Taking the Politics out of Satellite and Space-Based Communications Protocols

William D. Ivancic
NASA Glenn Research Center
Cleveland, Ohio
wivancic@grc.nasa.gov

Abstract—After many years of studies, experimentation and deployment, large amounts of misinformation and misconceptions remain regarding applicability of various communications protocols for use in satellite and space-based networks. This paper attempts to remove much of the politics, misconceptions and misinformation that have plagued space-based communications protocol development and deployment. This paper provides a common vocabulary for communications; a general discussion of the requirements for various communication environments; an evaluation of trade-offs between circuit and packet switching technologies and the pros and cons of various link, network, transport, application and security protocols. Included is the applicability of protocol enhancing proxies to NASA, DOD and commercial space communication systems.

Keyword—protocol, satellite, network, gateway, security, communications

I. INTRODUCTION

NASA is developing a new Space Communications and Navigation Architecture enabling NASA’s Exploration and Science programs to be executed between 2010 and 2030 [1]. NASA has long recognized that efficient, high-quality communication is an essential enabler for all space activities. NASA’s Space Exploration Initiative will require communication technology development efforts and protocols that fit into an evolving, dynamic space communication architecture. These communication protocols must match the needs of the emerging Exploration Program as it matures.

Space-based communications networks are often thought of as having unique characteristics that require special consideration. There is a measure of truth to this, but, only when applied appropriately. Volume, mass and power are at a premium in space – as well as on Earth, for mobile and ad-hoc communication. Intermittent connectivity is a common mode of operations for planetary relays – as it is for many military operations. Delay and latency have to be considered for space-based protocols – as they must be for terrestrial protocols, where delays via low-bandwidth, highly-processed, links may be in the order of seconds [2]. Deep-space communication has extremely long delays in the order of hours – terrestrially we use email and text messaging and are not terribly concerned about the delay so long as the message gets through eventually. Reliability and redundancy are of major concern – as they are in aeronautical, military and commercial networks. Finally, space hardware must withstand radiation effects – this one is relatively unique to space, though also present in military and high-altitude applications. Thus, although space-based networks have many interesting and complex characteristics, these characteristics are not necessarily unique, and potential solutions often already exist.

After many years of studies, there is still an amazing amount of misinformation and a number of misconceptions regarding communications protocols in satellite and space-based networks. This paper attempts to provide an unbiased presentation of what one needs to consider when determining what protocols to use – particularly for space-based networks. First, we develop a common vocabulary for communications (link, circuit, packet, frame, etc…). Next, we identify the major protocol suites. This is followed by a general discussion of trade-offs between circuit and packet switching technologies and the pros and cons of various link, network, transport, application and security protocols. Included is a discussion of the applicability of protocol enhancing proxies. NASA, DOD and commercial space communication systems are addressed. We then explore the characteristics of various communication environments (surface, near planetary, and deep space).

II. VOCABULARY

Many times information is misconstrued or misunderstood because the communicating parties are not using a common definition – unbeknownst to them. This is easily understood once one begins to look at the overlap of definitions. This overlap may be due to ones point of reference or area of expertise (e.g. radio communications, telecommunications or networking profession). For example, the following definitions for channel, link and circuit where taken from the Alliance for Telecommunications Industry Solutions (ATIS) [3]. Note the ambiguity.

“Circuit
1. The complete path between two terminals over which one-way or two-way communications may be provided.
2. An electronic path between two or more points, capable of providing a number of channels.
3. A fully operative communications path established in the normal circuit layout and currently used for message, Wide Area Telephone Service (WATS)
Channel
1. A connection between initiating and terminating nodes of a circuit.
2. A single path provided by a transmission medium via either (a) physical separation, such as by multipair cable or (b) electrical separation, such as by frequency- or time-division multiplexing.
3. A path for conveying electrical or electromagnetic signals, usually distinguished from other parallel paths.
4. Used in conjunction with a predetermined letter, number, or codeword to reference a specific radio frequency.

Link
1. The communications facilities between adjacent nodes of a network.
2. A portion of a circuit connected in tandem with, i.e., in series with, other portions.
3. A radio path between two points, called a radio link.
4. A conceptual circuit, i.e., logical circuit, between two users of a network that enables the users to communicate, even when different physical paths are used.

Note 1: In all cases, the type of link, such as data link, downlink, duplex link, fiber optic link, line-of-sight link, point-to-point link, radio link and satellite link, should be identified.

Note 2: A link may be simplex, half-duplex, or duplex.

In general, when referring to a circuit, one usually infers reserved capacity between points, where hard state (fixed software configuration) is used to set up that capacity. We talk of circuits being set up for use when the state is initialized and torn down once done, when that software configuration state is removed from the switching devices. When considering two hosts in a packet-switching network, the term circuit is used most frequently to describe a connection between the hosts that behaves as though it’s a direct connection even though it may physically be circuitous, a virtual circuit. In this case, the two hosts can communicate as though they have a dedicated connection even though the packets might actually travel across a number of separate links before arriving at their destination.

Virtual circuits are connections between two end points passing over a shared packet-switched network of some type. Two types of virtual circuits exist: permanent virtual circuits (PVCs) and switched “temporary” virtual circuits (SVCs). PVCs are always available, whereas SVCs are setup on demand.

Circuit switching is a process that, on demand, connects two or more end systems and permits the exclusive use of the complete path between two terminals over which two-way communications is provided until the connection is released. Here, dedicated capacity is reserved for communication by the endpoints during circuit setup and exists until released. Telephony voice communication at one time was the primary example of dedicated circuit switching; however, packet-switched voice-over-IP (VOIP) is now used throughout much of the telecommunication industry – often with the subscriber unaware of its use.

Packet switching is a technology that explicitly allows the capacity of a link to be shared by numerous users for voice, video and data services, via multiplexing of packets from different sources to different destinations. Quality or Service (QoS) is achieved in such systems by use of traffic shaping, policing at the edge of the network, marking packets with precedence bits and differentiated services as well as reserving capacity. In general, one can achieve far greater overall link utilization with a multiplexing packet-switched technology than with circuit-switching as unused capacity by one service becomes available for use by another.

Traditionally, space-based systems have used separate radio links and separate reserved channels for command and control communications. However, with today’s ability to provide priority to packetized communications, one could easily share a single radio link with many services and still obtain the desired QoS with command and control having precedence over other data types.

Patch switching provides much greater flexibility than circuit switching when there is a diversity of traffic types and burstiness of traffic. With circuit switching it is highly advantageous to know the type of data and amounts of data passing over various circuits in order to manage the bandwidth. Circuit-based switching manages capacity via a combination of connection management or other access control (e.g. manual configuration to allocate bandwidth). Packet-based switching manages data throughput via a statistical combination of queue management, traffic policing and the application of appropriate protocols.

Frames
In communications, a frame is a block of data transmitted as a single entity. Frames can be fixed length or variable length. A frame usually consists of a header and payload. The header provides the necessary information to determine the beginning of a frame (synchronization), the length of the frame, possible the source and destination of the frame and information on how to handle the payload.

In the early days of space communication, processing power was minimal and use of commutation was common. Data was placed into frame structures via commutation prior to transmission to the ground. “Traditionally, telemetry transmitted from the spacecraft was formatted with a Time Division Multiplexing (TDM) scheme, where data items were multiplexed into a continuous stream of fixed-length frames based on a predefined multiplexing rule. To design and implement a data system for spacecraft, each project was forced to develop a custom system used by that project alone [4].”

One applications of framing is to delineate and synchronize data at the media access layer, layer-2 of the ISO network layer concept. Frames can be fixed sized or vary in size. Ethernet, High Speed Serial Interface (HSSI) and ANSI Asynchronous are common data-link framing techniques.
found in commercial-off-the-shelf (COTS) network communication equipment interfaces. Satellite modems use a variety of proprietary and standard framing optimized for various media access techniques such as Time-Division-Multiple-Access (TDMA), Frequency-Division-Multiple-Access (FDMA), Code-Division-Multiple-Access (CDMA) and hybrid combinations (e.g. MF-TDMA, multi-frequency-TDMA). The Consultative Committee for Space Data Systems (CCSDS) protocols specify standard framing formats. For example, CCSDS transfer frames multiplex telemetry packets and Advanced Orbiting System (AOS) data into frames for transmission to the ground.

One can argue that framing at the data link may be the point that provides the greatest interoperability challenge between tradition space communications and today’s common networking communication – not the transport or application layer. There are two main reasons for this: (1) the natural evolution of space-based systems based on their heritage and (2) optimization for point-to-point communications versus optimization for network communications where flexibility is paramount.

Space-based systems evolved from computationally-dumb low-processing-rate systems, and have always had to address radiation requirements and operate in limited powered environments. In addition, traditional space-based systems were point-to-point, space/ground communications – not networked. Hence, systems were designed to optimize processing and power and close the RF link as efficiently as possible. In fact, the framing and coding were even merged to the point where portions of what some consider the physical layer coding are found in the data link layer [22] [Figure 1 and 2]. Thus, it is difficult to separate some of these layers using CCSDS protocols. As a result, a gateway is required to perform protocol translation when moving between CCSDS framing and COTS frames (HDLC, SONET, HSSI, etc…) [Figure 3].

It is extremely difficult to integrate COTS equipment into CCSDS communication systems as portions of the radio reside in the front end processor. For example, NASA integrated a special interface card into an IBM PC docking station to complete the CCSDS coding and framing of the radio system in order to run IP into NASA’s space shuttle over TDRSS. This device is part of the Orbital Communication Adapter (OCA). If the radio were a complete and separate unit with standard interfaces as shown in figure 1, a COTS router could have been utilized with all the additional features provided by a router such as buffering, debugging tools and traffic policing [5].

III. COMMUNICATION PROTOCOLS

A communication protocol is a set of rules governing the exchange of information between entities. In networking, a protocol is a set of formal rules describing how to transmit data, especially across a network. Low-level protocols define the electrical and physical standards to be observed, bit-and byte-ordering and the transmission and error detection and correction of the bit stream. High-level protocols deal with the data formatting, including the syntax of messages, the machine-to-machine dialogue, character sets, sequencing of messages etc.

Protocols are the tools of network communication. There are hundreds of communications protocols, each with own purpose – just like physical tools. Often one can choose from a number of protocols to get a specific job done albeit, one may work better for the specific job than another. However, the “best” protocol may be determined by factors other than simply performance. For example, one may choose a protocol that is readily available and gets the job done over a protocol that is optimal with regard to performance – particularly if the inefficiencies of the protocol are of little concern compared to the costs of design, testing, and implementation. One may also
There are a number of organizations that work on protocols concentrating on various layers of the 7-layer ISO/OSI network model and 4-layer TCP/IP network model [6]. The International Telecommunication Union (ITU), the American National Standards Institute (ANSI), the International Standards Organization (ISO), the Internet Engineering Task Force (IETF), and the Consultative Committee for Space Data System (CCSDS) are just a handful of organizations that work to develop specifications [7-10]. When considering communications and networking, the ITU is most prevalent in RF spectrum and modulation and coding, whereas the IEEE and ANSI activities are most prevalent in the physical and media access (radio systems and hardware specifications). CCSDS was formed in 1982 by the major space agencies of the world to provide a forum for discussion of common problems in the development and operation of space data systems. The IETF’s focus is on the network, transport and application layers of the Internet protocol suite. The IETF does not specify standards at the lower layers. Other bodies, such as the multiprotocol label switching (MPLS) Forum and asynchronous transfer mode (ATM) Forum, promulgate standards for their protocols and technologies.

Layering protocols as in the ISO and TCP models permits one to develop various protocols for each layer. This allows for a divide-and-conquer engineering approach to solving network communication problems and provides tremendous flexibility. However, one does not have to use a layered approach. By merging layers, one can often improve the overall efficiency of a system (size, mass, power, processing) at the expense of flexibility. Merging layers sometimes can simplify the design and reduce the overall processing requirements. Most of the first space missions merged layers as the entire communication systems was well defined and overall efficiency was of far greater importance than system flexibility. As we move from point designs to space-based networks, flexibility becomes paramount.

a. CCSDS Protocols

CCSDS protocols encompass all aspects of the communication network from modulation and coding up through application development. The following are just a small sample of the range of existing protocols that have been developed [4].

Telemetry Channel Coding – establishes a common framework and provides a common basis for the coding schemes used on spacecraft telemetry streams.

Packet Telemetry – establishes a common framework and provides a common basis for the data structures of spacecraft telemetry streams.

TM Synchronization and Channel Coding – specifications for synchronization and channel coding to be used on synchronous data channels.

Proximity-1 Space Link Protocol: Physical Layer – defines the Proximity-1 Space Link Protocol Physical Layer. The specification for the channel connection process, provision for frequency bands and assignments, hailing channel, polarization, modulation, data rates, and performance requirements

b. TCP/IP

Transmission Control Protocol/Internet Protocol (TCP/IP) is a suite of hundreds of network communications protocols dealing with packet formatting, routing, transport, services and applications.

The Internet Protocol version 4 and versions 6 (IPv4 and IPv6) specify packet formats used for routing packets through an IP network. Numerous routing protocols exist including: Routing Information Protocol (RIP), Open Shortest Path First (OSPF) and Border Gateway Protocol (BGP). Reliable transport protocols include the (TCP) and Stream Control Transport Protocol (SCTP). Reliable transport protocols tend to be delay-sensitive due to the fact that handshaking occurs to ensure data delivery: this behavior is worsened by congestion control algorithms ¹. The User Datagram Protocol (UDP) provides unreliable data delivery and forms the bases for many frameworks and provides a common basis for the data structures of spacecraft telemetry streams.

¹ Congestion control algorithms enable applications to share link capacity in a fair manner.
TCP, the Transmission Control Protocol, is probably the best-known and most widely-used networking transport protocol to date. TCP’s two major features are reliable transport and congestion control. The congestion control feature of TCP enables multiple users with different communication paths of different lengths to share a communication link and associated network in a fair manner. TCP assumes it knows nothing of the link and quickly probes the link to determine capacity by exponentially increasing its transmission rate. To date, in the vast majority of systems, packet loss is due to congestion. Thus, once a packet is lost, TCP assumes congestion and exponentially decreases its transmission rate then linearly probes for capacity from that point on. Thus, TCP is a poor tool for large file transfers over noisy paths – particularly if the path is not shared, which is currently the common mode of operations when communicating between spacecraft and ground. In addition, TCP uses a three-way handshake prior to beginning data transmission. As such, TCP is not a good choice for commanding over extremely long delays such as Earth to Mars. TCP has been used in space links and works well for small file transfers and commanding so long as the delays and small inefficiencies are not of concern [13, 14]. TCP performance over various links is very well understood and very well documented [15-21].

UDP is an unreliable transport protocol; packets that arrive are decoded reliably, but the arrival of those packets is not guaranteed. Many reliable and unreliable transport protocols written at the application layer have used UDP as the delivery mechanism for space-based applications. UDP has been used in space as the underlying transport mechanism for blind commanding as well as for the CCSDS file delivery protocol (CFDP), the Saratoga file delivery protocol and for the Multicast Dissemination Protocol (MDP) [22-25].

**BEWARE OF POOR TERMINOLOGY!**

The TCP/IP protocol suite is often referred to as simply TCP or simply IP. This has generated great confusion in the space community as TCP, the transmission control protocol, is not necessarily the transport protocol of choice for communication over long bandwidth-delay links, whereas numerous tools are available from the TCP/IP protocol suite that work well in space and are either delay-insensitive or delay-tolerant. The phase “TCP will not work in space” is incorrect in either case. TCP can work in space, but may not perform as well as it is designed for shared networks, not optimized for dedicated links. However, it is simply wrong and misleading to imply that the TCP/IP protocol suite will not work in space. Such statements are either made out of ignorance or with the intent to mislead.

**IV. SECURITY**

Security protocols exist at all layers of the OSI protocol stack. The most common are link-layer, network layer, transport layer and application layer. One may (and probably should depending on one’s risk assessment) implement security at multiple layers.

Some security mechanisms require the ability to communicate with certificate servers and key management systems in real-time. Such mechanisms are not appropriate for Moon, Mars, and interplanetary communication where access to a remote certificate server is difficult or intermittent.

Link-layer security using shared keys and perhaps even dynamic key updates is possible for near-Earth communications. However, the usefulness of dynamic key updates the need for such link-layer security for near-planetary communications – other than Earth – is questionable. Link-layer security often uses shared keys. Shared keys are relatively easy to manage for small networks such as those that would comprise the Space Exploration Initiative.

Both SCPS and IP security protocols (SCPS-SP and IPsec) can be used for space communications. However, one would generally utilize shared static keys in order to avoid a sophisticated key management infrastructure and as well as to alleviate performance problems associated with dynamic key updates over long delays. SCPS-SP is similar to IPsec transport mode [26]. Both SCPS-SP and IPsec can reside between the transport layer and the network layer or in the network layer between network segments. Both protocols provide integrity, confidentiality and authentication services. The SCPS-SP operates with the assumption that there exists a Security Association (SA) database that contains pertinent security information, for use between the communicating entities, such as the encipher key, the key expiration, the key length, the encipherment algorithm, the integrity algorithm. Use of IPsec for space-based applications would require the similar types of security databases – as would any other security protocol including bundling.

For Interplanetary internet, a bundling security mechanism is being devised. This bundling mechanism applies security more at the data and application layers [27]. As with IPsec and SCPS-SP, some type of pre-placed static keys and security association database configuration are necessary. This is not necessarily part of the security protocols, but rather, part of protocol configuration.

The United State Government has a National Information Assurance (IA) Policy for all U.S. Space Systems [28] which mandates various level of information assurance, particularly with regard to protecting the command and control links. Additional security guidelines are also provided in this administratively-controlled document. To summarize this policy, one should highly protect the command and control links and perform a due-diligence risk assessment on all systems to determine the necessary level of protection for those systems.

**V. GATEWAYS**

Gateways provide a translation interface between two different protocols at the same layer of the protocol stack. Gateways require maintenance when protocols change – and they do change! In addition, gateways can unintentionally
break some protocols as you move data between one protocol with one set of assumptions and semantics and another, different, protocol with different assumptions and semantics. This may be considered a minor inconvenience on the ground compared with getting two different systems to interoperate. In space-based systems, gateway maintenance is much more difficult. Custom gateways are relatively expensive, as they must completely implement and support more than one protocol at a layer. Gateways are a tool that is best avoided if possible. However, sometimes a gateway is a necessary evil.

a. Link-Layer Gateways

Figure 3 shows the need for gateway between the CCSDS telemetry, telecommand, or advanced orbiting systems (AOS) data-link protocols and commercial data-link protocols. This is necessary as large commercial suppliers of networking equipment have indicated that they have no intention of building special interface cards for a small space community. Thus, the space community has to either manage interfaces with special gateways, adopt commercial practices where practical, or has to remain completely separate from other networking worlds, attempting to fund and develop its own brand of networking

b. Performance Enhancing Proxies

Performance Enhancing Proxies (PEPs) are used to improve degraded TCP performance caused by characteristics of specific link environments [29]. PEPs may be employed in satellite, wireless wide area network (WAN), and wireless local area network (LAN) environments. For communication through geostationary satellites and for other space-links with large bandwidth-delay products, PEPs are deployed as middleware in an attempt to optimize control loops in TCP [Figure 4]. In general, PEPs are designed to optimize the transmission control protocol, TCP. PEPs usually have to be able to examine the transport layer of the protocol. Thus, if IPsec or some other layer-3 encryption is implemented, the PEP can not provide the desired improvement on the encrypted traffic. As such, a PEP must be placed before any network layer encryption function (or device). Great care should be taken to fully understand the type of data being sent through the PEP and the environment it is used in.

![Figure 4 - Performance Enhancing Proxy](Image)

2 Often call “link accelerators” for satellite and wireless applications.

c. Interplanetary Gateways (Deep Space)

There are obvious incompatibilities between terrestrial networks and interplanetary systems due to different requirements placed on the terrestrial and space-based networks and assumptions about propagation delay and sharing of links (congestion). Work is ongoing in this arena under the Delay Tolerant Network working group (formally the Interplanetary Internet working group [30]) within the Internet Research Task Force (IRTF). The incompatibilities are due to the nature of the various communication environments depicted in the next section. In order to overcome these incompatibilities, interplanetary gateways are necessary.

VI. COMMUNICATION ENVIRONMENTS

When considering space-based communications there are basically three operating environments: surface, near-planetary, and interplanetary. Each has its own characteristics which dictate what protocols are appropriate and inappropriate for the given environment. Obviously, robustness, reliability and the ability to withstand harsh environments are necessary for the communication equipment. However, this is independent of the actual protocols being deployed.

a. Surface (Terrestrial)

Systems operating on the surface of a planet have similar, if not identical, operating characteristics to those systems on Earth: delay, power and congestion, connectivity and mobility. The problem is actually much simpler regarding protocols as security may not be necessary between terrestrial nodes on the surface of the Moon or Mars. Furthermore, congestion may not be an issue for a small number of users.

b. Near-planetary

Near-planetary systems communicating with their corresponding terrestrial systems have similar characteristics to communication between Earth-based satellites and their corresponding ground-based systems. Regarding Earth-based communications, geostationary satellites (GEO) provide round trip time delays of approximately 500 milliseconds whereas low Earth orbiting satellites have delays in the 100s of milliseconds. GEO [areosync for Mars] provide continuous visibility and connectivity whereas LEO connectivity is on the order of minutes. One should expect similar characteristics for near planetary systems relative to each planet’s orbital dynamics. Thus, solutions that work well for Earth-based space communications should readily apply to other planets.

c. Interplanetary

Interplanetary communication is quite different than terrestrial or near-planetary communications. The general characteristics are: speed-of-light delays, intermittent and unidirectional connectivity, and error-rates characteristic of deep-space communication. One has to take into account when a system will be on and pointed and orbital dynamics in

3 Although not stated, the moon, sun, and any celestial body can be considered – each with its own orbital dynamics characteristics.
order to point-and-shoot to close the link at the proper time. Feedback is very limited due to an extremely long time-delay. Thus, communication methods must be developed that can accommodate such operational environments. Current thought are to utilize message switching [31] and bundling protocols somewhat analogous to email [32-34]. Interplanetary time-synchronization is also critical for interplanetary communication [35].

VII. SUMMARY

This document was generated to help dispel much of the misinformation and misconceptions regarding applicability of various communications protocols for use in satellite and space-based networks. The following key points should go a long way to help one decide on what protocols are appropriate for their particular applications:

- Vocabulary is very important when speaking of networking. Be precise.
- Protocols are simply tools for communication. One size does not fit all.
- Packet-based switching is generally simpler to configure, more flexible and often provides better bandwidth utilization than circuit-base switching.
- The operating environment heavily dictates what protocols can be used – particularly delay, bandwidth, and intermittent connectivity.
- Many protocols in the TCP/IP protocol suite operate well in space. Others, such as TCP or routing protocols are applicable only to surface and some near-planetary applications.
- CCSDS protocols have evolved over time as technology and processing power has improved. Originally designed to optimize power and processing on point-to-point links, CCSDS has begun incorporating networking capabilities with the advent of SCPS.
- Neither IPv4 nor IPv6 interoperate with SCPS-NP. A gateway is necessary.
- Current CCSDS data-link protocols are incompatible with COTS data-link protocols, requiring a data-link gateway for interoperability.
- Many CCSDS protocols – particularly legacy systems – merge layers and thus require application level gateways to operate with COTS protocols such as general Internet protocols. Such merging of layers results in one-off implementation and makes interoperability difficult.
- Great care should be taken when deploying PEPs. Understand their limitations.
- Gateways, including PEPs, must be maintained as protocols change. This can be an expensive proposition.
- Security is difficult anywhere. Sophisticated key management systems are not practical for space-based networks. Thus space-based security architectures should be as simple as policy will allow.

REFERENCES

[17] IETF RFC 3819, “Advice for Internet Subnetwork Designers”
[18] IETF RFC 3150, “End-to-end Performance Implications of Slow Links”
[20] IETF RFC 3366, “Advice to link designers on link Automatic Repeat reQuest (ARQ)"

4 The CCSDS Space Link Extensions (SLE) is designed to help alleviate some of this interoperability problem. However, as with any gateway, all variants of any protocol-pair mapping that must be accomplished, must be implemented. To date, each CCSDS implementation, although based on a standard, is a variant. Thus, implementing wide-spread interoperable SLE services is a daunting task.
[21] IETF RFC 3449, “TCP Performance Implications of Network Path Asymmetry” – give URLs. Non-IETF people won’t know where to find these.


[29] RFC 3135, “Performance Enhancing Proxies (PEPs) Intended to Mitigate Link-Related Degradations”  


