INTEROPERABILITY AND PERFORMANCE ISSUES IN HF E-MAIL

Eric E. Johnson
New Mexico State University and
Science Applications International Corp.

ABSTRACT

Electronic mail is particularly tolerant of the sometimes challenging HF radio communications medium. Its latency tolerance and store-and-forward delivery mechanism cope naturally with slow links and the occasional outage. This paper documents the results of a study of the interoperability and performance of leading standards applicable to e-mail service over HF networks, and recommends specific protocol suites for next-generation networks as well as directions for further research.

1. INTRODUCTION

The key challenges of the high-frequency (HF) radio medium that must be addressed in electronic mail (e-mail) and other data applications are listed below, along with closely corresponding mitigation approaches.

Challenge	Response
Low signal-to-noise ratio	Robust modem waveforms, with forward error correction (FEC) and interleaving
Multipath fading channels	
Fades and interference that overwhelm FEC	Automatic repeat request (ARQ) data link protocols
Propagation variation with hour, season, and sunspot cycle	Automatic link establishment (ALE) and related adaptive technologies
Limited channel capacity	Prioritization, flow control

These techniques employed to overcome the challenges of the HF medium pose problems in turn for the transport and application layer protocols that convey e-mail through the network. In general, the HF subnetwork provides low bandwidth channels (due to the usual restriction to 3 kHz or narrower RF channels). Error-free data service comes only at the cost of delays that greatly exceed those experienced in the wired Internet. Beyond the delays imposed by low data bandwidths, the interleavers used to cope with burst errors result in link turnaround delays on the order of seconds to tens of seconds. This makes protocols that require frequent link turnarounds unattractive for HF applications.

1.1 Layers that must interoperate

Interoperability of HF e-mail systems requires interoperability of all of the communications layers and sublayers needed to convey messages. The relevant layers and sublayers include the following:

- Radios and frequency assignments
- Automatic link establishment
- Cryptographic algorithms and keys
- Data modem waveforms
- Data link protocol
- Transport protocol
- E-mail protocol

The first three items are beyond the scope of this paper. For the remainder of this paper, it is assumed that a usable HF link is available, including communications security (COMSEC) when required. The remaining four items are the subject of this investigation. The specific standards considered are enumerated in the next section.

1.2 Overview of the study

The purpose of this investigation was to evaluate the interoperability and performance of the leading open standards for HF e-mail. The standards considered are listed below by layer:

Physical Layer (Data Modem)	MIL-STD-188-110B	75 – 9600 bps
	MIL-STD-188-141B	3G burst waveforms
Data Link Layer ARQ protocol	STANAG 5066	
	MIL-STD-188-141B 3G ARQ	
	none	
Transport protocol	TCP	
	none	
E-mail protocol	SMTP	
	HMTP	

E-mail protocol interoperability was evaluated by analysis, while the performance of the protocol stacks was evaluated by detailed simulations.

2. E-MAIL PROTOCOL INTEROPERABILITY

Only the protocols used to "push" e-mail messages through the network to the server nearest the recipient are considered here. (In many cases, the recipient then uses another protocol such as POP3 or IMAP to "pull" messages from that server, often over a wired network.)

- The Simple Mail Transfer Protocol **SMTP** [RFC-821] is the standard within the Internet. Its frequent link turnarounds are unattractive for networks with extended turnaround times, however.
- An HF-oriented variant of SMTP called **HMTP** reduces the number of link turnarounds by sending multiple commands or responses per transmission ("pipelining").

The two e-mail protocols, SMTP and HMTP, employ nearly identical messages and sequences to convey a message from one machine (the client) to another (the server).

- SMTP requires a lockstep exchange of messages from the client and acknowledgements from the server, resulting in 12 link turnarounds.
- HMTP allows a grouping of the commands and responses to reduce the number of link turnarounds to 2 (see Annex F of STANAG 5066 or Appendix E[1] of MIL-STD-188-141B Change Notice 1). HMTP requires fall-back to basic lockstep SMTP operation if this aggressive pipelining is not supported.

Analysis of the protocol requirements for SMTP and HMTP showed that all combinations of *correctly implemented* SMTP and HMTP clients and servers will interoperate, although an HMTP client sending mail to an SMTP-only server might have to send the message twice [1].

3. TRANSPORT LAYER INTERACTIONS

The Transmission Control Protocol **TCP** [RFC-793] is the default transport protocol for e-mail throughout the Internet. TCP provides end-to-end reliable delivery of application data using a sliding window ARQ protocol with adaptive timeouts and window size. It is well known that the adaptive algorithms used in the most widely implemented versions of TCP are oriented to congestion control in a reliable network. As a result, they "back off" very quickly after datagram timeouts expire, and are widely believed to be unsuitable for use over wireless links.

An earlier version of the simulator used in the present study was developed to identify approaches to successfully employ TCP on lossy links such as the HF medium. Concisely, the results are as follows:

- If TCP is used over HF links without an error-correcting data link protocol, TCP will be responsible for error correction, so its maximum timeout should be set to a value of 30 to 60 seconds.
- If an ARQ data link protocol is used to correct errors, the TCP maximum timeout should be set to a large value, effectively eliminating TCP retransmissions after TCP learns the round trip time statistics of the network.

The simulation results in section 5 suggest that TCP can indeed be used successfully in wireless applications, although the overall speed of an optimized protocol suite generally will be better without TCP than with TCP.

4. DATA LINK PROTOCOLS

This study included implementation of two data link protocol families, along with appropriate modems:

- The STANAG 5066 Subnetwork Interface/Channel Access/Data Transfer protocol, operating in ARQ mode after data link setup. This suite is collectively denoted **5066-ARQ**. It achieves its best ARQ performance with modem waveforms that "self identify" their data rate and interleaver. The 75–9600 bps waveforms in MIL-STD-188-110B have this property. (The same waveforms are used in NATO STANAG 4539.)
- A **3G ARQ** subset including the Traffic Manager and the HDL and LDL ARQ protocols. These protocols and the associated burst waveforms are specified in MIL-STD-188-141B Appendix C and NATO STANAG 4538.

The two families are not interoperable with each other.

This section provides some details of the implementations of these protocols and their associated modems in the simulator used for this investigation.

4.1 Channel and modem models

The time to deliver a message depends on the channel principally in how many times each PDU must be sent to be received correctly. The level of detail required in the HF channel model for this investigation is therefore satisfied if we know the probability of correct PDU reception as a function of waveform, PDU size, and SNR. The benefit of finer-grained fidelity in the specific order of good and bad frames is likely to disappear in the resulting total latency, although interactions with independent time-based processes (such as TCP retransmissions) require simulation rather than a simple analytical modeling.

For the simulations in this investigation, measured performance of the MIL-STD-188-110B and 3G burst waveforms in a Watterson-model 2-path fading channel (ITU-R "Poor") was coded as tables of error rates versus SNR. Interpolation between measured SNR values was used as required. The MIL-STD-188-110B HDR waveforms used the 2.16s interleaver.

The SNR for each simulation was constant throughout the simulation. The Rayleigh fading effects on the modems were included in the modem measurements, but

---

[1] A Compressed File Transfer Protocol (CFTP) is also specified in Appendix E, but this protocol is not interoperable with either SMTP or HMTP. It is discussed further in section 5.

longer-term SNR variations were not simulated (as they would be in a NetSim [2, 3] simulation, for example).

4.2 5066-ARQ implementation

Implementation Details. The 5066-ARQ protocol as specified in STANAG 5066 Annexes A-C v1.1, dated 19 April 1999 was implemented in the simulated environment. Specifics of the implementation are documented in [1].

Validation. File transfer measurements from DERA [4] were used to check the simulator. However, the prototypes measured for the DERA report did not adapt data rate, used ~100-octet C_PDU segments, and used a simple file transfer protocol rather than one of the client protocols implemented in this study.

The first two differences from the 5066-ARQ simulator were accommodated directly by modifying the simulator used for validation. The third was approximated by observing the simulator performance with an unacknowledged file transfer client, with and without TCP in the transport layer. Duplicate datagram filtering (see section 4.4) was disabled. The resulting simulator without TCP performed somewhat better than the measurements; with TCP it performed somewhat worse than the measurements. As the burden imposed by TCP is expected to be more onerous than that of the file transfer protocol used in the DERA measurements, this result lends plausibility (though not validation) to the 5066-ARQ simulator results.

4.3 3G ARQ implementation

Implementation Details. The TM, HDL, and LDL protocols specified in STANAG 4538 and MIL-STD-188-141B Appendix C were implemented as described in [1].

Validation. Measurements of a similarly capable LDL implementation were provided by Harris RF Communications. The measured performance of e-mail delivery and pings was well within the 90% confidence interval of the simulated performance of similar operations.

4.4 Optimizations

A key optimization implemented in both of these data link protocols was a simple filtering of duplicate client datagrams. Each time a client (e.g., TCP) submitted a datagram for transmission, the data link protocol compared it to queued datagrams awaiting transmission, and silently discarded it if it was identical. When the ARQ protocol was operating in “greedy” mode, i.e., sending all waiting datagrams at one station before reversing the direction of client data flow, client datagrams were also compared to all datagrams sent since the other station had a chance to send client acknowledgements.

Although a few duplicate datagrams were accepted that an algorithm with a complete history of traffic would have rejected, this simple technique was quite effective in eliminating many of the unnecessary retransmissions sent by TCP as it “learned” the network latency.

5. PERFORMANCE

We now explore the performance of the selected combinations of e-mail, transport, and data link protocols. As noted before, the complex interactions among the protocols preclude use of simple analytical models. Resource constraints made it infeasible to pursue a program of measured performance, leaving a simulation study as the only viable option.

Implementation of the e-mail protocols in the simulator was straightforward. In each case, scripts for the client and server were prepared, and the e-mail protocols simply stepped through them as the appropriate messages were received. Each simulation began with a call by the client to open a connection to the server. The server responded with a 220 message (see [RFC-821]), and a dialog was conducted over the transport connection to transfer a single 5000 octet e-mail message[2].

When TCP provided the transport service, it opened the connection using the usual SYN and SYN/ACK handshake. The default TCP timeout of 500 ms expired repeatedly during this initial handshake; each timeout resulted in a doubling of the next timeout duration. The TCP simulator fully implements sliding window flow control (including avoidance of the fabled Silly Window Syndrome), ARQ, segmentation and reassembly, and selectable maximum segment size and timeout upper bound. The simulations in this section all used a maximum TCP segment size of 1460 octets, corresponding to the usual Ethernet maximum transmission unit of 1500 octets. Unless otherwise noted, the maximum TCP timeout was set to one hour.

The simulations cover an SNR range from –10 to +30 dB. The lower limit is set by the most robust waveform evaluated (the 3G LDL burst waveform). The upper limit is set by radio receiver performance (e.g., the harmonic distortion specification in MIL-STD-188-141B).

The result of each simulation is the time required to deliver a single message using the selected protocol suite. The average latency for several independent simulations is converted to “messages per hour,” effectively assuming that a new message transfer begins immediately after the previous message is delivered. Note that the message throughput calculated in this way is extremely sensitive to message size. CFTP throughput in msg/hr for 5000 byte *compressed* files will be similar to that of HMTP.

The charts in the following sections show performance including all protocol overheads (except for ALE), and cannot be compared to charts that do not include all of the protocols necessary to transfer e-mail over the HF channel, or to results for different message sizes.

---

[2] Analysis of 311 e-mail messages recently received via the Internet showed a median size of 3618 octets. This was close to the 5000 octet size used in the testing of prototype 5066 ARQ systems, so the latter value was used here as well.

5.1 SMTP with TCP: HF as a transparent subnetwork

As a starting point for our investigation of e-mail performance in HF networks, let us consider the e-mail protocol structure least attuned to the HF medium: end-to-end SMTP and TCP, using an HF subnetwork as just another subnet in the Internet. We examine three cases:

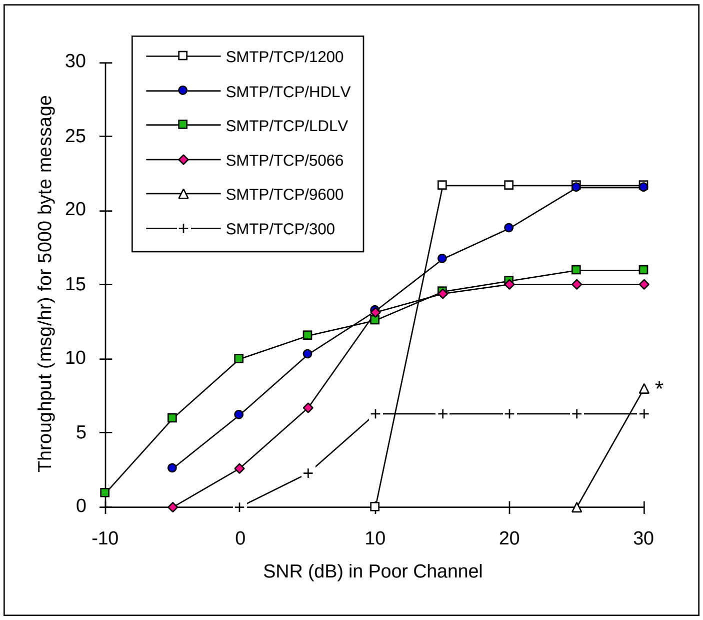
- SMTP and TCP using only a 110B modem with no data link ARQ protocol
- SMTP and TCP using 5066 ARQ with a 110B modem
- SMTP and TCP using 3G ARQ and the 3G burst modem

A system using a 110B modem with no data link protocol is presumed to have no adaptive data rate mechanism, and simply runs the modems at a fixed data rate. Three rates were simulated: 9600 bps, 1200 bps, and 300 bps.

The ARQ protocols also have no initial knowledge of the channel, but adapt their data rates during each simulation as described in the respective standards. The 5066 ARQ protocol starts at 300 bps and adapts its rate up or down based on experience with the channel. The 3G ARQ protocols select a frame size considering only the message size; data rate adaptation in the 3G protocols occurs as a result of code combining.

Several aspects of the results below are of interest:

- The unaided modems performed quite well when given a channel that resulted in very low error rates. When the BER exceeded 10^{-4}, however, it was necessary to modify the TCP timeout limit so that TCP would act as an adequate ARQ mechanism. As the BER approached 10^{-3}, normal-size TCP segments experienced nearly 100% frame error rates, and no throughput was possible.



* The throughput for 9600 bps at 30 dB reflects an artificial reduction of the TCP maximum timeout to 30 seconds. Without this modification, throughput was reduced to less than 1 msg/hr by the slow retransmission rate.

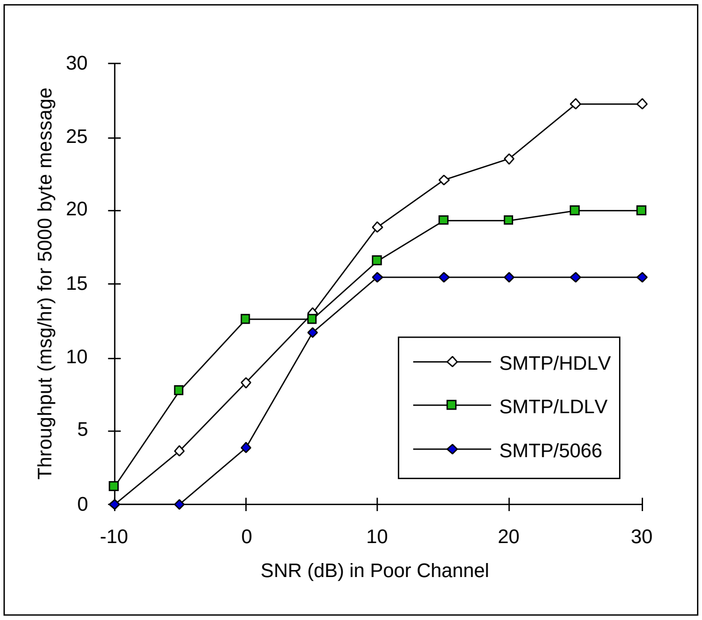
- The ARQ protocols coped with the entire range of SNR values by adapting their modem data rates, and by using smaller frame sizes than the TCP segment size.
- 5066 ARQ was unable to make effective use of its 9600 bps modem at high SNR because of the number of link turnarounds required to boost the modem data rate from its default initial value to the best rate the channel could support. It also suffers from sequence number starvation at the higher data rates due to the fixed frame size.
- The curious dip in the LDLV curve at +5 dB SNR seems to result from a threshold in its code combining ARQ. This anomaly is even more pronounced in later results.

5.2 SMTP without TCP

In order to isolate the effects of the various protocols on email system performance, we next eliminate TCP and run SMTP as a direct client of the data link ARQ protocols. (Use of SMTP with only a 110B modem is not feasible over the SNR range considered, due to the lack of error correction in the resulting system.)

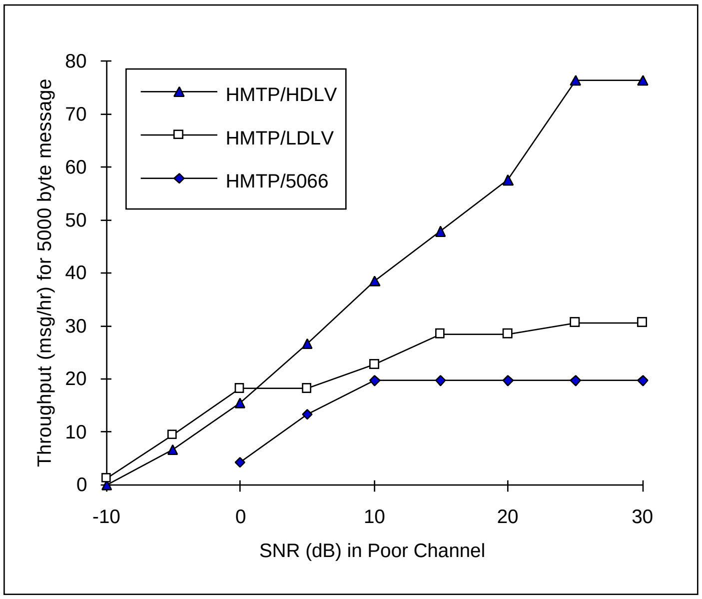
The throughput results for SMTP without TCP are plotted below. TCP clearly reduced message throughput somewhat, but the performance gained by employing HF gateways to eliminate TCP was typically only 25-30%. The extra message latency that TCP introduces may be acceptable in some applications which do not otherwise require a gateway at the interface with the Internet.

Again we see the plateau in 5066 ARQ performance. This suggests that setting the initial data rate and frame size to match the channel would be of great benefit, if the channel could be characterized reliably *a priori*.



5.3 HMTP without TCP

Finally, we examine HMTP with our HF data link protocols. Because HMTP will normally operate with dedicated gateways, TCP was not simulated. While both 5066 ARQ and LDL achieve somewhat higher performance when using HMTP, the most dramatic improvement is seen in e-mail throughput using HDL, which rises smoothly with SNR until it reaches the throughput limit of its 4800 bps modem. HDL throughput with HMTP (and CFTP) is more than double the throughput achieved with an SMTP client.



6. CONCLUSIONS AND FUTURE WORK

This study investigated the interoperability and performance of a small number of candidate modems and protocols for use in HF e-mail systems.

- Interoperability was evaluated through analysis of the relevant standards.
- Performance was studied using a simulator created for this project. The simulator was validated using a few measurements of similar systems.

The findings of this study are as follows:

- Any e-mail server software that follows the rules for SMTP command pipelining (RFC 2197 or RFC 1854) will be interoperable with both of the SMTP-based protocols studied (SMTP and HMTP).
- E-mail clients must fall back to standard SMTP operation when working with a server that cannot support pipelining so that universal interoperability is maintained. Sending a batch of SMTP commands that includes a message of significant size can result in sending that message twice if the server is not first interrogated for support of pipelining.
- The Internet Transmission Control Protocol TCP can be used transparently over HF radio subnetworks, although its presence on HF links can reduce overall performance. The link layer ARQ protocols studied (3G ARQ and 5066 ARQ) support TCP operation over their respective ranges of channel conditions without requiring modifications to host software (e.g., to specially tune TCP parameters).
- When channel conditions are not known *a priori*, the 3G ARQ protocols can provide significantly higher performance than the STANAG 5066 ARQ protocol in both throughput and SNR range.
- The 5066 ARQ suite offers a higher-speed modem than 3G ARQ, and an ability to carry client data in both directions during a session. Its high-SNR performance was limited here because it must “adapt up” to the channel from its initial settings while 3G instead “adapts down” via code combining in low SNR links.

These results suggest the following program of evolution of the HF e-mail standards and technology:

- HF e-mail systems in challenging environments should employ the 3G ARQ suite for the higher performance and increased robustness that it provides.
- Continued development of the 3G ARQ protocol suite should address duplex flow of higher-layer data and acknowledgements, code combining using a dense constellation (e.g., the 12,800 bps waveform from MIL-STD-188-110B), and multiple channel operation.
- STANAG 5066 ARQ protocol performance could improve dramatically in high-SNR channels if its initial data rate and frame size are matched to the channel [5]. This should be straightforward in applications with slowly-varying channels. Techniques to accurately characterize more challenging channels in advance should also be investigated.

REFERENCES

1. Eric E. Johnson, “E-Mail Standards for HF Radio,” Technical Report, NMSU, July 2001.
2. E. Johnson, R. Desourdis, and M. Rager, “Simulation of MIL-STD-187-721C automated HF networking,” *Proceedings of the Ionospheric Effects Conference IES ‘96*, 1996.
3. Eric E. Johnson, “Fast Propagation Predictions for HF Network Simulations,” *Proceedings of MILCOM ’97*, IEEE, 1997.
4. “STANAG 5066 Prototype Development and Testing: Results (Cont.),” DERA CIS1.3 Portsdown West, UK. December 1998.
5. W. L. Beamish and W. N. Furman, “Performance of HMTP Based Email using Second and Third Generation HF Data Links,” *Proceedings of HF ’01,* August 2001.