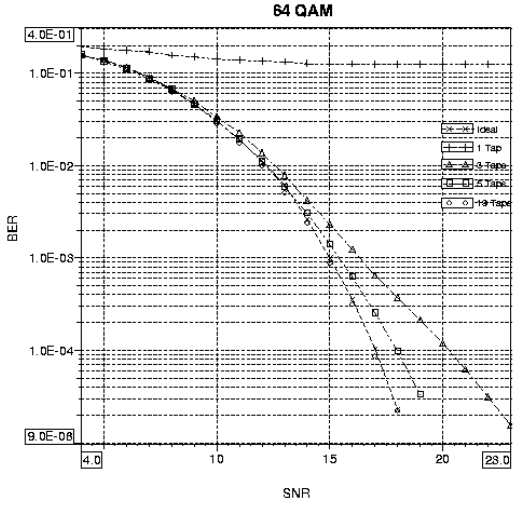


Figure V-6 64-QAM BER curves.

This chapter has shown that the overall system performance decreases significantly by introducing low-cost analog filters. Depending on the Euclidean distance a 64-QAM signal is much more sensitive against filter inaccuracies than 16-QAM signal. Finally a 19-coefficients pre-equalizer can remove the imperfections and high signal accuracy at the transmitter's output can be reached.

VI. Conclusion

This paper introduces a cost-reduction model for multi-antenna transmitters. Because it is important to provide low-cost analog front-ends for multi-antenna systems there is a need to remove the imperfections of the low-cost analog base band filters. This can be done by an IQ sample estimation algorithm, which calculates the IQ symbols at the transmitter's antenna output without an extra down-modulation process. Feeding the estimates to the digital pre-equalizer the imperfections can be removed completely. By employing a software-based IQ estimation and pre-equalization setup a low gate count implementation for multi-antenna systems can be reached. IQ estimation and pre-equalizer achieves significant improvements for the BER. This leads to high system reliability although low-cost analog filters have been used.

Acknowledgment

Parts of this work have been carried out within the European founded IST-project STRIKE (**S**pec**TR**ally efficient f**I**xed wireless networ**K** bas**E**d on dual standards). We would like to thank the European Commission and the partners involved in the project STRIKE for the support of this work.

References

[1] Multiple-Input Multiple Output (MIMO) Systems, H. Bölcskei, A.J. Paulraj, The communication Handbook, FWF-grant J1868-TEC

[2] Multiantenna Digital Radio Transmission, Massimiliano "Max" Martone, Artech House, 2002

[3] High Precision Analog Front-End Transceiver Architecture for Wireless Local Area Network, Edmund Coersmeier, Yuhuan Xu, Ludwig Schwoerer, Ken Astrof, 6th International OFDM-Workshop 2001, Hamburg

[4] Multicarrier Techniques for 4G Mobile Communications, Shinsuke Hara, Ramjee Prasad, Artech House, 2003

[5] Adaptive Filter Theory, Simon Haykin, Prentice Hall, Third Edition, 1996

[6] Comparison between different Adaptive Pre-Equalization Approaches for WLAN, E. Coersmeier, E. Zielinski, IEEE PIMRC 2002, Lisabon, Portugal

[7] Adaptive Pre-Equalization in Analog Heterodyne Architectures for Wireless LAN, Edmund Coersmeier, Ernst Zielinski, IEEE RAWCON 2002, Boston, USA

[8] Frequency Selective IQ Phase and IQ Amplitude Imbalance Adjustments for OFDM Direct Conversion Transmitters, Edmund Coersmeier, Ernst Zielinski, ISART 2003, Boulder, USA

[9] Software IQ Sample Estimation for Multi-Antenna Systems, Edmund Coersmeier, Ernst Zielinski, Klaus-Peter Wachsmann, IEEE RAWCON 2003, Boston, USA

Satellite Communications using Ultra Wideband (UWB) Signals

Yoshio KUNISAWA Hiroyasu ISHIKAWA Hisato IWAI and Hideyuki SHINONAGA
KDDI R&D Laboratories
YRP Center No. 3 Bldg., Hikarinooka 7-1, Yokosuka, Kanagawa, 239-0847, JAPAN
Phone : +81 46 847 6350; Fax : +81 46 847 0947
E-mail: {kuni, ishikawa, iwai, shinonaga}@kddilabs.jp

Abstract *This paper considers the satellite communication systems using the multiband UWB signal format. For terrestrial short-distance high-speed communications, the multiband UWB scheme have been proposed in IEEE 802.15 TG3a and discussion is ongoing at the standardization body. In the multiband UWB scheme, frequency hopping is adopted over 3.1 - 10.6 GHz, which is regulated by the FCC (Federal Communications Commission) of the U.S.A., and the bandwidth of one hopping spectrum (subband) is about 500 MHz. Multiband technology inherently has suitable characteristics for the terrestrial UWB such as applicability to variable transmission rates, avoidance of harmful interference to other systems, simple localizability of frequency allocations, and so forth. This paper presents a satellite communication downlink employing the multiband UWB signal transmission. The total bandwidth is assumed to be 500 MHz in the allocation of the satellite downlink and it is divided into multiple subbands. We report the initial results of the study on the link budget calculation and the estimation of the signal transmission speed assuming the multiband UWB signal transmission from a GSO (GeoStationary Orbit) satellite to the earth's surface.*

1. Introduction

The FCC (Federal Communications Commission) has defined the characteristics of the UWB devices to promote the commercial use [1]. According to the FCC regulations, the emission level is restricted to as low as -41.3 dBm/MHz in 3.1 GHz – 10.6 GHz as described in the next section. Thus, although the occupied bandwidth of the UWB signal is very wide, the spectrum power density is very small. Therefore, the UWB system is capable of suppressing interference to and from other narrowband systems. It is said that the UWB signal can be overlaid on frequency currently used by the other narrowband systems, and the effective use of frequency may be realized.

Although discussion of the UWB has mainly focused on the terrestrial short distance communication, the UWB could be radiated from satellites to the earth as one type of satellite services ("Satellite UWB"). In a satellite communication that overlays the UWB signal on a frequency band currently used by the existing satellite communications, a new communication channel can be added without an assignment of new frequency to the existing satellite communications.

Some UWB systems use pulse communication technology, and these systems have a simple configuration of a transmitter and a receiver. So it is expected a terminal consumes relatively low power in comparison with other wireless communication systems, and a small-sized terminal powered by a small battery could be developed. The devices for the terrestrial UWB system would be low cost due to mass-production. By incorporating the same system in satellite communication, it is possible to use the devices of the terrestrial UWB system, and the cost of the terminal for the satellite UWB system is reduced.

The IEEE 802.15 High Rate PHY Task Group (TG3a) for Wireless Personal Area Networks (WPANs) is working to define a project to provide a higher speed PHY enhancement for applications including imaging and multimedia communications. The multiband UWB scheme has already been proposed in the working group.

In this paper we examine whether sufficient transmission speed is obtained or not by the satellite UWB, assuming mono-pulse type PAM(Pulse Amplitude Modulation) modulated signal and the multiband UWB scheme proposed in the standardization body. As the result of the discussion we show the satellite UWB has preferable properties for the development of new satellite communications.

2. FCC Regulation on UWB

U.S. FCC has already regulated the UWB system, including the operating restrictions, authorizing the use of UWB devices on an unlicensed basis. Various applications have been considered, such as communications, measurements, radar systems, and so forth. The followings are the spectrum and emission limitations of the regulations for the handheld UWB devices, which are typical communication devices using the UWB signal.

Bandwidth :

Fractional bandwidth equal to or greater than 0.2, or bandwidth equal to or greater than 500 MHz.

Radiated emissions :

0.96 - 1.61 GHz	<	-75.3 dBm/MHz
1.61 - 1.99 GHz	<	-63.3 dBm/MHz
1.99 - 3.1 GHz	<	-61.3 dBm/MHz
3.1 - 10.6 GHz	<	-41.3 dBm/MHz
10.6 GHz -	<	-61.3 dBm/MHz

Peak level of emissions :

The peak level of emissions contained within a 50 MHz bandwidth centered on the frequency at which the highest radiated emission is 0 dBm EIRP.

3. Satellite UWB

The satellite UWB system in this paper is a fixed satellite system, which employs a UWB type signal for the downlink transmission. Figure 1 shows the conceptual view of the satellite UWB system using the Ku-band. The UWB signal is usually characterized by transmitting very short monocycle wavelets or pulse-modulated carrier. As presented in Section 2, the signal bandwidth of the terrestrial UWB device is very wide such as more than 500 MHz. The assumed bandwidth used in the satellite UWB system is 500 MHz.

The satellite UWB has suitable characteristics for exploring new satellite services from the following perspectives :

- The UWB signal can be overlaid on the existing narrowband spectrum. This is expected to contribute to increasing spectrum efficiency of satellite systems.

- The terrestrial UWB device can be utilized for the satellite UWB application, which would reduce the cost of the satellite system. The terrestrial UWB device is expected to become very popular, and mass-production of the terminal will greatly reduce the production cost of the hardware.

4. Link budget

In the satellite UWB system, if the power transmitted from a satellite to the earth is at the same level as the terrestrial UWB device, the received signal on the earth is very low, and the transmission speed is limited very low. Therefore, higher power, which is comparable to that with existing satellite transponder, is assumed to be transmitted from the UWB satellite. This paper assumes the satellite transmission power as 108 W (20.3 dBW) and the transmitting satellite antenna diameter as 1.27 m. When these transmission characteristics are adopted by the satellite to transmit the UWB signal to the earth, radiated EIRP from the satellite is much greater than that of the terrestrial UWB. But the signal power density received at the earth's surface is assumed comparable to or smaller than that of the terrestrial UWB device as described below.

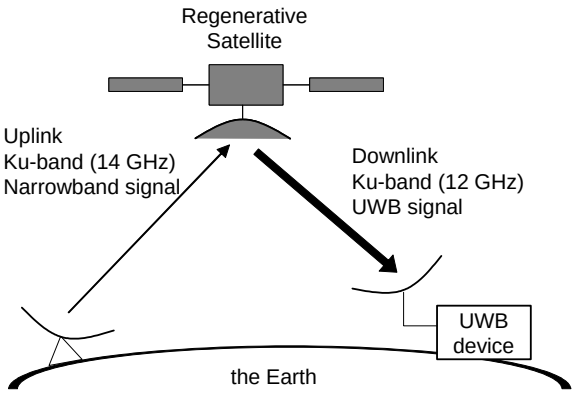


Fig. 1. Conceptual view of satellite UWB system using Ku-band.

The link budget of the downlink are estimated in the case where 500 MHz in the Ku-band is assumed as the downlink spectrum. Table 1 summarizes the downlink link budget of the system. The free-space path loss for the distance of 3 m at the center frequency of 6.85 GHz, a typical value for the terrestrial UWB device using the

Table 1. Downlink link budget.

Center frequency	12	GHz
Bandwidth	500	MHz
Transmission power	20.3	dBW
Satellite antenna diameter	1.27	m
Satellite antenna gain (efficiency = 60%)	41.8	dBi
EIRP	65.1	dBm/MHz
Link margin	5	dB
Rain margin	3	dB
Path loss to the earth's surface (at 12 GHz)	205.2	dB
Power density at the earth's surface	-148.1	dBm/MHz

3.1 - 10.6 GHz spectrum, is around 60 dB. In the terrestrial UWB device, the power density, which is given by [EIRP]-[Path Loss] in dB scale, at a distance of 3 m from a transmitter is -101.3 dBm/MHz. The table shows that the power level of the satellite UWB signal received at the earth's surface (-148.1 dBm/MHz) is much smaller than the signal level at the distance of 3 m of the terrestrial UWB. Therefore, the other service would not be affected by the satellite UWB system.

5. Throughput analysis

5.1. *M*-ary PAM UWB

The *M*-ary PAM(Pulse Amplitude Modulation) is a modulation scheme where information is modulated with +/-*M* amplitude variations. The pulse has a short duration, and its energy concentrates within the bandwidth of the satellite downlink, in the satellite UWB (Fig. 2).

Results of the research have been reported for the communication performance of the terrestrial UWB device. Here, the performance is discussed using the approach presented in Ref. [2].

A coherent detection is assumed as the demodulation scheme. The symbol error probability P_M of the *M*-ary PAM is given by

$$P_M = \frac{M-1}{M} erfc\left(\sqrt{\frac{3}{M^2-1}\times\frac{E_s}{N_0}}\right). \quad (1)$$

And the probability of a bit error P_b is [3]

$$P_b = \frac{1}{k} P_M, \quad (2)$$

where k is the number of bits, which are transmitted in one symbol, i.e. $k=\log_2 M$. Using Eqs. (1) and (2), the required E_s/N_0, a signal power per symbol to noise power density ratio, can be calculated. Table 2 shows the required E_s/N_0 for the bit error rate of 10^{-3}.

Table 2. Required E_s/N_0 for *M*-ary PAM.

M	Required E_s/N_0 [dB]
2	7
4	13.75
8	19.77
16	25.5

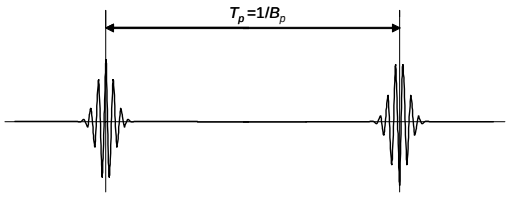


Fig. 2. Symbol of *M*-ary PAM.

On the other hand, E_s/N_0 is also presented by the following equation.

$$E_s/N_0 = P_{ave}T_p/N_0 = [P_{sd}/N_0]\times[B_s/B_p], \quad (3)$$

where,

P_{ave}: Average received power,
T_p: Pulse repetition period,
P_{sd}: Average power spectral density,
B_s: Equivalent occupied bandwidth, and
B_p: Pulse repetition frequency.

Equation (3) indicates that the pulse repetition period T_p becomes larger as required E_s/N_0 becomes larger. Taking the receiver noise figure N_F into consideration, the pulse repetition frequency B_p can be written as

$$B_p = [P_{sd}/N_0]\times B_s / N_F /[E_s/N_0]. \quad (4)$$

Table 3. Achievable throughput of *M*-ary PAM UWB [bit/s].

	2-ary	4-ary	8-ary	16-ary
0 [dBi] (Isotropic antenna)	9.96k	4.21k	1.58k	563
5.0 [dBi] (Patch antenna)	31.5k	13.3k	4.99k	1.78k
19.8 [dBi] (10 cm dish)	951k	402k	151k	53.7k
33.7 [dBi] (50 cm dish)	23.3M	9.87M	3.70M	1.32M
39.8 [dBi] (1 m dish)	95.1M	40.2M	15.1M	5.37M

Using $\log_2 M$ equal to the number of bits transmitted by one pulse, the achievable throughput R can be calculated as

$$R = B_p \times \log_2 M \,. \tag{5}$$

Assuming free-space propagation between a satellite UWB transmitter and a receiver, and also assuming P_{sd}=-208.1 [dBm/Hz], B_s=500 [MHz] from Table 1, N_0=-174 [dBm/Hz] at room temperature (17[°C]), and N_F=6 [dB], the achievable throughput can be calculated from Eqs. (4) and (5). Table 3 summarizes the achievable throughput of the M-ary PAM UWB transmitted from the satellite using the Ku-band.

5.2. Multiband UWB

In the terrestrial UWB, multiple transmission schemes adopting frequency-hopping over 3.1 - 10.6 GHz have been proposed at IEEE 802.15 TG3a. The mission of the standardization body is to define the physical layer specification for WPANs. Multiband UWB is a frequency-hopping scheme and has the feature of bit rate scalability with the occupied frequency.

Figure 3 shows an example of the symbol structure of the multiband UWB [4]. The symbol pulse consists of subpulses. And the subpulses are hopping over multiple frequency bands. Data is encoded into the sequence pattern of bands and phase information of the subpulses. The number of bits (N) transmitted by one symbol is

Table 4. Required E_s/N_0 for S—bands UWB.

S	Required E_s/N_0 [dB]
4	14.5
8	18

$$N = \log_2({}_S C_B \times {}_T P_B \times 2^{BP}), \tag{6}$$

where,

S: Number of frequency bands,
T: Number of subpulse time slots in a pulse,
B: Number of non-zero entries, and
P: Number of polarity bits.

${}_S C_B$ and ${}_T P_B$ indicate combination and permutation, respectively. In Eq. (6), data of $\log_2({}_S C_B)$ and $\log_2({}_T P_B)$ bits are transmitted by the sequence pattern, and data of $\log_2(2^{BP})$ (=BP) bits are transmitted by the phase information of the subpulses.

Assuming S=T=B, Eq. (6) can be written as follows;

$$N = \log_2(S!) + SP \,. \tag{7}$$

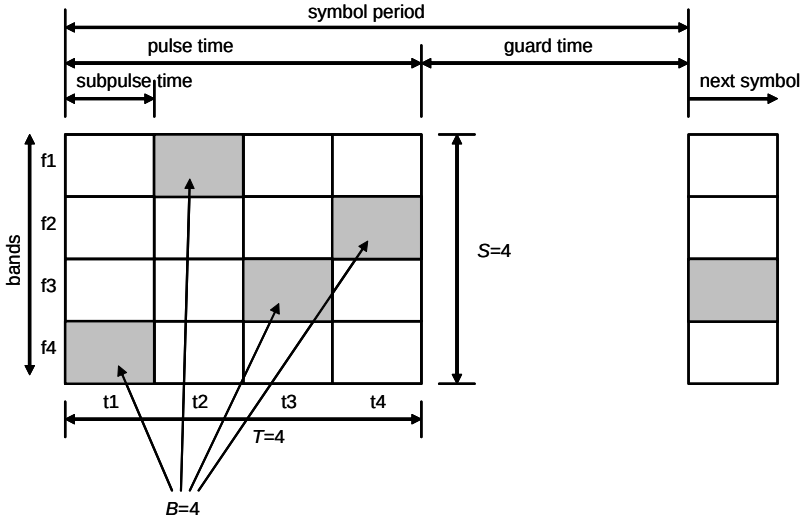


Fig. 3. Example of symbol structure of multiband UWB.

Upper bound of the subpulse error probability P_s of multiband UWB, which uses S bands, is given by

$$P_s = 4(S-1)Q\left(\sqrt{\frac{2E_{sp}}{N_0}}\right), \quad (8)$$

where,

S: Number of frequency bands,
E_{sp}: Energy per subpulse, and
N_0: Noise spectral density.

The relation between the energy per subpulse E_{sp} and the energy per symbol E_s is

$$E_s = E_{sp} \times S . \quad (9)$$

Using Eqs. (8) and (9), the required E_s/N_0 can be calculated. Table 4 shows the required E_s/N_0 for the subpulse error rate of 10^{-3}.

Similar to the M-ary PAM, the pulse repetition frequency B_p can be written as

$$B_p = [P_{sd}/N_0] \times B_s / N_F / [E_s/N_0]. \quad (10)$$

Because the number of bits in one symbol is expressed by Eq. (7), the achievable throughput R can be calculated as

$$R = B_p \times \left[\log_2(S!) + SP\right]. \quad (11)$$

Table 5. Achievable throughput of S-bands UWB [bit/s].

	4-bands		8-bands	
	BPSK	QPSK	BPSK	QPSK
0 [dBi] (Isotropic antenna)	15.2k	22.3k	18.4k	28.4k
5.0 [dBi] (Patch antenna)	48.1k	70.5k	58.3k	78.3k
19.8 [dBi] (10 cm dish)	1.45M	2.13M	1.76M	2.36M
33.7 [dBi] (50 cm dish)	35.6M	52.3M	43.2M	58.1M
39.8 [dBi] (1 m dish)	145M	213M	176M	236M

Using the same assumptions as the M-ary PAM, P_{sd}=-208.1 [dBm/Hz], B_s=500 [MHz], N_0=-174 [dBm/Hz] and N_F=6 [dB], the achievable throughput can be calculated from Eqs. (10) and (11). Table 5 presents the achievable throughput of the S-band UWB transmitted from the satellite using the Ku-band.

5.3. Analysis

Table 3 shows that the transmission speed of the binary PAM up to 950 kbit/sec can be realized using a very small user antenna such as 10 cm. Moreover, when a larger antenna is utilized, considerably larger throughput is realized.

In Table 4, by adopting the multiband UWB scheme, the satellite UWB transmission speed of over 1 Mbit/sec can be achieved using a 10 cm dish antenna. Throughput over 100 Mbit/sec is realized by utilizing a 1 m dish antenna.

In the process of conducting the transmission speed, the bit error rate of 10^{-3} is used at M-ary PAM, and the subpulse error rate of 10^{-3} is used at the multiband UWB. In the multiband UWB scheme, data is transmitted by the sequence pattern of bands and the phase information of the subpulses, so the relation between the subpulse error rate and the bit error rate is difficult to determine. As described above, in calculating the transmission speed, the error rate assumption of the M-ary PAM and the multiband UWB differs, so it is difficult to compare the transmission speeds directly. However, the satellite UWB using the M-ary PAM or the multiband UWB offers sufficiently high transmission speed, which means that these schemes are effective to fixed satellite communications.

6. Conclusion

Technical consideration and performance analysis are conducted for the satellite UWB system. The system could realize sufficient receive signal strength and throughput with a small antenna in addition to its inherent suitable characteristics to widely broadcast information to many users simultaneously. Satellite communication plays an important role in public communications. The satellite UWB enables new services, and is expected to open new markets.

ACKNOWLEDGMENT

We would like to express our appreciation to Dr. Toru ASAMI, president of KDDI R&D Laboratories for his advices on this research.

REFERENCES

[1] Federal Communications Commission, First Report and Order in ET Docket 98-153, “In the matter of Revision of Part 15 of the Commission’s Rules Regarding Ultra-Wideband Transmission Systems,” Apr. 2002.

[2] Jeff Forster, et. al., “Ultra-Wideband technology for short- or medium-range wireless communications,” Intel Technology Journal, 2nd quarter, 2001.

[3] E.A. Lee and D.G. Messercshmitt, Digital Communication, Kluwer Academic Publishers, 1988.

[4] Naiel Askar, “Overview of General Atomics PHY Proposal to IEEE 802.15.3a,” IEEE 802.15-03/105r1.

[5] S. Benedetto, E. Biglieri, V. Castellani, “Digital Transmission Theory,” Prentice-hall, 1987.

[6] J. Alper, J.N. Pelton, “The INTELSAT Global Satellite System,” American Institute of Aeronautics and Astronautics, 1984.

Spectrum Agile Radio: Detecting Spectrum Opportunities

Kiran Challapali, Stefan Mangold, Zhun Zhong
Wireless Communications and Networking Department
Philips Research Laboratories
Briarcliff Manor NY 10510, USA
{kiran.challapali | stefan.mangold | zhun.zhong}@philips.com

***Abstract* — The opening up of the unlicensed bands for commercial use has been a tremendous success. Wireless communications in computing, mobile, medical and consumer electronics market segments have grown rapidly in the past few years. Due to this success, radio resources in the unlicensed bands are progressively becoming scarce. Recently, the Spectrum Policy Task Force (SPTF) within the FCC has recommended that the FCC regulate spectrum allocation based on market principles. Such regulation implies radio networks wherein radios sense their environment and make opportunistic use of available radio resources while not interfering with the operation of existing licensed networks. In this paper, we focus on a key component of such Spectrum Agile Radio (SARA) systems, namely, the detection of spectrum opportunities. We present results of simulation studies of the use of Hough Transform and autocorrelation function for the detection of spectrum opportunities.**

***Keywords* — Spectrum Agile Radio, Opportunity Identification, Hough transform, IEEE 802.11k Radio Resource Measurement**

I. Introduction

The increasing popularity of radio communication networks over the last years, and of wearable, hand-held computing and communicating devices, as well as consumer electronics, indicates that there will be an ever increasing demand for radio communication networks providing high capacity communication. Considering this increase in demand, it is clear that the necessary radio spectrum will not be available in the future, due to the limited nature of radio resources. Today, consumer electronics radio communication systems operate mainly in unlicensed bands. Radio resources in the unlicensed bands are therefore often efficiently used [1]. However, most of the radio spectrum is allocated by traditional licensed radio services, and often not used at all. With the current FCC approach to regulation, radio spectrum resources are often not efficiently used.

This problem is approached by Spectrum Agile Radio (SARA) systems. SARA makes use of the licensed radio spectrum in an opportunistic way, controlled by SARA policies. A SARA device seeks opportunities, i.e. unused radio resources prior to communicating, and then communicates using the identified opportunities without interfering with the operation of licensed radio networks. Therefore, a key mechanism of SARA is to identify opportunities to communicate, and to identify other, competing radio systems. SARA systems will work with evolving FCC regulations for radio spectrum allocation that are based on Spectrum Policy Task Force (SPTF) recommendations [2].

Approaches to SARA are discussed in the context of Next Generation (XG) framework [3].

To facilitate the rollout of SARA, it should be built on top of existing radio communication standards such as IEEE 802.11 with its recent extensions for radio resource management [4]. Therefore, we discuss the emerging supplement standard to the popular IEEE 802.11 wireless Local Area Network (LAN) for radio resource measurements, namely IEEE 802.11k [5]. We discuss measurements based on the carrier sensing, i.e., Clear Channel Assessment (CCA), and approaches for spectrum opportunity identification from the obtained measurement results.

II. Radio Resource Measurement in IEEE 802.11k

IEEE 802.11 Task Group k (TGk) was formed in January 2003 to develop extension to IEEE 802.11 wireless LAN specification for radio resource measurement. This extension will specify the types of radio resource information to measure and the request/report mechanism through which the measurement demands and results are communicated among stations.

The goal of TGk is to provide tools by which a radio station can measure and assess the radio environment and take corresponding actions. To fulfill this goal, the current TGk draft defines seven types of measurements [5]:

- In Beacon report, a measuring station reports the beacons or probe response it receives during the measurement period.
- In Frame report, a measuring station reports information about all the frames it receives from other stations during the measurement period.
- In Channel Load report, a measuring station reports the fractional duration over which CCA indicates the channel is busy during the measurement period.

- In Noise Histogram report, a measuring station reports non-802.11 energy by sampling the channel only when CCA indicates that no 802.11 signal is present.
- In Hidden Node report, a measuring station reports the identity and frame statistics of hidden nodes detected during the measurement period.
- In Medium Sensing Time Histogram report, a measuring station reports the histogram of medium busy and idle time observed during the measurement period.
- In Station Statistic report, a measuring station reports its statistics related to link quality and network performance during the measurement period.

The measurements in TGk enable an IEEE 802.11 radio network to collect information of neighboring access points (via Beacon report) and information on link quality to neighbor stations (via Frame report, Hidden Node report and Station Statistic report). The tool set also provides ways to find out interference level (via Noise Histogram report) and medium load statistics (via Channel Load report and Medium Sensing Time Histogram report).

Those are useful information for a station to collect when assessing its radio environment. However, none of the measurement enables the station to identify future opportunities to use the medium. Ways to identify spectrum opportunities, and other interfering radio systems, are therefore discussed in the following.

III. Spectrum Opportunity Identification

As indicated in earlier sections, when *radio networks* encounter *other devices* that emit energy (and therefore use shared radio resources) in their vicinity, it is desirable to characterize the radio resource usage patterns of these other devices. Such a characterization of the usage patterns results in the identification of opportunities for the radio networks.

Other devices referred to previously includes radars, which are primary emitters, or other radio networks, which are secondary emitters.

III-1 Autocorrelation

A classical approach to determine periodic occurrences of spectrum opportunities, or radar pulses is based on the autocorrelation function. The sequence of CCA events obtained through listening to the channel, is processed with the autocorrelation function. Periods in the channel conditions are indicated by local maxima in the resulting function.

III-2 Hough Transform

In this section we will examine the use of Hough Transform for the detection of radar pulses as an example for any type of radio signals that create periodic patterns. We will use a version of the Hough Transform, known as Randomized Hough Transform (RHT) to detect the parameters of helixes wrapped around cylinders, as explained later in the section. The Hough Transform [6] has been studied in image processing literature

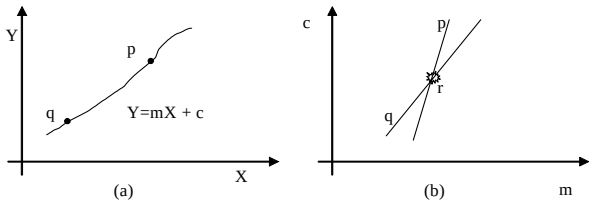


Fig. 1: Hough Transform used to detect straight lines (a) image space and (b) parameter space.

for detection of patterns such as lines, circles and ellipses in binary images. The effectiveness of Hough Transform in detecting patterns in data with many overlaying patterns and random noise is proven in [6]. In the presence of outliers, the Hough Transform is more robust than least squares estimation.

In brief, the Hough Transform is used to transform data from image space to an accumulator (or histogram) in parameter space, as shown in Fig. 1.

The image space is represented by (x, y), whereas, the parameter space is represented by (slope, intercept), that is (m, c). For each point in the image space (e.g. p and q), a line is generated in the parameter space as shown. The parameter space can be seen as a two dimensional histogram. A peak, r, in the parameter space corresponds to a line in the image space. The Hough Transform is robust because in the image space, a collection of collinear points is enough to result in a peak in the parameter space. However, it has the drawback that the parameter space could require large amount of memory in the computer. To address this drawback the RHT was developed [7]. The RHT as applied to straight-line detection, results in randomly picking pairs of points and computing and accumulating a parameter (for instance, slope). When enough confidence in the peak is achieved, the process stops, thus reducing both memory and processing time.

The use of Hough Transform for radar pulse detection was first studied in [8]. The original radar pulse train is a 1-D signal.

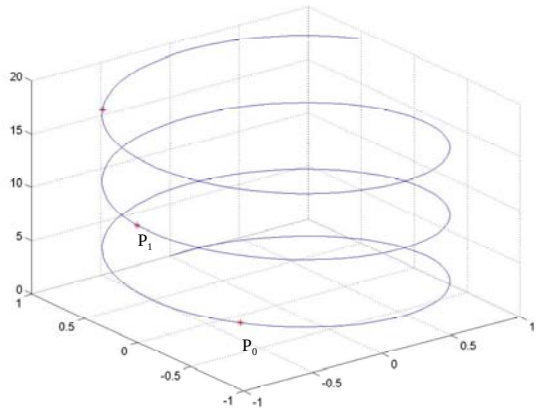


Fig. 2: A helix given by the Eq. (1), for $\varpi = 1$.

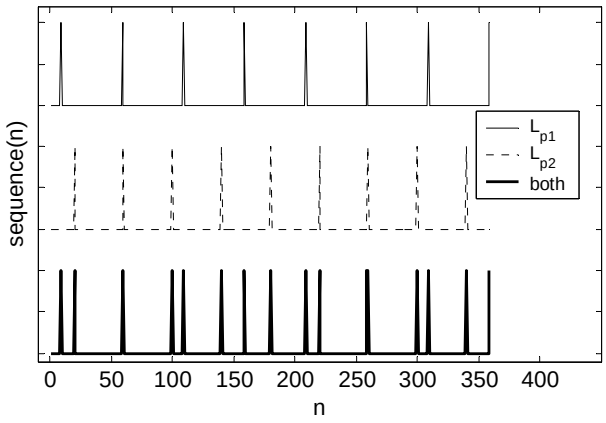


Fig. 3: Discrete sequence vector used for evaluation.

The authors have used 1-D to 2-D transformation (like a raster scan) and then applied the Hough Transform to detect straight lines, which correspond to pulse trains. Furthermore, they have computed the noise floor. We extend their work by first transforming the 1-D signal to a 3-D helical signal, and apply RHT to it. A helix may be represented by the following parametric equations:

$$\begin{aligned} X(t) &= \sin(\omega t) \\ Y(t) &= \cos(\omega t\ ) \\ Z(t) &= t \end{aligned} \tag{1}$$

This helix is cylindrical (as opposed to the more general elliptical) and has unit radius. Based on the parameter ϖ a new helix can be generated that wraps around the cylinder more slowly as ϖ decreases. In Fig. 1, the points (marked with *) on the helix themselves form a helix, with an ϖ value less than one. Given two points on the helix P_0 (x_0, y_0, z_0) and P_1 (x_1, y_1, z_1), the parameter ϖ can be given by the following equation:

$$\omega = \frac{a\tan\left(\frac{y_1}{x_1}\right) - a\tan\left(\frac{y_0}{x_0}\right)}{z_1 - z_0} \tag{2}$$

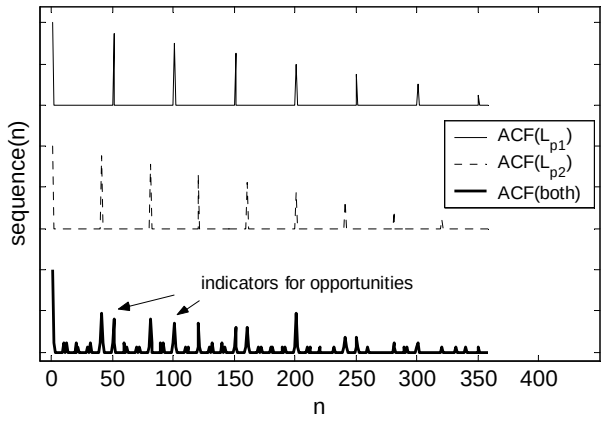


Fig. 4: Autocorrelations (right hand side) of the three sequences.

If the two points P_0 and P_1 are inside one whorl of the helix, then ϖ works out to be 1. The length of the line segment given by one twirl of helix is given in Eq. (3).

$$l = 2\pi \cdot \sqrt{1 + \left(\frac{1}{\varpi}\right)^2} \tag{3}$$

IV. Evaluation and Basic Concepts

We discuss the RHT and the autocorrelation approach separately in the following.

IV-1 Randomized Hough Transform

Let us represent the location (time-of-arrival) of the radar pulse train with the discrete sequence vector L_p. The sequence L_{p1} = [9, 59, 109, 159, 209, 259, 309, 359] is shown in Fig. 3, top sequence. The corresponding right-hand side of the autocorrelation function is indicated in Fig. 4, top sequence. For this sequence, the ϖ histogram as indicated in Fig. 5 is obtained by the Hough Transform. For this case, $\varpi = 0.116$. Now let us consider the case where there are two pulse trains that are multiplexed and represented by L_{p2} = [9, 20, 59, 60, 100, 109, 140, 159, 180, 209, 220, 259, 260, 300, 309, 340, 359], as illustrated in Fig. 3, bottom sequence.

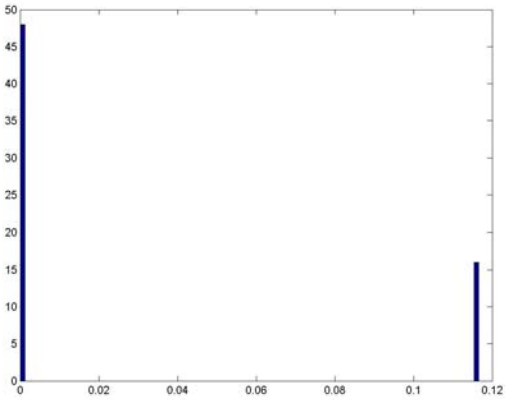


Fig. 5: Histogram of ϖ for Lp1

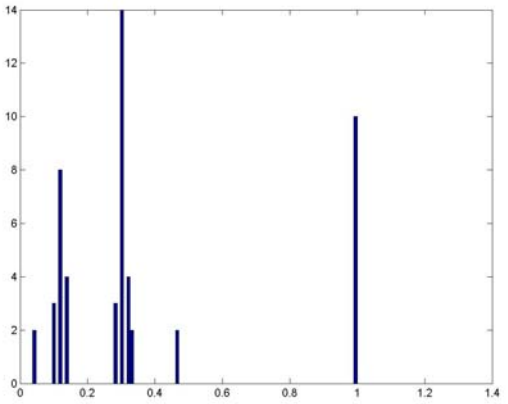
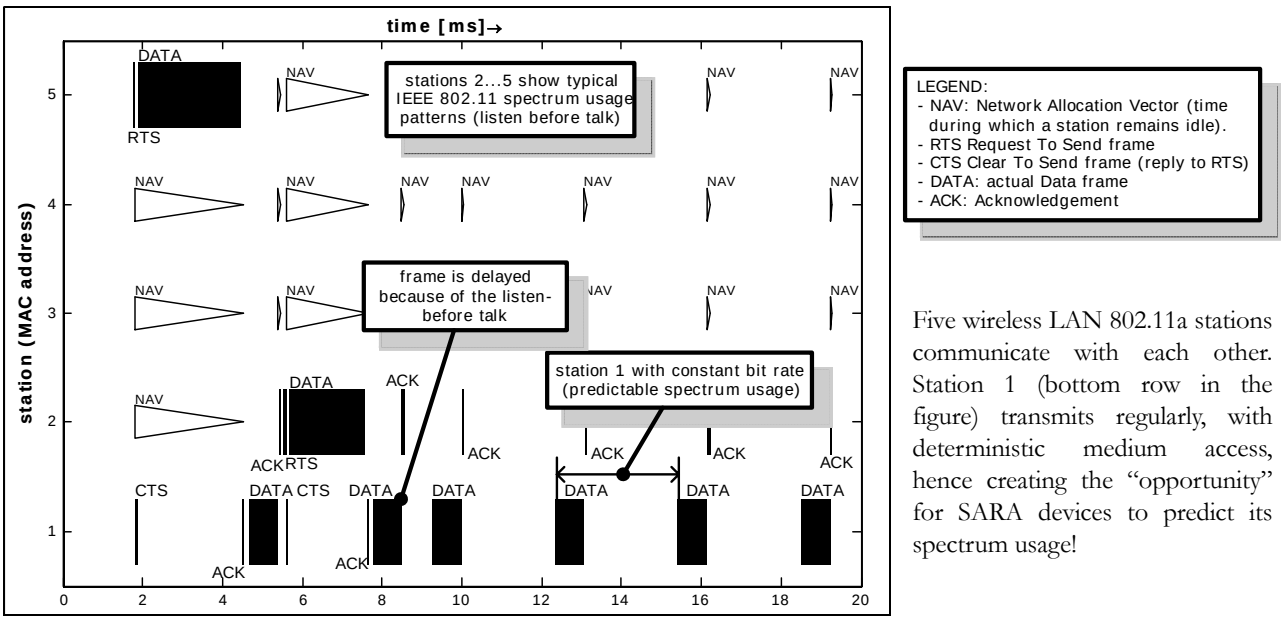


Fig. 6: Histogram of ϖ for Lp2



Five wireless LAN 802.11a stations communicate with each other. Station 1 (bottom row in the figure) transmits regularly, with deterministic medium access, hence creating the "opportunity" for SARA devices to predict its spectrum usage!

Fig. 7: Spectrum usage pattern for wireless Local Area Network (LAN) IEEE 802.11a (5GHz unlicensed band).

The corresponding right-hand side of the autocorrelation function is indicated in Fig. 4, bottom sequence. The ϖ histogram as indicated in Fig. 6 is obtained by the Hough Transform of this multiplexed sequence. Note that $\varpi = 1$ corresponds to points on the helix within one whorl.

The advantage of the autocorrelation function, namely, its notional simplicity has to be balanced with its disadvantage, namely computational complexity. Similarly, for the Hough Transform, its advantage of computational simplicity and robustness has to be balanced with its disadvantage namely possible dependence on the choice of parameters.

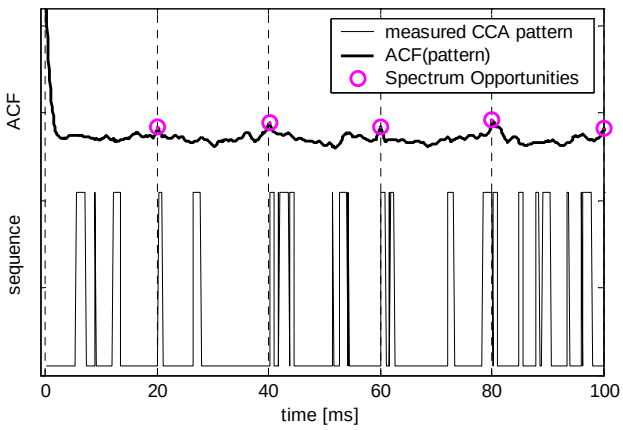


Fig. 8: The first *100ms* of the spectrum usage pattern (bottom) and corresponding ACF, for a measurement duration of *1000ms*. The "Spectrum Opportunities" indicate the detection of deterministic spectrum usages.

IV-2 Autocorrelation Function

Fig. 7 illustrates a typical spectrum usage pattern of IEEE 802.11a, when five stations communicate. The dark solid fields illustrate frame transmissions, the triangles illustrate timers that are set by the individual stations. Station 1 carries a traffic that offers a constant bit rate, and hence produces a deterministic spectrum usage pattern, because the intervals between consecutive frame exchange attempts that are initiated by station 1 do not change over time. However, the medium is busy when other stations transmit, and during busy times, station 1 does not access the medium, because of the nature of the listen-before talk based medium access control protocol in IEEE 802.11. Apparently, the autocorrelation function is suitable to determine the deterministic medium accesses, and to assess what the period of the medium access is. This is illustrated in Fig. 8 and Fig. 9. In these figures, a spectrum usage pattern similar to the one in Fig. 7 is illustrated (bottom graph in the two figures, the deterministic medium access occurs every *20 ms* and is embedded in other random frame exchanges), and the corresponding autocorrelation functions (top graph in the two figures). It can be seen how the deterministic medium access is identified. The difference between the figures lies in the length of the measurement duration: whereas for the identification of spectrum opportunities, in Fig. 8 the measurement duration was *1000 ms*, the measurement duration for Fig. 9 was only *100 ms*. Spectrum opportunities, i.e., significant local maxima in the autocorrelation function, are indicated. For better comparison, in both figures the first *100 ms* of the measured CCA patterns are shown. When comparing the two figures, it can be seen that with the longer measurement durations, spectrum opportunities are more reliably identified, at the cost of higher computation effort, and longer measurement durations.

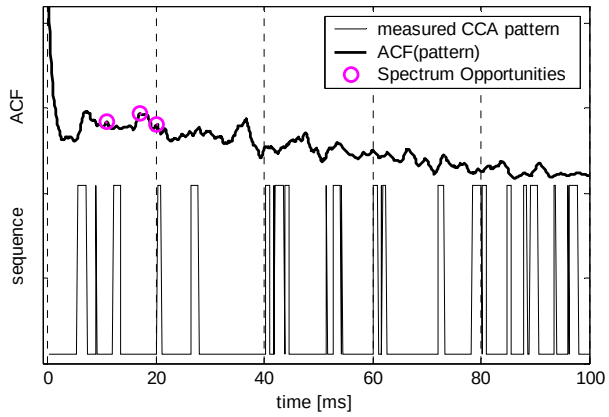


Fig. 9: Spectrum usage pattern (bottom) and corresponding ACF. Measurement duration: *100ms*.

V. Conclusion

We have outlined and compared two approaches for spectrum opportunity identification, and radio system identification, based on the CCA mechanism of IEEE 802.11. We use the autocorrelation of the sequence of CCA events as well as the random Hough transform of the data. We have shown that introducing this type of measurement into IEEE 802.11 (for example as part of 802.11k), provides a first step towards Spectrum Agile Radio. In the future, the two methods described to identify periodic accesses to the radio spectrum may be associated with each other in order to increase the precision and accuracy. We expect that both alternatives show advantages and disadvantages in different scenarios, and a combination of both may therefore result in the most precise identification of other radio systems.

References

[1] MANGOLD, S. AND CHALLAPALI, K. (2003) Coexistence of Wireless Networks in Unlicensed Frequency Bands. In: *9th Wireless World Research Forum*, Zurich Switzerland 1-2 July 2003.

[2] Spectrum Policy Task Force web site as of October 2003, http://www.fcc.gov/sptf/

[3] XG WORKING GROUP, "The XG Vision." Request For Comments, version 1.0. Prepared by: BBN Technologies, Cambridge, Massachusetts, USA. July 2003.Available from: http://www.darpa.mil/ato/programs/XG/rfcs.htm [Oct 03].

[4] MANGOLD, S. AND CHOI, S. AND HIERTZ, G.R. AND KLEIN, O. AND WALKE, B. (2003) Analysis of IEEE 802.11e for QoS Support in Wireless LANs. *IEEE Wireless Communications*. Dec 2003, vol. 10 no. 6, pp. 40-50.

[5] IEEE 802.11 WG, Draft Supplement to STANDARD FOR Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Specification for Radio Resource Measurement, IEEE 802.11k/D0.7, Oct 2003.

[6] ILLINGWORTH, J., AND KITTLER, J., "A survey of the Hough Transform," *Computer Vision, Graphics and Image Processing (CVGIP)* 43, 1988.

[7] XU L., OJA E., AND KULTANEN P., "A new curve detection method: Randomized Hough Transform (RHT)," *Pattern Recognition Letters*, 11(5), 1990.

[8] PERKINS, J., AND COAT, I., "Pulse train deinterleaving via the Hough Transform," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Volume: iii, April 1994.

Alternative Communication Networking in Polar Regions

Abdul Jabbar Mohammad, Nandish Chalishazar, Victor Frost, Glenn Prescott
Information and Telecommunication Technology Center, University of Kansas
Phone: 785-864-7747, Fax: 785-864-0387 E-mail: jabbar@ittc.ku.edu

Abstract – *Research is being conducted in the Polar Regions that generates significant quantities of important scientific data. Real-time exchange of this information in the field and access to the Internet is crucial. This paper presents a reliable, truly mobile, lightweight, and relatively inexpensive integrated data communications system to provide wireless Internet access in remote regions. It describes the work done as part of the Polar Radar for Ice Sheet Measurements (PRISM) project to support data communication requirements of science expeditions in the harsh climactic and technologically challenged regions of Greenland and Antarctica. An inverse multiplexed, multi-channel Iridium based system integrated with a long-range 802.11b network is developed to provide wireless Internet access at moderate speeds. Results of field experiments conducted at the North GRIP site in Greenland to evaluate the overall performance of the system are presented. The system has an average throughput of 9.26 Kbps and efficiency greater than 90%. The average time interval between call drops is observed to be 100 minutes with modem uptimes as high as 95%, which means the system is suitable for autonomous operation. Experiments conducted using the Wi-Fi system showed reliable communications over a distance of 10 Km with 802.11b throughputs varying from 4.8-0.23 Mbps depending on the signal to noise ratio at the receiver. The integrated communication system proved to be a reliable, lifeline alternative data/Internet connection in Polar Regions.*

1. Introduction

Modern telecommunication facilities have grown tremendously over last few decades. While the local area network speeds exceed hundreds of Mbytes/sec, wide area networks too have advanced from dial-up 56Kbytes/sec connections to T1 lines operating at 1.5 Mbps to fiber links at several Gigabits per second. But, these state-of-the-art technologies have evolved mainly in developed areas. To this day, there are places where these technologies have not penetrated for one reason or another. Some are developing nations; where as other regions are geographically remote. Arctic, Antarctic and other remote regions are such places where data and Internet access still remains an issue.

Though commercial broadband satellite systems have helped to solve the problem in some of the populated regions, they offer intermittent coverage in oceans and the Polar Regions. Such coverage ceases to exist beyond 70^0 N/S latitudes. On the other hand, research in Polar Regions involving data collection and telemetry has grown significantly over the past few years. Numerous field expeditions are being conducted at various locations year round. The telecommunication requirements of the Polar science community are continuously increasing. In 1999 the bandwidth requirement of the South Pole station alone is estimated at 14Gbytes/day [5]. Since the NASA satellites (ATS3, LES9, GEOS, TDRS1 and MARISAT2), currently providing broadband access to these regions [7] is either geostationary or geosynchronous, they have limited visibility window at poles as seen in figure 1. Further, they have a very low elevation angle (about 1-4 degrees) from Poles, which combined with high altitude of the satellite results in extremely large field equipment [3] (10-meter radii antennas) that had to be properly pointed towards the satellite. Hence, these systems cannot be used for small field camps and science expeditions

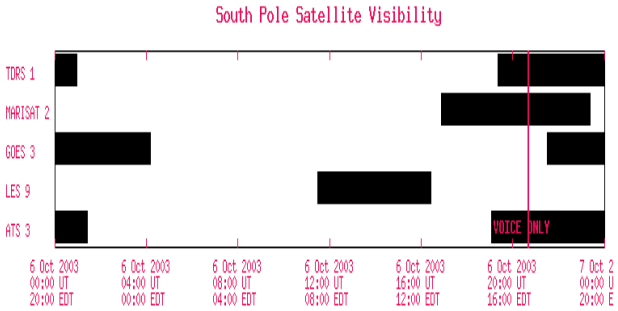


Figure 1 Satellite visibility at South Pole (Source: [8])

Thus, the need for a compact, easily portable and field deployable data communication system with round the clock coverage is clearly evident.

In this paper, we present a mobile lightweight integrated data communication system of moderate capacity based on Iridium satellite system, which is the only commercial satellite system with truly global

coverage. In order to increase the otherwise low capacity (2.4 Kbps) of the Iridium system, packet level inverse multiplexing is implemented using Multi-link point-to-point protocol (MLPPP) [4]. This mechanism combines multiple channels to obtain a seamless data connection with a capacity approximately equal to the sum of the individual link rates providing a reliable round the clock data communication system with *scalable capacity*.

A local area network (LAN) with an extended range is also required to share data and Internet access among the field participants and sensors, scattered over a wide area in the filed camps. Field camps in the past have known to use wired Ethernet LANs, which does not provide ubiquitous connectivity around the camp and suffers from the disadvantage of laying cables on the ice. A system of modems and a 400 MHz UHF radio telephone (Opti-phone) used for networking with other field camps has a low baud rate (9600 bps), requires Line of Sight and is not intended for mobile platforms. These issues could be addressed with a wireless LAN (WLAN) of a reasonable range in the Polar Regions. Long-range 802.11b installations in the past used parabolic antennas [6], which are unsuitable for mobile applications and do not provide ubiquitous coverage. Thus, an 802.11b system with external amplifiers and vertical collinear omni directional antennas is used here to provide wireless Internet and data access over a range of 10 Km for land-mobile and mobile-mobile systems.

2. System architecture and operation

2.1 Multi-Channel Iridium System

The remote subsystem system consists of 4 Motorola-Iridium modems connected to a rugged laptop with an USB-to-Serial converter as shown in figure 2. The antennas of these modems are installed on a metal plate of 1 sq ft that forms the ground plane. These four antennas are then mounted on a frame such that they are separated from each other by 2 ft in order to reduce the effects of interference. The PPP daemon (PPPD) on the remote terminal is configured as a PPP client so as to connect to the local terminal. The computer terminal at the local end has four PSTN modems connected to it via a multi port serial card and an octopus cable. This terminal is configured as a PPP server to receive call from the remote Iridium system through 4 PSTN phone lines.

The developed link management software configures the remote terminal as a PPP client, dials the Iridium modem to establish the serial connection with the mainland terminal and handles PPP negotiation to complete a point-to-point data connection. Once the basic connection is established, it dials the remaining modems and seamlessly attaches them to the first PPP connection, forming a single higher bandwidth pipe. Standard Internet protocols, TCP/IP are then used to provide end-to-end connectivity. This MLPPP system was developed in Linux and also monitors the satellite connection, detects any call drops during satellite hand-offs and immediately reconnects dropped modems. Further, it handles power failures and system resets providing a reliable and fully autonomous data link.

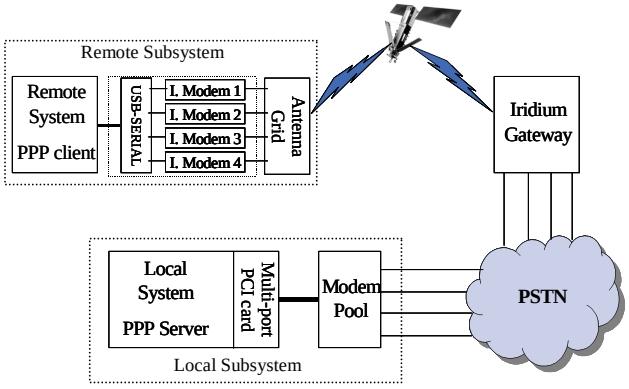


Figure 2: Four-channel Iridium communication system

The IP packets of a single application from the remote system are fragmented at the MLPPP layer into smaller segments depending upon the packet size, the link Maximum Transmission Unit (MTU) and the availability of the links. These segments are then sent simultaneously over multiple satellite links. The MLPPP layer at local system on the other end combines the received segments into the original data packet and checks for errors or segment loss. If the packet is successfully reconstructed, it is presented to the IP layer; else a packet error is reported. TCP/IP layer handles any errors or packet losses that occur. The same procedure is followed for packets going in the other direction (local system to remote system).

2.2 Wi-Fi System

The Wi-Fi system consists of a central base station that serves as an access point (using Orinoco AP-500) and is interfaced via Ethernet to the Iridium system. In order to increase the range of the access point and hence the wireless network, it is required to amplify the signal strength to overcome the propagation losses. Thus, the access point is connected to a 1-Watt bi-directional amplifier, which is connected directly to a 9-dBi vertical collinear antenna using a male-to-male UHF adapter and mounted on a mast 3 meters above the ground level. Cable connection from the amplifier to the antenna is avoided in order to minimize the losses and also reduce the noise figure of the receive system. The bi-directional amplifier in the receive mode acts as

a low noise amplifier (LNA) directly connected to the antenna and thus minimizes the noise figure of the receive system. The vertical collinear antenna has a horizontal beamwidth of 360 degrees and a vertical beamwidth of 7 degrees.

The basic WLAN setup shown in figure 3 extends the range of the access point to users and sensors situated within a few hundred meters of the base station. The users are provided with 802.11b wireless client cards that can be plugged into their laptops to enable Internet as well as data access. An extended 802.11b network is also provided to extend the coverage to mobile vehicles and other users situated at distances as far as 10 Km from the base station. The mobile vehicles consist of rugged laptops with 802.11b wireless clients connected to a 9-dBi vertical collinear antenna via a 1-watt bi-directional amplifier.

The choice of the amplifier and antenna are made based on a two-ray propagation model [1] that predicts a fourth power loss with distance. The height of the antenna mounted on the mobile vehicle also influences the received signal strength as predicted by the model and increases by 6 dB on doubling the height of the antenna. Based on this fact, the vertical collinear antenna is mounted on the mobile vehicle at a height of 3m from the ground to overcome the propagation losses and ensure reliable data communications over 10 Km.

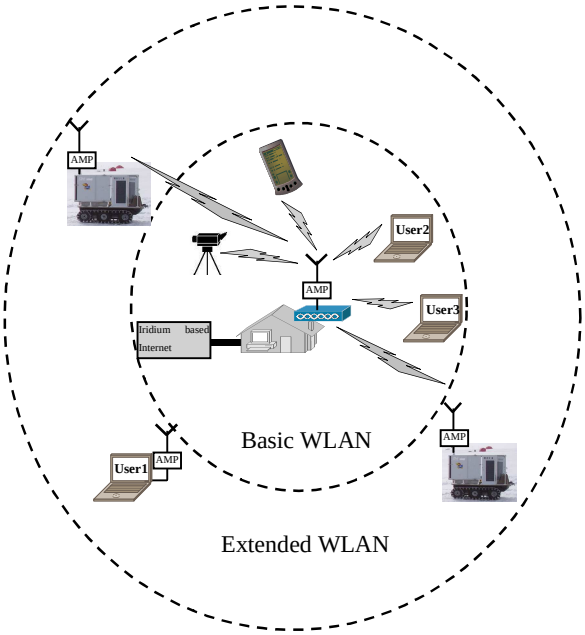


Figure 3: Long range wireless LAN

3. Network Architecture

The end-to-end network architecture that was used to provide data and Internet access to Polar field research camps/sites is shown in figure 4. A specific implementation of the system between the Polar (Greenland) Field camp and University of Kansas is used to illustrate the network architecture. This architecture could be generalized to provide Internet access to any remote field site from a mainland facility or to provide data connection between two remote field sites.

The remote field site is configured as a subnet of the mainland local network; like University of Kansas in the above example. The PPP client (system with multi-channel Iridium system) is configured as the default gateway of the Greenland subnet. Data from wired (Ruser4) and wireless users (Ruser1-Ruser3) of figure 4 is routed by the PPP client over the satellite link (through the PPP0 interface) to the PPP server in the University. PPP server being a part of the university's network routes the data packets through the University router to the World Wide Web.

Similarly, PPP server is configured as the default gateway to forward packets going from the University network and Internet to the remote (Greenland) network over the satellite channel. A static route on the University router forwards all the traffic intended for the remote subnet to the PPP server.

4. Field Experiments and Results

A goal of the summer 2003 Greenland field experiments was to determine the performance of integrated Iridium/Wi-Fi based data communication system in a polar environment. The field experiments were conducted at the NorthGRIP ice core drilling camp in Greenland (75° 06' N, 42° 20' W) from June 23-July 17, 2003.

4.1 Iridium System Performance

Field experiments were conducted to evaluate the performance of multi-link Iridium point-to-point communication system. Further, the goal was to obtain quantitative network performance data from the field, including 24hr access, call drops, packet loss and delay; this data would be used to evaluate the throughput and reliability of the system. The objectives also included evaluating the suitability of the link for the transfer of large files, e.g. non-real time video, and real time video/audio communications

4.1.1 Delay and Loss Performance

Ping measurements were done at various times during 24-hour period to determine the round trip delay of the multi-channel Iridium system. The experiment was repeated on several days to obtain the average delays. Table 1 shows the system round trip time (RTT) and packet loss observed.

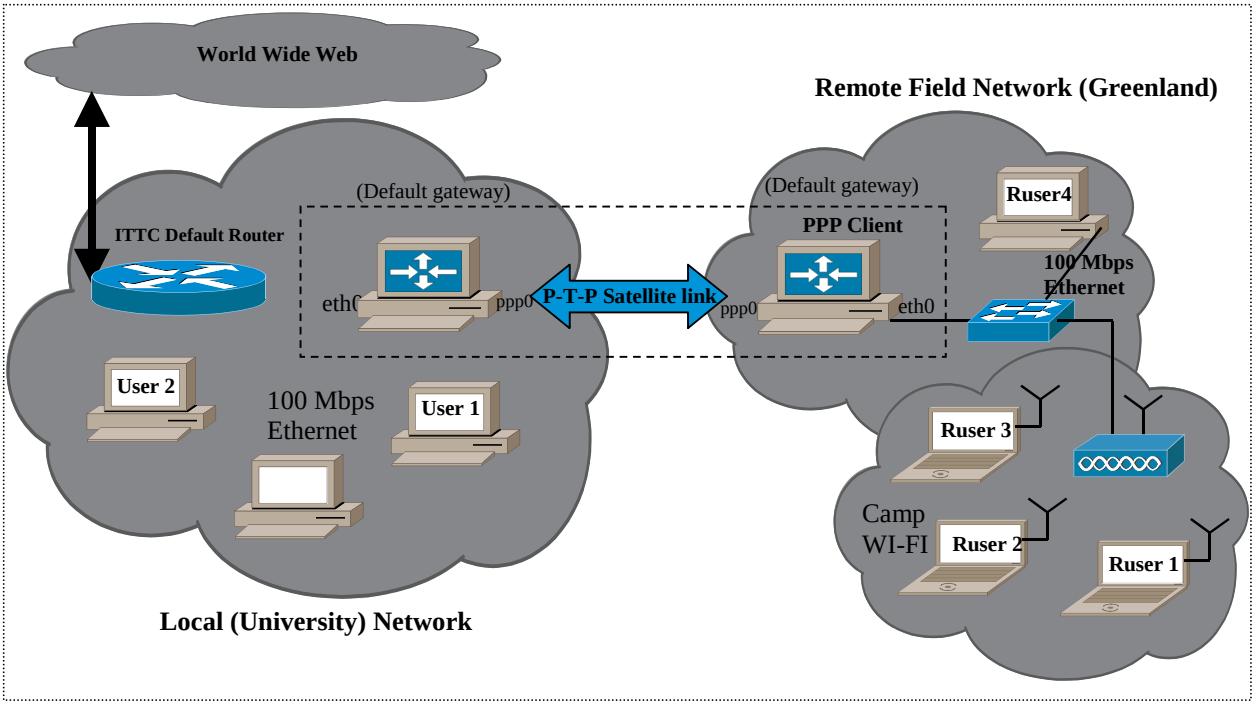


Figure 4: Network architecture to support data and Internet access to Polar Regions

Table 1: Round trip time and packet loss of system

Packets Sent	Packets Received	% Loss	RTT (sec)			
			Avg	Min	Max	Mdev
50	100	0	1.835	1.347	4.127	0.798
100	100	0	1.785	1.448	4.056	0.573
100	100	0	2.067	1.313	6.255	1.272
200	200	0	1.815	1.333	6.228	0.809

In order to understand the experimental results, first consider the theoretical end-to-end delay observed by a 64byte packet between Greenland and University of Kansas.

The Propagation segments are: Satellite uplink (from Greenland) – Iridium satellite channel – downlink at Hawaii gateway – PSTN link Kansas.
Distance traveled =800+8000+800+6000 =15600 Km
Propagation time = distance traveled/speed of light = 15600 Km/ (3e5) Km/sec= 52msec
Transmission time for a 64 bytes@2.4Kbps = 64*8/2400=213msec
Unknown parameters = inter satellite switching time + processing time at gateway + additional overheads
Theoretical end-to-end delay = 265 msec + unknown value

Though the theoretical RTT is 530 msec, the average RTT during field experiments, as seen in Table 2, was observed to be about 1.8 seconds. The additional delay could be attributed to the inter-satellite switching, processing at the gateway, call hand-offs and the constantly varying satellite constellation.

4.1.2 Throughput Performance

In order to determine the efficiency of the multi-link system, throughput measurements were done with increasing number of modems. Again, the experiments were repeated several times. Figure 5 displays the average throughput as a function of number of modems. These results were obtained using the TTCP and IPERF bandwidth measurement tools.

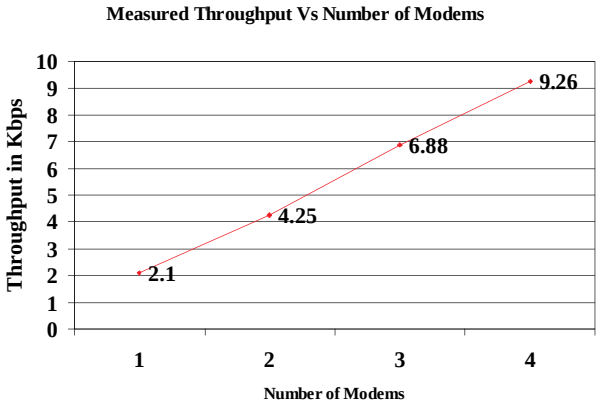


Figure 5: Throughput of the multi-link system

The maximum throughput observed with 4 modems was 9.7 Kbps with an average throughput of 9.26 Kbps. The system on an average was thus about 96% efficient. The system was also used to upload large video files from the field, ranging in size from 0.75 MB to 3.2 MB. The results of these file transfers using FTP are shown in Table 2.

Table 2: File transfer throughputs

File Size (MB)	Upload Time (min)	Throughput (bits/sec)
0.75	11	9091
3.2	60	7111
1.6	23	9275
2.3	45	6815
1.5	28	7143
2.5	35	9524

Modem call drops during the transfers resulted in the throughput being less than expected in some cases. However, it is important to note that these large files were successfully transferred automatically even in the presence of call drops, indicating that the link management software operated properly.

4.1.3 Reliability – Modem Call Drops

Initial studies on Iridium call drops [2], [6] have reported a call drop rate of 6-18% based on call duration of 10-15 minutes. This means 6-18% of the calls were dropped with in the first 15 minutes of the call. But, it should be noted that if the remaining 94-82% of the calls were to be continued for long periods of time, it is likely that they would eventually experience a call drop. Hence the performance criterion is not how many calls are dropped, but the interval between the call drops.

To determine the reliability of the system we conducted two 24-hour tests. During these tests the management software controlled the 4-channel communication system. The management software detects any call drops/link failures, logs the event, automatically redials the dropped link and attaches it to the multilink bundle. Figure 6 shows the call drop pattern on the first modem, which defines the basic connection itself. Since calls drop on the first modem results in the termination of the entire session, these call drops represent the system failures. During the 24-hour test 13 call drops were recorded. The average connection time between the call drops was observed to be approximately 100 minutes. The overall percentage up time on the first modem was about 96%. The longest up time without a call drop was observed as 618 minutes. The typical time to make a connection is 1 minute while on an average it takes 2 retries to reconnect after a call drop.

4.1.4 Reliability – Modem Up times

It should be noted that a drop on the first modem results in a complete loss of the communication link, whereas call drops on the other modems result in a brief reduction in the available bandwidth before the management software reconnects the dropped link and attaches it to the bundle. In order to determine the available bandwidth during the same 24-hour test period, the number of online modems vs. time is plotted in figure 7. The statistics obtained from this graph are shown in Table 3. It is seen that during 80% of the test time all the 4 modems were running providing full bandwidth of 9.6 Kbps. On the other hand we had at least one modem connected for 96% of the time.

Table 3: Statistics of the 24-hour test

Number of online modems	Up time (min)	% Up time
All the 4 modems	1161	80.6
At least 3 modems	1323	91.8
At least 2 modems	1365	94.7
At least 1 modem	1395	96.8

4.1.5 System Performance under Motion

The system was tested while the communications system, the four modems and their antennas, were in motion. The platform was transported at speeds up to 20 m/h. The system performance with and without motion was comparable. However, the number of attempts to complete a connection increased with the system in motion.

4.1.6 Qualitative Performance

The Iridium communication system was used for the transfer of large files to support the general activity of the NGRIP camp. Files as large as 7.2 MB were downloaded from various Internet hosts. In combination with a modified Wi-Fi deployment, the system provided Internet access for the entire NGRIP camp; this was the first time the camp had such a capability and it was very well received. The NetMeeting software was used to test the real time video/audio capabilities of the system. As expected, the long delays made real time interactions difficult.

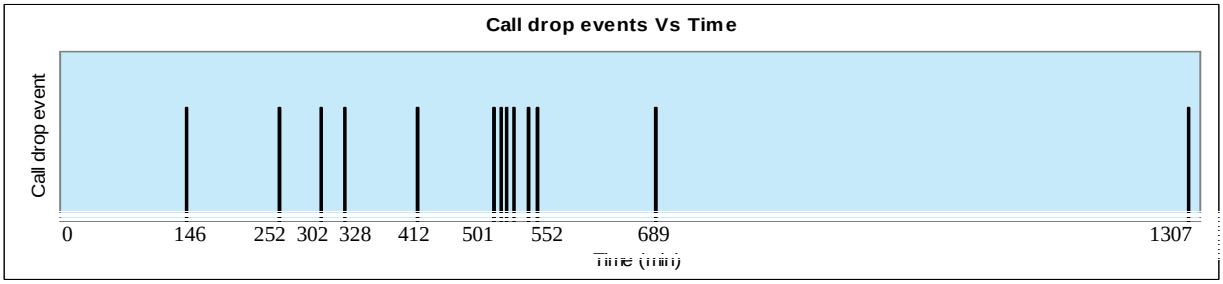


Figure 6: Call drop pattern of the first modem during the first 24 hour test

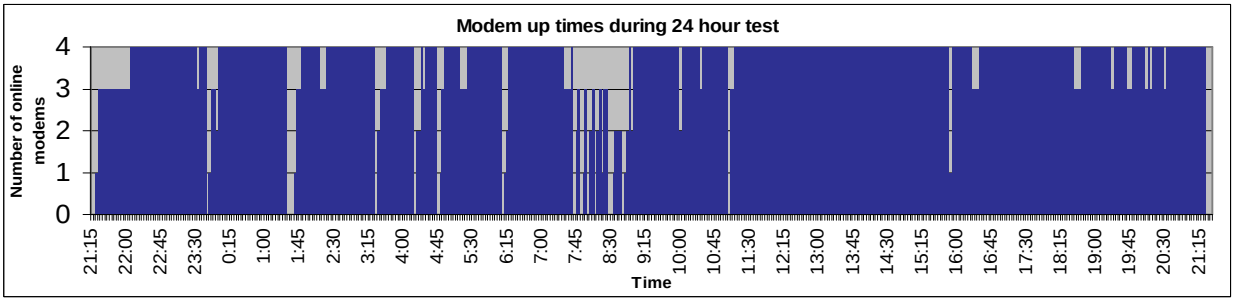


Figure 7: Availability of modems during the first 24 hour tests

4.2 802.11b system performance

Basic WLAN and Extended WLAN tests were carried out with the Wi-Fi system installed at the NorthGRIP site in Greenland. It was required to determine variation of the received signal strength with distance, which in turn determines the range and throughput achievable in both these networks.

4.2.1 Basic WLAN measurements

The hardware setup for the basic WLAN measurements consisted of the base station with the access point connected to external amplifier and antenna as shown in Figure 4. Measurements involved plugging an Orinoco 802.11b wireless client into a rugged laptop and moving to different locations in the vicinity of the base station. The signal to noise ratio (SNR) and throughput are measured at each location. Experiments are conducted at a time when no other user is allowed to access the Internet, and all the four modems of the Iridium system are operational at the start of the experiments. Results of the basic WLAN measurements shown in figure 8 reveal a range of around 1 Km and throughputs varying from 8.9-9.67 Kbps over this range.

4.2.2 Extended WLAN - Received signal strength and radio propagation model

In an extended WLAN, peer-to-peer field experiments are carried out between a base station and a mobile vehicle as shown in figure 3. The access point in the base station is replaced with a Orinoco 802.11b client plugged into a rugged laptop and connected to the external 1 watt amplifier and 9 dBi vertical collinear antenna. A Garmin GPS 12 receiver is connected to the RS-232 port of the rugged laptop on the mobile vehicle to log the latitude and longitude along a traverse.

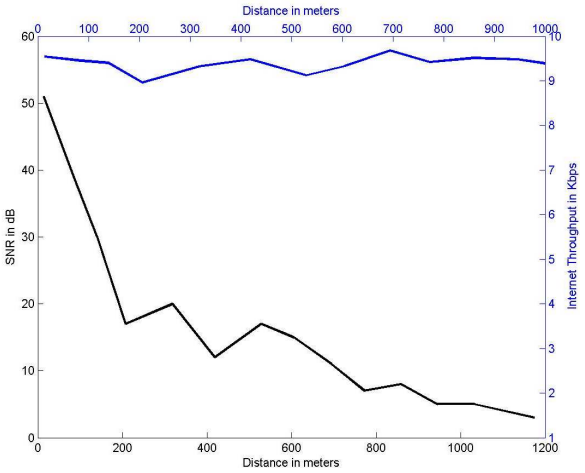


Figure 8: SNR and throughput in a basic WLAN

Once the hardware is installed and antenna fixed at a particular height, the mobile vehicle is used to traverse along the ice to measure the received signal strength, SNR and throughput up to a distance of 8 km from the base station. The Orinoco client manager software logs the received signal strength, noise level, SNR and the corresponding time stamp measured every 2 seconds. The latitude, longitude and altitude information from the GPS with the corresponding time stamp is also logged every 2 seconds.

Measurements were carried out for different combinations of antenna height at the base station and the mobile vehicle to gain a better understanding of the radio propagation model for communication over ice. The antenna at each end could be raised to one of the four different heights of 1.4,2,3 and 5m using a variable length mast on which the antenna is mounted. Further, the above-mentioned tests were repeated along different tracks, each of length 8 km from the base station.

The variation of the received signal strength with distance for six different combinations of antenna heights at the base station and mobile vehicle was analyzed and compared with the theoretically predicted received signal strength The effects of using a multi-element vertical collinear antenna that is used during the field experiments is included in the two-ray propagation model to obtain a realistic prediction of the received signal power. The measurements and the corresponding theoretically expected results for two specific cases (base station antenna height=3m, mobile antenna height=3m and base station antenna height=3m, mobile antenna height=1.4m) are shown in figures 9-10.

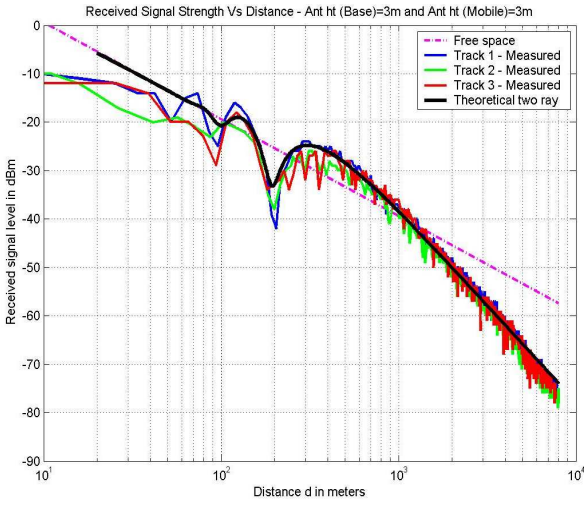


Figure 9: Received signal strength variation with distance for base station and mobile antenna heights of 3m with a GPS error of +10m

It is seen from these plots that the received signal strength variation matches very well with theoretical two-ray propagation model. The lobing pattern observed in the measured results before the Fresnel break point is well accounted by the effects of using a multi-element antenna over the flat ice surface and the results confirm the validity of using the two-ray propagation model for communication over the flat ice sheets in Polar Regions.

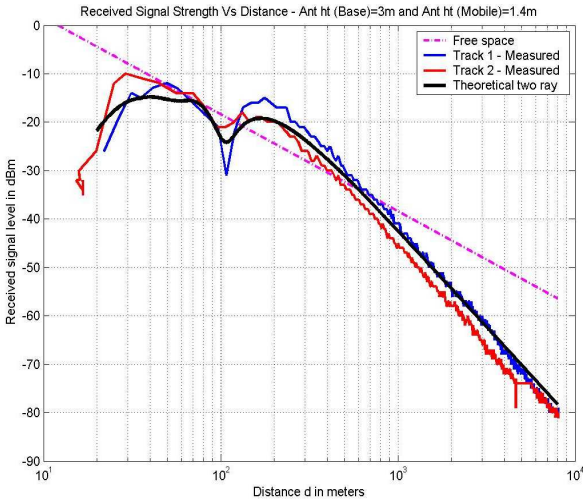


Figure 10: Received signal strength variation with distance for base station and mobile antenna heights of 3 and 1.4m respectively with a GPS error of -10m

4.2.3 Throughput performance of extended WLAN

Peer-to-peer throughput and signal to noise ratio (SNR) measurements are made every 0.5 km along a particular track and the results for four specific cases (equal antenna heights on the base station and mobile vehicle) along track1 are plotted in figures 11-12. Theoretical 802.11b data rates of 11 Mbps are not achievable in practice due to the packet overhead, Request to send / Clear to send (RTS/CTS), acknowledgement times etc.

Measured data rates vary from 0.2-4.9 Mbps depending on the SNR, which in turn depends on the distance of separation and the antenna height. It is seen in figures 11-12 that the throughput does not monotonically decrease with distance as may be expected to occur due to a general drop in signal to noise ratio with distance. There are data points where the throughput may be higher for larger distances compared to a smaller distances of separation. This is primarily attributed to the large variation in the round trip times (RTT) that are encountered in 802.11b networks and also the response of the underlying TCP protocol to random packet errors

that may occur even at high signal to noise ration (SNR) and that are predominant at low values of SNR.

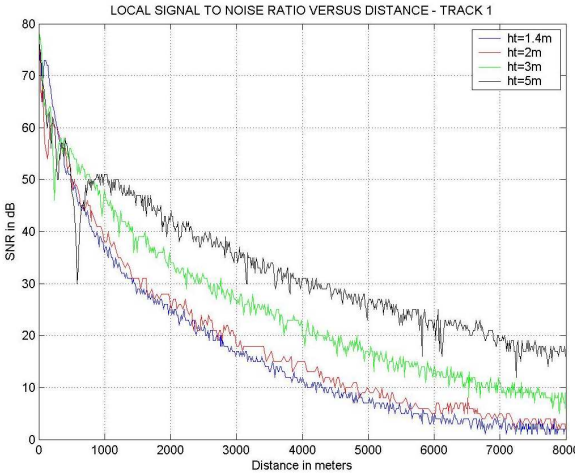


Figure 11: SNR variation along track1 for antenna heights of 1.4, 2, 3 and 5 meter on the base station and mobile vehicle

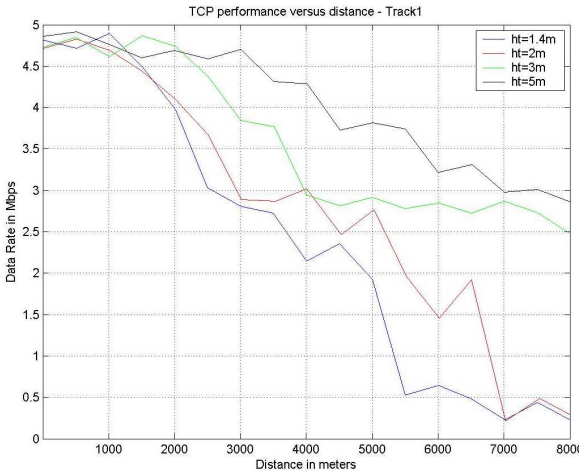


Figure 12: TCP Throughput variation along track1 for antenna heights of 1.4, 2, 3 and 5 meter on the base station and mobile vehicle

However, the bandwidth available is still high enough to exchange video data between the mobile vehicle and the base station. Seamless communication between the two peers implies a highly reliable link, and that wireless Internet would be continuously available over this long range because replacing the base station with an access point would make this an infrastructure network with Internet access available to any node such as the one installed on the mobile vehicle in these experiments.

5. Conclusions

In this paper, we presented an integrated Iridium/Wi-Fi communication system as a reliable, truly mobile and lightweight alternative to the data communication and Internet needs in remote regions.

The system was implemented at NGRIP; Greenland during the Summer 2003 field season and the performance was studied. Using MLPPP technology, a 4-modem Iridium system provided Internet access and data access to home institutions at speeds up to 9.8 Kbps from the field camp, with efficiencies over 90%. The system had an up time of 94% with at least one modem and 80% with all the four modems. Further, we have developed management software that handles call drops, system and power failures providing a fully autonomous operation. This can provide a reliable lifeline data/Internet connection to all the polar field camps. Further research is being conducted to understand and possibly reduce the round trip time (1.8 sec) of the Iridium system, which impairs real-time communications.

We have demonstrated two different wireless local area networks. Using external amplifier and antenna at the access point, a network of approximately 1 km radius is formed serving wireless laptops with widely available 802.11b cards. An extended network of 10 km radius was achieved by using similar antenna and amplifier on the mobile user end too. We obtained LAN data rates of 4.828 Mbps at close separation distances and 0.2 Mbps for distances up to 10 Km from the base station.

6. Acknowledgements

This work was supported by the National Science Foundation (grant #OPP-0122520), the National Aeronautics and Space Administration (grants #NAG5-12659 and NAG5-12980), the Kansas Technology Enterprise Corporation, and the University of Kansas.

7. References

[1]. K. Bullington, *"Radio Propagation Fundamentals"*, The Bell System Technical Journal, Vol. 36, No.3. 593-625, 1957

[2]. Frost and Sullivan, *"Satellite telephone Quality of Service comparison: Iridium Vs. Globalstar"*, 2002

[3]. Nicolas S. Powell, *"South pole satellite communications update"*, United States Antarctic Program, July 2002

[4]. K. Sklower, B. Lloyd, G. McGregor, D. Carr, T. Coradetti,*"RFC 1990 - The PPP Multilink Protocol"*, 1996

[5]. *"Proceedings of National Science Foundation United States Antarctic Program Communication Workshop"*, March 1999

[6]*." Recommendations of the South Pole Users committee meeting (SPUC)"*, Colorado 2002

[7]. Satellites utilized by the South Pole station, http://amanda.wisc.edu/data/comms-summary.shtml

[8]. South Pole satellite visibility, http://adelie.harvard.edu/spole/satellite/riseset_table.html

Signal Capacity Modeling for Shared Radio System Planning

Gary Patrick, NTIA/OSM, 202-482-9132, gpatrick@ntia.doc.gov
Charles Hoffman, NTIA/OSM, 202-482-3456, choffman@ntia.doc.gov
Robert Matheson, NTIA/ITS, 393-497-3293, matheson@its.bldrdoc.gov

Abstract. *Almost every major Federal agency operates an independent mobile radio system in the 162-174 MHz band to provide critical radio communications with its own agents. Last year, NTIA began a joint OSM/ITS Spectrum Efficiency Initiative, which includes a study of whether a shared radio system (e.g., a trunked system) could functionally and advantageously replace most of the specialized single-agency radio systems. The first phase of this work is to understand the amount of service provided by the current single-agency radio systems. A "signal capacity model" was developed, which uses Federal Government Master File (GMF) license data to calculate the number of independent radio signals that could be received by a mobile user at 1-mile increments within a 100-mile radius of Washington, D.C. Since various radio network architectures transmit the same signal from multiple sites, different algorithms were used to calculate the signal capacity for different types of mobile networks. Peak and average signal capacity maps were produced, based on different assumptions about the probable location of users. This data will form the basis for the design of possible alternatives for future shared radio systems.*

1. NTIA's Role in System Planning

The National Telecommunications and Information Administration (NTIA) is the Executive Branch agency responsible for developing and articulating domestic and international telecommunications policy. NTIA's responsibilities include establishing polices concerning Federal spectrum assignments, allocation in use, and providing various Federal departments and agencies with guidance to ensure that their conduct of telecommunications activities is consistent with these polices. One of these policies is to ensure that Federal use of the spectrum is as efficient as possible.

There are many well-known ways to improve the efficiency of mobile radio systems used by Federal agencies [1]. These methods include decreasing bandwidth, increasing geographical frequency reuse, increasing the amount of time that a given frequency is in use (Erlang efficiency), using higher gain directive antennas, and more. The Federal government has expended considerable effort in the past on ensuring that their use of the spectrum is as technically efficient as practicable by adopting many of these spectrum efficient technologies including: narrow-banding, sharing, overlaying, relocating, and applying new spectrum distribution analytical and planning techniques. However, although these methods are well known and their adoption has been strongly encouraged, they often have not been widely applied, because they sometimes also have substantial disadvantages. Sometimes the disadvantage is merely a matter of higher cost (including the requirement to buy new equipment to replace older, less-efficient systems). Sometimes the disadvantages actually entail a decrease in performance or convenience – in addition to high cost.

Therefore, a more realistic objective regarding new mobile radio technologies has been lumped together under the title of "effectiveness." Effectiveness includes a wider range of factors than merely being spectrum efficient; effectiveness also includes usability and cost factors. Usability factors might include larger or heavier equipment with shorter battery life, poorer intelligibility, shorter operational range, more latency (delay time) between a push-to-talk command and the distant user hearing the message, less interoperability with other users, more complexity of operation, lack of needed special features like encryption and digital messages, etc. A new technology with improved spectrum efficiency – by itself – has little merit for the Federal user, if the adoption of the new technology would result in a decrease of effectiveness. A decrease in effectiveness would mean that substantial cost or usability factors outweigh the improved spectrum efficiency.

By its very definition, spectrum effectiveness means "comparing apples and oranges." How much should a 25% improvement in spectrum efficiency weigh against a 60% increase in cost or a 10% decrease in intelligibility? It is likely that the effectiveness of a proposed change would be evaluated differently by different users, for whom the relative benefits of various factors are evaluated differently. Therefore, we will not be able to derive a single formula that will serve to calculate the relative effectiveness of a proposed change as seen by all Federal

users. For example, urban users competing for unused channels in a crowded urban environment will probably count spectrum efficiency as more valuable, compared to rural users, where unused spectrum is relatively more plentiful but where too many locations do not have adequate coverage from widely-spaced base stations.

Nevertheless, "effectiveness" is an important concept (even if it can't be calculated precisely), because it forces planners to evaluate proposed changes against a much broader and more complete set of criteria. Proposed changes that benefit some users – but which place other users at a disadvantage – are less likely to be pushed through on the basis of a single factor. Effectiveness-based changes are more likely to benefit a wider range of users and, therefore, such changes may actually be adopted and implemented by a larger number of users.

A study of effectiveness has led us to consider broader questions on how the Federal government is using the radio spectrum and to consider whether larger-scale structural and organizational changes – such as shared radio systems – could improve both the technical efficiency and mission effectiveness of Federal radio systems. Implementing a shared Federal radio system would be a much more challenging project than merely causing Federal users to switch to a more spectrum efficient radio technology (and it might not be worth the trouble, based only on spectrum efficiency gains). However, a large number of related factors are part of the "effectiveness" equation, including large required expenditures to independently solve problems of interoperability between public safety agencies (e.g. SAFECOM), Homeland Security (e.g., HSD and IWN), spectrum efficiency (e.g., the narrowbanding deadlines), improved radio capabilities (e.g., data, encryption, emergency capacity, greatly expanded wireless Internet systems, etc.), and very powerful economy-of-scale and complexity arguments.

In an effort to better understand how Federal agencies are using the spectrum and how we can improve the effectiveness of this use, NTIA has embarked on a multi-phased study of spectrum efficiency and effectiveness within the Federal government land mobile bands. The first phase of the study (described in this paper) will analyze the present Federal use of the radio spectrum in the 162-174 MHz band within a 100-mile radius of Washington, DC by developing a quantitative model of the "signal capacity" of current Federal use of the radio spectrum. The second phase will be based on the quantitative model results developed in the first phase and will explore various modern radio system alternatives to current Federal systems such as shared trunked systems. Depending on the apparent overall benefits identified by the results of this phase, one or more concepts may be selected for further studies, detailed engineering, and/or eventual large-scale or small-scale implementation.

The 162-174 MHz band in the Washington, DC area was selected for this study for the following reasons. The 162-174 MHz band is the most intensely used Federal mobile radio band, and the Washington, DC area is one of the most congested geographical areas for Federal users. Therefore, this selection of frequency band and location should provide the best of opportunities for studying Federal system operations in highly congested areas. Selection of a congested area such as Washington, D.C. provides an opportunity in terms of: 1) investigating spectrum efficiency, 2) maximum opportunity for economic savings, 3) utility for advanced technology solutions, and 4) a maximum familiarity with the territory on the part of Federal spectrum management personnel. The initial decision to include only the 162-174 MHz band in this study was intended to help obtain initial results more rapidly.

When considering the possible advantages of replacing many current smaller individual-agency radio systems with a larger shared radio system, it is necessary to have realistic information on the services provided by the current systems. This information is needed for various reasons, with the main reason being that it serves as a starting point for the design of the new system. It is the intention that any proposed new system design could include many additional factors (e.g., projected growth, etc.), but as a minimum it must match levels of service provided by current systems. Unless previous levels of service are known, there is no way to know whether any replacement system would provide equivalent service.

In addition, many advanced radio systems provide relative advantages or disadvantages that depend substantially on certain economy-of-scale factors. It is often true that the bigger the system, the more efficiently it operates. This is especially true for trunked radio systems, but such results also affect many aspects of calculating peak loading factors, etc. To compare new designed systems with current systems, it is necessary to actually have numeric values for some basic parameters, such as how many channels are required for current systems. Since these systems cannot be realistically designed or compared without this data, the first phase of this study must obtain quantitative information on current system operations.

2. Possible Sources of Current Usage Data

There are various techniques for obtaining the information needed to estimate the level of services for current systems. For example, a detailed survey could be submitted to Federal users to obtain their estimates of current and projected levels of service from their radio networks. However, a survey of this type would be quite difficult and time consuming for many of the agencies since agencies may have to survey each independent bureau and coordinate the results. Furthermore, the individual agencies probably would each have to develop suitable models to describe their own current levels of service, and NTIA would need to convert the results of these models to a common overall model.

Another method would be using the NTIA/ITS Radio Spectrum Measurement System (RSMS) to conduct measurements within the Washington, DC area, measuring the actual amount of traffic on specific radio channels for all frequencies in the band. However, the major concern with the RSMS data would be the time required to measure and analyze such data at many sites over a wide geographical area. Such a concern would be further complicated in measuring low power and intermittent applications. Furthermore, such measurements would probably not be able to collect accurate information on multi-site systems using the same frequencies, and the collected data would require much interpretation by NTIA personnel. Some limited measurements would be useful and may be considered in the future to supplement data in some aspects of this study.

The Government Master File (GMF) is another possible source of information. It contains records of the frequencies assigned to all U.S. Federal Government agencies in the U.S. Although the information contained in the GMF information is somewhat limited, it provides information on Federal radios over a wide geographic range and it is easily accessible. A cursory look at Federal licenses in the 162-174 MHz band showed 1945 specific assignments (licenses) within 100 miles of the center of Washington, DC., as shown in Figure 1.

After considering various possible methods of obtaining quantitative information on current Federal mobile radio service levels, it was decided to estimate current levels of service by analyzing the data that is already available in the GMF, using extensive computer modeling to generate maps showing a quantity called "signal capacity" (as described below). When necessary, the existing GMF data was augmented by consulting with agency representatives to provide any required additional operational data.

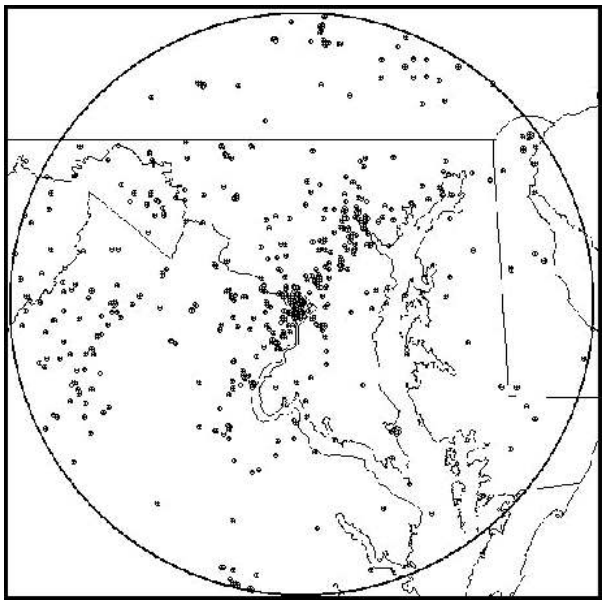


Figure 1 - Federal radios in 162-174 MHz band within 160 km of Washington, D.C.

Several characteristics should be reflected in any model that is used to characterize the amount of service that is being provided to Federal users by current Federal mobile radio systems in the 162-174 MHz band. The model should calculate a quantity that is closely related to the service delivered to Federal mobile radio users. Since Federal radio systems use a wide range of radio technologies, the model should be able to reduce these services to a "lowest common denominator," under which all mobile services could be added together and summarized. Ideally, the model should use only data that is easily available, as provided within the GMF. The model results should also be additive or otherwise allow easy manipulations to include different agencies or other user populations. The model should be transparent, in the sense that a given user population can identify and verify its own contribution to the total model results. The model should allow easy determination of anticipated results from prototype or imaginary alternative radio systems, to facilitate comparisons between existing and alternative systems. Finally, the model should require a minimum number of assumptions that could drastically change the model results.

3. The Signal Capacity Model

Based on the above considerations, we selected a "signal capacity" model to estimate the amount of radio service currently available to Federal users. The signal capacity model assumes that most Federal radio users operate in a

mobile or portable environment and depend on a two-way voice channel created by a Federal-owned base station. Therefore, service can be quantified by merely counting the number of independent voice channels that are available at any given geographical location.

Signal capacity (SC) is defined as "the number of independent 2-way voice radio channels that can be received by a typical mobile radio user at a given geographical location."

The SC model considers only the geographical distribution of the signal field strength from base station transmitters that could be received by mobile users. Field strength values higher than a selected threshold are considered to provide "radio coverage." The SC model includes no information about base station receivers, or any receivers whatever. However, any future radio system designed to meet the signal capacity specifications produced by the model would obviously also need to meet suitable base station receiver functions (possibly including receive-only sites) and (probably) be capable of mobile-to-mobile operation.

This SC definition is particularly useful for our purposes for several reasons:

1. The major service provided to Federal users in the 162-MHz band is two-way voice channels to mobile users. Therefore, this definition captures most of the use in the 162-MHz band.

2. The SC is additive. This means that the SC values produced by individual transmitters can be simply added to give a cumulative SC value for the group of transmitters.

3. The SC can be calculated from data that is available in GMF license records and technical models, including transmitter data, terrain/ground cover data, and propagation models.

However, the SC definition makes many simplifications that might limit its usefulness in specific circumstances. These include:

1. There is no data on whether a given channel is lightly or heavily used. However, the actual existence of a radio channel suggests that the system was needed, and that need is presumed to exist when a future system is designed.

2. All voice channels are considered identical, whether analog or digital, narrowband or wideband, simplex or duplex, high priority or low, etc. Equally important, signal capacity considers only the radio link, with no distinction as to what capabilities are available via that link: data, encryption, database access, telephone access, wide area access, etc. For SC purposes, "a channel is a channel is a channel."

In summary, although the SC model may not provide an exact measure of the service that a radio produces, it provides a useful measure of service that can be calculated fairly easily. The SC is a useful model because radio systems can be designed to match or exceed given SC values with a reasonable assurance that a new full-featured radio system (e.g., a trunked system) would match or exceed most performance measures of the old systems.

4. Signal Capacity Analysis Program

The signal capacity analysis program (SCAP) performs its analysis in several stages. The first stage includes reading a modified version of the GMF, which has been sorted to include assignments (licenses) in the 162-174 MHz band within a 100-mile radius of Washington, DC. Each assignment record has been augmented with a function code that tells SCAP how each assignment record is to be analyzed. The function code includes a determination of function and a network identification. SCAP uses the function code to identify base station transmitters, which are used with a terrain-based Longley-Rice propagation program to compute predicted field strengths from that transmitter. The field strength predictions are used to predict a coverage area for each transmitter; all field strengths higher than a certain threshold will be assumed to provide coverage (service) for mobile users. Each independent transmitter that provides coverage at a given location adds to the SC value.

However, the definition of SC counts only *independent* voice channels. It is necessary to determine that multiple radio signals received at a given location are actually independent, since some networks transmit identical signals from multiple sites or prevent multiple sites from transmitting simultaneously. SCAP uses the function and network identification codes to identify transmitters that might not be independent (for example, signals from the same channel of multiple simulcast sites). A network is defined as any related set of transmitters that are part of an integrated system (or network) following a uniform set of signal capacity rules. Because of the need to determine independence, the transmitters belonging to each network must be analyzed as a single unit. A minimal network includes a single transmitter, but other networks could

include systems with many sites and many transmitters at each site.

SCAP uses the basic coverage information to calculate a pair of peak and average signal capacity maps for each network, following the specific algorithms for each function code. Specific algorithms have been developed for various technologies, including simulcast, trunked, various repeater networks, and more. The peak SC and average SC maps for each network can be combined to give similar paired peak and average SC maps for multiple networks by simply adding the corresponding elements representing peak or average signal capacity at respective geographical locations. The peak and average SC maps can continue to be added respectively to obtain peak and average SC maps for various larger groups of networks. Continuing the process of addition, coverage maps for groups of networks for individual agencies can become coverage maps for whole Federal departments, and finally for the Federal Government as a whole.

The SC map values are calculated at 1-mile intervals for all locations over a 200-mile square area. However, the GMF database used for these calculations included only GMF records for systems located within a 100-mile radius circle, centered on Washington, DC. Therefore, there can be some substantial "edge" effects in this modeling. A transmitter located just outside the 100-mi radius circle would not be included in the model at all, even though it could have a substantial portion of its coverage area inside the area of the map. A transmitter just inside the circle would be included in the model, but part of its coverage area could lie outside the calculated area of the square map. This reduction in the apparent size of its coverage area could affect the numbers in the average signal capacity maps.

5. Use of Peak and Average SC Maps

Although the signal capacity (SC) was defined initially at a given location, it is convenient to summarize the SC at many adjacent locations as a map, using shading to identify areas that have certain ranges of SC values. Figure 2 is an example of such a map from a single transmitter, calculated for a 200-mi square.

This map shows geographical areas of coverage from a single transmitter. Although this map follows the earlier definition of signal capacity, it is called a "peak" signal capacity (PSC) map to distinguish it from an "average" signal capacity (ASC) map described later. The light-shaded areas of the map indicate coverage from the associated transmitter. Since only one transmitter is present in this example, every point on the map will have

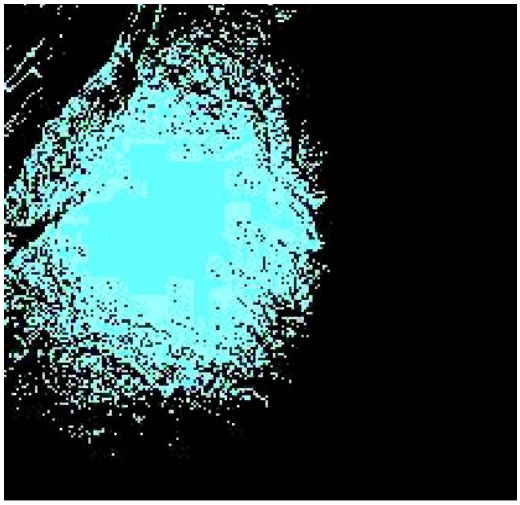


Figure 2 - PSC map for single transmitter

coverage from either 1 or 0 transmitters, giving SC values of either 0 or 1 (no coverage, or coverage by one transmitter). The peak (PSC) map can be understood as showing the maximum number of independent users that could be simultaneously served at any given location. Note that this does not imply that only one individual user can actually be served by the system at a given time. It means that users are served by only one independent signal; the "user" could be a talk group with a hundred members that are receiving a common message. Or, it could be an individual user.

An "average" signal capacity (ASC) map can also be defined, which is different from the earlier definition for signal capacity. The ASC map shows the number of independent users *per square mile* that could be served at one time if users were distributed evenly across the coverage area of a transmitter. The ASC map for an individual transmitter is determined by calculating the PSC map, totaling the number of square miles of coverage, and normalizing the PSC values by dividing by the total coverage area. The ASC values are expressed in terms of the number of independent users per square mile to which the transmitter could provide service, assuming that users are evenly spaced across all of the coverage area. Integrating the ASC values across the entire coverage area of one transmitter will always give a total of "one user." The ASC map of a single transmitter looks identical to the PSC map for that transmitter, except that the numeric value of coverage areas is different. Specifically, Figure 2 showed radio coverage of about 6,500 square miles (out of a total of 40,000 sq mi included in the map). Therefore, the ASC map shows that this transmitter would provide coverage to about .00016 independent users per square mile, if the users were evenly distributed over the whole 6,500 sq mi area of

coverage. (6500 mi sq x .00016 users/mi sq = 1 user.) Note that a smaller transmitter coverage area provides larger values on an ASC map.

The PSC map and the ASC map provide two different ways of looking at the problem of providing comparable service, based on different assumptions about how mobile users are distributed geographically. The PSC map assumes that all users might be concentrated at the same geographical location. The ASC map assumes that users are evenly distributed across the coverage areas of their respective base station transmitters. In analyzing the signal capacity that is provided by an existing transmitter, one does not know whether that transmitter was intended to serve users who are statistically evenly distributed across the coverage area or users that are sometimes located in one small portion of the coverage area. Lacking this specific insight for each transmitter, the analysis covered the extreme cases by calculating both the PSC and ASC maps. In many ways, the PSC and ASC maps represent the worst and best cases of user geographical distributions, respectively. Real-world user distributions presumably must lie somewhere between these two extremes, but it is not necessarily clear exactly where. Nevertheless, the peak and average SC maps place bounds on the effects of user location.

If transmitters with one coverage area are replaced with future transmitters having a different coverage area, the assumptions about whether replacement transmitter requirements scale proportionally to coverage area or whether they do not become very important, since these two different assumptions give much different results. ASC values scale proportional to coverage area; designing a new system with microcells having 10% of the coverage area of standard cells would imply designing the microcells to handle 10% of the traffic. PSC values do not scale with coverage area. If PSC rules hold in the real world, a 10%-sized microcell would still need to handle all of the traffic of the standard cell, since it is possible that all of the users from the standard cell might sometime be crowded into that single micro-cell. Therefore, the availability of both sets of maps is important as a more complete basis for designing a range of possible alternative radio systems to match the current capabilities of Federal users in the 162-174 MHz band.

6. How Function and Network Codes Define SC Algorithms

The lowest level of a radio system that can be analyzed to produce peak or average SC maps is not a single transmitter but is instead a "network." A network is a group of transmitters (one or more) designed to cooperatively produce a specific type of service over a given area. The concept of network must be used to compute SC, because SC is defined only for independent signals. However, sometimes signals received from different transmitters are not independent. For example, signals received at the same frequency from different simulcast sites cannot contain different information. The network (as defined here) is the smallest set of transmitters that must be included in the SC calculations to obtain a correct value. Once the ASC and PSC maps for a network have been properly calculated, these values are independent quantities and SC maps from one network can be freely combined with other SC maps.

The Signal Capacity Analysis Program (SCAP) uses the GMF license database as a source of technical information. A function code (F, N) is added to each GMF assignment record. The parameter F shows SCAP what type of technology the network uses. Different algorithms are employed for independent base stations, repeaters and multi-site repeater systems, simulcast systems, trunked base stations and multi-site systems. N shows which other transmitters to include in this analysis (all transmitters with the same N belong to the same network).

At present, the function codes are determined by NTIA staff, following a study of the GMF records and (usually) consultation with frequency management staff at the respective Federal agencies. These consultations provided very useful insights into how specific systems were used. In some cases, systems were identified that provided services that could not reasonably be supplied by a shared Federal system.

As an example of the SCAP analysis of a basic network, showing how different types of networks give different peak and average SC map results, consider a network that consists of a group of four base stations, each station having one channel. The individual coverage areas of these four base stations are shown as Figure 3. These four transmitters will be analyzed as a simulcast network (Figure 4) and as a network of single-channel trunked radio stations (Figures 5 and 6).

The PSC map for the simulcast system is shown in Figure 4. This map is calculated by first calculating the coverage areas of each base station and placing a "1" at each location where coverage is available. In locations where coverage is available from multiple base stations, the peak signal capacity for the simulcast system is still "1," since all base stations transmit an identical message, so only 1 independent message can be received at any location.

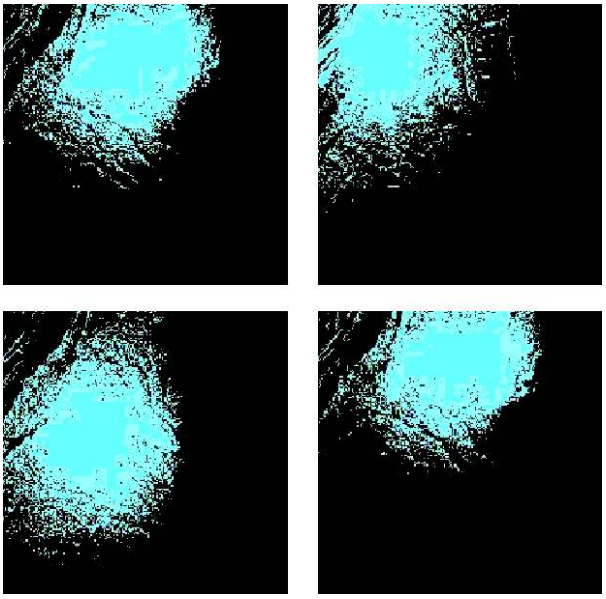


Figure 3 - Coverage areas of four transmitters

The average signal capacity is determined by dividing the peak signal capacity at a specific location by the area over which each transmitter provides coverage. In this case, the entire 4-site coverage area acts like the coverage area of a single transmitter that covers a very large area. Therefore, the simulcast ASC map looks identical to the simulcast PSC map, except for a scaling factor that shows a single relatively low average number of independent

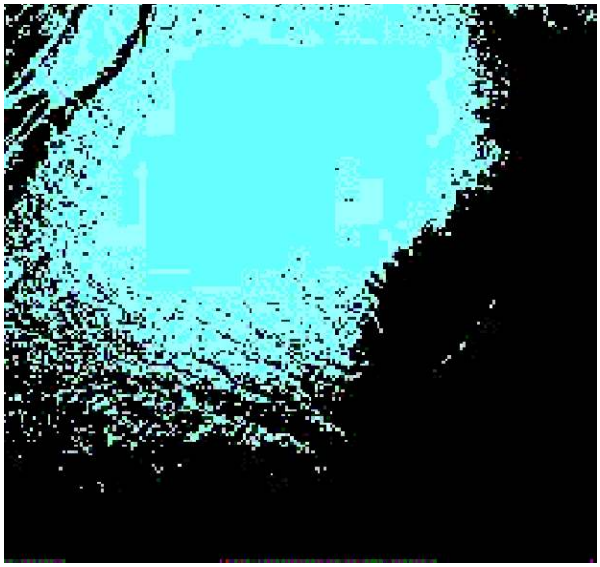


Figure 4 - PSC (and ASC) map for 4 simulcast sites

users per square mile that can be served by the simulcast system.

The PSC map for a 4-site trunked radio system using independent frequencies to provide coverage around each site (Figure 5) shows that each site could provide independent service to a user (serving as many as 4 independent users at some locations). PSC values add together in areas where coverage is available from multiple sites.

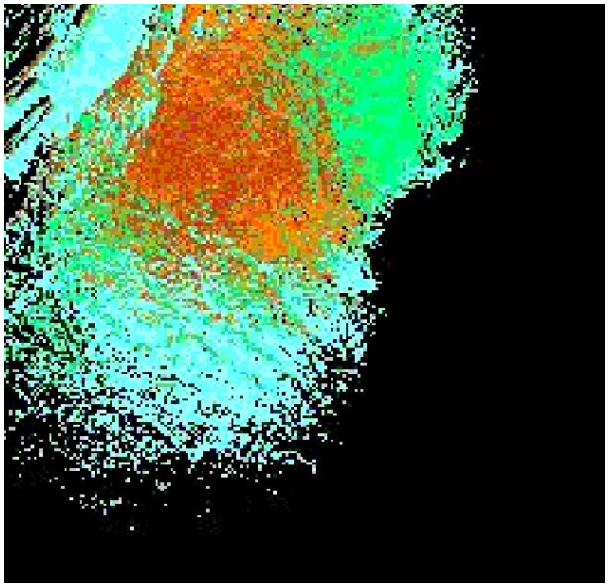


Figure 5 - PSC map for 4 trunked radio sites

The ASC values for the 4-site trunked network are calculated by finding the ASC values for each base station on an individual basis, and adding the ASC values at points where coverage is available from multiple stations (Figure 6). In this case, the transmitters operate independently, so the coverage area is the respective area for each separate transmitter. Since each transmitter can serve a different population of independent users, the ASC values add together where coverage is available from multiple transmitters.

Although not visible in these maps, the trunked ASC and PSC maps differ by more than a simple scaling factor, since the coverage areas of the 4 sites are different and the corresponding ASC numeric values for each site are therefore somewhat different.

The maps are quite different for the 4-site simulcast network and the 4-site trunked network. The peak SC per simulcast channel would be one independent user per channel over the entire 4-site coverage area. The peak SC per trunked channel would be one independent user per channel, but there would be many locations that might

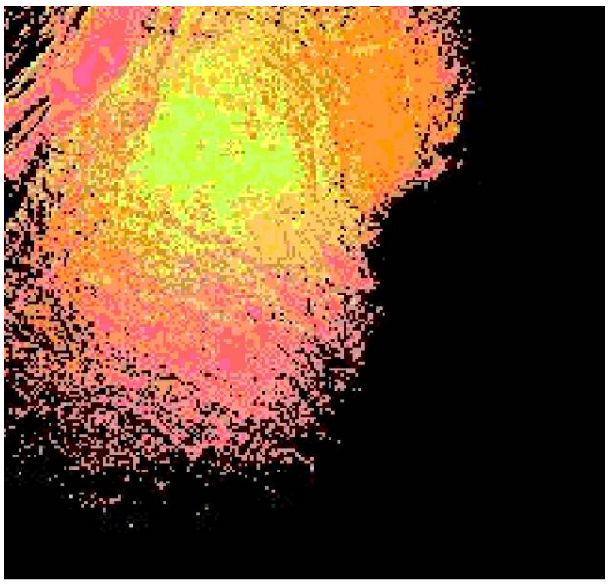


Figure 6 - ASC map for 4 trunked radio sites

provide service to 1-4 independent users, due to overlapping coverage areas from adjacent sites. The average independent usage per square mile that the 4-station simulcast system can support would be smaller than the average independent usage per square mile for the 4-site trunked system for two reasons. The coverage area per channel is larger for the simulcast system, giving a smaller numerical value for the one simulcast channel. In addition, the (individually larger) trunked ASC values add together in overlap areas. Therefore, the maximum average number of independent users per square mile is considerably larger at some locations for the trunked system.

The above examples were included to illustrate some of the principles involved in calculating PSC and ASC maps for systems using various radio technologies. It should be noted that some other technologies involve more complex calculations than the examples shown here.

7. PSC & ASC Maps for Federal Agencies

Using the techniques described above, PSC and ASC maps were computed for each major Federal department in the Washington, DC area. Figure 7 shows an example of a PSC map for some Federal agencies having a small number of radios in the 162-174 MHz band. In this example, the coverage areas for these radio systems are easily distinguished, and there are many areas on the map that have no coverage at all for these agencies.

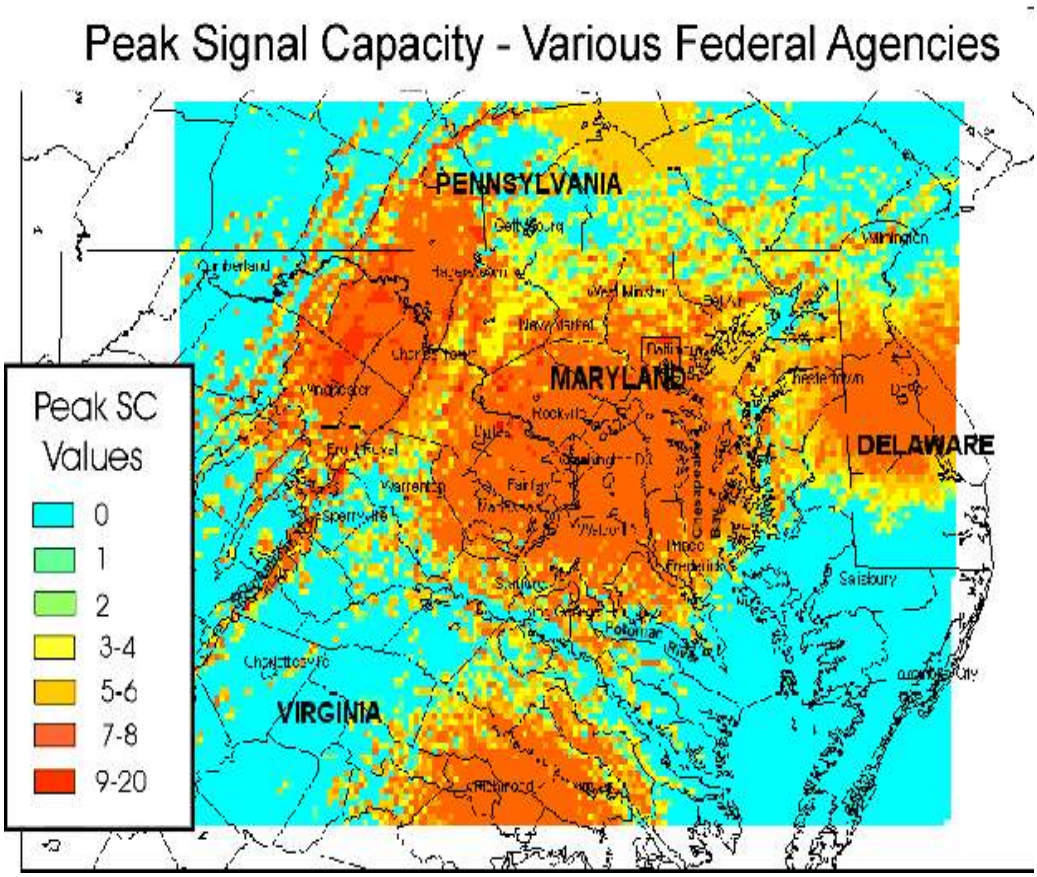


Figure 7 - PSC map for several Federal agencies

Maps produced for individual agencies can be further combined to give maps for larger groupings of agencies, or even for all agencies in the Federal government. Figure 8 shows the PSC map for all Federal agencies in the 162-174 MHz band. Some features of this map should be noted. Partly because of the large number of radio licenses used to produce the map, features associated with individual radio systems have largely disappeared. Therefore, the overall map features appear mostly as a large "bullseye," with the highest concentration of radios located near the center of Washington, D.C.

The highest PSC value on the map is "337." This means that as many as 337 independent radio channels could be received by a mobile user at some Washington locations. An additional observation is that there appear to be very few locations where radio coverage is not available from multiple radio systems.

The corresponding ASC map for all Federal agencies is shown in Figure 9. The actual ASC values in this map have been multiplied by 10,000, to give them values that are closer to the ASC values, for purposes of facilitating easier comparisons between the two maps.

Corresponding ASC and PSC values for identical locations differ by a ratio of about 5,000. This suggests a typical site coverage area is about 5000 square miles, which is equivalent to a circle about 80 mi across. Although such large areas would be more than expected for a single site, many of these radio systems involve multiple sites.

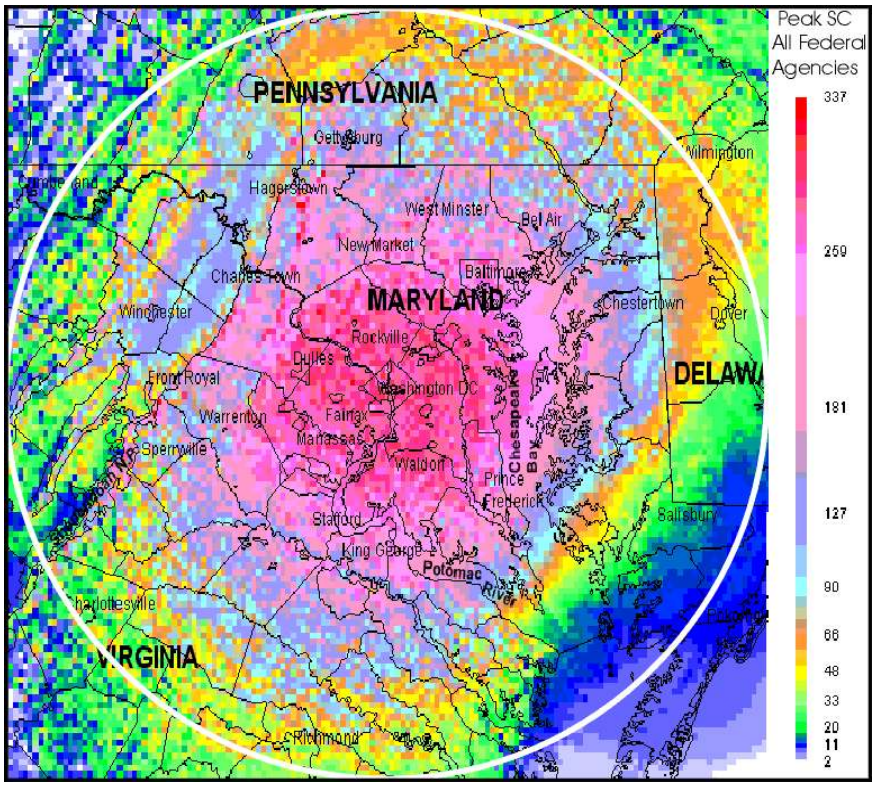


Figure 9 - PSC map for all Federal agencies in the 162-174 MHz band

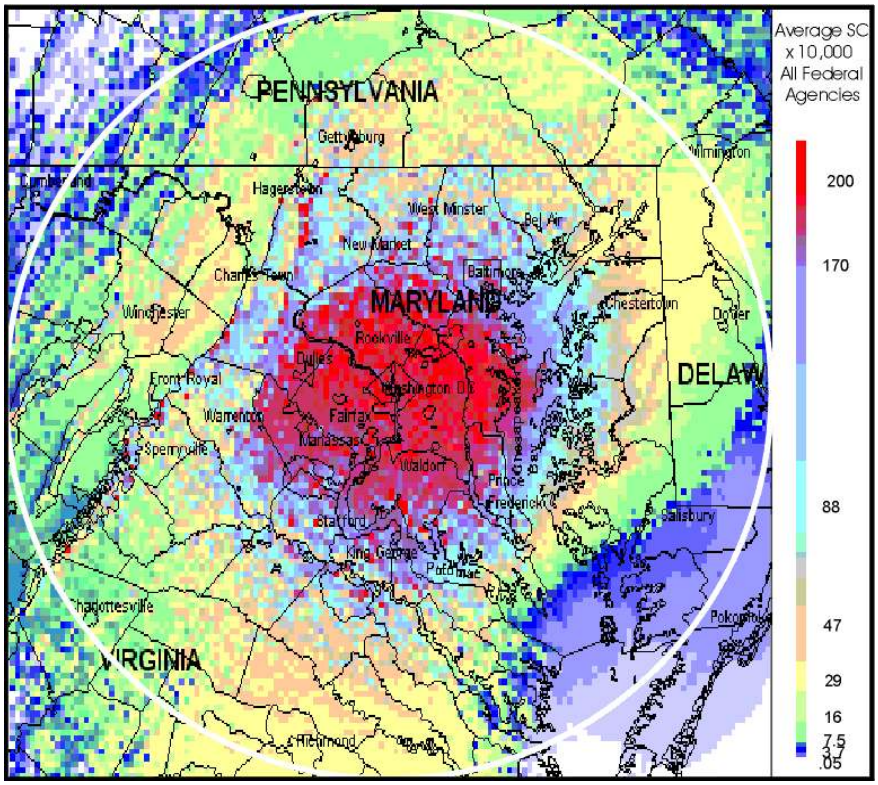


Figure 8 - ASC map for all Federal agencies in the 162-174 MHz band.

8. Conclusion

In summary, this first phase of a multi-phase effort has made a careful effort to understand the radio services provided by many different Federal agencies in the Washington, D.C. area, who are using a wide range of mobile architectures and technologies. Throughout the analysis, we had substantial interaction with many agencies to learn how each radio system was used and to confirm the correct interpretation of GMF data. We believe that we have developed a useful conceptual and quantitative model of the capabilities of current Federal agency mobile radio systems.

We believe that a proper application of modern technologies might provide opportunities for very substantial improvements to the effectiveness and efficiency of Federal mobile radio systems. This signal capacity model and the Washington, D.C. data were developed to serve as a realistic foundation for exploring these opportunities by designing alternative future shared radio systems for Federal agencies. The payoff strategy for all shared systems is essentially as follows:

1. Most current mobile radio systems provide services that could be duplicated or improved by modern shared radio systems. (The signal capacity model shows the quantitative aspects of this conversion.)

2. Trunked radio systems, for example, can carry 3-10 times more traffic (Erlangs) over each channel, compared to most single-channel radio systems and still provide very high channel availability. However, this improvement in channel capacity is present only for trunked systems with a large number of channels (ten or more). Most agencies don't use many channels, so trunked systems have fewer advantages unless multiple agencies combine their requirements within a single shared system.

3. If many agencies share a trunked system, it will require relatively fewer channels to carry the combined traffic. Moreover, multisite trunked networks could provide large coverage areas, fast priority access for selected users, very effective interoperability solutions, flexibility to reconfigure for emergencies (including additional channels and preemptive access when crowded), and many other advanced features.

4. Because of #3, shared trunked systems can require fewer frequencies, give better service and coverage, and even cost less money. However, many of these advantages disappear if there are not enough users on the trunked system. Therefore, a realistic study of trunked system advantages needs good information on the number of users, the total amount of radio traffic, and more.

Over the next couple years we intend to use this data to design multiple systems using different assumptions about how many agencies might be participating in a shared system and how fully these agencies would integrate their systems, varying the coverage area of typical sites, the use of receive-only sites versus small-cell full-feature sites, alternative ways to handle emergencies, use of advanced non-trunked technologies, and different ways to achieve interoperability.

The signal capacity model will facilitate many aspects of these studies. The SC maps show where base station coverage is needed and provide information on how many signals are needed at each location. This information can be easily modified, depending on assumptions about which agencies are to be included in a particular version of some future shared system. Some agencies will need coverage in some geographical areas, while others need coverage in different locations. Therefore, changing among different combinations of agencies will generally change both the magnitude of the highest PSC values and the shape of the coverage area on the map. Similarly, the ASC maps provide usage on a "per square mile" basis. This data allows the design of alternative future radio systems having base stations that provide different sized coverage areas. Finally, after a radio system is designed, the signal capacity model can be used to compare the current SC values with corresponding SC values derived from the new system design. Thus, the SC model will be useful at several stages in the process of designing shared radio systems.

Although none of these studies is aimed at a specific deployment of new systems, we hope that such data and studies will help provide realistic and useful insights on Federal and public safety spectrum requirements, how the benefits of shared systems and interoperability solutions scale with size or the number of users, new interoperability solutions based on shared systems, identification of the best possibilities for additional sharing of facilities and frequency bands, and more.

In current and future studies, we look forward to working closely with other agencies and organizations to help us with planning our work, performing our studies, and critiquing our results.

References

[1]. R. J. Matheson. "A survey of relative spectrum efficiency of mobile voice communication systems," NTIA Report 94-311. 1994.

Rapidly Deployable Broadband Communications for Disaster Response

Charles W. Bostian, Scott F. Midkiff, Timothy M. Gallagher, Christian J. Rieser, and Thomas W. Rondeau
Center for Wireless Telecommunications
Virginia Tech
Blacksburg, Virginia 24061-0111
Phone 540-231-5096 Fax 540-231-3004 e-mail bostian@vt.edu

Abstract: We present the design of a rapidly deployable backbone communications system for disaster response. Although work on the system began in 2000, it is intended for disasters like those that occurred on September 11, 2001. It illuminates the disaster site with RF that supports high-capacity (100 base T or Gigabit Ethernet) links to the outside world. Operating initially at 28 GHz with a 5 GHz version now under construction, it uses non-line-of-sight (NLOS) "bounce paths" of opportunity to provide coverage at shadowed locations. Since at these frequencies most building walls and terrain features are electro magnetically rough, the radio paths are highly dispersive and require careful characterization for optimum radio performance. In order to identify such paths of opportunity and study their characteristics, we developed an impulse channel sounder based on ultra wideband technology. When perfected, it will allow our system to identify the best path and set the radio parameters for optimum quality of service. Since the channels encountered in a given disaster situation may differ significantly from those anticipated or previously experienced, our goal is to use the sounder output to drive a cognitive engine that will control the radios. We present our genetic algorithm based implementation of a cognitive radio and outline our plans for implementation and testing a cognitive version of our system in the 5 GHz band.

1 Introduction

In 2000 our group in collaboration with colleagues at SAIC began developing a broadband communications system for use in response to disasters like those that occurred on September 11, 2001. It is intended for deployment at a disaster scene where all communications infrastructure has been destroyed in an area several kilometers on a side – exactly what occurred at the World Trade Center. The system operates by using one or more hub stations on the periphery to flood the disaster scene with RF that supports high-capacity (100 base T or Gigabit Ethernet) backbone links to portable remote stations within the disaster area. See Figure 1. The remote stations in turn support a mix of individual workstations, routers, 802.11 wireless LANs, etc. The hub connects to the outside world through surviving optical or copper circuits available at the edge of the disaster area or by satellite or microwave point-to-point links to such circuits. The overall goal is to provide high-capacity Internet access to all workers on the disaster scene and to do it quickly. Fire and rescue personnel depend on this access in order to use a variety of management tools, and it provides an excellent way to link the diverse land mobile radio networks whose lack of interoperability seriously hampers response to major disasters. See [1] for full details.

At the time we began the project, broadband fixed wireless was in its infancy. IEEE 802.11 systems operated at 11 Mbps and were primarily used for office LANs. The best possibility for obtaining 100 Mbps service was to adapt satellite modem technology for use in the 28 GHz LMDS band. This approach was convenient for us, since Virginia Tech owns the LMDS spectrum in our part of Virginia. In 2000-2001, fixed wireless technology advanced rapidly and the commercial LMDS market collapsed, and 155 Mbps point-to-point operation in the unlicensed 5 GHz band became commonplace. While this paper presents data for our 28 GHz system, we are now moving forward with a 5 GHz demonstration system based on commercial radio technology modified to include significant cognitive radio functions. See [2] for a full discussion of the technological and social changes that followed 9/11/2001 and their influence on the project.

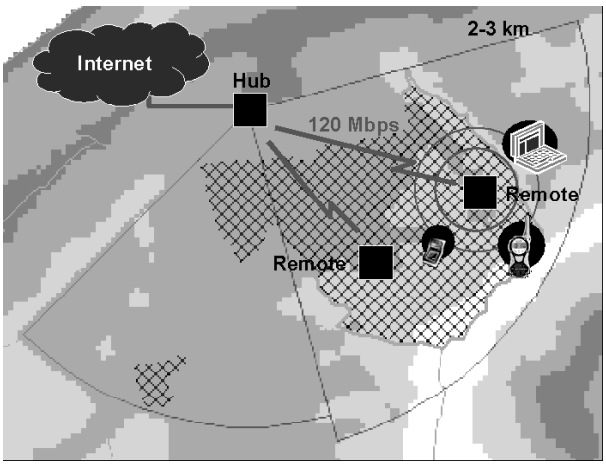


Figure 1. Disaster Communications System Concept

While 28 GHz operation normally implies point-to-point line-of-sight (LOS) operation, coverage may be extended by taking advantage of diffracted signals and particularly of "bounce paths" incorporating reflections from building walls and terrain features. The combination of short wavelengths and highly directional antennas associated with 28 GHz systems makes the propagation environment rather different from that in the 1-2 GHz spectrum well known to engineers who work with cellular telephone and IEEE 802.11 systems. Almost all reflecting surfaces appear rough at 28 GHz, and diffuse scattering (rather than specular reflection) dominates. Received signals do not consist of discrete multipath components, and signal amplitudes are not Rayleigh or Ricean distributed. In a shadowed area (one in which no clear LOS path to the transmitter exists), a receiver with a directional antenna can often find one or more good bounce paths that take advantage of scattering by building walls. Because of their rumored short lifetimes and dispersive characteristics, such "paths of opportunity" are of little commercial interest, but they may be invaluable in disaster situations. We discuss their characteristics in Section 3.

We expect diffuse scattering to be less important at 5 GHz, but at this time experimental evidence is lacking. Rough surface scattering at 28 GHz is of interest from an electromagnetics perspective, while 5 GHz effects are more important for near-term system deployment. We are investigating path characteristics in both bands.

In order to identify paths of opportunity and study their characteristics, we developed an impulse channel sounder based on ultra wideband technology. When perfected, it will allow our system to identify the best path and set the radio parameters for optimum quality of service. Since the channels encountered in a given disaster situation may differ significantly from those anticipated or previously experienced, our goal is to use the sounder output to drive a cognitive engine that will control the radios. Section 2 discusses the sounder and Section 4 presents our genetic algorithm based implementation of cognitive radio. In Section 5 we outline our plans for implementation and testing a cognitive system in the 5 GHz band.

2 The Virginia Tech Channel Sounder

One of the key attributes of the system is the integration of a novel, low-cost, broadband channel sounder developed at Virginia Tech [3]. In essence, the channel sounder takes a snapshot of the channel impulse response and passes it on to the system. This information can then be used to make intelligent estimates of possible link performance and intelligent decisions regarding system configuration. It is possible to use the digitized sounder output as input to a genetic algorithm that can determine the best system configuration for the given environment.

The channel sounder implementation is a combination of ultra-wideband technology at the transmitter and very precisely controlled sampling at the receiver. A train of short RF pulses is transmitted at a rate that is very precisely controlled by GPS clock information. The receiver samples the incoming waveform at a slightly lower rate than the transmitted pulse repetition rate. One sample is recorded for each transmitted pulse, but the effective sampling time for each successive pulse is slightly later in each transmitted pulse period. In this way, it is possible to reconstruct a single channel response based on the transmission of multiple pulses. For a more complete description of the channel sounder operation, see [1].

We have integrated the channel sounder in a pair of 28 GHz radios. Fig. 2 shows the transmitted pulse and Fig. 3 is a typical channel sounder output captured during our experiments. Note that Fig. 3 shows the effects of diffuse scattering and multipath on the received pulse.

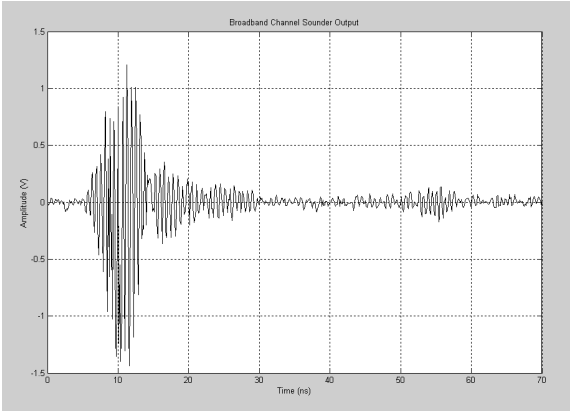


Figure 2. Pulse transmitted by the broadband channel sounder. The plot shows amplitude in volts versus time in ns.

By comparing the transmitted pulse in Fig 2 to the received waveform in Fig. 3, it is possible to extract the channel impulse response. The channel impulse response contains information regarding the distortion introduced by the channel. In the case of Fig 3, it is clear that the receiver first sees about 20 ns of energy that does not look like the transmitted pulse followed by a relatively undistorted pulse at around 30 ns. The initial energy seen at the receiver is a combination of multiple pulses reflected from a rough surface (diffuse scattering) while the pulse seen at around 30 ns is a specular reflection from a relatively smooth surface. Both phenomena contribute to changes in the

communication link performance and both need to be characterized in order to drive the genetic algorithms necessary to control the cognitive radio process.

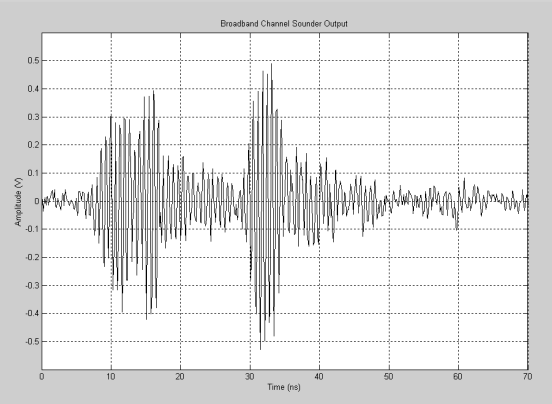


Figure 3. Typical broadband channel sounder output. The plot shows amplitude in volts versus time in nS.

3 Characteristics of Bounce Paths

For many applications at relatively low frequencies, it is generally accepted that the reflections causing multipath are primarily specular. That is, bounce paths from a single surface consist of a single reflection. The signal may be attenuated, but the energy is not spread over time. However, as the frequency increases to the level where the signal wavelength is on the order of the height of the surface irregularities, significant signal spreading may occur. Experiments conducted at Virginia Tech utilized radios operating at 28 GHz where the wavelength is approximately 1 cm. At such small wavelengths, many surfaces look rough and are therefore more likely to produce diffuse scattering.

In [4], Dillard, *et al*. showed that at 28 GHz, limestone walls exhibit significant diffuse scattering with energy arriving at the receiver as long as 75 ns after the specular reflection was received. The severity and duration of the diffuse scattering also depended on the path geometry and the transmitter and receiver placement. When the transmitter was closer to the surface than the receiver, the excess delay was larger. Both results show that the assumptions made in existing lower frequency systems (specular multipath and symmetric link performance) are not valid for rough surfaces – *i.e.*, for 28 GHz bounce paths. New methods need to be devised to characterize what the absence of these assumed conditions means to the quality of the communications link.

Very little has been published concerning the effect of diffuse scattering on communication link performance. In [5], Miniuk collected several channel impulse responses using equipment similar to [4], modeled the channel in software, and ran several Monte Carlo simulations to quantify the effect different diffuse scattering channels had on communication links. The results were interesting in that they showed that increasing the symbol rate in some channels does not necessarily decrease the performance. This suggests that the standard practice of estimating a channel's coherence bandwidth and using that value as the maximum permissible data rate may be inadequate for fixed broadband wireless communications, at least over bounce paths.

Work is continuing at Virginia Tech to determine what metric or metrics are more suitable to characterizing these channels. Early results show that amplitude and group delay variation across the band of interest could be used to better describe the quality of the link [6]. It may also be possible to calculate the actual coherence bandwidth of the channel rather than estimate it using established empirical formulas that may only be valid for other, more well-behaved channels.

4 The Genetic Algorithm Approach to Cognitive Radio

In [7], Rieser *et al.* propose a cognitive radio architecture based on genetic algorithms. Most traditional radios have their technical characteristics set at the time of manufacture. More recently radios have been built that self adapt to one of several several preprogrammed RF environments that might be encountered. Cognitive radios go beyond preprogrammed settings to operate both in known and unknown wireless channels. Most cognitive computing systems to date have been based on expert systems and neural networks. Such systems can be quite brittle in the face of unknown environments or else they require extensive training.

The model in [7] is based on biologically based models of cognition inspired by child development theories of two-way associative learning through play. Our cognitive model imitates the ability of young minds to adapt rapidly to new situations. We found genetic algorithms well suited for this task because of their ability to find global solutions to changing solution spaces that are often quite irregular. Genetic algorithms are (a) able to synthesize best practices through the crossover operation and (b) enable spontaneous inspiration and creativity through the mutation operation. We devised a multi-tiered genetic algorithm architecture that allowed sensing of a wireless channel at the waveform or symbol level, on-the-fly evolution of the radio's operational parameters, and cognitive functions through use of a learning classifier, meta-

genetic algorithm, short and long term memory and control. Fig. 4 shows the architecture.

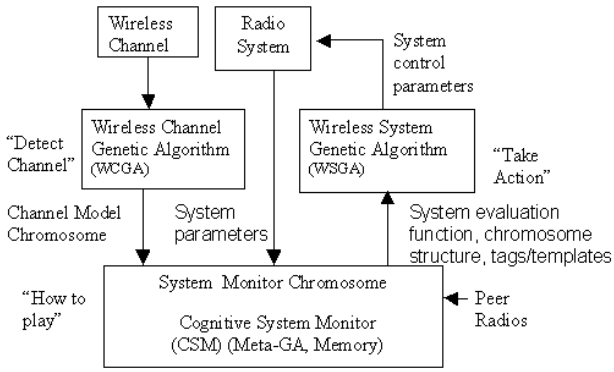


Figure 4: Biologically Inspired Framework for Cognitive Radio Based on Genetic Algorithms [7]

The Wireless Channel Genetic Algorithm (WCGA) allows modeling of any wireless channel error stream using the compact form of a Hidden Markov Model (HMM). For more discussion of HMM modeling of wireless channels please see [8].

Several chromosome structures were devised that allows the representation of wireless channels. An example of an HMM and the equivalent WCGA chromosome is shown in Figures 5 and 6 below.

The HMM of Figure 5 has N = 3 states and M = 2 possible outputs from any state.

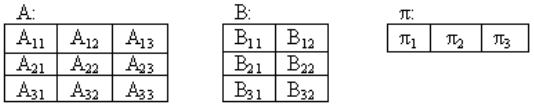


Figure 5: A Generic HMM [7]

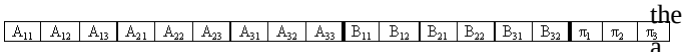


Figure 6: Chromosome Structure for a HMM [7]

The WCGA algorithm works as follows:

1. Initialize population of HMM chromosomes
2. Repeat until stopping criterion
   - *Choose* parent HMM chromosomes
   - *Crossover* parent chromosomes to create new HMM chromosome
   - *Mutate* new HMM chromosomes
   - *Replace* HMM chromosomes
   - *Evaluate* statistics of output sequence produced by new HMM chromosome population
3. Choose best HMM chromosome from final generation

The genetic algorithm (GA) essentially searches for the best HMM of a given observed symbol level error stream and generates a channel model that is statistically similar to the observed wireless channel. In [8], we showed that the WCGA could produce a wireless channel model of an error stream derived from a measured channel impulse response that closely matches the actually measured bit error rate (BER) behavior of a broadband wireless channel. This experiment showed that the "sensing" portion of the cognitive radio architecture matched real world tests.

The WCGA uses are error stream for the input, which is a train of symbols representing the number of bit errors per symbol. For the WCGA to produce an accurate model, many thousands of error symbols must be collected, which would require a long training sequence, taking both time and bandwidth. A more compact and efficient approach to channel modeling is to utilize the information collected by the channel sounder.

While the channel sounder response can provide an immediate understanding of the channel, the data received from the sounder is large and bulky. By using the channel sounder response, a model of the channel is derivable by simulating the channel as a filter with an impulse response derived form the channel sounder. A random bit sequence passed through the simulated channel will produce an error sequence. The WCGA can now receive an error sequence without the required overhead of a training sequence.

Because we are interested in a statistical model of the channel, we can use the simulated channel instead of the true error sequence. The Hidden Markov Model of the channel developed by either a true error sequence or a simulated error sequence is still a statistical representation of the channel. However, this representation is very small compared to the channel sounder data and is capable of representing the channel equally well.

Fig. 7 shows how well-matched the simulated channel is to a theoretical AWGN channel. The WCGA was then used with the simulated error stream to develop a channel model with the statistics represented in Fig. 8, which shows a histogram of the number of errors of a certain burst length over the channel. Fig. 9 then shows how well matched the HMM representation of the channel is compared to the simulated channel, and therefore, how well matched the HMM representation is compared to the actual channel (via Fig. 7).

Figs. 7, 8, and 9 are the subject of a paper we submitted to the Microwave Theory and Technique Society's 2004 International Microwave Symposium that has not yet been published [8]

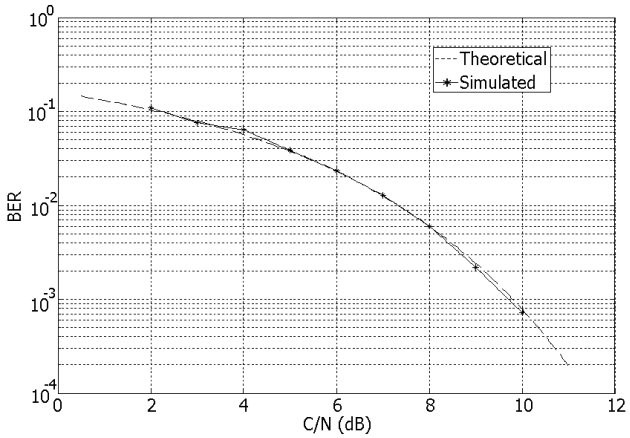


Fig. 7. A simulated model of an AWGN wireless channel versus the theoretical channel.

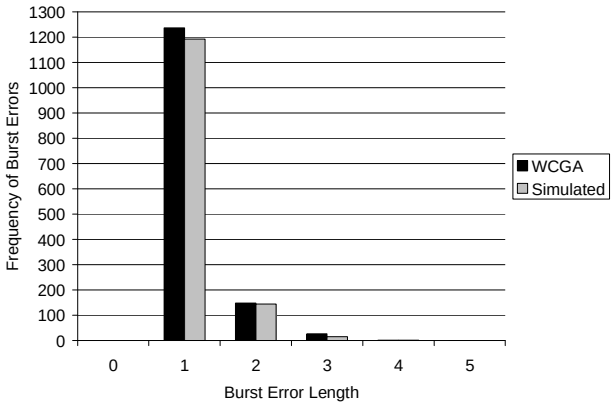


Fig. 8. Burst error statistics of the simulated channel versus the HMM channel. [7]

The Wireless System Genetic Algorithm (WSGA) operates in a similar manner as the WCGA in that it uses a chromosome structure that represents the parameters of the radio under test. The WSGA is given a fitness function, or set of goals, by the Cognitive System Monitor (CSM) module and continuously adapts the radio based on these goals. Example goals could be providing a desired balance of BER, power, frequency, modulation, and data rate behavior for a given Quality of Service and wireless spectrum band.

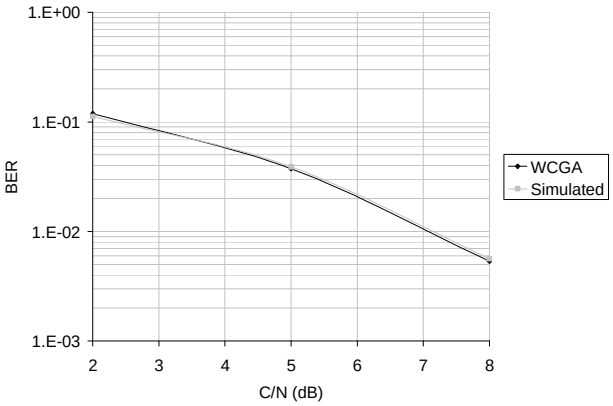


Fig. 9. BER curves of simulated channel versus the HMM channel. [8]

The CSM genetic algorithm consists of a learning classifier function that classifies the observed channel model received from the WCGA or broadband channel sounder and a meta-genetic algorithm that determines the appropriate fitness function, chromosome structure, and tags and templates using the crossover operator based on knowledge from its short and long term memory as well as the creative new solutions generated from its mutation functions.

The GA approach to adapting a wireless radio provides many benefits. First, it is a chaotic search with controllable boundaries that allow it to seek out and discover unique solutions efficiently. In unknown channels, chaotic behavior could produce a solution that is absolutely correct but counter-intuitive. By being able to control the search space by limiting the number of generations, crossover rates, mutation rates, fitness evaluations, etc., the cognitive system can ensure legal and regulatory compliance as well as efficient searches.

Another major benefit of the GA approach is the versatility of the cognitive process to any radio. While a software radio is an ideal host system for a cognitive processor, any legacy radio with the smallest amount of adaptability can benefit from our cognitive processes. The cognitive system defines the radio by a chromosome, where each gene represents a radio parameter such as transmit power, frequency, modulation, etc. The adaptation process of the WSGA is performed on the chromosomes to develop new values for each gene, which is then used to adapt the radio settings. If a radio cannot adjust a particular parameter, then the adaptation process will ignore the gene representing that parameter. Also, if there are certain parameters unique to a particular adaptable radio, we have left a few genes unused so as to be used for such proprietary purposes.

Because each radio will have a unique method of adapting the radio parameters and each parameter will mean something different, a small hardware interface module is required to connect the WSGA to the radio. The interface module will take the chromosome from the WSGA and use the gene values to properly update the radio. The interface is a small piece of software required for each radio while the cognitive processing engine remains system-independent.

While the independence of the WSGA and cognitive processor to the radio allows any radio to become a cognitive radio, it should be clear that the more adaptable a radio is, the more powerful the cognition becomes.

5 Plans for Implementation and Testing

We are implementing the WSGA and CSM for a set of 5 GHz Proxim *Tsunam*i radios. The WSGA module of the cognitive radio test bed will enable the radios to change a number of their operating parameters based on the sensed behavior of the wireless channel. These parameters include power, modulation index, forward error correction (FEC), and TDD mode. We have written a hardware interface module that provides a uniform interface between the WSGA and the Proxim radio. We plan several tests of the CSM and WSGA algorithms in which the cognitive radio test bed algorithms are supplied with different stored error streams representing several known and unknown wireless channels and the algorithms are run to show their performance in these channels on the fly. This experiment will validate the WSGA's ability to evolve the radios based on changing wireless channel conditions and spectrum availability as well as the CSM's cognitive ability to discern how best to direct the radios to operate in the given wireless spectrum environment. Future tests will enable live measurement of the wireless channel using impulse responses captured by the broadband channel sounder or symbol level error streams recorded by the receiver.

As the project matures, we envision the use of the cognitive processes on a more adaptable software defined radio, which would enable us to show the increased power of the GA based cognitive radio.

6 Conclusions

Our work represents early steps in using an impulse sounder to characterize a wireless channel and pass appropriate information along to the genetic algorithms that will allow them to effectively configure a broadband disaster communication system and realize a functioning cognitive radio.

References

[1] C.W. Bostian, S. F. Midkiff, W. M. Kurgan, L. W. Carstensen, D. G. Sweeney, and T. Gallagher, "Broadband Communications for Disaster Response", *Space Communications*, Vol. 18 No. 3-4 (double issue), pp. 167-177, 2002

[2] Charles W. Bostian and Scott F. Midkiff, "Demonstrating Rapidly Deployable Broadband Wireless Communications for Emergency Management," dg.o.2002, Redondo Beach, CA, May 19-22, 2002.

[3] Christian Rieser, "Design and Implementation of a Sampling Swept Time Delay Short Pulse Wireless Channel Sounder for LMDS", Master's Thesis, Virginia Polytechnic Institute and State University, July 2001.

[4] C. L. Dillard, T. M. Gallagher, C. W. Bostian, D. G. Sweeney. "28 GHz scattering by brick and limestone walls," *2003 IEEE Antennas and Propagation Society International Symposium*, Vol. 3, pp. 1024-1027.

[5] Mary Miniuk, "Channel Impulse Response and its Relationship to Bit Error Rate at 28 GHz," Master's Thesis, Virginia Polytechnic Institute and State University, December 2003.

[6] Tim Gallagher and Mary Miniuk, "Methodology and preliminary findings toward the characterization and evaluation of non-line-of-sight paths for fixed broadband wireless communications for emergency and disaster response," *National Conference on Digital Government Research*, May 2003.

[7] C. J. Rieser, T. W. Rondeau, and C. W. Bostian, "Cognitive Radio Architecture Based on Genetic Algorithms: A Proposed Architecture and Some Intitial Results," *IEEE Trans. on Wireless Communications*, submitted, 2003.

[8] T. W. Rondeau, C. J. Rieser, T. M. Gallagher, and C. W. Bostian, "Online Modeling of Wireless Channels with Hidden Markov Models and Channel Impulse Responses for Cognitive Radios," *IEEE International Microwave Symposium*, submitted, 2004

Acknowledgments

This work was supported by the National Science Foundation under awards 9983463 and DGE-9987586.We acknowledge the contributions of our colleagues W. Michael Kurgan and Richard Klobuchar at SAIC. The thesis research of Cindy L. Dillard and Mary T. Miniuk contributed significantly to this paper.

Trends in Telecom Development Globally: A Perspective from Washington

By Diane E. V. Steinour
Office of International Affairs
National Telecommunications and Information Administration (NTIA)
U.S. Department of Commerce
Washington, D.C. USA
Ph: (1-202) 482-1866
Fx: (1-202) 482-1865
Email: dsteinour@ntia.doc.gov

Abstract

Both in the United States and overseas, wireless technologies are playing a significant role in development, in many cases where wireline technologies have literally fallen short. At the same time, decision-makers no longer need to be convinced of the benefits that access to and use of Information and Communications Technologies (ICTs) provide. Rather, the shift is to find the best paths to universal access and developing principles for success to share as best practices. The U.S. Government is pursuing several paths to promote more universal access to ICTs, both in the United States and abroad. The outcomes of the recent World Summit on the Information Society (WSIS) and the launch of the U.S. Digital Freedom Initiative underscore these efforts.

In early January 2004, the People's Republic of China announced that there are now more mobile customers in China than there are people in the United States of America.(1) In late January 2004, the GSM industry proudly crowed that it expects the one billionth GSM customer in the First Quarter of 2004.(2) Up near Nome Alaska, the Bering Straits natives organized a non-profit tribal consortium called Kawerak, Inc. Twenty Inuit tribes have pooled their knowledge and resources to develop the Wireless WALRUS project – Web Access Links for Remote User Services. These native peoples have found another way to preserve their 3,000 year old presence in the Bering Strait region by using wireless connectivity over 26,000 square miles. (3)

In Tajikistan and Armenia, institutions and commercial users are using an "air bridge," or radio modem for Internet Service Provision connectivity. (4, 5) Over in Chicago, the Center for Neighborhood Technology (CNT) is a low-income community-based network. Early on, they learned the need to develop a scalable, replicable, and self-sustaining method to deliver high-speed, low cost Internet service that was revenue generating. The CNT now operates the Wireless Community Networks, focused on capacity-building using WiFi connections (802.11b) between four urban, suburban, and rural Illinois communities. (6)

These are key and current examples of recent and varied approaches using wireless technologies to promote greater Information and Communications Technologies (ICTs) development. They focus on partnership, on novel uses of current and new technologies, and in some cases, on mammoth commitment of national resources as in China.

There is no doubt the world is growing fonder of innovative wireless applications. But each country has to foster an economic and social environment that will allow technological innovation to flourish. Human capacity-building is just as important as siting new cellular towers or matching venture capital to new entrepreneurial efforts.

By exploring some of the current trends in wireless telecommunications development, we can establish the context for current USG development policy goals and initiatives in this area.

I. The Case for Development of the ICT Sector

A question often heard in development circles is why should a poor or developing country focus limited resources on development of the ICT sector? Why should developing economies pursue this path?

In the new global economies, ICT capabilities and skills – or their lack --- help to determine a nation's ability to compete, its economic growth, and most important, its standard of living. Exhibit number one and two: the two largest world economies, the United States and China. As of December 2003, according to Dr. John Marburger, the Director of the United States Office and Science and Technology Policy in the White House, the U.S. ICT industry represents only 8 percent of all American

enterprises. However, this 8 percent produces 29 percent of U.S. exports, generates high-quality jobs, and contributes strongly to productivity growth across all economic sectors. It is estimated that 40 percent of U.S. productivity growth between 1995-2002 can be attributed to ICT. (7) As of January 2004, China's Ministry of Information Industries (MII) Vice Minister Lou Qinjian notes that ICT growth in China has generated 6 percent of China's Gross Domestic Product (GDP) growth. (8)

A growing recognition of the importance of ICT for economic and social growth led to the United Nations' (UN) call for two World Summits on the Information Society (WSIS). The "First Phase" of the WSIS took place December 10-12 in Geneva, Switzerland. There, participants from over 175 nations agreed on "the pressing need for universal ICT access and the widespread infrastructure on which it is founded. It [the need for universal access] also points to enabling environments as essential for wider technology access and use and underscores that strong capacity building efforts are needed to achieve universal access. The widespread availability of low-cost applications plus respect for multilingual, diverse and culturally appropriate content are endorsed as well." Such issues as intellectual property rights, the control and management of Internet infrastructure, ICT development financing, human rights and freedom of expression were also addressed. (9)

The Summit agreed on specific goals, such as "connecting all villages, schools, hospitals and governments with ICT by 2015 and ensuring that half of the world's people are within reach of ICT." The goals are linked to pursuit of the UN's Millenium goals, an effort to combat "poverty, disease, homelessness, environmental degradation and gender inequality." Summit participants recognized a "pressing need" for universal access to ICT and related infrastructure, while noting that "strong capacity building efforts are needed to achieve universal access." (10) Participants will meet again in 2005 to measure progress.

At the WSIS, the U.S. head of delegation, Dr. John Marburger noted three key principles that reflect the U.S. Government's broad ICT developmental goals. These goals are aimed at stimulating and cultivating science, skills, and business infrastructure. The first is that domestic policies must encourage investment in research and innovation. Supporting goals emphasize movement toward privatization of ICT services supply, that is elimination or reduction of government ownership, and introduction of competitive supply models. To stimulate greater investment in infrastructure, governments must strive to create a stable and positive business and social environment.

Second, governments and the private sector must strive to invest in human capacity-building efforts, to best utilize ICTS and to share in their benefits. Workforces must be trained and well-educated to let the promise of ICTs flourish. Dr. Marburger noted that, "A vital communications infrastructure expresses the full range of cultural imagination, without the divisive barriers of censorship," striking a blow for freedom of expression and the lifting of control of Internet content worldwide.

Third, the intellectual property of innovators, content producers, and generally, consumers, must be protected or there will be insufficient trust in ICT products and services. Network security on a global scale informs part of this need – if no one can trust ICT products and services, there is no reason to keep building and producing them. (11)

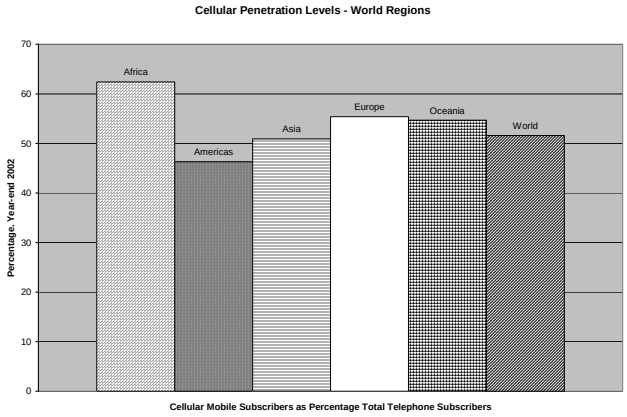
These general policy goals inform all of our ICT development efforts.

II. The Case for Wireless

Text messaging, fixed wireless infrastructure, growth of third-generation technologies and applications, and now Wi-Fi connectivity—these are all trends found in the developing economies. In recent years, the influx and importance of wireless technologies for worldwide ICT growth is often measured by the percentage of fixed versus mobile subscribers in any given country. The International Telecommunication Union (ITU) closely tracks these figures. Since 2001, mobile subscribers have passed the 50 percent mark, with cellular mobile subscribers now standing at 51 percent of all telephone subscribers worldwide at year-end 2002. (12)

What is very revealing is to see what geographical regions have high cellular penetration levels: those regions with a high percentage of developing economies. See Table I. Some of the growth figures are staggering. China, considered a developing economy by many, expects 400 million mobile subscribers by 2005. (13) Uganda had 5,000 cellular subscribers in 1997, and 393,000 in 2002, for a compounded annual growth rate (CAGR) of 139.4 percent. Paraguay had 84,000 cellular subscribers in 1997, which grew to 1.7 million in 2002, at a CAGR of 81.7 percent. (14)

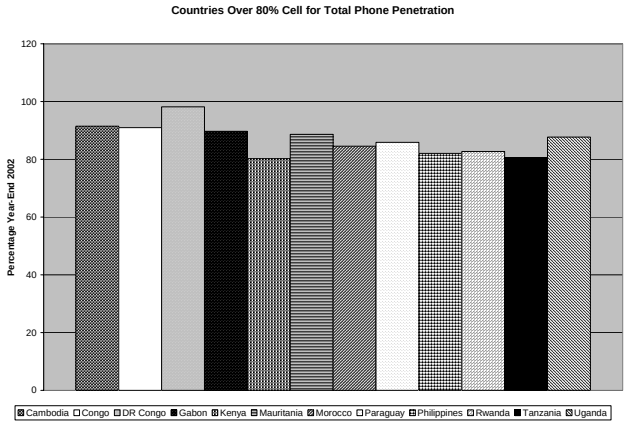
TABLE I: *Cellular Penetration Levels – World Regions*



Source: ITU, December 2003.

Twelve countries have surpassed the 80th percentile for cellular subscribership as a percentage of total telephone subscribers. See Table II.

Table II: *Countries Over 80 Percent Cellular Subscribers as Percentage of Total Telephone Subscribership*



Source: ITU, December 2003.

But everything is not roses and poetry when it comes to deployment of wireless technologies in developing countries. In the case of WiFi, there are many grey areas in domestic laws in overseas markets. Some countries interpret new technologies such that, if it isn't banned, go ahead and pursue it.

More likely, though are cases such as Kenya (15) where if it a technology is not specifically allowed, or a frequency is not specifically allocated for licensing, it is forbidden. Such policies are often adopted by governments that are trying to protect legal monopoly and state-owned services providers. The U.S. Embassy in Albania notes that in Albania, "wireless technology is used only by those who can afford it, such as banks, government agencies, and international organizations. Today there are fewer than 50 clients that use this technology…." (16) Since wireless is usually a less costly solution for connectivity needs, these examples underlies the issue of scalability, the need for regulatory reform and pro-competitive supply models.

In the case of Senegal, it is a grey area. An interagency team including technical experts from ITS, NIST and State have been working in recent months on a WiMAN (Wireless Metropolitan Access Network) project to see if it might take root in Senegal. However, under Senegalese law and policy, there are no legal specifications yet for WiFi or WiMAN provisioning. The U.S. Government and its U.S. AID contractors are now working with the new Senegalese regulatory body to address how best to pursue such technology-neutral development policies as the regulator oversees the introduction of competition into telecom services provisioning in 2004.

However, more and more regulators and policy-makers see the potential for wireless connectivity to promote Internet uptake. WiFi, WiMAN, whatever label you choose, the pundits foresee strong prospects. In July 2003, Pyramid Research forecast there will be 700 million WiFi users worldwide by 2008. (17) Clearly, countries will have to find novel ways to address the WiFi situation through appropriate policy and regulatory approaches.

III. Some Principles & Best Practices

The U.S. Government pursues a variety of paths on the pro-development agenda for ICTs. There are the traditional aid organizations and funding initiatives. Also, NTIA joins our colleagues in the State Department, the Federal Communications Commission, the U.S. Trade Representative's Office and the International Trade Administration to promote pro-competitive policy and regulatory reform around the world. We work in partnership with recipient countries, directly on a bilateral level, and jointly through multilateral and regional bodies,

establishing joint principles for action and best practices resources.

Following the 1998 ITU World Telecommunication Development Conference, NTIA became the rapporteur for a joint public-private sector effort to promote Internet access in developing countries. Participants performed a technology review and made recommendations to develop pro-competitive telecommunications policies and regulations. The result was the 2001 release of "*Promotion of Infrastructure and Use of the Internet In Developing Countries,*" better known as Study Question 13/1.

The report noted that many factors limit access to and use of the Internet, particularly in developing countries. These include restriction of ISPs and public Internet access points, restricted access to international gateways, insufficiency of Internet points of presence in rural and disadvantaged communities, inadequacy in advanced networking techniques, and regulatory policies that favor telephone monopolies. The technology review included such wireless options as: 1) VHF and UHF radio systems using narrow packet radio technology; 2) Global System for Mobiles (GSM400) using packet switching technology; 3)Time Division Multiple Access (TDMA) based on Point-To-Point (PTP) or Point-to- Multipoint (PMP) Radio Systems; 4) Code Division Multiple Access (CDMA) 450 MHZ; 5) Multipoint Multichannel Distribution System (MMDS); 6) Local Multipoint Distribution System (LMDS); 7) Very Small Aperture Terminals (VSAT),;and 8) Satellite Based Internet Access.

The U.S. Government has endorsed the policy recommendations advanced under Question 13/1, which include:

- Make leased lines available at reasonable cost and access charges for dial-up services affordable
- Enable submarine cable operators to obtain backhaul at competitive rates
- Promote satellite interconnection between ISPs
- Allow network providers to sell capacity directly to ISPs
- Lower custom tariffs and taxes on telecommunications equipment
- Promote private investment in telecommunications and Internet infrastructure
- Establish a consortium of public service institutions to contribute to Internet access, use and development
- Encourage the development of information strategies and models that facilitate community access
- Develop national programs to promote capacity building in Internet development and use, and the creation and dissemination of multicultural and multilingual Internet content. (18)

At the same time that the U.S. Government was pursuing Internet development policies at the ITU, we also worked with the 21-member Asia Pacific Economic Cooperation forum, or APEC. NTIA helped steer the APEC's Telecommunications and Information Working Group's (TEL) efforts to develop a Digital Divide Blueprint for Action, adopted in 2001. Approximately 12 APEC economies are considered developing economies (depending on one's view regarding China). As part of that effort, the TEL gained consensus on six policy principles to help improve Internet uptake and investment in new ICT technologies. These focus on:

- Leadership – noting governments should create national, regional, and local initiatives to create a vision and to develop institutions and structures to address issues;
- Partnerships – economies should work to create partnership between and among business, education, civil society, and government;
- Policy Coherence - governments should ensure that all policies (macroeconomic, social, educational, etc.) are working seamlessly to create the desired economic and social environment;
- Market Focus – governments are encouraged to promote pro-competitive equipment provisioning and services supply environments, to foster demand that can justify the investment required.
- Sustainability – all parties should work to ensure the continuation of initiatives and services beyond the seed money stage; and
- Scalability – designers of initiatives and projects should work to ensure that these

can be remodeled and replicated for other applications and geographic areas, especially under-served communities. (19)

In the Western Hemisphere, the U.S. Government works primarily with the Inter-American Telecommunications Commission, or CITEL, to promote a pro-development agenda. Under the auspices of the Organization of American States, the 35-member states of CITEL strive to make telecommunications a catalyst for the dynamic development of the Americas.

One of the CITEL's prime development activities is the release of "The Blue Book," or "*Telecommunication Policies for the Americas Region."* Released jointly by the ITU's Development Sector, the Blue Book gives a fresh perspective on best practices and provides a baseline for discussion on the impact of convergence and the Internet. There is a strong focus on legislative, policy and regulatory environments. Countries can avail themselves of the Blue Book as they think fit, in accordance with their own national public policy and juridical, administrative and social framework. The Bluebook is a dynamic instrument, subject to periodic review; a third edition is now under development. CITEL has also performed detailed studies on best practices, conducted jointly with the ITU's Development Sector, on Tele-Education, on Tele-Medicine, and on Universal Service. (20)

IV. The Digital Freedom Initiative

The U.S. Government's most recent effort to promote ICT development worldwide is through the Digital Freedom Initiative, or DFI. We are fast approaching the first anniversary of the DFI. In fact, it coincides with the ISART Conference, on March 4. The goal of the DFI is to promote economic growth by transferring ICT benefits to entrepreneurs and small businesses in the developing world. The DFI approach leverages the leadership of the U.S. Government with the creativity and resources of America's leading companies and the vision and energy of entrepreneurs. The first DFI effort was launched mid-year 2003 in Senegal, with new efforts underway in Peru in November 2003, and in Indonesia in January 2004. We expect an additional five countries to be named in 2004, and up to 20 countries by 2008.

DFI's implementation depends on an alliance between U.S. private sector companies, the U.S. Government, the host government, and the host private sector. The private sector is a key pillar of the DFI's sustainability. We strive to incorporate private sector strategic thinking and business savvy to develop replicable and scalable solutions.

The U.S. Government and especially NTIA are drawing upon its experiences in the ITU, APEC, CITEL and through bilateral experiences to inform the DFI process. In addition to the traditional aid agencies such as US AID, the Peace Corps, and the State Department, new DFI partners include the Federal Communications Commission, NTIA, and the International Trade Administration, all under the leadership of the U.S. Commerce Department's Technology Administration.

While we are in early days for Peru and Indonesia, we have done extensive design work in Senegal for three pilot projects. These pilots focus on improving productivity in Telecenters/Cybercafes, improving access to markets for Small and Medium-Sized Enterprises (SMEs) using ICT tools, and creating a supportive environment for micro-finance in a region where banking is centralized in a neighboring country. Recent successes include the inauguration of a Cisco Networking Academy in December 2003, and the formation of a new association to represent the views of the users community before the government, called SITSA (French acronym), the Senegalese equivalent of the Information Technology Association of America.

In terms of policy and regulatory development, U.S. agencies have worked closely with senior Senegalese policy officials, and a new Senegalese regulatory team, to complete draft decrees and undertake the legal legwork needed to bring a new legal framework to fruition. This new framework will guide the introduction of full competition into the Senegalese telecommunications market, where a privatized former state-owned enterprise still controls most telecommunications market segments. The United States also sponsored capacity-building workshops for the new regulatory team. The onus has now shifted to Senegal to announce its next steps to implement the new legal framework.

And as mentioned before, U.S. Government technical experts participating in the Senegal initiative are developing a feasibility study plan for next-generation wireless technology, to include 802.16 WIMAN technologies. U.S. participants note how well-suited wireless technologies are to accomplish the goals of DFI, offering greater connectivity at greatly reduced prices to both the builders and consumers of the service. The point-to-multipoint capabilities of wireless technologies over

diverse geographical areas and long distances make them attractive in Senegal. Our experts' objective is to develop costing models and a testbed that could then be replicated in additional DFI candidate countries in the future. First steps in Senegal include the use of WIMAN technology to connect an existing Wireless Internet Service Provider (WISP), or a traditional ISP or Internet café entrepreneur. Project participants would provide equipment and training as needed with the dual goals of extending connectivity while developing a more dense user base. Technical configurations and a business model will be developed jointly with local entrepreneurs to ensure sustainability and consistent service levels. The team will also work with the local regulator and incumbent operators to ensure there is truly a competitive environment and a supportive regulatory environment to allow WIMAN to flourish. Timeframes to implement the study are still under development.

To celebrate the DFI's first anniversary, the U.S. Department of Commerce's Under Secretary for Technology, Phillip Bond, plans to lead a delegation to Dakar, Senegal between March 8-12, 2004. Members of the DFI Business Roundtable will accompany him, demonstrating the joint public-private nature of the initiative.

V. Going Forward

Aside from further pursuit of myriad new activities under the DFI, the U.S. Government has other new ICT development activities.

At WSIS, the United States pledged US$400 million in grant money to support ICT development in developing countries. The U.S. Government's Overseas Private Investment Corporation (OPIC) has established a "support facility" to encourage U.S. investment in the sector, at a time when capital expenditures are down globally for ICT development. The grants will fund joint ventures between the public and private sectors in the 152 countries where OPIC operates. (21)

Also, collaboratively, the NTIA is chairing a new ITU effort to develop a best practices resource entitled "IP Policy Manual." The manual will advise ITU Member States on a variety of Internet issues such as domain names management. It has a particular focus on the needs and questions of developing countries.

NTIA continues to provide assistance to the reconstruction of Afghanistan and now Iraq. We have met with senior Afghani ministry officials to provide advice on implementation of their new development plans. In Iraq, we will shortly detail an NTIA wireless policy expert to assist in general telecommunications policy reform and reconstruction efforts. He will supplement the efforts of a current detailee from our spectrum management office, who is assisting on spectrum activities for the Coalition Provisional Authority. They are only an email away. Let's hope their Blackberries work over there.

End Notes

1. Ministry of Information Industry (MII) of China, at U.S.-China ICT Seminar, January 13, 2004.
2. GSM Association, January 26, 2004.
3. Technology Opportunities Program (TOP), National Telecommunications and Information Administration.
4. U.S. Embassy Dushanbe, Tajikistan.
5. U.S. Embassy Yerevan, Armenia.
6. Technology Opportunities Program
7. Marburger, John, Director, Office of Science and Technology Policy, Executive Office of the President, USA.
8. Ministry of Information Industry.
9. "ITU Secretary-General Opens First Global Information Summit," December 10, 2003.
10. *Ibid.*
11. Marburger.
12. International Telecommunication Union, *World Telecommunication Indicators.*
13. China Unicom, at U.S.-China ICT Seminar, January 13, 2004.
14. *World Telecommunication Indicators.*
15. U.S. Embassy Nairobi, Kenya.
16. U.S. Embassy Tirana, Albania.
17. Pyramid Research, "Pyramid Predicts 700 Million WiFi Users by 2008," July 23, 2003.
18. International Telecommunication Union, *Promotion of Infrastructure and Use of the Internet In Developing Countries.*
19. Asia Pacific Economic Cooperation (APEC) Telecommunications and Information Working Group (TEL).
20. Inter-American Commission for Telecommunications (CITEL).
21. U.S. State Department, http://www.state.gov/e/eb/cip/wsis/.

Resources and Bibliography

Asia Pacific Economic Cooperation (APEC) Telecommunications and Information Working Group (TEL). At www.apectelwg.org.

Digital Freedom Initiative, www.dfi.gov.

GSM Association, "GSM on Target to Connect Billionth Customer in Q1," GSM Association Press Release, January 26, 2004, at http://www.gsmworld.com/news/press_2004/press04_06.shtml.

Inter-American Commission for Telecommunications (CITEL) (http://citel.oas.org).

- *The Blue Book: Telecommunication Policies for the Americas Region (* http://www.intu.int)
- *Tele-Education in the Americas* http://citel.oas.org/Tele-Education/Table%20of%20Content.asp
- *Universal Service in the Americas* http://citel.oas.org/pubs/universal_service.asp

International Telecommunication Union, Geneva, Switzerland.

- *IP Policy Manual,* at http://www.itu.int/ITU-T/special-projects/ip-policy/index.html
- "ITU Secretary-General Opens First Global Information Summit," December 10, 2003. http://www.itu.int/wsis/geneva/newsroom/press_releases/wsisopen.html.
- *Promotion of Infrastructure and Use of the Internet In Developing Countries,* ITU Development Sector, Document 1/185(Rev.1)-E, 24 October 2001 at www.itu.int.
- *World Telecommunication Indicators*, December 2003. http://www.itu.int/ITU-D/ict/statistics/at_glance/cellular02.pdf

Marburger, John, Director, Office of Science and Technology Policy, Executive Office of the President, USA. "Information and Communication Technology is a Key to the Future Prosperity of All Nations." Statement before the World Summit on the Information Society, December 11, 2003. http://www.state.gov/e/eb/rls/rm/2003/27670.htm.

Ministry of Information Industry (MII) of China, at U.S.-China ICT Summit, January 13, 2004, Washington, D.C. Author's personal notes from attendance.

Pyramid Research, Cambridge, MA. "Pyramid Predicts 700 Million WiFi Users by 2008," July 23, at http://www.pyr.com/info/press/release_030721.asp.2003.

Technology Opportunities Program (TOP), National Telecommunications and Information Administration, Washington, D.C. At http://ntiaotiant2.ntia.doc.gov/top/2003/index.cfm.

U.S. Embassy Dushanbe, Tajikistan. "Tajikistan: Wi-Fi Survey Response." Cable number 2633, May 21, 2003.

U.S. Embassy Nairobi, Kenya. "Kenya: Wireless Internet Survey." Cable number 1848, May 6, 2003.

U.S. Embassy Tirana, Albania. "Responses to Wireless Internet Survey." Cable number 717, May 5, 2003.

U.S. Embassy Yerevan, Armenia. "Armenia: Response to Wireless Internet Survey." Cable number 996, May 16, 2003.

World Summit on the Information Society, First Phase, December 10-12, 2003, Geneva. At http://www.itu.int/wsis/. See also at U.S. State Department, http://www.state.gov/e/eb/cip/wsis/

Mesh Networks: The Next Generation of Wireless Communications

Jason Melby
Loop Start Consulting Group
Phone (703) 779-7970
Fax (703) 779-5745
melbyj@loop-start.com

Conceived by the U.S. Military, mobile ad hoc networks, commonly known as mesh networks, provide end-to-end Internet Protocol (IP) communications for broadband voice, data, and video service combined with integrated geographical location logic designed to function in a mobile wireless environment. Unlike 802.11 wireless local area networks (WLANs) and point-to-multipoint digital cellular networks, mesh networks accommodate a more dynamic operational environment where their radio frequency (RF)-independent, self-forming, and self-healing properties meld the best of both worlds between WLAN and cellular systems. This paper examines the concept of mesh networks with a look at recent commercial and military development of what some consider a disruptive, next-generation wireless communications technology.

1. Introduction

Loosely speaking, mesh networks form a wireless Internet where any number of host computing nodes can route data point-to-point in an intricate web of decentralized IP links built upon many of the routing features first employed by earlier packet radio networks [4]. Borne from a heritage of 1960s and 1970s packet data radios designed to provide reliable communications for connectionless, non-real-time traffic, today's mesh networks have evolved to provide multicast IP traffic with real-time requirements [1]. In essence, mesh networks extend the concept of packet data radio communications by using sophisticated digital modulation schemes, traffic routing algorithms, and multi-hop architectures that challenge the laws of physics by using minimal transmission power to increase data throughput over greater distances. With mesh networks, any node within the network can send or receive messages and can relay messages for any one of its hundreds or thousands of neighboring nodes, thus providing a relay process where data packets travel through intermediate nodes toward their final destination. In addition, automatic rerouting provides redundant communication paths through the network should any given node fail [2]. This ability to reroute across other links not only provides increased reliability but extends the network's reach and transmitting power as well. This resilient, self-healing nature of mesh networks stems from their distributed routing architecture where intelligent nodes make their own routing decisions, avoiding a single point of failure. Because mesh networks are self-forming, adding additional nodes involves a simple plug-and-play event [3]. And because mesh networks don't rely on a single access point for data transmissions, users of this technology can extend their communication reach beyond a typical WLAN. Furthermore, mesh networks and their low power, multi-hopping ability allow simultaneous transmissions to reach nearby nodes with minimal interference [17]. Achieving this self-forming, self-healing utopia with minimal power and signal interference involves the implementation of sophisticated routing logic within the software and hardware to account for minimum latency, and maximum throughput, as well as provide for maximum security and reliability [7]. Figure 1 depicts a mesh network configuration with a single wireless access point connected to a wireline backbone that provides end-users with Internet access. If so desired, the four end-nodes could function as a self-forming independent service set capable of sending and receiving voice, video, and data between themselves without a wireless access point.

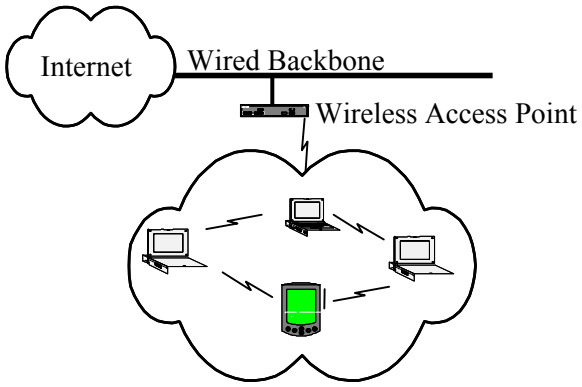


Figure 1. Mesh Network with Single Access Point

As with all radio frequency (RF) communication systems, mesh networks must contend with noise, signal fading, and interference; however, unlike other RF systems, mesh networks deal with noise, signal

fading, and interference through an air interface protocol originally designed to provide reliable battlefield communications. Known as quad division multiple access (QDMA), this air interface provides the driving force behind mesh network capabilities. Conceived by Military Commercial Technologies (MILCOM) and a communications division of ITT Industries, QDMA allows mesh networks to facilitate higher throughput without sacrificing range - or extending transmission range without sacrificing throughput. QDMA supports low-power, high-speed broadband access in any sub 10 GHz frequency band, providing non-line-of-sight node linking to dramatically increase signal range without sacrificing throughput. Geared toward wide area mobile communications, QDMA compensates for wild fluctuations in signal strength with powerful error correction abilities and enhanced interference rejection that allows multi-megabit data rates – even from a mobile node traveling at 100 mph and beyond. And with shorter distances between network nodes, the resulting decrease in interference between clients provides for more efficient frequency reuse. Furthermore, QDMA offers highly accurate location capabilities independent of the satellite-based global positioning system (GPS) [2], [4], [5], [6].

2. Commercial Deployments

Since the inception of QDMA and the subsequent commercialized version of this technology, venture capital firms have invested more than $100 million since 2001 for continued design and development of mesh networks that could ultimately compete with IEEE's 802.11b [3]. One firm, appropriately named MeshNetworks, has adopted the QDMA technology with direct sequence spread spectrum (DSSS) modulation in the 2.4 GHz industrial, scientific, and medical (ISM) band, providing 6 Mbps burst rates between two terminals. Backed by almost $40 million in venture funding from 3Com Ventures, Apax Partners, and others, MeshNetworks signed its first customer, Viasys Corporation, in November 2002. Eventually, MeshNetworks plans to offer their networking capability in the 5 GHz unlicensed national information infrastructure (UNII) band [8]. For now, MeshNetworks, headquartered in Maitland, Florida, is testing a 2.4 GHz prototype in a five-square-mile test network around its Orlando suburb with an FCC experimental license to build a 4000-node nationwide test network [6]. To maintain Internet connectivity, MeshNetworks relies on multi-hop routing between nodes mounted on buildings, light poles, vehicles, and end-user devices [17]. Aside from designing prototype routers, relays, and PDA-size client devices, MeshNetworks plans to offer a software overlay solution for 802.11b clients in existing networks, effectively extending the range and link robustness of existing Wi-Fi networks through mesh-style multi-hopping [6]. Furthermore, MeshNetworks recently announced a deal with auto-parts manufacturer Delphi to test the feasibility of mesh networks in a telematics environment [9]. MeshNetworks competitors include FHP Wireless, which recently announced its formal launch date in March of 2003, and Radiant Networks from Cambridge, U.K., which has deals in place with British Telecom, Mitsubishi, and Motorola [3].

Interestingly, each of these potential mesh network providers will face a similar network coverage dilemma, a sort of catch-22 where the ability to expand network coverage hinges on the deployment of new subscribers whose mobile nodes will act as router/repeaters for other nodes. In this scenario, requirements for expanded coverage dictate the need for more subscribers – but the service provider can't solicit new subscribers until the coverage extends to the new subscribers' area. To resolve this, MeshNetworks and Radiant Networks supply 'seed nodes' mounted on telephone poles or streetlights for initial coverage and redundancy with the level of required seeding determined by specific business objectives [10], [12].

3. Military Perspective

Aside from efforts to tame mesh network technology for commercial deployment, the U.S. Government has spent significant time, money, and resources on the research, development, and field deployment of mesh networks for tactical military operations. With any mesh network deployment, the addition or deletion of network nodes can alter the dynamic network topology, emphasizing the need for efficient network organization, link scheduling, and routing to contend with varying distance and power ratios between links. A military environment, however, imposes additional complications by enforcing low probability of intercept and/or low probability of detection requirements, which in turn pose stringent power and transmission requirements on every network node [4].

Tactical military operations must also contend with varying degrees of mobility that occur within the military's echelon of four Divisions per Corp, four Brigades per Division, three Battalions per Brigade, four Companies per Battalion, and three Platoons per Company [13]. In this particular hierarchy, the often unpredictable nature of battle can dictate the need to merge and reconfigure sections of missing forces, disrupting the communication paths from node to node within Battalions, Companies, or other command structures. And while some engineers argue that alternatives to mesh networking exist to support communications in these battlefield conditions, others

highlight the mesh network capability for instantly configurable, decentralized, redundant, and survivable communications in frontline battle areas or during amphibious or airborne operations where a clustered, ad hoc network configuration might consist of people, planes, ships, and tanks. In this military environment, mesh networks must contend with the military's requirement for preservation of security, latency, reliability, intentional jamming, and recovery from failure [1], [4].

The Joint Tactical Information Distribution System (JTIDS) provides one example of a repeater-based, full mesh military network architecture that uses airborne relay to perform base station functions such as routing, switching, buffering multiple packet streams, and radio trunking. Developed for air-to-air and air-to-ground communications, JTIDS consists of up to 30 radio nets each sharing a communications channel on a time division multiple access (TDMA) scheme with most nodes in the network containing minimal hardware and processing power. In this configuration, the loss of any node within a radio net would have no negative impact on communications connectivity [1].

In another example, the Army's Communications Electronics Command oversees ITT Industries' development of the Soldier Level Integrated Communications Environment (SLICE). Designed for voice communications and troop mapping functions, SLICE represents the latest in military mesh network capabilities. Originally conceived as the DARPA Small Unit Operations Situational Awareness System, SLICE supports simultaneous networking of voice, video, and data transfer with a waveform and media access protocol that yields effective communications in urban canyons and dense jungle environments. In its present form, SLICE consists of a backpack-size computer with a headset display and built-in microphone. By 2005, ITT expects SLICE to shrink to the size of a PDA. With respect to SLICE, JTIDS, or any other military radio architecture, the theme of digitized battlefield communications describes the war fighter landscape with requirements for wearable, ruggedized personal computers capable of flawless performance under harsh conditions [14], [15, [16].

4. Final Thoughts

With low transmission power requirements and a multi-hop architecture, mesh networks increase the aggregate spectral capacity of existing nodes, providing greater bandwidth across the network. And since mesh networks transmit data over several smaller hops instead of spanning one large distance between hops, mesh network links preserve signal-to-noise ratios and decrease reliance on bandwidth-pinching forward error correction techniques [17]. In terms of scalability, mesh networks can accommodate hundreds or thousands of nodes with control of the wireless system distributed throughout the network, allowing intelligent nodes to communicate with one another without the expense or complication of having a central control point. Furthermore, these networks can be installed in a manner of days or weeks without the necessity of planning and site mapping for expensive cellular towers. As with other peer-to-peer router-based networks, mesh networks offer multiple redundant communications paths, allowing the network to automatically reroute messages in the event of an unexpected node failure. Thanks in part to standards efforts underway in the Internet Engineering Task Force (IETF) MANET Working Group, the design and standardization of algorithms for network organization, link scheduling, and routing will help facilitate the commercial acceptance of mesh network technology.

Despite their potential to provide a more sophisticated WLAN alternative, mesh networks must effectively address security issues with end-device and router introduction, user data integrity, device control and authentication, and network authentication. Aside from security issues, the RF-independent, self-forming, and self-healing characteristics these networks display come at the expense of complex and power intensive computer processing. Even in static environments with all nodes stationary, mesh network topologies remain dynamic due to variations in RF propagation and atmospheric attenuation. With mobile nodes, a mesh network's constantly shifting topology dictates the need for dynamic routing allocation, resource management, and quality of service management – all of which must be precisely choreographed to ensure optimum performance and reliability. Other skeptics contend that as ad hoc multi-hop networks grow, performance tends to deteriorate due in part to excessive traffic control overhead required to maintain quality of service along a path with multiple hops besieged by inconsistencies in routing and connectivity as nodes are added and dropped. Also, the network must handle multiple access and collision problems associated with the broadcast nature of RF communications. Regardless of these technical hurdles, researchers at Intel continue to push the research and development envelop in an effort to design a 100 Mbps mesh network where every network element (PC, PDA, mobile phone, etc.) could act as a data relay and link itself to all the devices in an intelligent network [10], [12], [17], [19].

With the ability to deploy a wide-spread coverage network without towers, mesh networks pose a viable alternative to traditional cellular architectures. Labeled as a potentially disruptive fourth-generation technology,

QDMA-based mesh networks aren't alone in their quest for the ultimate radio communications system capable of operating in unlicensed spectrum. Though technologically disparate from QDMA-based networks, ultra wideband (UWB) mesh networks present one alternative to MeshNetworks, Inc. proprietary QDMA-based software, thanks in part to recent FCC rulings approving limited usage of UWB devices. Several companies are championing the development of UWB networks, which promise data rates of 100 Mbps at very low power levels over a wide bandwidth from 1 to 10 GHz. By employing time-modulated digital pulses in lieu of continuous sine waves, mesh networks with UWB technology can send signals at very high rates in wireless communication environments that suffer from severe multipath, noise, and interference. Whether UWB mesh networks or QDMA-based mesh networks will prevail remains to be seen. Some analysts give the edge to UWB as an open standard, which is steadily gaining support in commercial and military markets. Either way, the continued development of mesh networks for military and commercial markets holds promise for a radical shift in the way we view the world of wireless communications [18], [20].

5. References

[1] "Alternative Architectures for Future Military Mobile Networks," Obtained April 7, 2003 from: www.rand.org/publications/MR/MR960/MR960.chap3.pdf

[2] Poor, Robert, "Wireless Mesh Networks," *Sensors [on-line]*, February 2003. www.sensorsmag.com/articles/0203/38/main.shtml

[3] Braunschweig, Carolina, "Wireless LANs Could Turn Into a Big Mesh," *Private Equity Week [on-line]*, February 3, 2002. www.ventureeconomics.com/vec/1031551158703.html

[4] "Project: Wireless Ad Hoc Networks," National Institute of Standards and Technology. Obtained April 8, 2003 from: http://w3.antd.nist.gov/wctg/manet/

[5] "QDMA and the 802.11b Radio Protocol Compared," *MeshNetworks: Technology, [on-line]*, April 9, 2003. www.meshnetworks.com/pages/technology/qdma_vs_80211.htm

[6] Blackwell, Gerry, "Mesh Networks: Disruptive Technology?" *802.11 Planet [on-line]*. Obtained April 8, 2003 from: www.80211-planet.com/columns/article.php/961951.

[7] Black, Uyless, *Computer Networks: Protocols, Standards, and Interfaces*, 2nd ed. New Jersey: Prentice Hall, 2003.

[8] Stroh, Steve, "MeshNetworks – From the Military Battlefield to the Battlefield of Modern Mobile Life," *Shorecliff Communications [on-line]*, Vol. 2, No. 2, February 2001. www.shorecliffcommunications.com/magazine/print_article.asp?vol=10&story=85

[9] Morrissey, Brian, "The Next 802.11 Revolution," *Internet News [on-line]*, June 13, 2002. www.internetnews.com/wireless/article.php/1365611

[10] Rubin, Izhak, and Patrick Vincent, "Topological Synthesis of Mobile Backbone Networks for Managing Ad Hoc Wireless Networks," Electrical Engineering Department, University of California Los Angeles, 2001.

[11] Krane, Jim, "Military Networks Trickling into Civilian Hands," *The Holland Sentinel [on-line]*, December 8, 2002. www.thehollandsentinel.net/stories/120802/bus_120802072.shtml

[12] Fowler, Tim, "Mesh Networks for Broadband Access," *IEE Review*, January 2001.

[13] Graff, Charles, *et. al.*, "Application of Mobile IP to Tactical Mobile Internetworking," *IEEE Magazine*, April 1998.

[14] "ITT Industries Awarded $44 Million to Develop Advanced Soldier Communications System," *PR Newswire [on-line]*, November 25, 2002. www.cnet.com/investor/news/newsitem/0-9900-1028-20696617-0.html

[15] "Mesh Networks Keep Soldiers in the Loop," *Associated Press [on-line]*, January 27, 2003. www.jsonline.com/bym/Tech/news/jan03/113806.asp

[16] Omatseye, Sam, "The Connected Soldier," *RCR Wireless News*, March 17, 2003.

[17] Krishnamurthy, Lakshman, *et. al.*, "Meeting the Demands of the Digital Home with High-Speed Multi-Hop Wireless Networks," *Intel Technology Journal [on-line]*, Vol. 6, Issue 4, November 15, 2002. http://developer.intel.com/technology/itj/index.htm

[18] Smith, Brad, "Smell the Coffee: Disruptive Technologies on the 2002 Horizon," *Wireless Internet Magazine*, January 7, 2002. www.wirelessinternetmag.com/news/020107/020107_opinion_brad.htm

[19] Ward, Mike, "Promise of Intelligent Networks," *BBC News [on-line]*, February 24, 2003. http://news.bbc.co.uk/2/hi/technology/2787953.stm

[20] Barr, Dale, "Ultra-Wideband Technology," Office of the Manager, National Communications System Technical Notes, Vol. 8, No. 1, February 2001.

Performance Analysis Of Dynamic Source Routing Using Expanding Ring Search For Ad-hoc Networks

V.Malathi * **Dr.A.M.Natarajan**** **S.Venkatachalam^**

**** Lecturer*** *****Professor*** ***^ Assistant Professor***

Department of CSE ***Department of ECE***

Kongu Engineering College, Perundurai, TN, India

E-mail: sv@kongu.ac.in

Abstract:

This paper presents a protocol for routing in ad hoc networks that uses dynamic source routing (DSR). DSR uses a route discovery mechanism to dynamically discover routes when needed. This is done by broadcasting a Route Discovery packet with a hop limit of one and if no reply is received for this packet then broadcast a packet with a hop limit of a predefined maximum value. This ends up in disturbing almost all nodes in the network with a considerable number of routing overhead packets. Expanding Ring Search for DSR gradually increases the hop limit in the route discovery packet resulting in a gradual search for the destination thereby reducing this routing overhead. The protocol was implemented using the ns-2 network simulator. In this paper we have provided a summary of our simulation results, comparing the routing overhead generated by the basic DSR route discovery mechanism and DSR with Expanding Ring Search mechanism.

Index Terms: Ad-hoc Network, Expanding Ring search

1.Introduction

Mobile users will want to communicate in situations in which no fixed infrastructure is available. Ex: Disaster recovery (flood, earthquake). In such situations a collection of mobile hosts with wireless network interfaces may form a temporary network without the aid of any established infrastructure or centralized administration. Currently there are two types of mobile wireless networks-

a.Infrastructured Networks.

b. Infrastructureless Networks**.**

This is commonly known as ad hoc network. An ad hoc network is a dynamically changing network of mobile nodes that communicate without the support of a fixed infrastructure. In such a network, each mobile node operates, as a host as well as a router. A key protocol in ad hoc networks is routing. Routing protocols used for ad hoc networks must deal with the typical limitations of these networks, which include- limited wireless transmission range, packet losses due to transmission errors, mobility induced route changes, mobility induced packet losses, battery constraints, potentially frequent network partitions, ease of snooping on wireless transmissions, low bandwidth.

Routing Protocols for Ad Hoc Networks:

Routing protocols for ad hoc networks are generally categorized as-

a. Table Driven Routing:

These protocols require each node to maintain one or more tables to store routing information and nodes propagate routing updates throughout the network in response to changes in the network topology.

Disadvantages: Frequent broadcasts of the routing table will degrade the throughput of channel access and increase the overhead as the population of mobile hosts increases.

b. Source-Initiated On-Demand Routing:

The source node initiates a route discovery process within the network only when it needs a route. Once a route has been established, it is maintained by a route maintenance procedure.

2.Dynamic Source Routing Protocol

The Dynamic Source Routing protocol (DSR) [1]-[3] is a simple and efficient routing

protocol designed specifically for use in multi-hop wireless adhoc networks of mobile nodes. Using DSR, the network is completely self-organizing and self-configuring, requiring no existing network infrastructure or administration. The use of source routing allows

- Nodes forwarding or overhearing packets to cache the routing information in them for their own future use.
- Packet routing to be trivially loop free.
- Avoids the need for up-to-date routing information in the intermediate nodes through which packets are forwarded.

Network nodes cooperate to forward packets for each other to allow communication over multiple "hops" between nodes not directly within wireless transmission range of one another. As nodes in the network move about or join or leave the network, and as wireless transmission conditions such as sources of interference change, all routing is automatically determined and maintained by the DSR routing protocol. Since the number or sequence of intermediate hops needed to reach any destination may change at any time, the resulting network topology may be rapidly changing. The DSR protocol allows nodes to dynamically discover a source across multiple network hops to any destination in the ad hoc network. Each data packet sent then carries in its header the complete ordered list of nodes through which the packet must pass, allowing packet routing to be loop free and avoiding the need for up-to-date routing information in the intermediate nodes through which the packet is forwarded. By including this source route in the header of each data packet, other nodes forwarding or overhearing any of these packets may also easily cache this routing information for future use.

The following assumptions [2] are made regarding the way computers are situated with respect to each other in an ad hoc network -

- All nodes wishing to communicate with other nodes within the ad hoc network are willing to participate fully in the protocols of the network.
- The diameter of the ad hoc network is assumed to be often small.
- Nodes within the ad hoc network may move at any time without notice, and may even move continuously, but we assume that the speed with which nodes move is moderate.
- Wireless communication ability between any pair of nodes may at times not work equally well in both directions

DSR Protocol Description:

The DSR protocol is composed of two mechanisms that work together to allow the discovery and maintenance of source routes in the ad hoc network:

a. DSR Route Discovery:

Route Discovery allows any host in the ad hoc network to dynamically discover a route to any other host in the ad hoc network, whether directly reachable within wireless transmission range or reachable through one or more intermediate network hops through other hosts. A host initiating a route discovery broadcasts a route discovery packet, which will be received by those hosts within wireless transmission range of it. The route request packet contains-

- The address of the original initiator and target node of the route request.
- A route record, in which is accumulated a record of the sequence of hops taken by the route request packet as it is propagated through the ad -hoc network during its route discovery.
- Each route request packet also contains a unique request id.

When any host receives a route request packet, it processes the request according to the following steps:

1. If this packet id is found in this host's list of recently seen requests, then discard the route request packet and do not process it further.
2. Otherwise, if this host's address is already listed in the route record in the request, then discard the route request packet and do not process it further.
3. Otherwise, if the target of the request matches this host's own address, then the route record in the packet contains the route by which the request reached this host from the initiator of the route request. Return a copy of this route in a route reply packet to the initiator.
4. Otherwise, append this host's own address to the route record in the route request packet, and re-broadcast the request.

The route request thus propagates through the ad hoc network until it reaches the target host, which then replies to the initiator. Only those hosts within wireless transmission range of the initiating host receive the original route request packet, and each of these hosts propagates the request if it is not the target and if the request does not appear to this host to be redundant. Discarding the request because the host's address is already listed in the route record guarantees that no single copy of the request can

propagate around a loop. Also discarding the request when the host has recently seen one with the same id removes later copies of the request that arrive at this host by a different route.

In order to return a route reply packet to the initiator of the route discovery, the target host must have a route to the initiator. If the target has an entry for this destination in its route cache, then it may send the route reply packet using this route in the same way as is used to send any other packet. Otherwise, the target may reverse the route in the route record from the route request packet, and use this route to send the route reply packet. This however requires the wireless network communication between each of these pair of hosts to work equally well in both directions, which may not be true in some environments or with some MAC-level protocols. An alternative approach is to piggyback the route reply packet on a route request targeted at the initiator of the route discovery to which it is replying.

All source routes learned by a node are kept in a route cache, which is used to further reduce the cost of route discovery. A node may learn of routes from virtually any packet the node forwards or overhears. When a node wishes to send a packet, it examines its own route cache and performs route discovery only if no suitable source route is found.

b. Route Maintenance:

This is the mechanism by which a source node is able to detect, while using a source route to a destination, if the network topology has changed such that it can no longer use this route. Route Maintenance is used only when a node is actually sending packets to a destination.

3. *Problem Formulation*

Existing Route Discovery Operation:

The initiator of the route request packet has the ability to specify in the route request packet, a "hop limit". This ability is used during route discovery as follows:

1.**Nonpropagating Route Request:** This route request packet has a hop limit of one and is the initial route discovery packet.

2. **Propagating Route Request:** If no route reply is received from the previous route request within a small timeout period, a new request is sent with a hop limit set to a predefined "maximum" value.

This procedure uses the hop limit on the route request packet to inexpensively check if the target is currently within wireless transmitter range of the initiator or if another host within range has a route cache entry for this target. Since the initial request is limited to one network hop, the timeout period before sending the propagating request can be quite small.

Disadvantages: Potentially every node in the network will be disturbed whenever a request packet is created.

4.Solution

Route Discovery Using Expanding Ring Search:

An expanding ring search can be implemented for the route discovery, in which the hop limit is gradually increased in subsequent retransmissions of the route request for this target.

Implementation:

1. A non-propagating route request (hop limit = 1) is broadcast first and the initiator waits until the timeout interval to receive a reply.
2. If no reply is received within the timeout interval the node initiates another route request with a hop limit of five.
3. For each route request initiated, if no route reply is received for it, the node increases the hop limit used on the previous attempt by five. This is done until the predefined maximum value of hop limit is reached. In actual use, it is expected that hosts communicate mostly with a small common subset of the available hosts (such as servers), which would reduce the number of route discoveries required.

The hop limit is implemented using the *Time-to-Live (TTL)* field in the IP header of the packet carrying the route request.

Table I show the constants used in the simulation of the ad hoc network using the DSR protocol with Expanding Ring Search.

Table I: Constants used in the Simulation

Parameter	Value
Non-propagating route request time out	100msec
Maximum route request period	10sec
Time to hold packets awaiting routes	30sec
Route request time out	500msec
Packet forwarding jitter	1-2msec

5. Performance Analysis and Results

Parameter for Analysis:

The main goal of this project was to measure the routing overhead of the DSR protocol using the Expanding Ring Search in comparison to the basic DSR route discovery mechanism that does not use an expanding ring search to discover routes.

Routing overhead:

It is the total number of routing packets transmitted during the simulation. For packets sent over multiple hops, each transmission (each hop) of the packet counts as one transmission. Routing overhead measures the scalability of the protocol, the degree to which it will function in congested or low-bandwidth environments. Protocols that send large number of routing packets: Consumes more battery power, Consumes bandwidth and may delay data packets in network interface transmission queues.

Simulation results:

Fig.1 shows the routing overhead (packets) generated by DSR without Expanding Ring Search and the routing overhead generated by DSR with Expanding Ring Search that was measured for 10 and 20 sources for pause time values ranging from 0 to 500 seconds using the ns-2 simulator [4], [5].

The simulation results (Fig.1) shows that the DSR using Expanding Ring Search mechanism for route discovery performs better than the basic DSR route discovery mechanism. The Fig.1 clearly shows a decrease in the number of routing overhead packets, by a margin of 1000 packets. This is because, the basic DSR route discovery mechanism results in the route discovery packet being propagated throughout the network when the nonpropagating route request is not replied, whereas the Expanding Ring Search limits the propagation of the route discovery packet by including a hop limit, which results in a gradual search for the destination node. The Expanding Ring Search performs extremely well at higher values of pause times, when node mobility is minimal and routes once discovered can be used for a longer time.

6. Conclusion

The previous works on DSR has shown that it delivers over 95% of data packets regardless of mobility rate and has the least overhead compared to other routing protocols used for ad hoc networks. In this paper we have provided the results of implementing the same DSR protocol, but using an Expanding Ring Search of route discovery that has further reduced the routing overhead.

References

1.David B. Johnson “Routing in ad hoc networks of mobile hosts”, Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications, December 1994.

2.David B. Johnson and David A. Maltz “Dynamic Source Routing in ad hoc wireless networks” , Mobile Computing, edited by Tomasz Imielinski and Hank Korth, pp. 153-181. Kluwer Academic Publishers, 1996.

3.David B. Johnson David A. Maltz and Yih-Chun Hu “The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)”, Internet-Draft, draft-ietf-manet-dsr-00.txt, February 2002. Work in progress.

4.Kevin Fall and Kannan Varadhan, editors. NS notes and documentation. “The Vint project”, UC Berkeley, LBL, USC/ISI, and Xerox PARC, November 1997. http://www-mash.cs.berkeley.edu/ns/.

5.The Network Simulator – ns-2. http://www.isi.edu/nsnam/ns/

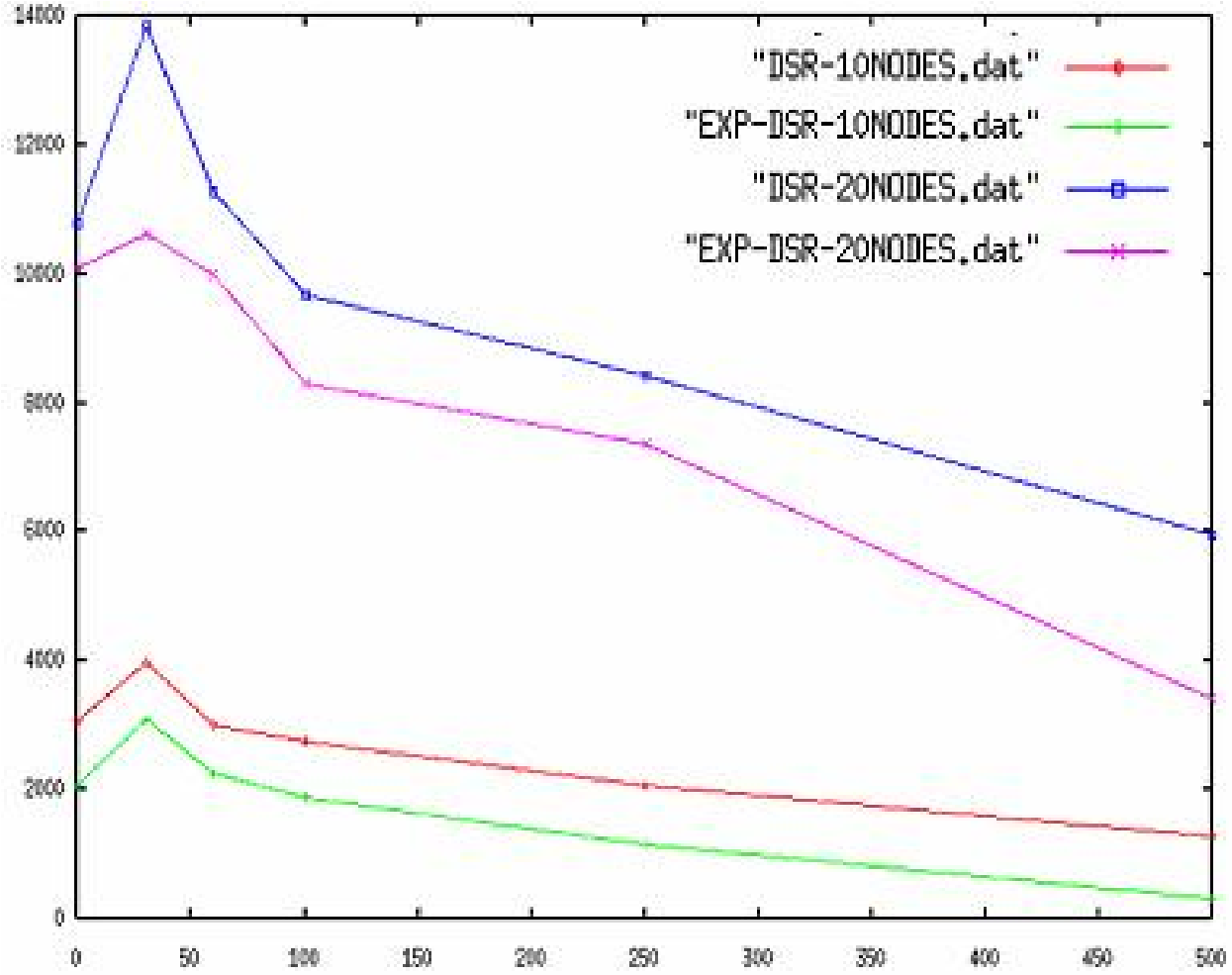


Fig.1 Routing overhead Vs Pause Time (10 and 20 nodes)

Label-based Multipath Routing (LMR) in Wireless Sensor Networks

Xiaobing Hou, David Tipper and Joseph Kabara
Department of Information Science & Telecommunications
University of Pittsburgh, Pittsburgh, PA 15260
{xiaobing, dtipper, jkabara}@mail.sis.pitt.edu

Abstract

Current wireless sensor network routing protocols are still struggling to find valid paths between source and destination, and multipath routing for fault tolerance is quite a new research area and studied insufficiently. The multipath routing techniques designed for ad hoc network do not apply to the sensor network due to the lack of global ID in sensor networks. In this paper, we propose a novel approach called Label-based Multipath Routing (LMR) using only localized information. LMR can efficiently find a disjoint or segmented backup path to provide protection to the working path.

I. Introduction

A sensor network consists of a large number of densely deployed sensor nodes. The position of the sensor nodes is not usually predetermined, as the network may be deployed in inaccessible terrains or disaster relief operations. Therefore, the topology may be random. Some of the application areas of sensor networks are medical care, military, and disaster recovery/relief. Due to the large size of such networks compared to the transmission range of individual devices, routing protocols are necessary for end-to-end communication. Compared to ad hoc networks, sensor networks have some unique feature and application requirements [1]. First, they normally have more nodes, higher density, more limited power supply and computational capacity than nodes in mobile ad hoc networks. Second, sensor networks can be characterized as data centric networks, where users are interested in querying an attribute of the phenomenon, rather than querying an individual node. Third, sensor networks are application-specific in that the requirements on the network change with the applications. As an example, some applications require delay sensitive transmission, e.g., fire monitoring, whereas others do not, e.g., tempature control in an office building. Fourth, adjacent nodes might have similar data; therefore, sensor networks should be able to aggregate similar data to reduce unnecessary transmissions and save energy. Last, assigning unique IDs may not be suitable in sensor networks because these networks are data centric – routing to and from a specific node is not required. In addition, the large number of nodes requires long IDs and must be minimized to conserve power.

Presumably, the sensor network application requires reliable data disseminations. Given the unreliable nature of the wireless channel and the high failure rate of individual sensors [1], multiple paths are required to maintain reliability. Specifically, with current single-path routing protocols, fault tolerance can not be provided because the continuity of end-to-end communication can not be maintained without routing protection and restoration techniques. Studies done for ad hoc networks may not be applicable to the energy constrained multipath routing in sensor networks. We propose a novel approach, namely Label-based Multipath Routing (LMR) for sensor networks.

LMR broadcasts a control message throughout the network for a possible alternate path. During the process, labels are assigned to the paths the message passes through. The label information is used for segmented backup path search if a disjoint path is not achievable. Our analysis and simulation show that this label information can reduce the routing overhead and backup path setup delay.

The remainder of this paper is organized as follows. We present a brief review of sensor network routing in section II. Various of multipath routing techniques in ad hoc networks and sensor networks are surveyed in section III. In section IV, Label-based Multipath Routing (LMR) is proposed. The performance evaluation of LMR is presented in section V. We then conclude our paper in section VI.

II. Sensor Network Routing

Basically, there are two types of sensor network routing protocols in the literature, cluster-based and flat. Cluster-based routing schemes divide the network into clusters and utilize a sleep mode to save energy and prolong the network lifetime. Flat routing schemes try to reduce the routing overhead directly by using localized information only.

In cluster-based routing protocols, all nodes are organized into clusters with one node selected to be cluster-head for each cluster. This cluster-head receives data packets from its members, aggregates them and forwards data to a data sink. Examples of cluster-based routing protocols are LEACH [2], TEEN [3], and APTEEN [4].

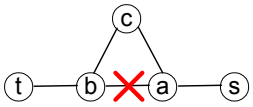


Fig. 1. Route repair

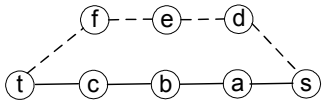


Fig. 2. Alternate routing

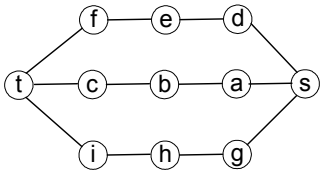


Fig. 3. Redundant routing

Low-Energy Adaptive Clustering Hierarchy (LEACH) [2] is designed for proactive networks, in which the nodes periodically switch on their sensors and transmitters, sense the environment and transmit the data. Nodes communicate with their cluster-heads directly and the randomized rotation of the cluster-heads is used to evenly distribute the energy load among the sensors. Threshold sensitive Energy Efficient sensor Network protocol (TEEN) [3] is designed for reactive networks, where the nodes react immediately to sudden changes in the environment. Nodes sense the environment continuously, but send the data to cluster-heads only when some predefined thresholds are reached. Adaptive Periodic Threshold sensitive Energy Efficient sensor Network protocol (APTEEN) protocol [4] combines the features of the above two protocols by modifying TEEN to make it send periodic data. The cluster-based routing protocols can arrange the sleep mode of each node to conserve energy at the cost of a high computational complexity and control overhead.

There are three types of flat routing schemes, namely, flooding, forwarding and data-centric based routing. Flooding is an old routing technique that can be used in sensor networks. In flooding, every node repeats the data once by broadcasting. It doesn't require costly topology maintenance and complex route discovery algorithms. But it has several deficiencies [1]:

- Implosion: duplicated messages are sent to the same node. A node with multiple neighbors may receive multiple copies of the same message.
- Overlap: if two sensors share the same observation region, both of them may sense the same stimuli at the same time. As a result, neighbor nodes receive duplicated messages.
- Resource blindness: flooding doesn't take into account the available resources, e.g. the remaining energy stored in the sensor node.

Forwarding schemes utilize local information to forward messages. Unlike the traditional routing protocols, forwarding doesn't maintain end-to-end routing information. Instead, intermediate nodes maintain only neighbor information. One example is the gossiping protocol [5], a node only forwards data to one randomly chosen neighbor, so it doesn't maintain any routing information or we can say it uses randomness to forward data. Best Effort Geographical Routing Protocol (BEGHR) [6] employs position information to forward data, and therefore requires GPS or other positioning service. Field based Optimal Forwarding employs cost field to forward data [7]. A cost field is the minimum cost from a node to the sink on the optimal path. The sink node is the destination of all of the data in the network.

In data-centric based routing, an interest message is disseminated to assign the sensing tasks to the sensor nodes and data aggregation is used to solve the implosion and overlap problems [1]. There are two types of data-centric based routing based on either the sink broadcasts the attribute for data, e.g. Directed Diffusion [8], or the sensor nodes broadcast an advertisement for the available data and wait for a request, e.g. Sensor Protocols for Information via Negotiation (SPIN) [9].

III. Related Work

In wireless ad hoc and sensor networks, nodes may be weakly connected or damaged, so that links may be asymmetric or broken for some period time. Battery-powered nodes may die out or go to sleep to save energy. A fault tolerant routing protocol must expect and overcome these problems.

Fault tolerant routing mechanisms for ad hoc networks include *route repair* [10]. After detecting a break in link *a*-*b*, node *a* can repair the route by finding another node *c* so that *a*-*b* can be replaced by *a*-*c*-*b* as shown in Fig. 1. If node *a* can not repair the route, it sneds an error message to the source(s). However, the repaired route may be suboptimal and after only a few repairs, the route may be very long and inefficient. Second, it may result in loops unless a source routing protocol (e.g. DSR [11] [12]) is used.

Alternate routing is a scheme where the source searches for a full alternate route after a failure [10]. As shown in Fig. 2, if the working route (solid line) is broken, the source receives the notification of route unavailability from the intermediate nodes and establishes a new route (dashed line). Although compared to the route repair the new path is optimal, establishing the path requires even more time and overhead. Basic AODV [13] and DSR [11] [12] protocols are using this scheme.

Redundant routing establishes alternate paths before the failure happens [10]. In Fig. 3, multiple paths are created between the source *s* and the destination *t*.

Compared with alternate routing, this approach is able to reduce the rerouting overhead since finding multiple paths at the same time (called *multipath routing* in literature) is cheaper than finding them one by one. Also the rerouting delay is smaller since the alternate path is available before the failure happens. But multiple paths having the same age may be similarly unreliable at the same time for a mobile network.

Preemptive routing proposed in [14] can improve the alternate routing by discovering an alternate path before a working path breaks. When a path is likely to be broken, a warning message is sent to the source indicating the likelihood of a disconnection. The source then initiates path discovery early, potentially avoiding the disconnection altogether. With alternate routing, when a path break occurs, the connectivity of the flow is interrupted and a hand-off delay is experienced by the packets that are ready to be sent. Preemptive routing switches a traffic flow to an alternative good path *before* a break, minimizing both the latency and jitter. Mechanisms used in cellular networks, such as the signal strength, can be used to trigger path discovery. Other warning criteria such as location/velocity and congestion can also be used as the preemptive trigger [14]. This scheme may increase the routing overhead of alternate routing protocols since some path discoveries are being carried out proactively.

Neighborhood aware Source Routing (NSR) [15] reduces the effort required to fix working routes by using alternate links available in the two-hop neighborhood of nodes. The two-hop neighborhood information is maintained by exchanging link-state information among neighboring nodes proactively. The repair delay can be alleviated since the alternate links are known before the failure occurs. Of course, extra overhead is required to maintain the proactive two-hop link state updates.

The techniques discussed above are examples of multipath routing. In each case, as is common for ad hoc networks, a global ID system is assumed so that every node has a unique ID and different paths can be easily recognized. However, this may not be the case in a sensor network. The great number of the nodes and the very low data rate make a global ID an unbearable overhead. Therefore, a multipath routing using localized information only is desirable for sensor networks. Two schemes have been proposed employing Directed Diffusion. *Disjoint Multipath* tries to find a disjoint path by randomly pick a neighbor to ask for a backup path to the sink. The request is otherwise rejected [16]. *Braided Multipath* follows the same idea but tries to form a braid around the working path. Both employ a brute force search technique, so that if a disjoint path can not be found, there is no information left for Braided Multipath to take advantage of. In the next section, we propose a new scheme to better utilize the localized information.

IV. Label-Based Multipath Routing (LMR)

Wireless sensor networks typically consist of a large number of nodes and work at a very low data rate. Therefore, assigning globally unique IDs may be extremely expensive in terms of bandwidth and power consumption. Additionally, it's not necessary because these networks are data-centric – routing to and from a specific node is not required. Similar to Disjoint Multipath and Braided Multipath [16], LMR is designed to use only the localized information to find disjoint paths or segments to protect the working path. With one flooding, LMR can either find disjoint alternate paths or several segments to protect the working path. The flooding overhead is reduced by the associated schemes used by the underlying routing protocols, e.g., location information or cached data in Directed Diffusion [8]. LMR can work with different data-centric routing protocols, e.g., SPIN and Directed Diffusion. For clarity, we introduce it over Directed Diffusion and we assume there is no mobility.

Multipath routing has been widely studied in wireline networks [17], and one of the difficulties, which also arises in wireless sensor networks, is *trap topology* [18]. In a trap topology, the working path may block all the possible disjoint paths. For example, the working path *s-a-b-c-t* in Fig. 4a has no disjoint backup path, although two disjoint paths exist between s and t. There are two solutions. One is to route the working and the back paths simultaneously. This is very difficult in a network without global ID. The second is to select multiple partially disjoint path segments to protect the working path and that's the one we are using in LMR.

A. Label

In Directed Diffusion [8], the sink node broadcasts the attributes for data, termed *interest*. The intermediate nodes create a *gradient* directed to the node from which the interest is received. After the source receives the interest, it sends an *exploratory* data message to each neighbor for whom it has a gradient at a low data rate as shown in Fig. 4b. After the sink starts receiving the exploratory data, it *reinforces* one particular neighbor by sending a *positive reinforcement* message in order to "draw down" the data at a higher data rate as shown in Fig. 4c. Similarly, a *negative reinforcement* message is used to remove a link from a path. Multiple paths may be reinforced. But this is different from the multipath routing we are studying. Firstly, there is no way we can guarantee that for each node failure we have an alternate path to protect it. Secondly, requiring every node receive data from two or more upstream nodes may result in the prohibitively high total overhead.

In LMR, after the nodes on the working path reinforce one of their links as the link to form a working path, they broadcast a *label message* to the rest of their neighbors. Both the reinforcement and label messages

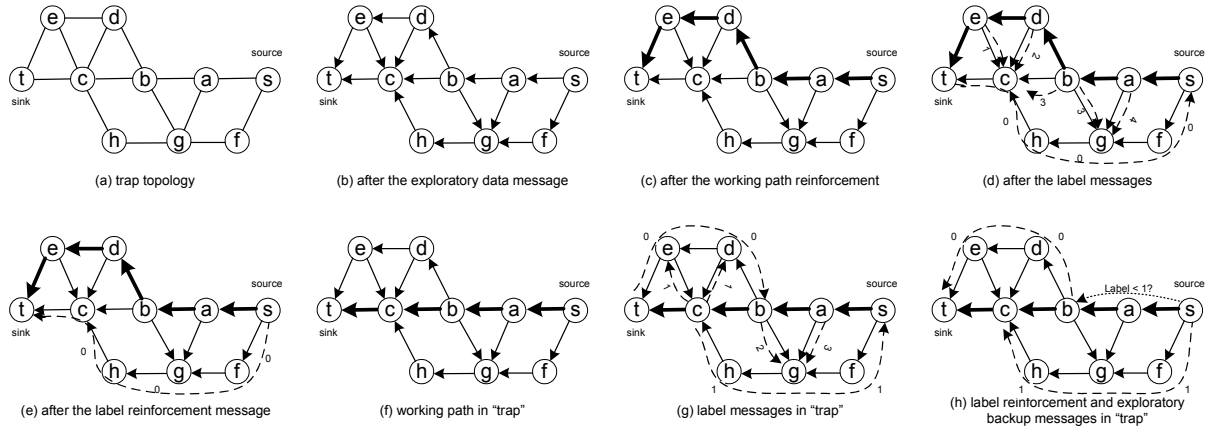


Fig. 4. Illustration of different aspects of LMR.

take an integer, termed *label*. The value of the label is increased by 1 by each working node which then broadcasts a new label message. Every working node should remember this value as its own *node label*. The label messages are forwarded towards the source along all the paths which the exploratory data messages pass through. A node receiving two or more label messages will forward the one with smaller label value only. The idea is to make the label message from the node closer to the sink go as far as possible so that the disjoint paths are possible to be found. The working nodes do not forward the label messages from any other nodes. Every node should remember all labels it has seen and the associated neighbors they are coming from. If a node receives multiple label messages with same label value from different neighbors, only the first one is recorded to find a shortest backup path. This process is shown in Fig. 4d.

B. Backoff algorithm

To avoid the excessive label message flooding, nodes must forward the smallest label message only. Therefore, a backoff algorithm is necessary to increase the probability that nodes receive the smallest label message before they start forwarding. A new label message should be delayed long enough so that a label message with a smaller label can go beyond this working node. Then the smaller label message will reach every node before the larger one if there are paths for them. However, if the delay is not long enough, the larger label message may reach the node first even with the delay. If the delay is long enough for a message to cover the entire network, we can guarantee that all nodes receive the smaller label first, but the setup delay of the backup paths may be long. So a tradeoff is necessary.

In LMR, if delay t_d is used, the working node with label w_i should broadcast a new label message after a backoff delay shown as follows,

$$T_i = w_i \times t_d \tag{1}$$

where, i=0, 1, ..., is the working node which has a new label message to broadcast and 0 is the sink. Another way to generate a new label is to make every working node increase the label by 1 no matter if it's necessary to broadcast a new label message or not. By this way, the node label $w_i = i$, and

$$T_i = i \times t_d \tag{2}$$

C. Label reinforcement

After the source receives a label message, it can immediately start label reinforcement process, since the backoff algorithm makes the smaller label message arrive first. A smaller label means we have a disjoint path segment to a working node closer to the sink. If a label 0 is received, that means we find a disjoint backup path. The source then sends a *label reinforcement* message to the node originating the label. The reinforcement continues with that node checking its memory to see which node this label comes from and then reinforcing that node. The process is a reversed reinforcement process of Directed Diffusion until the sink is reached, resulting in two disjoint paths (Fig. 4e). If the label received by the source is not 0, that means we may fall in a "trap" as shown in Fig. 4f. The label messages in this case are shown in Fig. 4g. Besides reinforcing a path segment, the source should send another message along the working path, called *backup exploratory* message. This message takes the label the source received. Any working node receiving this message whose node label is larger than this label either starts reinforcing a new backup path segment or forwards it. The new backup path segment should have a label smaller than the one the source received so that more working nodes can be protected. If the label of this new segment is not 0, a new backup exploratory message is initiated with the new label. The process is repeated until either a backup segment with label 0 is reinforced, or no new segments with smaller labels can be found, i.e. not all of the working nodes can be protected. This process is

shown in Fig. 4h.

After the backup path has been established, LMR may be repeated to find a third path. LMR can recursively find the *n* paths treating the first *n-1* paths as working paths.

V. Performance Evaluation

A. Complexity

To find the possible alternate paths, LMR incurs overhead, a flooded label message, and a label reinforce message and a backup exploratory message. We represent the sensor network as a graph $G = (N, E)$ with a diameter d in term of hops (i.e., the longest path between two nodes) and the average node degree is D. L_w represents the average length of a working path, and L_b the average length of a backup path. We consider the overhead of LMR based on two cases, local unicast, i.e. each node can only communicate with one of its neighbors at any time, and local multicast, i.e. each node can send a message to all of its neighbors at the same time. If the sensor network doesn't support local multicast but local unicast only, a node must send label messages to its neighbors individually and the number of the messages is in the order of D. If the network is not partitioned, almost all the nodes are involved except the source, therefore the label message overhead is $D \times |N|$. Since the label reinforcement message is disseminated along the backup path only, the total packet generated is L_b. Similarly, the backup exploratory message is sent along the working path only and the overhead is at most L_w. So the total overhead of LMR without multicast is $D \times |N| + L_b + L_w = D \times O(|N|)$. If the local multicast is supported, the label messages can be reduced by a factor of D. Therefore, the total overhead is $O(|N|)$ provided $N >> L_b + L_w$.

Disjoint Multipath and Braided Multipath try their neighbors one by one for the backup paths, so they can not benefit from local multicast and the complexity is same for two cases. or one failed try, two messages are involved, positive reinforcement and negative reinforcement [16]. Therefore the best case overhead is $L_b = O(d)$ and the worst case overhead is $2D \times O(|N|)$. It's worth noting that these two schemes are independent. If Disjoint Multipath fails, Braid Multipath must start over and double the overhead. LMR is efficient with local multicast and is reducing the average number of messages by $1/2D$.

Another measure of the performance of a multipath routing protocol is the delay to setup a backup path. We represent the link delay for transmission of one packet with t_p. LMR requires one round trip to set up a backup path and one of them may incur backoff delay. At the best case, a disjoint path can be found by the label message starting from the sink and no backoff delay is incurred, the total delay is $2L_b t_p = 2t_p \times O(d)$. At the worst case, all backoff delays occur at every hop and the total delay is $2t_p L_b + t_d L_w = (2t_p + t_d) \times O(d)$. Although Disjoint Multipath and Braided Multipath don't have a backoff delay, they may incur more link delay due to the brute force search. Obviously, their best case is half of LMR's. Their worst case may need to search every node in the network, so the total delay is $2t_p \times O(|N|)$. In a large network, since $O(|N|)/O(d) \gg t_d/t_p$, LMR can outperform the other two schemes in term of backup path setup delay. The above analysis is summarized in Table I.

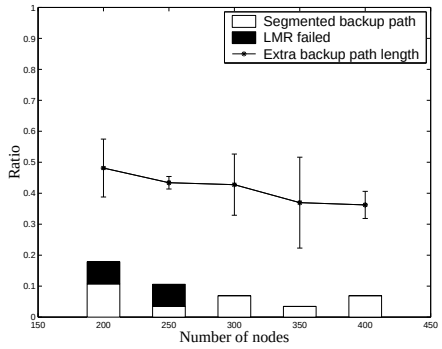


Fig. 5. Ratio of extra backup path length, segmented backup path and failure of LMR.

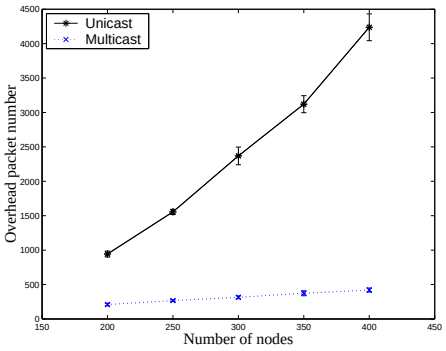


Fig. 6. Overhead of LMR.

B. Simulation

We utilized ns-2 network simulator [19], with CMU Monarch Project wireless and mobile ns-2 extensions, to study the characteristics of LMR. The distributed coordination function (DCF) of IEEE 802.11(b) for wireless LANs is used as the MAC layer. It uses Request-to-send (RTS) and Clear-to-send (CTS) messages and virtual carrier sensing for data transmission to reduce the impact of the hidden terminal problem. The radio model is similar to Lucent's WaveLAN, which is a shared media radio with a nominal bit rate of $2Mb/sec$ and a nominal radio range of 250 meters.

LMR is implemented over Directed Diffusion available in ns-2. The simulation results presented in this paper are based on scenarios randomly generated by CMU ns-2 extensions. We use 200 to 400 static nodes to study the density effects and nodes are randomly placed

TABLE I
COMPLEXITY COMPARISON
(B: BEST CASE, W: WORST CASE)

	LMR	Disjoint Multipath	Braided Multipath
Overhead(unicast)	$D \cdot O(\|N\|)$	B: $O(d)$ W: $2D \cdot O(\|N\|)$	B: $O(d)$ W: $2D \cdot O(\|N\|)$
Overhead(mcast)	$O(\|N\|)$	B: $O(d)$ W: $2D \cdot O(\|N\|)$	B: $O(d)$ W: $2D \cdot O(\|N\|)$
Setup delay	B: $2t_p \cdot O(d)$ W: $(2t_p + t_d) \cdot O(d)$	B: $t_p \cdot O(d)$ W: $2t_p \cdot O(\|N\|)$	B: $t_p \cdot O(d)$ W: $2t_p \cdot O(\|N\|)$

within a $2500m \times 2500m$ area. Besides these nodes, we put two nodes working as source and sink at the location (500, 1250) and (2000, 1250). Theoretically, at least 6 hops are needed for them to communicate. In a random topology generated by the above method, around 11 hops on average are used. For a given density, more than 30 topologies are used to get a 95% confidence interval.

Fig. 5 shows that, in most simulations, LMR can successfully find a backup bath, especially when the density is higher. In some cases, LMR cannot find a disjoint path and segmented paths are created. The ratio of extra backup path length is also shown in Fig. 5. Similar to the length of a single disjoint backup path, the length of a segmented backup path is the total hops on all the segments. This ratio is calculated as follows,

$$(L_b - L_w)/L_b \tag{3}$$

From the figure, we can see that, at lower densities, the backup paths are relatively longer since fewer alternate paths exist in the topologies and LMR has to pick up a longer one.

Fig. 6 shows the overhead of LMR in term of packets. Both local unicast and local multicast are simulated and the results match the analysis in the last subsection closely. The average node degree can be estimated by the following equation,

$$D = \pi(250)^2/(2500)^2 \times |N| - 1 \tag{4}$$

which is approximately the average number of nodes within the transmission range of a node. For example, with a 400 node network, the average degree D is about 11.6, which is close to the simulation result, i.e. 4430/420=10.6.

VI. Conclusions

In this paper, we present a review of current research on multipath routing in ad hoc networks and sensor networks. While a rich body of literature exists for ad hoc networks, few methods are appropriate for sensor networks due to the lack of global IDs. We proposed a novel approach called Label-based Multipath Routing (LMR), which employs localized information only. Analytical and simulation results show that LMR can find disjoint or segmented backup paths more efficiently compared to the Disjoint and Braided Multipath methods [16]. The label information in LMR can be used for segmented backup path search if a disjoint path is not found, reducing overhead and delay. Furthermore, LMR can take advantage of local multicast, significantly reducing the routing overhead.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, Aug. 2002.

[2] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *IEEE Hawaii International Conference on System Sciences*, 2000.

[3] A. Manjeshwar and D. P. Agrawal, "TEEN: A routing protocol for enhanced efficiency in wireless sensor networks," in *IEEE International Parallel Distributed Processing Symposium*, 2001.

[4] A. Manjeshwar and D. P. Agrawal, "APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks," in *IEEE International Parallel Distributed Processing Symposium*, 2002.

[5] S. Hedetniemi, S. Hedetniemi, and A. Liestman, "A survey of gossiping and broadcasting in communication networks," *Networks*, vol. 18, 1988.

[6] C. M. Okino and M. G. Corr, "Best effort adaptive routing in statistically accurate sensor networks neural networks," in *Proceeding Of IJCNN*, 2002.

[7] F. Ye, A. Chen, S. Lu, and L. Zhang, "A scalable solution to minimum cost forwarding in large sensor networks," in *Proceedings of the International Conference on Computer Communications and Networks*, 2001.

[8] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: A scalable and robust communication paradigm for sensor networks," in *Proceeding Of ACM MOBICOM*, 2000.

[9] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination protocol for wireless sensor networks," in *Proceeding Of ACM MOBICOM*, 1999.

[10] S. Chakrabarti and A. Mishra, "Qos issues in ad hoc wireless networks," in *IEEE Communications Magazine*, Feb. 2002.

[11] D. Johnson, "Routing in ad hoc networks of mobile hosts," in *Proc. IEEE Workshop on Mobile Computing Systems and Applications*, Dec. 1995.

[12] D. Johnson and D. Maltz, "Dynamic source routing in ad hoc wireless networks," 1996.

[13] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Proc. IEEE Workshop on Mobile Computing Systems and Applications*, 1999.

[14] T. Goff, N. B. Abu-ghazaleh, d. S. Phatak, and R. Kahvecioglu, "Preemptive routing in ad hoc networks," in *Proceeding Of ACM SIGMOBILE*, July 2001.

[15] M. Spohn and J. Garcia-Luna-Aceves, "Neighborhood aware source routing," in *Proceeding Of ACM MOBIHOC*, 2001.

[16] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin, "Highly-resilient, energy-efficient multipath routing in wireless sensor networks," in *ACM Mobile Computing and Communications Review*, vol. 5, no. 4, 2001.

[17] J. W. Suurballe, "Disjoint paths in a network, networks," no. 4, pp. 125–145, 1974.

[18] W. D. Grover, *Distributed Restoration of the Transport Network, Telecommunications Networks Magament in the 21st Century, Techniques, Stanstards, Technologies and Applications*. IEEE Press, 1994.

[19] K. Fall and K. Varadhan, *The ns Manual*. http://www-mash.cs.berkeley.edu/ns/, 2002.

Low Cost Broadband Wireless Access – Key Research Problems and Business Scenarios

Jan Markendahl, Jens Zander
Wireless@KTH, Royal Institute of Technology, Electrum 418, S-164 40 Stockholm-Kista, Sweden
email: jens.zander@wireless.kth.se

***Abstract*–The most prominent problem in providing anywhere, anytime wideband mobile access is the towering infrastructure cost as it is basically proportional to the bandwidth provided. In this paper, we provide a simple, initial, analysis of the various infrastructure cost factors. This analysis shows that, contrary to what one may expect, the infrastructure cost is not dominated by electronic equipment, but rather by other deployment related costs (towers, wiring, building, network connections) and maintenance costs. In the paper some novel architectural approaches for future wideband mobile access focusing on these dominant cost factors are described and the related key research issues are discussed.**

I. INTRODUCTION

Today's mobile communication systems are primarily designed to provide cost effective wide-area coverage for users with moderate bandwidth demands (voice and low rate data). In contrast to the traditional mobile systems, wireless local area networks (WLAN) are designed for higher bandwidth demands, while the area coverage is significantly limited. What the consumer of mobile telecommunication services of tomorrow will expect to receive, besides some vague notion of the "Wireless Internet", is not that clear. However, to obtain a widespread demand for wireless services, they have to be widely available, simple to purchase and access, and they must be affordable to large numbers of consumers. We expect that providing higher bandwidths that enable the use of truly new and innovative multimedia services is not sufficient: the users' communication cost per month must be similar or even lower than in second and third generation cellular systems.

Providing cost effective, affordable wireless bandwidth (almost) everywhere is one of the key success factors for future wireless systems. As the success of the Internet is largely attributed to the fact that it is virtually free of (incremental) charges (such as flat rate, independency of traffic volume), it is generally perceived that mobile data communications has to provide services in a similar way.

The challenge of providing flat rate, wireless access at the cost of fixed Internet access is indeed hard. The conventional cellular concept does not scale in bandwidth in an economical sense. The cellular systems include both the radio access network (RAN) and the core network (CN) components, which have different cost and capacity performance. The more decentralised WLANs have a slightly shifted RAN/CN performance relation due to short range and high access capacity. The cost of the wireless infrastructure (C_{system}) can (for a given allocated spectrum) basically be broken down into the following factors[1]:

$$\frac{C_{system}}{N_{user}} \approx \frac{c_{AP} N_{AP}}{N_{user}} \approx c' B_{user} A_{service} f(Q) \qquad (1)$$

Where

N_{AP} the number of access points (base stations)
N_{user} the number of users
B_{user} the average data rate of the users
$A_{service}$ the service area covered (volume indoors)
$f(Q)$ is a function of the required Quality of Service.

We here assume that cost of the core network part (wiring, switching nodes, servers and gateways etc.) is proportional to the number of access points and can thus be included in the factor c_{AP}. The cost factors generally depend rather weakly on the basic radio technology (e.g. the air interface) employed. This is mainly due to the fact that current modulation and signal processing technologies are quite advanced and so close to the Shannon limits that a radical improvement in signal processing capabilities alone will not significantly improve the performance. An abundance of spectrum may to some extent

Clearly, mobile telephony users have got used to large coverage areas with relatively good coverage and service availability (anytime, anywhere). This has been feasible since the bandwidth B has been low. Maintaining $A_{service}$, N_{user} and $f(Q)$ constant, it is clear that the cost is directly proportional to the user data rate, or equivalently, the cost per transmitted bit is the same. The classical telecommunication approach is to provide strict Quality–of–Service (QoS) guarantees at very low levels in the network hierarchy, corresponding to a high $f(Q)$. Sacrificing some Quality–of Service would thus be one way to significantly lower costs, but this has to be done in a way so that we can still provide interesting and desirable end-user services. Packet access techniques without absolute delay guarantees, e.g. the new HSDPA standardization effort in 3GPP is one

example of looking into more flexible resource utilization.

Other critical issues are related to financial and business aspects of the deployment of such a wideband wireless infrastructure. In traditional telecommunication world, a monopoly operator could make large investments in infrastructure, expecting to recover these in 20, 30 years. In a rapidly changing industry, this seems no longer to be an option. Today's vertically integrated market with operators "owning" the customers, providing most of the services and also owning and operating the network. Evolving network technology now enables that all functionality for customer care & billing as well as all network infrastructure may be offered on a disintegrated market by many different companies. Other players are service providers and Mobile Virtual Network Operators (MVNO), operators entirely without their own access network. We can also expect that the network access can be provided by specialized network providers and by private persons or enterprises. Mechanisms for sharing the cost (and risk!) for the deployment of new wireless infrastructure among these players are no longer obvious. One may well envisage infrastructure solutions that, at the aggregate level over a long time horizon, have the potential to provide reasonable costs for the end-user, but where markets and the cost-sharing mechanisms are not properly working. This would, in turn, prevent an effective take-up. Clearly, solutions have to be sought in the intersection of infrastructure business models, regulation and wireless technology.

The engineering challenge is to find technical designs that reduce costs significantly. We will see that the traditional approach, to provide cheaper and cheaper equipment is not alone going to solve the problem. To explore what kind of solutions that need to be sought, we will in the following take a more detailed look into the cost structure of wireless infrastructure in order to find the dominant cost. Based on this analysis, we will propose some alternative infrastructure concepts and architectures for future wireless systems, which focus on these dominant cost factors. Of particular interest are system concepts that allow simple and cheap deployment of infrastructure and concepts that allow efficient sharing of infrastructure resources.

II. Cost structure And Cost Drivers in Wireless Infrastructure systems

. Figure 2 shows an example of a cost structure for a typical cellular operator. The large grey sections, are related to marketing and administration and constitute 55% of the total cost, whereas the remainder, the colored slices are related to the network and infrastructure.

It can be noted that the annualized equipment (i.e. the depreciation cost for the equipment investment), e.g. base-station and switching equipment, is only a small part (15%) of the total infrastructure cost.

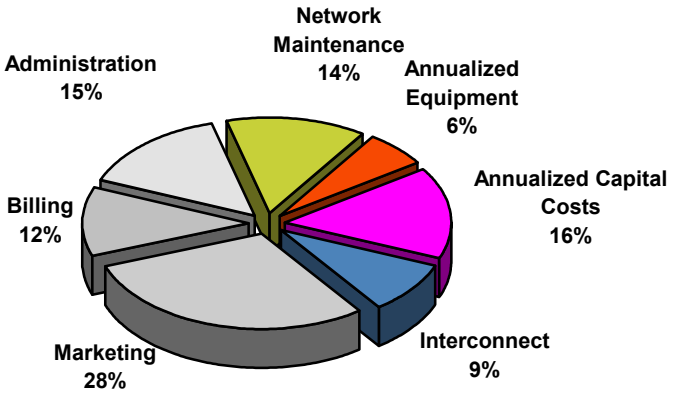


Figure 1 Cost structure for US mobile operators in the mid 90`s(FCC).

The other investment related costs are mainly due to site construction, antenna tower and similar items. The trend is that equipment related costs are becoming even smaller in future systems (with the possible exception of temporary "glitches" when new technology is introduced, before it has "gone down the learning curve"). In the rest of this paper we will mostly focus on the costs related to the network and infrastructure.

Operator costs for the network are often expressed as Capital expenditure (CAPEX) and Operating expenditure (OPEX). CAPEX are costs related to investment in equipment and the costs for the design and implementation of the network infrastructure; site acquisition, civil works, power, antenna system and transmission. The equipment includes the base stations (AP´s), the radio controllers, BSC´s and RNC´s, and all core network equipment. An example of CAPEX and of the relations between different types of implementation costs are shown in Figure 2.

In Figure 3 the relative costs from figure 2 are shown as a comparison. The same conclusion can be drawn from the both estimates; costs for base stations sites are much higher that costs for the base station equipment. We can also expect this difference to be larger over time due according to Moore´s law and the learning curve

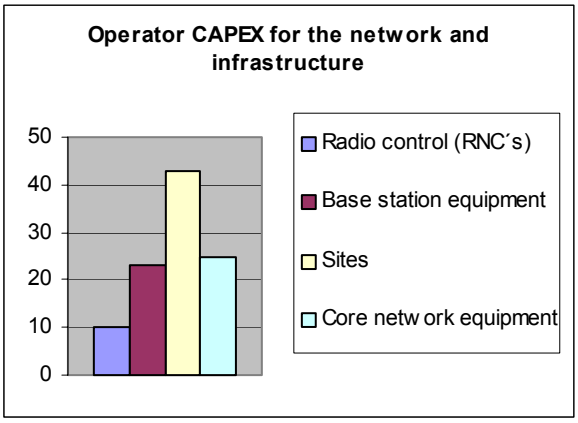


Figure 2 Example from 3G networks in Germany, estimates of cumulated CAPEX for the first 9 years [5]

.

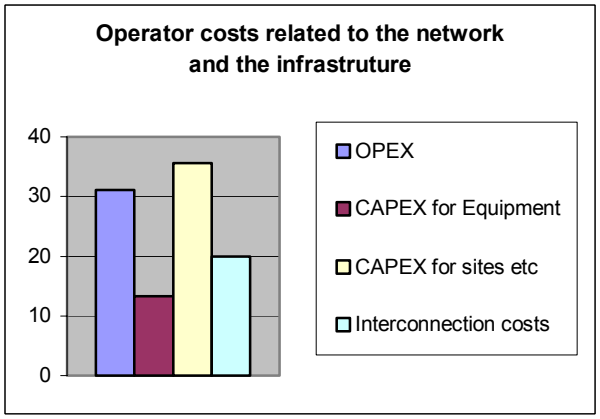


Figure 3 Network costs from figure 2 shown as comparison to figure

The OPEX are made up of three different kinds of costs

- **Customer driven,** i.e. costs to get at customer, terminal subsidies an dealer commissions
- **Revenue driven,** i.e costs to get a subscriber to use the services & network or costs related to the traffic generated; service development, marketing staff, sales promotion, interconnection.
- **Network driven**, costs associated with the operation of the network; transmission, site rentals, operation and maintenance.

Our current knowledge indicates that the dominating factors are related to customer acquisition, marketing, customer care and interconnection.

The fraction of OPEX to the overall cost is of course changing over time; in the "mature" phases the OPEX is the dominating factor. However, an estimate indicates that the network related OPEX are roughly 25 % of the total costs for the full life cycle.

Some general conclusions we may draw based on this simple analysis are:

A. Equipment cost is not the dominant part of the overall network CAPEX or OPEX.
B. The fraction of equipment cost to total infrastructure cost is likely to be reduced over time
C. Site construction & deployment costs and rents are the a major part of the network costs.
D. Network maintenance costs are a significant

III. Some Possible Research Directions

Based on the conclusions in the previous section, we can now identify some of the potential technology components of a cost effective solution. These are illustrated in Figure 4. Combining these technologies leads to a number of distinct research directions ("road maps") as indicated in the figure.

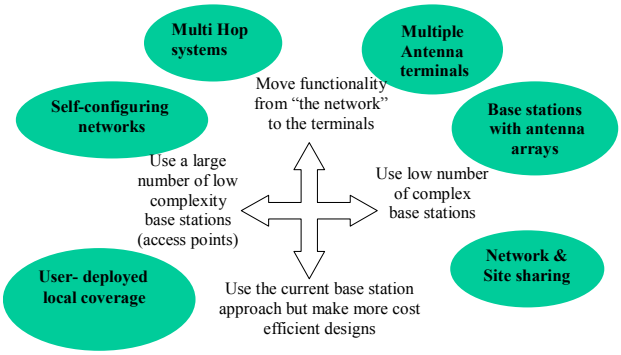


Figure 4 Possible solution components & research directions to provide low cost infrastructure .

As can be seen in the figure each of the directions correspond to approaching one over several of the cost drivers in our infrastructure cost model. Lowering the number of access-points by increasing their efficiency is a obvious approach. Using terminals to forward messages to other in a multi-hop mode, can also reduce the number of base station sites. Self-configuring technologies allows for reducing planning and deployment costs as well as O&M costs is also an interesting paths. Finally reusing and sharing infrastructure between operators and user is also reducing the number of required new sites. In the following we will present a number of candidate architectures illustrating how these technologies could be applied. Fig 5 shows how these candidates are related.

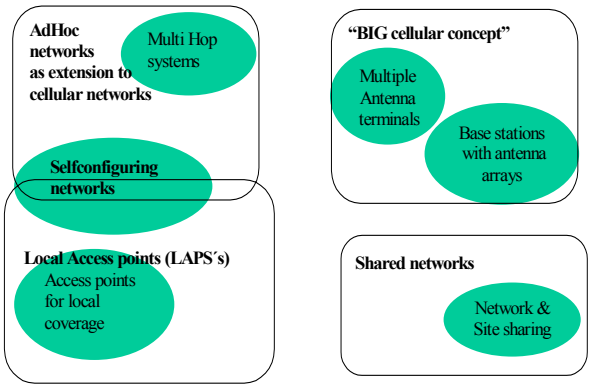


Figure 5 Proposed architectures and solutions for low cost infrastructure

IV. SOME CANDIDATE ARCHITECTURES AND KEY RESEARCH ISSUES FOR LOW COST INFRASTRUCTURES

"BIG cellular" concept: Highly efficient base stations

Since the infrastructure CAPEX is dominated by site related cost, and not by equipment, one obvious philosophy for decreasing the cost is to keep the number of base-stations low. This of course requires that the capabilities in terms of range and capacity of each base-station has to be high. Measures that can be taken increase the capability of each base-station is the use of high power, smart antenna arrays and high-gain antennas, high masts and use of low frequency bands. These measures make the base-station "big" in different aspects.

The working assumption is that a "traditional operator" owns and operates the network due to the need for large investments, centralized control and high competence to build and operate a "BIG cellular" network.

Among the key research issue for this concept we find

- Design and integration of efficient spatial RRM
- MIMO performance in hot spots, city centers
- Life cycle cost structure for the base stations
- Mobile terminal feasibility and cost
- Potential with "relayed approach", i.e MIMO technology for a more advanced transceiver acting as a relay port for a small local area network

Easily deployed Local Access Points (LAP´s)

In this strategy, we aim to bring down costs for network planning, deployment and maintenance by deployment of "many" low cost "base stations". These LAP´s should be possible to deploy "anywhere" where power and wire line connections are available and thus they require no specific "sites". Deployment & planning and O&M should be quick, simple and automatic, thus implying built-in functionality for auto-tuning and self-diagnostics.

Some key research issues that would need to be considered for this concept

- Cost efficient design of multi radio access LAP
- Principles for RRM and frequency & channel allocation of licensed bands used by LAP´s
- Design of LAP support functionality for "self deployment", configuration and auto-tuning
- Principles for support functionality for self-diagnostics, failure reporting and re-configuration
- Cost efficient design of multi radio access terminals with additional bands for voice communication using "local wireless access"

A related key issue is if it possible to provide interesting services over what can be seen as a combination of cellular systems and an organically growing infrastructure where individual users, companies and operator contribute to provide capacity and coverage - "the internet way".

Shared infrastructure

Present cellular solutions where (traditional) operators co-operate using national roaming or operate a common shared network. National roaming will be sufficient in rural areas where coverage is the main issue, in urban areas where high capacity is to be provided shared common network is more cost efficient. To reduce the number of sites, these should be utilized with high degree of efficiency; in the case of common shared networks one solution can be multi-operator base stations with all the frequency bands licensed to the operators.

The total number of sites is reduced, resulting in benefits due to higher efficiency in the usage of network equipment, transmission and sites. Costs for the site acquisition, operation & maintenance and interconnection, will be reduced due to fewer sites. On the other hand, planning need may be increased in areas where the operators will use most of the allocated spectrum.

This is clearly an evolutionary strategy and no changes will be needed for the mobile terminals

The current sharing solutions allows reductions of CAPEX and OPEX up to 40-50% [[3][5]. We believe there exists a large potential for further savings for the operators, however today no real incentive for the vendors exits today.

Among the key research issues we find

- Efficiency in deploying and operating the shared network (how much can be gained?)
- Fair sharing of resources (RRM)

- Pooling of resources from different operators
- Design of multi-operator base station
- On line tracking and monitoring of resources as input to traffic statistics, billing and planning.
- Principles for generalized roaming where many network providers contribute to the coverage

AdHoc networks extensions to cellular networks

To extend the coverage of the present cellular systems without adding more base stations one option is to use self- organizing adhoc networks based on terminals and /or repeaters with multihop, routing and buffering capability. The total number of sites is reduced or maintained. No planning will be needed since the network is self-organizing. More functionality will be added in the terminals such as network control and routing and buffering, also the physical design itself must allow for more memory and power consumption.

The key question is whether it is at all possible to provide interesting QoS guarantees in ad-hoc systems with little or no control over system resources and where propagation conditions, user location and radio interference may be unknown or hard to predict quantities. Resource management in the wider sense would be a critical issue.

V. BUSINESS SCENARIOS

Why business scenarios?

In order to evaluate the cost–performance characteristics of the proposed candidate solutions a wide range of use cases, user needs, business models and deployment strategies have to be considered. The candidate architectures target different sets of use cases and requirements, i.e. all solutions are not applicable everywhere.

Below some scenarios are described to highlight where the main benefits of a specific solutions can be expected. The presented scenarios are neither "user centric" nor "operator centric", they can be characterized by being more "deployment centric" in order to illustrate and highlight

- the need and use of new business models
- the different sets of requirements and working assumptions needed to evaluate the candidates.

For the "proof of concept" phase in the research, the scenarios will provide a common set of system and user requirements.

The proposed candidate architectures will be mapped onto an overall "market space " with local and wide area on one axis and "type of access provider " on the other axis.

"BIG cellular" concept

The "BIG cellular" solution targets requirements and scenarios with wide area coverage and high capacity, i.e. high performance cellular systems.

The scenarios of interest include both the traditional mobile operator business model and new business models where user owned equipment may contribute to the "infrastructure" e.g. by moving gateways in cars, buses and trains.

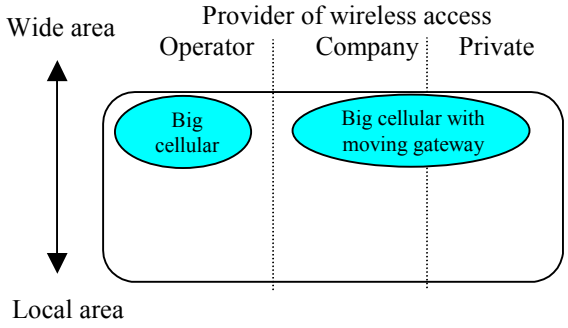


Figure 6 Big Cellular Scenario

Easily deployed Local Access Points (LAP´s)

The LAP´s are owned by the house or facility owner or by a specific local access provider. LAP´s used in domestic and corporate environments are intended for "own" users, i.e family and employees of the own company. Access for visitors can be an optional feature.

When maintenance is needed this is most likely provided by specific service companies.

For this local area (mainly indoor) candidate solution, where "other" market players than mobile operators (private persons, companies, facility owners, hotels, shopping malls) own and /or operate the wireless access infrastructure, we will consider two different business scenarios:

- Privately owned access networks mainly intended for the own users, but public access is possible. The main drivers are cost reductions and service performance for own users, i.e. not the possible revenues for the public access. An important part of the scenario is the re-use of existing fixed infrastructure and (in the case of companies) also "re-use" of staff for operation and maintenance.

- Local access providers for public use, i.e. no "own" users" exist. The main driver is to make money on the access itself and possibly to support the "core business", e.g fast food or coffee shops.

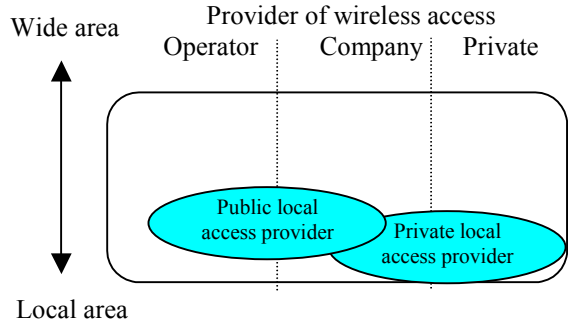


Figure 7 Local access provisioning scenario

These business scenarios may include regulatory issues as "local licenses" of spectrum and franchising solutions, e.g. to operate the network in area X "on behalf of " operator Y

For the first scenario we also have to consider the types of agreements the customer most likely will have with a phone company, a mobile operator or an ISP with a "common" subscription and terms & conditions for the usage of coverage & capacity offered by the owner of the LAP.

AdHoc networks extensions to cellular networks

One set of scenarios for adhoc networks includes fast rollout, rapid deployment and/or intermediate solutions for capacity & coverage expansion.

One driver is the possibility to provide customers with some degree of access in the near future, compared to the case where full quality access is provided when the full network rollout or expansion is finalized.

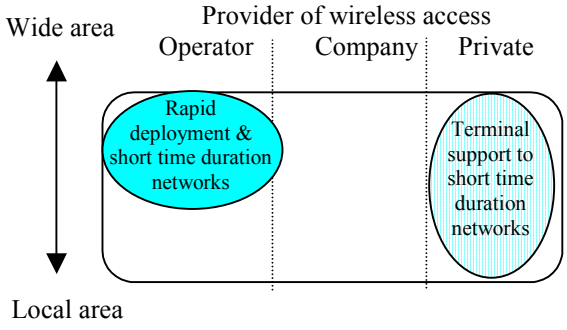


Figure 8 Ad-hoc/Rapid deployment scenario

This is believed to be important for market entrants (greenfield operators) in order to acquire customers, to get some revenue and in order decrease the number of customers for competitors.

Other scenarios include self-organizing networks to handle hot spots with short duration (e.g. traffic jam). One driver for this kind of solution is to cut traffic peaks without any need for "over-dimensioning" of the network.

Shared networks and use of common resources

One scenario focus on co-operation between operators in order to save costs and to make more efficient use of "all" available resources in an area including strategies as co-location of sites, pooling of telecom equipment and/or frequency bands. In this scenario cost reduction for OPEX and CAPEX is the main driver.

Another scenario, which is similar to the previous one with terminals forming adhoc networks, can be identified in areas with sparse infrastructure. Here co-operation between different "competing" network providers, including the users "belonging" to different operators, may be required in order to be able to provide ANY access at all in an area. Coverage and capacity can be offered at lower "total" cost and/or more early compared with parallel full capacity and coverage networks.

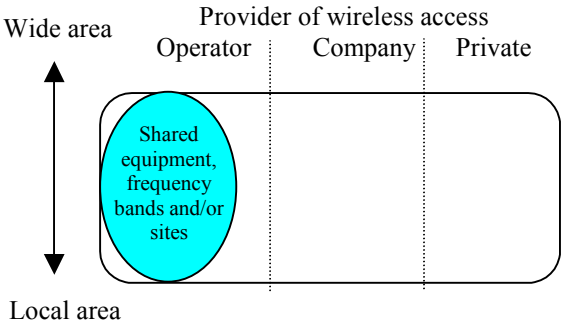


Figure 9 Shared network scenario

VI. DISCUSSION

In this paper we have identified the infrastructure cost as one of the main barriers en-route to pervasive wireless mobile access. We provided a brief analysis of the key cost factors in the wireless infrastructure and concluded that a scaled-up version of the traditional cellular concept is not by itself a viable solution to any-time anywhere broadband wireless access. Our analysis points at that the cost of equipment is like to be only a small fraction of the total infrastructure cost, whereas the bottlenecks lie mainly in the planning, deployment and maintenance of the infrastructure. Some new architectural concepts targeting these bottleneck costs were given. Most of these concepts in themselves provide non-trivial technical bottlenecks that could provide interesting new directions for engineering research. It is by no means obvious which of these

concepts (if any) is providing the most effective solution to our challenge. It is on the contrary likely that good solutions can be found in combining some of the features of the archetypical system designs outlined above.

VIII. ACKNOWLEDGEMENT

The contributions from our partners and colleagues in the Low Cost Infrastructure project, in particular Dr. Tim Giles, MSc Klas Johansson and Dr. Per Zetterberg at the Royal Institute of Technology, Dr. Bertil Thorngren and Econ Lic. Jonas Lind at the Stockholm School of Economics, Dr. Göran Malmgren at Ericsson Research, Dr. Jan Nilsson at the Swedish Defense Research Institute and Dr. Peter Karlsson at TeliaSonera AB, are gratefully acknowledged. The Low-Cost Infrastructure is part of the Affordable Wireless Services and Infrastructures program, supported by the Swedish Strategic Research Foundation.

IX. REFERENCES

[1] Zander, J, "On the cost structure of Future Wireless networks", IEEE Veh Tech. Conf, VTC97, Phoenix, AZ, May 1997

[2] Giles et al, "Cost Drivers and Deployment Scenarios for Future Broadband Wireless Networks – Key research problems and directions for research", to appear at VTC04, Milano, May 2004.

[3] Christensen, Clayton M. "The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail", Harvard Business School Press 1997

[4] Zander, J., "Affordable Multiservice Wireless Networks- Research Challenges For The Next Decade". PIMRC 2002, Lisbon, Sept 2002.

[5] Ericsson, "*White Paper - Shared Networks*", See Ericsson's Internet Home Page: www.ericsson.com.

[6] Siemens, "*3G Infrastructure Sharing - The Siemens Perspective*",see www.siemens.com.

[7] Oftel, Office of Telecommunications "Review of the charge control on calls to mobiles", 26 September 2001, http://www.oftel.gov.uk/publications/mobile/ctm0901.pdf

[8] TONIC (TechnO-ecoNomICs of IP optimised networks and services) is a project within the IST Programme (Information Society Technologies). http://www-nrc.nokia.com/tonic/

[9] Bria et al, "Wireless Foresight, wireless scenarios of the mobile world in 2015", Wiley, 2003