Spectral Efficiency Considerations for Packet Radio

Phil Karn, KA9Q

ABSTRACT

Radio spectrum is a lot like land; with the possible exception of the Dutch, nobody's making any more of it. So the enormous growth in demand for spectrum means that existing users, especially radio amateurs, either have to find ways to make do with less, be displaced by new users considered more worthy by regulatory agencies, or both.

This paper qualitatively discusses the spectral efficiency of packet radio from several angles, ranging from antenna design, RF modulation and channel access methods to network protocols, routing algorithms and data encoding methods. Maximizing the useful carrying capacity of a spectrum assignment requires a comprehensive look at all of these, factors and more. Digital techniques finally make it possible to exploit these gains, but so far amateur radio has been very slow in adopting them. I hope that this paper will stimulate some work in this direction.

1. Measuring Spectral Efficiency

One seemingly basic problem with spectral efficiency is this: How do you measure it? For example, in the debate over the codeless license, CW was frequently asserted to be more efficient than all other operating modes because it (usually) uses less bandwidth. In effect, the units for measuring spectral efficiency were claimed to be l/Hz; the narrower the signal, the better.

But this is a naive measure of efficiency because it doesn't take into account the *rate* at which information is being transmitted. A better measure is the ratio of the data rate to the occupied bandwidth; this has units of "bits per second per Hz of bandwidth". Modems are frequently evaluated in this way; indeed, the FCC sets minimum requirements for this figure for commercial digital microwave systems.

Unfortunately, even this figure does not give the complete picture. Very few (if any) individual radio transmitters have exclusive, worldwide use of their channels. Somewhere else that same spectrum is almost certainly in use by other transmitters, and hopefully they are all far enough away from each other to avoid mutual interference. This is commonly known as "geographic spectrum reuse". A combination of directional antennas and/or physical separation (terrain blocking or propagation losses) isolates the transmitters in different geographic areas to avoid harmful interference between them.

So a better measure of spectrum efficiency would have units of "bits per second per Hz of occupied bandwidth per square kilometer". This is the total useful data rate, summed across all of the transmitters in a given geographic area that are sharing a given piece of spectrum.

Yet another factor that we should take into account is the distance involved Just as the work performed by a freight company is the product of cargo weight times the distance it is moved, the measure of useful "work" performed by a data network should be bits times distance. This prevents any apples-and-oranges comparisons of spectral efficiency between wide- and local-area networks. So our final measure of network spectral efficiency has units of "bits per second times distance moved, per Hz of occupied bandwidth, per square kilometer of geographic area":

$$E = \frac{rD}{BA}$$

where:

E = network spectrum efficiency figure of merit

r = total network traffic capacity in bits/sec

D = average distance between source and destination nodes in km

B = total RF bandwidth allocated to the network in Hz

A = geographical area occupied by the RF allocation in square km

The value *E* is the figure that we should maximize.

2. Directional Antennas

One effective way to increase the carrying capacity of spectrum is by using highly directional antennas, especially on the higher UHF and microwave bands. Glenn Elmore, N6GN, has made this point quite eloquently. [Elmore90] I completely agree with Glenn that amateurs should deploy microwave links with directional antennas whenever and wherever they are practical. Because microwave antenna patterns can be made so narrow, the design of the associated modem hardly matters (as long as it works). Even with a "low efficiency" modem (in a *bitslsec /Hz* sense), the resulting system spectral efficiency is still far larger than on VHF with omnidirectional antennas because so little geographic area is covered by each microwave beam. This allows many different links to share the same spectrum in very close proximity without interference.

However, there are still many situations where point-to-point microwave links are not yet practical, such as portable and mobile stations (particularly in emergency situations), and where line-of-sight microwave propagation is obstructed (e.g., by trees and hills). So VHF and UHF packet radio with conventional low-gain antennas will still be with us for quite some time, and the resulting interference problems must still be dealt with.

3. Interference Limited Systems

This brings up a vitally important issue in the calculation of spectral efficiency that is only now getting the attention it deserves. The single ***most important factor in the efficient use of spectrum is the minimum geographic spacing required to avoid harmful co-channel interference between transmitters.*** The closer the allowable co-channel transmitter spacing, the more traffic the channel can support in a given geographic region.

When spectrum is "reused" in this manner, it is ***not*** necessary that there be no interference (i.e., that any co-channel interference be well below the noise floor of the receivers). It is ***only*** necessary that a desired signal be sufficiently stronger than the undesired signals at a given receiver so that the demodulator can work properly even in the presence of this weak interference. This is referred to as operating in an "interference limited" (as opposed to "noise limited") environment. [1]

The interference rejection capability (aka "capture effect") of the modulation method in use therefore becomes a prime factor in the efficiency calculations. However, low (good) "capture ratios" are inherently associated with wide band modulation methods, while narrow band modulation schemes are inherently much more sensitive to interference. Yet low capture ratios are so vital to spectral efficiency that they almost always "pay back" far more than they cost in extra bandwidth by allowing much closer co-channel transmitter spacing. This leads to the somewhat paradoxical fact that by going to a wider (and seemingly less efficient) modulation method, overall spectrum efficiency can often be greatly increased! [Costas59] This is why, for example, current cellular telephone systems use FM rather than SSB; FM has a capture effect while SSB does not, so using FM more readily allows the reuse of frequencies in other nearby cells. The bottom line is that a FM cellular radio system is more spectrally efficient than one using SSB, despite the much wider FM channel. [2] [Lee89] gives this approximate for-

---

[1] *Unfortunately, many* **amateur repeater owners insist on a very wide protection area for their frequency assignment to preclude the accidental triggering of their systems by the weak, distant users of other repeaters sharing the same assignment. They refuse to implement tone-coded squelch (PL) even when it could totally solve the problem because of the advantage given to local users by the FM capture effect. As we will see, this attitude is *extremely* wasteful of spectrum.**

[2] **It is most unfortunate that the FCC did not understand this paradoxical connection between modulation bandwidth and spectrum efficiency when they were convinced to reallocate 220-222 MHz to the Land Mobile Service for use with a supposedly more efficient modulation method, SSB.**

mula for the frequency reuse factor as a function of the required carrier-to-interference ratio in a hexagonal cellular radio system that uses omnidirectional antennas:

$$K=\frac{\left[6\frac{C}{I}\right]^{\frac{2}{\gamma}}}{3}$$

where
K = frequency reuse factor.
$\frac{C}{I}$ = required carrier-to-interference ratio (expressed as a power ratio, not dB)
γ = propagation slope factor, 2 for free space (inverse square), 4 for a typical terrestrial mobile environment

A given channel. can be used in only $1/K$ of the cells. E.g., if $\frac{C}{I}$ is 18 dB (as it is in a FM cellular radio system), then $K=7$ and a given channel can be used in only 1 of every 7 cells. If the required $\frac{C}{I}$ ratio can be decreased to 10 dB, then K would be less than 3; this would more than double the number of transmitters that could share each channel. The next generation of cellular telephony will substantially improve on FM's capacity by going digital, where by the proper choice of modulation method and with forward error correction coding (FEC), the capture ratio can be as low as 7 dB. The most promising scheme, not coincidentally, also has the widest signal bandwidth: CDMA (Code Division Multiple Access, more commonly known as "spread spectrum").[3] [Gilhousen91]

[3] The big win of spread spectrum for mobile communications is its ability to handle multipath fading. In analog FM, 8 dB of system margin is required to account for fading, over and above the 10 dB C/I ratio required on a nonfading channel. But spread spectrum allows the separation of multipath components, avoiding the rapid "mobile flutter" so characteristic of multipath fading in narrow band FM. Automatic power control easily compensates for the slow propagation variations that remain. This avoids the need for a big fading margin and allows co-channel transmitters to be much more closely spaced. In a network of fixed stations, multipath fading is not as much of a problem, so the important factor in system capacity calculations is just the C/I ratio required by the modulation (and coding method, if any). In both the fixed and mobile cases, the extra bandwidth required by adding FEC usually more than pays for itself in the closer co-channel transmitter spacing it allows.

4. **Power Control and Routing**

Simply using RF modems capable of good capture ratios isn't enough, however; automatic transmitter power control is also required to take full advantage of them. With automatic power control, each transmitter ensures, on a continuous basis, that a sufficient signal-to noise ratio (actually $\frac{E_b}{N_0}$ ratio) exists at its intended receiver *and no* more. Running more power than necessary to yield good performance is like buying higher octane gasoline than your car needs -- *it* doesn't work any better, and you only squander money and natural resources. The whole purpose of designing modems with good capture ratios is to allow transmitters sharing the same channel to be placed more closely together, and this is not possible unless each transmitter uses only the minimum power required to reach any given receiver.

The transmitter power required to reach a given receiver in a network of packet ***stations*** can vary widely depending on the distance between them, the presence of obstacles, nonideal antenna patterns, etc. So an interesting question appears: given the choice of relaying a packet to one of two intermediate stations between the sender and the destination, one of which is close and the other farther away (but closer to the destination), which one should be chosen?

The answer? It depends. If the packet *must* be delivered with the absolute minimum delay, then sending the packet to the relay station that is farther away (and that much closer to the destination) will clearly get il. there faster, since each relay hop takes time. But the additional transmitter power required to reach the station farther away means that a larger geographical area must be blanketed by the transmission, thus denying the simultaneous reuse of the channel to that many more stations. So if maximizing total network capacity is the goal, then the routing algorithm must minimize something other than simply the number of hops taken (or even the total distance traveled) to reach a destination.

Dave Mills, W3HCF has reported on a study of this problem done for the DARPA Multiple Satellite System (MSS). [Mills87] The study concluded that the proper metric to be minimized by an MSS routing algorithm is the sum of the squares of the distances between the nodes (relay satellites). Because of the inverse-square

receiver signal levels, e.g., with an A/D conlaw in free-space radio propagation, this effectively minimizes the total RF energy, summed over all of the transmitters involved, needed to relay a bit of information to its destination. This is true even when more nodes are used than if the routing algorithm simply picked the least number of hops to the destination.

In a terrestrial store-and-forward network, propagation losses usually increase with distance much faster than the square of the distance because of obstructions, scattering and multipath; the fourth power of the distance is generally used in analysis. But if each node measures the actual transmitter power required to reach a given destination and reports that as its routing metric for that link, then the propagation effects are automatically taken into account when the routing algorithm minimizes the sum of the transmitter powers used in reaching a given destination.[4]

5. Channel Access Methods

The algorithms that determine when a station transmits are another important factor in a network's overall spectral efficiency. Because of the need to operate in an interference-limited environment, carrier-sense multiple access (CSMA) schemes won't work very well if they always inhibit transmission whenever a signal is heard on channel, no matter how far away that other transmitter may be. Such systems are analogous to the repeater operator who refuses to use PL and still complains about remote triggering of his repeater, as mentioned in an earlier footnote. Schemes that rely on receiver feedback (as opposed to channel sensing at the transmitter) to avoid collisions would seem to be the only practical approach to this problem. See [Karn90] for a discussion of one possible approach.

[4] This tradeoff between delay and network efficiency suggests an interesting use for the long-ignored type-of-service (TOS) bits in the IP header. By default, packets would be routed using the minimum-total-power criteria, but IP datagrams that have the "low delay" bit set would use a different set of routing criteria that minimize delay at the expense of network capacity. This could be quite useful for emergency or priority traffic.

6. Protocols

From the point of view of the packet subnetwork designer, the upper layer protocols used are irrelevant; they are simply part of the user data that is to be moved. However, from the user's point of view, everything but his data is overhead. Therefore the cost of these protocols could be considered as part of the overall network spectral efficiency equation.

When properly implemented and tuned, the overhead taken by the protocols used in a computer network is usually a second-order factor in the overall efficiency of the system. Even the overhead of a "heavy" protocol like TCP/IP is easily minimized by using sufficiently large data packet sizes or by compressing headers [Jacobson90]. But this applies only when the protocols are used as intended, e.g., providing reliable point-to-point transfers with TCP. Unfortunately, it is a common practice to use multiple point-to-point protocol connections to emulate a broadcast or multicast service; the data is sent N times to N receivers, even when omnidirectional antennas are being used and the receivers could easily have shared a common, single transmission. The biggest offender in this regard is the *DX Cluster,* which *in this* author's opinion comes close to beiig a criminal abuse of the **AX.25** protocol. Another is the common practice of multiple users individually re-reading the same public bulletin from a BBS when the bulletin could have been broadcast to everyone at once. Fortunately, protocols designed specifically for the efficient broadcast of digital information have been designed and are now being deployed. [Price90] Given that much of the information carried by the amateur packet radio network is of general interest, such protocols should *significantly* enhance the effective capacity of the network. In our network efficiency equation, a broadcast protocol in use by N receivers would effectively multiply overall efficiency by approximately N.

7. Compression

Another higher-level issue in spectrum efficiency is the use of data compression. A network doesn't care about the values of the bits it carries; it "costs" just as much to send a million "0" bits as a million-bit text document, even though the useful information content of the former is probably quite a bit less. Users should therefore try to use the network's capacity in the most efficient way possible by compressing their data before transmission.

Data compression has been well studied and is widely used in the computer field. Public domain and shareware utilities (such as PKZIP) are quite common, and they typically yield 50-80% reductions in the size of English text and computer program files. Users can run these utilities manually before sending their files over the network, or they could use the automatic LZW stream compression features built into the NOS TCP/IP package by Anders Klemets. [Klemets91]

Data compression does not increase the capacity of the network per se, it simply uses it more efficiently. But the bottom line is the same: the network can do more useful work (moving user data) with the same spectrum resources.

8. Conclusion

It is the author's belief that an efficient, self-configuring, single-channel half-duplex store-and-forward amateur packet radio network would be quite practical if the design principles discussed here were pursued. The much-maligned "digipeater network" is so bad only because the modulation methods, channel access and routing algorithms are all so sub-optimal, and because there is no power control at all. Properly designed, a collection of "digipeaters done right" would have a lot of practical advantages because of its decentralized nature. All the nodes would be equal., so the failure of any one node need not bring the entire network down, as would happen if the hub or repeater in a centralized network were to fail. (This network model was used for the original DARPA packet radio experiments precisely because of this inherent robusmess.) And the system capacity could actually increase as additional nodes were added, because the average inter-nodal distance would decrease, allowing the min-power routing and automatic power control algorithms to reduce average transmitter powers.

As a first step toward such a network, I urge the manufacturers of digital radios to include the "hooks" for automatic power control. What's urgently needed is a way for the packet controller CPU to quickly vary the power of the transmitter in discrete steps, e.g., with a D/A converter, and a way to measure incoming receiver signal levels, e.g., with an A/D converter on the AGC line. Once we have these hardware features, we software types can do the rest.

9. Acknowledgements

I would like to thank my new colleagues at Qualcomm, Inc, particularly Klein Gilhousen, WT6G and Mike Brock, WB6HHV, for some very enlightening discussions.

10. References

[Costas59] John P. Costas, K2EN, ***Poisson, Shannon and the Radio Amateur,*** Proceedings of the IRE, December 1959.

[Elmore90] ***Glenn*** Elmore, N6GN, ***Physical Layer Considerations in Building a High Speed Amateur Radio Network,*** ARRL Computer Networking Conference, 1990.

[Gilhousen91] Klein S. Gilhousen, WT6G, et al, ***On the Capacity of a Cellular CDMA System,*** IEEE Transactions on Vehicular Technology, Vol. 40 No. 2, May 1991.

[Jacobson90] Van Jacobson, ***Compressing TCP/IP Headers for Low-Speed Serial Links,*** Internet RFC 1144, February 1990.

[Karn90] Phil Karn, KA9Q, ***MACA - A New Channel Access Method for Packet Radio,*** ARRL Computer Networking Conference, September 1990.

[Klemets91] Anders Klemets, SMORGV, ***LZW Compression of Interactive Network Traffic***, ARRL Computer Networking Conference, 1991.

[Lee89] William C. Y. ***Lee, Mobile Cellular Telecommunication Systems,*** McGraw-Hill, 1989.

[Lee91] William C. Y. Lee, ***Overview of Cellular CDMA, IEEE*** Transactions on Vehicular Technology, Vol. 40 No. 2, May 1991.

[Mills87] David L. Mills, W3HCF, ***Advanced Topics on the DARPA Internet System,*** Interop '88 tutorial notes, September 1988.

[Price90] Harold Price, NK6K, and Geoff Ward, K8KA, ***Pacsat Broadcast Protocol, ARRL*** Computer Networking Conference, September 1990. (See also other related papers in the same issue).