Interactions Among Ionospheric Propagation, HF Modems, and Data Protocols

Eric E. Johnson
New Mexico State University

Abstract

The well-known challenges in using the ionospheric skywave channel for data communications have been increasingly well-addressed by recent generations of HF data modems and link protocols. However, the means used to cope with the skywave channel present unusual challenges to Internet-style network- and application-layer protocols, which have been designed with reliable, high-bandwidth channels in mind.

In this paper, we explore the interactions among the skywave channel, HF data modems, data link protocols, and the higher-layer protocols, and identify the promising features that will form the basis of reliable Internet operation via next-generation HF radio networks.

Introduction

The ionospheric channel is well known for exhibiting temporal effects over a wide range of time scales [1], including multipath spreads on the order of milliseconds that produce intersymbol interference, various types of fading on the order of seconds to minutes, hourly diurnal variations, and so on up through the 11-year sunspot cycle. However, the value of beyond-line-of-sight wireless communications is such that technologies have been developed to deal with each of these challenges, so that we now contemplate using high-frequency (HF) radio to carry Internet traffic.

The techniques employed to provide reliable data communications over ionospheric channels exhibit traits that vary substantially from the “wired” Internet. In this paper, we investigate the interplay among skywave channels, HF data modems, automatic repeat request (ARQ) protocols developed specifically for HF radio channels, and key protocols from the Internet.

HF Radio Channels

An HF skywave channel conveys signals beyond line of sight via refraction from the ionosphere (and possibly intermediate “bounces” off the earth) to one or more distant receivers. The refractive and absorptive characteristics of the ionospheric layers depend strongly on radio frequency, so the first requirement for an automated HF radio system is the identification and coordination of a usable frequency. In current systems, data from various measurement and prediction techniques may be combined to select a frequency, and an Automatic Link Establishment (ALE) protocol coordinates the tuning of participating radios to the channel and the transition to data transfer.

The audio channel provided by an HF radio is nominally 3 kHz in bandwidth, and exhibits low and fluctuating signal-to-noise ratio (SNR), bounded by treaty limits on radiated power and by galactic, atmospheric, and man-made noise. Signals reach the receiver via refraction from one or more ionospheric layers, each of which may be in motion. The received signal is often a composition of multiple signals having independent, time-varying path losses and phase shifts. Thus we may expect multipath interference, deep fades, and impulsive (non-Gaussian) noise, all superimposed on an SNR trend that may eventually require a fallback to ALE to select a new frequency and set up a new channel.

HF Data Modems

Modems for HF radio channels apply a suite of techniques to address channel problems ranging up to a few seconds. Historically, the intersymbol interference resulting from multipath was addressed by using a symbol period longer than the maximum expected multipath spread. With the resulting low baud rate, multiple “parallel” tones could be modulated within the audio passband to boost the data rate. However, recent advances in digital signal processing technology have made adaptive equalizers practical, permitting current high-performance HF data modems [2, 3, 4, 5] to use serial-tone phase-shift keying (PSK) or even quadrature amplitude modulation (QAM).

The forward error correction (FEC) used in the current HF data modems employs a wide range of code rates to cope with a correspondingly wide range of SNR conditions. During fades, of course, more symbols may be lost than the FEC can correct, even though the average SNR suggests that the error rate should be manageable. Interleaving is therefore employed to spread burst errors over longer symbol sequences so that the resulting error density is suitable for FEC. The US/NATO modems cited above use block interleavers ranging up to 8.64 seconds. When the connection to these modems is a synchronous serial interface, the end-to-end delay through the sending and receiving modems is at least two times the interleaver depth. Link turnaround times are at least twice that long, so ARQ systems use the shortest interleaver possible.

Because of the high variability of the HF skywave channel, HF data modems need to be able to change data rate rapidly so that throughput is maintained near the highest rate that the channel can support at each instant. Two approaches are used in the current US/NATO modems:

- MIL-STD-188-110 [2] / STANAG 4539 [3] modems encode in the common synchronization preamble that precedes each transmission the data rate and interleaver depth that will be used in that transmission. Thus, data rate change can be made unilaterally at the transmitting station, and the receiving station will automatically shift to the new waveform upon receipt of the sync preamble.
- The burst waveforms used for “third generation” ALE and ARQ in the US [4] and for ARCS in NATO [5] are even more agile. They employ code combining for data transmissions: complete channel coding is computed for each data block before transmission, but only a subset (one half or one quarter) of the code bits are sent in each transmission. If a packet is received with uncorrectable errors, the soft decisions are saved and additional code bits are requested in a retransmission of the packet. After each new reception, the additional received signal is combined in the FEC decoder with the earlier reception(s) until an error-free result is obtained. Since the retransmission of additional code bits is requested on a packet-by-packet basis, the code rate (and therefore the effective data rate) of each packet is reduced from the initial high rate only so far as is necessary for correct reception. Thus, with no more overhead than is already required for ARQ operation, data rate can adapt as required for each individual packet in a message.

ARQ Data Link Protocols for HF Radio Links

When the combination of FEC and interleaving is insufficient to recover error-free data at the receiver, we must either tolerate data with errors or request retransmission. In the former case, we use a simple one-way (or “broadcast”) data link protocol, whereas the latter requires an ARQ protocol. Several ARQ protocols have been developed specifically for use in HF radio channels. Three from current US and NATO standards will be discussed here:

- STANAG 5066 ARQ [6]: a classic selective repeat ARQ protocol with some special features such as an end-of-transmission announcement to simplify link turnaround timing. This protocol is frequently used with the serial-tone modems mentioned above [2, 3] but can in principle be used with any modem.
- Low-latency Data Link (LDL) protocol [4, 5]: a stop-and-wait ARQ protocol, tightly integrated with a very robust burst modem. Uses code combining to dynamically adapt FEC code rate frame-by-frame. Provides useful throughput at –10 dB SNR.
- High-throughput Data Link (HDL) protocol [4, 5]: a selective repeat ARQ protocol, tightly integrated with a code-combining burst modem that emphasizes throughput rather than low-SNR performance.

Note that the reliable packet delivery provided by ARQ protocols comes at the cost of variable delays due to retransmissions.

Internet Protocols

Internet application protocols (e.g., the hypertext transfer protocol HTTP used in the worldwide web, the file transfer protocol FTP, and the simple mail transfer protocol SMTP) exchange a number of short commands and responses before each large file transfer (web page, file, or email message). For example, in the course of transferring a single email message, an SMTP client will send at least four short SMTP commands (typically 10 to 20 bytes each) and receive a similar number of short replies from the SMTP server.

Internet application protocols such as these use the Transmission Control Protocol (TCP) to provide a reliable stream transport service to peer entities at distant hosts. TCP breaks long application messages into segments with a fixed maximum size, and employs sliding-window flow control and ARQ with adaptive timeouts to reliably transport these segments among application entities. The adaptive timeout mechanism of TCP maintains an estimate of the round-trip time (RTT) on each end-to-end path, and retransmits segments when acknowledgements fail to arrive within a small multiple of the RTT estimate. When a retransmission is required, TCP responds to the possibility that the RTT has increased by increasing its RTT estimate. Of course it is also possible that the segment or its acknowledgement was lost due to network congestion, so TCP also responds to segment timeouts by reducing the rate of sending segments into the network.

Both TCP and the User Datagram Protocol (UDP) call upon the Internet Protocol (IP) for routing their data through tandem connections of subnetworks, ranging from high-speed local area networks to low-bandwidth wireless networks. IP provides a “best effort” datagram service, and is generally insensitive to latency in the subnetworks; IP is not discusssed further here, but always accompanies TCP in this study.

Overview of the Study

The purpose of this study was to characterize the interactions among the varying behaviors of the ionospheric channel, several current data modems, ARQ protocols, and key Internet protocols. The need to control experimental conditions precisely and to analyze in detail the activities of the protocols made a simulation study the most practical approach.

The simulation model is described in the next section of this paper, including the validation of the channel and protocol models that provide some confidence in the results. The sections that follow present results of investigations of the effects on the ARQ protocols of varying the parameters of the channel model, and the effects on Internet protocols of latency in the HF subnetwork and certain optimizations in the HF subnetwork for support of Internet traffic.

Simulation Model

The simulators used for this study use NetSim technology, which has been used in numerous previous investigations in support of a wide range of HF networking applications. NetSim is a modular, discrete-event simulation framework. The modeling approach for key modules is described below.

Protocol Models

Protocols are implemented as stand-alone state machines with formal interfaces (e.g., PACKET_UP and PACKET_DOWN primitives to protocols at next higher and lower layers) as though intended to run in an actual system. The simulation framework provides a run-time environment for the protocols (e.g., event scheduling and timer services) analogous to what might be provided by a real-time operating system (RTOS)[1]. The protocols modeled in this thorough fashion for the present investigation include STANAG 5066 ARQ, HDL, LDL, and TCP.

Application protocols (e.g., SMTP) were not modeled as full state machines, but only as scripts that provided typical Internet loading on the lower layers.

Modem Models

Modems are implemented in NetSim to a level of detail appropriate for each investigation. When the modem itself is of interest, it is implemented in full detail as if running on a DSP. However, this is computationally expensive, so the modem is usually modeled as a statistical transducer that corrupts protocol frames as a

1 In at least one case, such protocol software was actually ported to a RTOS by others and used in a commercial product.

function of channel conditions, in accordance with measurements of an actual modem. This study used statistical models of the MIL-STD-188-110B waveforms from 75 bps through 9600 bps, and for the burst waveforms used by the HDL and LDL ARQ protocols.

Channel Model

Skywave channels are individually modeled using the "Walnut Street" approach[2]:

- The shortest-term channel effects (as in the Watterson model) are included in the modem model via measurements of the modem using a Watterson-model channel simulator.
- Hourly- and longer-term effects are introduced using prediction programs (e.g., ICEPAC), often using precomputed tables to remove these computations from the running time of simulations.
- Furman and McRae [7] noted that between the short-term Watterson model effects and the long-term variation described by prediction programs lies an intermediate-term regime characterized by lognormal SNR fluctuations of several dB with time constants on the order of seconds.

Validation

A version of NetSim that was developed to support the systems engineering of high-power HF radio networks was independently validated by the US Defense Information Systems Agency (DISA) Joint Interoperability Test Command (JITC) [8]. The Walnut Street channel model was validated using measurements from radios aboard Air Force One to various ground stations during an extended trip abroad by the U.S. President. The second-generation ALE protocol was validated in a parallel effort by JITC.

The models of the modems, the ARQ protocols, and TCP were validated by the author by comparison with published results of implementations of these systems (see [9]).

Channel Interactions with Link Protocols

As discussed above, the model of the ionospheric channel comprises three regimes: short-term effects that interact with the modem, long-term effects that are addressed by ALE and automatic link maintenance, and the intermediate-term fading that prompts use of an ARQ protocol. In this section, the focus is on the interaction between the ARQ protocol and the intermediate-term fading.

The measurements of intermediate-term SNR variations by Furman and McRae [7] on a north-south link between Melbourne, FL and Rochester, NY exhibited lognormal variation with a standard deviation of around 4 dB and a time constant of about 10 seconds. Goodman [1] also discusses such fluctuations, and plots the statistics for a range of standard deviations from 0 to 20 dB. Fade rates are known to vary over a wide range. The simulations in this section identify the limits of fade rate and SNR standard deviation beyond which the current ARQ protocols do not operate well.

Nominal Performance

Since the focus of this section is on the ARQ protocols, a lightweight workload (HMTP [10] without TCP) is used. We count the number of 5000 byte messages that are sent per hour, and run a continuous session for at least an hour so that startup transients do not dominate the results. Figure 1 shows the messsage throughput versus SNR for the nominal channel described by Furman and McRae. (The curve labeled "5066 turbo" is a specially tuned version of the STANAG 5066 ARQ protocol that includes some optimizations under consideration for the next revision of the standard.)

[2] Named for an establishment in Boulder, Colorado where the model was informally agreed by representatives of the HF Industry Association

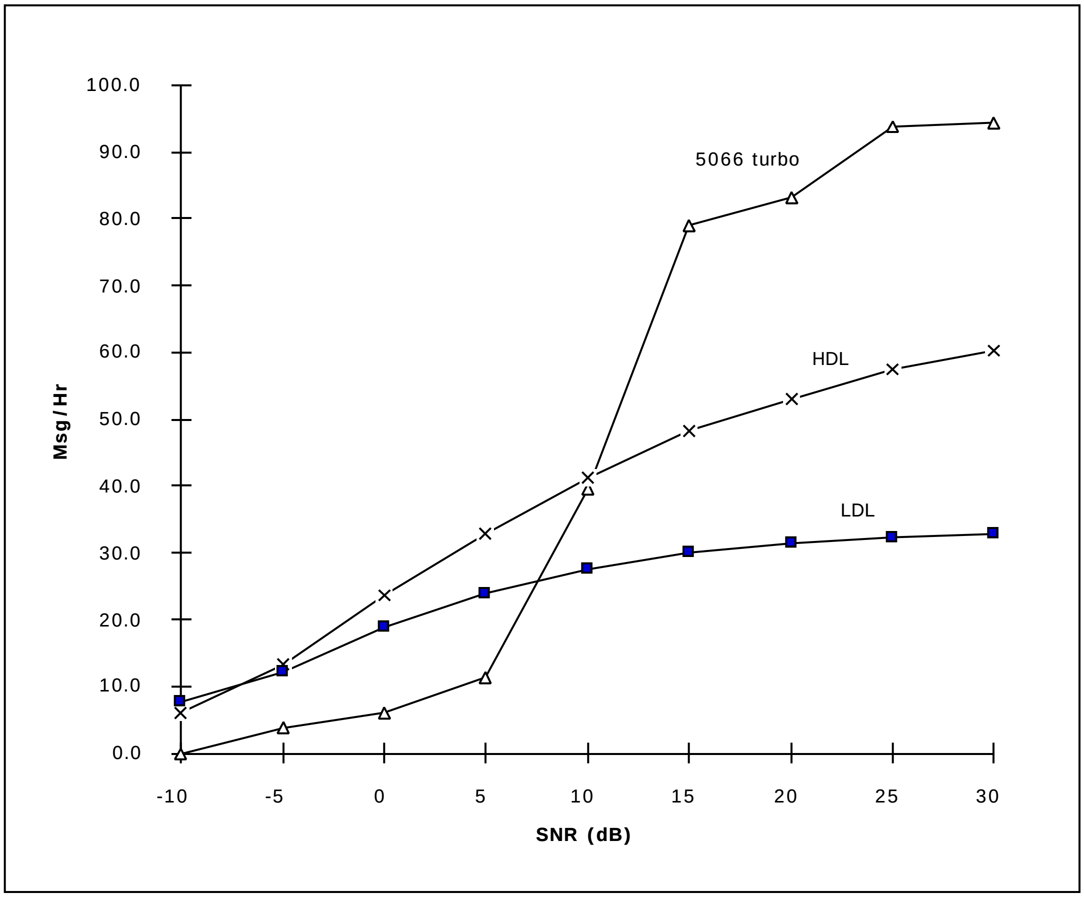

Figure 1: Throughput of 5000-byte messages in channel with 4 dB SNR standard deviation and 10-second autocorrelation time constant

This "nominal" channel has relatively slow fading with rather small amplitudes. Analysis of the simulation results shows that this allows the serial-tone modem and 5066 ARQ combination to settle in to a steady state (e.g., 75 bps at 0 dB SNR, 3200 bps at 10 dB, 8000 bps at 25 dB and 9600 bps at 30 dB). The code-combining systems likewise settle into a steady state, with a relatively smooth progression of packet retransmission rate with SNR. With these results as a baseline, we now proceed to analyze the interactions of less benign channels with our modem/protocol combinations.

STANAG 5066 ARQ Interactions

During the data phase of each data-ack cycle, the 5066 ARQ protocol sends multiple small data packets in a transmission of up to two minutes. At data rates below 2400 bps, the packet size is fixed at 200 bytes; above 2400 bps, packet size is increased to avoid a sequence number starvation effect (an optimization). The entire transmission is sent at a single data rate. The data rate change algorithm raises the data rate for the next transmission when all packets in the previous transmission were received error free, and lowers the data rate when more than half of the packets require retransmission.

It is intuitive that increasing the rate of fades will increase the fraction of packets in a transmission that experience error bursts. When this fraction reaches 50%, the data rate will be lowered, resulting in a more robust waveform. Unless the number of bytes in each packet is also reduced, however, a data rate reduction will result in longer on-air time for each packet, increasing the fraction of packets with errors unless the more heavily coded waveform is sufficiently robust to work through the fades. Thus, at a given median SNR and fade depth, we expect to find some fade rate above which throughput suddenly drops to a rate commensurate with an SNR equal to the median SNR reduced by the median fade depth.

Such an effect is seen in Figure 2, which plots ARQ throughput at 20 dB median SNR and 20 dB fade depth versus the autocorrelation time constant (which is inversely proportional to fade rate). Examination of the simulation results confirms the intuitive analysis of this effect: when frequent deep fades are damaging many of the packets in each transmission, the STANAG 5066 ARQ protocol reduces its data rate until the waveform is sufficiently robust to withstand most of the fades.

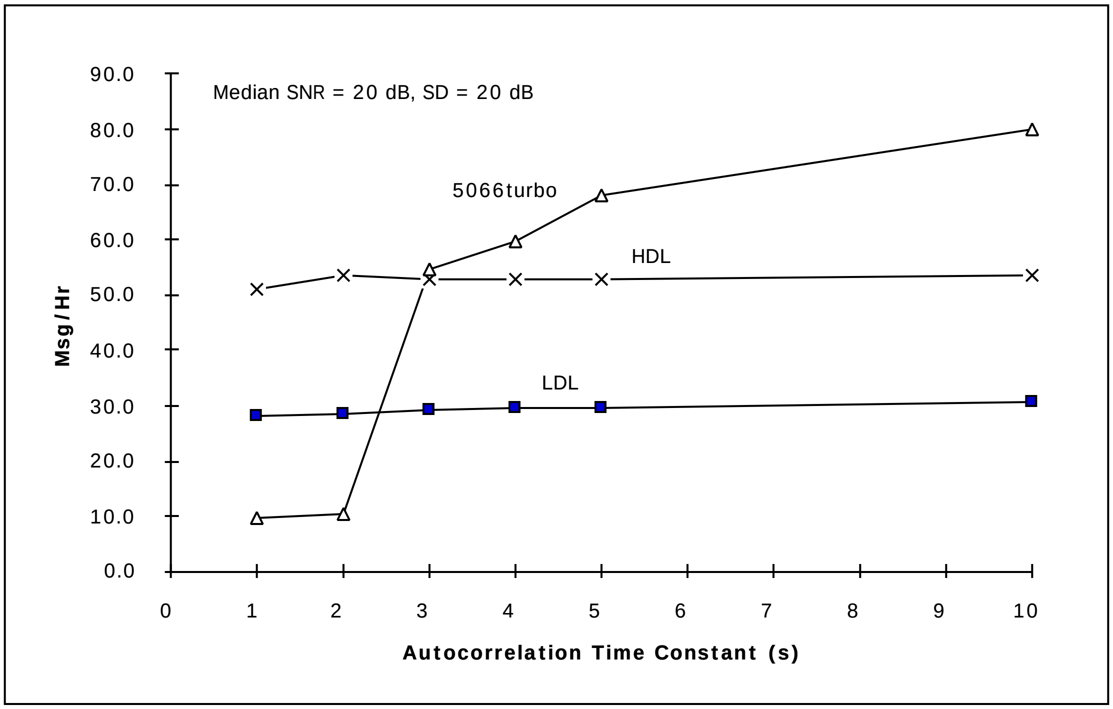

Figure 2: Effect of Fade Rate on ARQ Protocol Throughput

Third-Generation ARQ Interactions

The code-combining third-generation ARQ protocols (HDL and LDL) always send data at full speed (e.g., 4800 bps for HDL) and adapt code rate via retransmissions as described earlier. This avoids the need to determine future data rates from the recent history of the channel, so we do not expect to find abrupt changes in behavior in these protocols as the fade rate is increased. This expectation is borne out in Figure 2: the third-generation protocols do not exhibit the sudden drop in throughput as the fade rate is increased that was observed with the combination of a classic selective repeat ARQ protocol and a separate modem.

Impact on Internet Protocols

Although both TCP and link-layer ARQ protocols achieve reliable packet delivery through retransmissions, the impact of fading channels on TCP behavior is even more pronounced than upon ARQ protocols. The 5066 ARQ protocol was observed to "back off" (reduce its data rate) when it experienced frame error rates exceeding 50%. In contrast, TCP sharply reduces its rate of sending packets in response to a single lost packet.

Early experiments in using wireless links in the Internet sometimes omitted a link-layer ARQ protocol, using only TCP to provide reliable data delivery. The following example is illustrative of the results that may be expected in such a situation. This simulation experiment uses SMTP and TCP to transfer a message over our nominal channel with a 10 dB SNR.

- TCP directly controls an HF data modem set for 1200 bps with a 600 ms interleaver.
- TCP is using a segment size of 1500 bytes, corresponding to the maximum allowed segment size on the usual LAN connection from a user computer to the rest of the Internet. Efficient operation with 1500-byte retransmission units requires a bit error rate (BER) on the order of 10^{-5}. This is easily achievable in wired local area networks, as well as over wireless links with a suitable link-layer ARQ

protocol. However, the BER of the 1200 bps waveform on a 10 dB SNR fading channel is close to 10^{-5}, so we should expect frequent packet losses due to errors.

Transfer of a single 5000-byte message is discussed in three phases: setup, transfer, and conclusion:

1. It takes 37 seconds to complete TCP session establishment and the initial SMTP handshakes (HELO, MAIL, RCPT, and DATA commands). All of these transmissions are short (40 to 86 bytes), resulting in a segment loss rate of only 11%. As transmission of the message begins, TCP has improved its round-trip time estimate from a default of 0.5 s to 2.3 s and is using a timeout of 11.8 s. The window size is greater than 5000 bytes, so the entire message is queued at once.
2. The 1500-byte segments used for message transmission result in a 50% loss rate. Furthermore, the 11.8 s timeout occurs just as the first segment is arriving at the destination, leaving no chance for return of an acknowledgement. Both of these effects ensure that every message segment is sent at least twice. By the time the message is acknowledged (0:02:14 after the start of the simulation), the timeout at the sending TCP has grown to 756 s.
3. A fade occurs just as the nodes are trying to exchange the short messages that complete the SMTP transfer. The retransmission timeouts are already set to large values, and each time a message is lost, the TCP sending it doubles its timeout again (up to a maximum of 1 hour). As a result, the final handshake of the message transfer protocol is not completed until 1:30:31. Thus, a possible throughput of over 20 messages per hour has been reduced to less than one message per hour.

This example is typical of the effect of fading channels on TCP when a link-layer ARQ protocol is not used to insulate TCP from channel effects.

Interaction of Internet Protocols with Data Link Protocols

Performance of the Internet protocols SMTP and TCP *with* link-layer ARQ support is depicted in Figure 3.

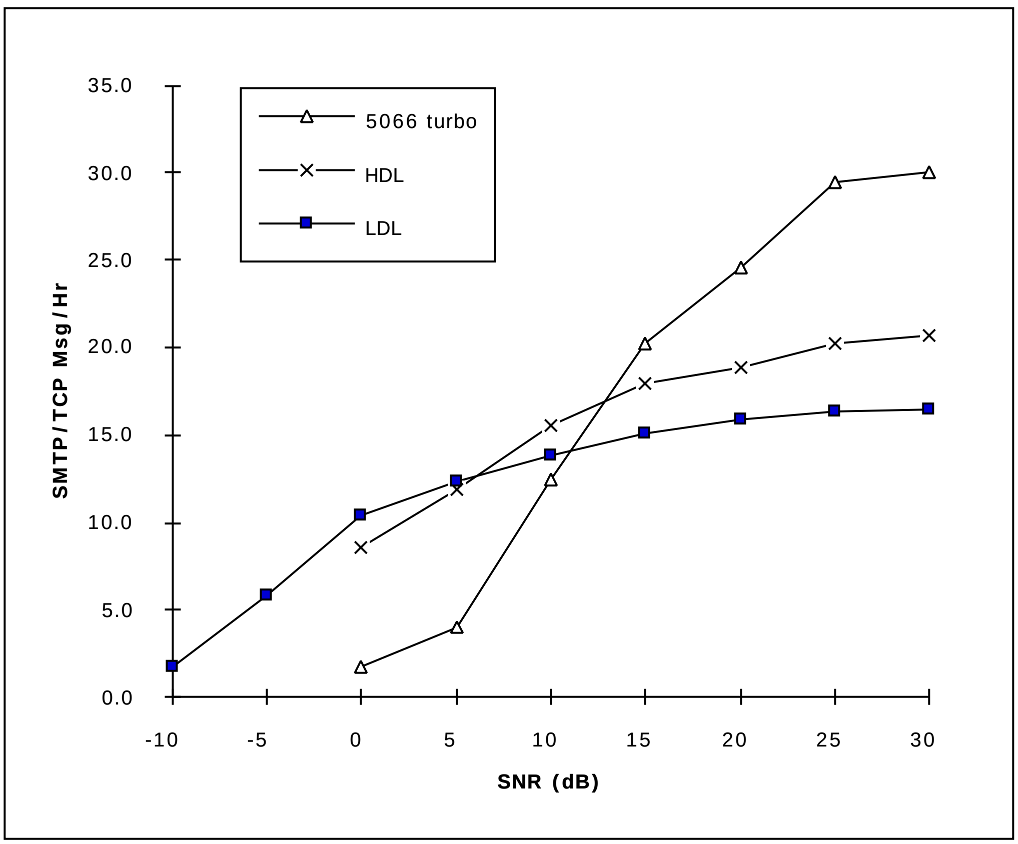

Figure 3: SMTP/TCP/ARQ Throughput of 5000-byte messages in nominal channel

Compared with the throughput results of an HF-friendly application client (Figure 1), the asymptotic throughput of the selective repeat protocols (5066 and HDL) drops by about two-thirds and that of LDL by about one-half [3]. Low-SNR performance of all of the ARQ protocols is insensitive to the use of SMTP and TCP.

The selective repeat protocols are seen to be especially sensitive to the use of Internet protocols at high SNR. This is the result of the frequent link turnarounds inherent in SMTP and TCP as well as the segmentation of messages into 1500 byte units by TCP:

- As discussed in the modem section, the standard HF data modems employ sizable interleavers, which must be filled and emptied for each transmission, which introduces a cost of a few seconds for each link turnaround in STANAG 5066 systems.
- This effect is absent in the burst modems used in HDL, but HDL (and LDL) do not support duplex client data and must perform a Traffic Manager handshake for each link turnaround.

In high-SNR channels, the selective repeat protocols can send an entire 5000-byte message in one transmission. Each link turnaround, as well as segmentation of the message into smaller transmission units, adds several seconds to the message delivery time, directly reducing their throughput.

Conclusions

The simulation results presented here indicate the capacity of current HF radio technology to support Internet applications over challenging skywave channels. In particular, code combining ARQ/modem combinations appear insensitive to even rapidly fluctuating channel conditions.

The adaptive mechanisms in TCP are not optimum for wireless networks, in which link error rates are both higher and burstier than in the wired Internet. Absent an appropriate link-layer ARQ protocol to hide link errors from TCP, throughput can drop dramatically in response to packet loss-induced timeouts.

The frequent link turnarounds implicit in TCP and SMTP reduce the performance that current HF ARQ protocols can achieve in high-SNR channels. Where the highest performance is desired, HF-friendly application and transport protocols should be used over the HF subnetworks, with application-layer gateways employed to translate between Internet-native and HF-native protocols at HF subnetwork boundaries. However, the HF subnetwork is also capable of providing transparent, low-bandwidth support for Internet protocols when needed.

References

1. J. Goodman, *HF Communications Science and Technology*, Van Nostrand Reinhold, New York, NY, 1992.
2. MIL-STD-188-110B, *Interoperability and Performance Standards for Data Modems*, 27 April 2000.
3. NATO Standardisation Agreement (STANAG) 4539, *Technical Standards For Non-Hopping HF Communications Waveforms.*
4. MIL-STD-188-141B, *Interoperability and Performance Standards for Medium and High Frequency Radio Systems*, Notice 1, 31 August 2001; Appendix C “Third-Generation HF Link Automation.”
5. NATO STANAG 4538, *Technical Standards For An Automatic Radio Control System For HF Communication Links*.
6. NATO STANAG 5066, *Profile For High Frequency (HF) Radio Data Communications*.
7. W. Furman and D. McRae, “Evaluation and Optimization of Data Link Protocols for HF Data Communication Systems,” *Proceedings of 1993 IEEE Military Communications Conference*, Boston, MA, October 1993, pp. 67-72.
8. R.K. Bullock, “SCOPE Command High Frequency (HF) Network Simulation Model (NetSim-SC) Verification and Validation Report,” Joint Interoperability Test Command, Ft. Huachuca, AZ, January 1997.
9. E. Johnson, “E-Mail Standards for HF Radio,” Technical Report, NMSU, July 2001.
10. MIL-STD-188-141B (ibid.), Appendix E.

[3] Note that these results apply to link-layer ARQ protocols that filter out duplicate datagrams resulting from TCP timeouts, as discussed in [9].