AREx System Architecture

June 17, 2020

This is a draft. Comment and critique welcome and encouraged.

Dr. Chris Bridges M0IEB
Michelle Thompson W5NYV
Thomas Parry

Front Matter

Table of Contents

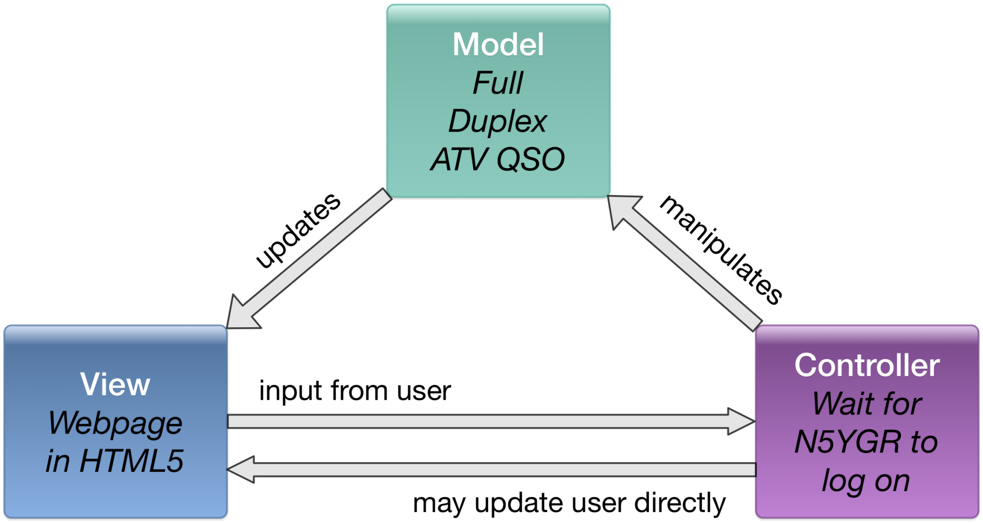
Introduction and Scope

This System Architecture document defines the models, controllers, and views of microwave band communications for Amateur Radio Experiment (AREx).

The **model** defines system structures. These include hardware, software, firmware, and data.

The **controller** defines the logical behavior of the structures defined by the model.

The **view** defines the displays or interfaces. Views connect the user to the controller and allow the user to see updates from the system model.



AREx proposes to design, build, test, and deliver amateur radio satellite service equipment for use in space. Communication services include routine amateur contacts and operation, student contacts, classroom and laboratory educational activities, communications experiments, scientific payload support, the ability to provide emergency or backup communications, and more. In order to provide these services, AREx must ensure reliable digital microwave communications across a variety of mission opportunities. These opportunities include high earth orbits (HEO), geostationary earth orbits (GEO), the near rectilinear halo orbit of Lunar Orbiting Platform - Gateway, and other interplanetary opportunities in the future.

Reliable means messages are guaranteed to reach their destination complete, uncorrupted, and in order. For the downlink, quasi-error free (QEF) is the standard. This means less than one uncorrected error-event per transmission hour, before the application layer. This can be approximated as a bit error rate of 10 to the negative tenth.

System architecture must achieve the following high-level goals.

- It must define a model and control system that prioritizes reliable amateur communications.
- It must define a control system that detects and adapts to digital subsystem failures.
- It must define views that are open, accessible, and documented.

The design has a frequency division multiple access (FDMA) uplink at 5 GHz. The received channelized signals are amplified, filtered, and digitized. The digital data streams are processed, multiplexed, and controlled. The data is packed into Digital Video Broadcast for Satellite version 2 and Satellite version 2 eXtension (DVB-S2/X) frames which are transmitted on a single 10 GHz time division multiplexed (TDM) downlink.

Summary of DVB-S2/X Downlink Protocol

DVB-S2 is the most popular protocol for commercial satellite television broadcast. The forward error correction (FEC) used is nearly ideal, approaching physical limits.

The protocol has a large number of code rates and modulation schemes. Why so many? The protocol was designed to maximize the number of bits transmitted across a wide variety of channel conditions. For an introduction to the basics of coding and modulation and a walk-through of design decisions for an amateur DVB-S2/X downlink, please read Adaptive Coding and Modulation for Phase 4 Ground. It's in Appendix A.

The general approach in DVB-S2/X is to select a symbol rate that corresponds to the available bandwidth, minus a roll-off factor. That rate is fixed for the life of the system. Fixing this value controls the costs of radio frequency hardware and reduces complexity. As different codes and modulations are used, the data throughput goes up or down. Heavier coding, to counteract noise, means that more bits per fixed-length frame are used for redundancy than are used for the data. Simpler modulations transmit fewer bits per symbol than the more complex constellations. Since the symbol rate is fixed, a more complex constellation can transmit more bits per symbol, but requires higher signal to noise ratio.

There are three methods for adjusting to channel conditions to maximize data throughput. First, there is constant coding and modulation (CCM). In general, a designer or operator selects the code and modulation for worst case conditions and leaves them there. This is simple and straightforward, but costs data throughput whenever signal to noise conditions improve above the worst case and limits design re-use. Second, there is variable coding and modulation, (VCM) where the decisions to change modulation and coding are an input to the system. They may be from a sensor or they may be from an operator or control station. Finally, there is adaptive coding and modulation (ACM), which automatically adjusts the code rates and modulation schemes to extract maximum bandwidth given channel conditions. The decision about which coding and which modulation to use can be made as fast as per frame in ACM.

The transport stream, or container format, in DVB-S2/X is MPEG. There are a variety of specifications within the MPEG standard. MPEG defines the digital codes used to represent video and associated audio for digital storage and streaming. The default transport stream of MPEG is replaced by Generic Stream

Encapsulation (GSE) on AREx in order to reduce overhead, allow for maximum flexibility of data types, and to enable a wide variety of experiments and applications.

DVB-S2, DVB-S2X, and GSE are open standards from the DVB organization that require no intellectual property agreements to use.

System Model

Channel Environment

The Earth-Moon channel for AREx is dominated by tracking requirements and path loss. Tracking requirements are explored in the GMAT AREx document.

Mean size of lunar dust is 19 microns but the size distribution is diverse. Lunar dust is 45% SiO2 and 15% Al2O3. It's magnetic, with an Fe patina. The particles are jagged and have a high porosity. Simulated dust causes high friction on tools and equipment.

Will traffic to the lunar surface cause enough dust around the moon to negatively affect the radio environment at the frequencies of interest for AREx?

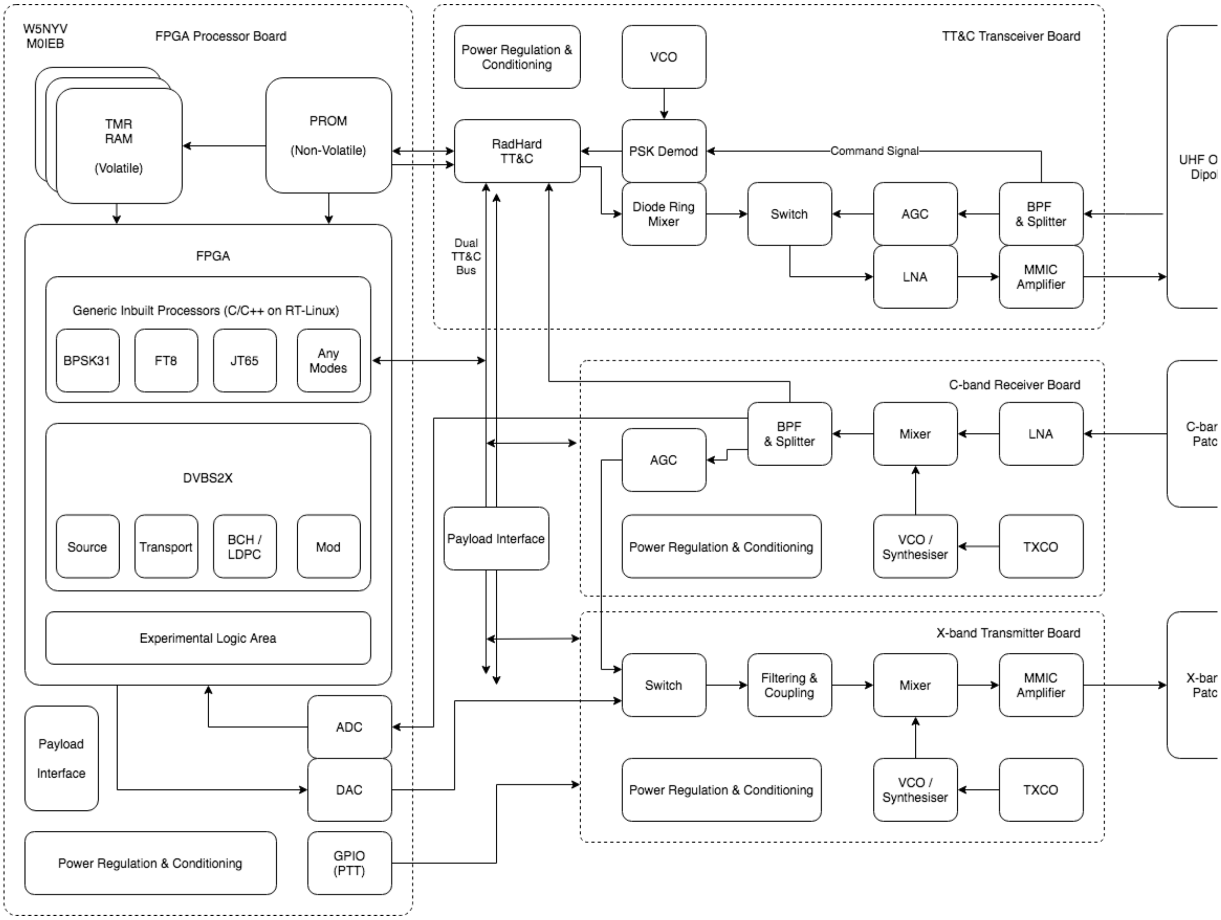
Link Budget

Several link budgets have been drafted. Attitude Determination and Control (ADAC) modes and phases are required to produce accurate link budgets.

System Block Diagram

The high level hardware block diagram is below. It has four sections, TT&C

transceiver, C-band receiver, X-band transmitter, and FPGA processor board.



There are four divisions, called “boards”, in the system design.

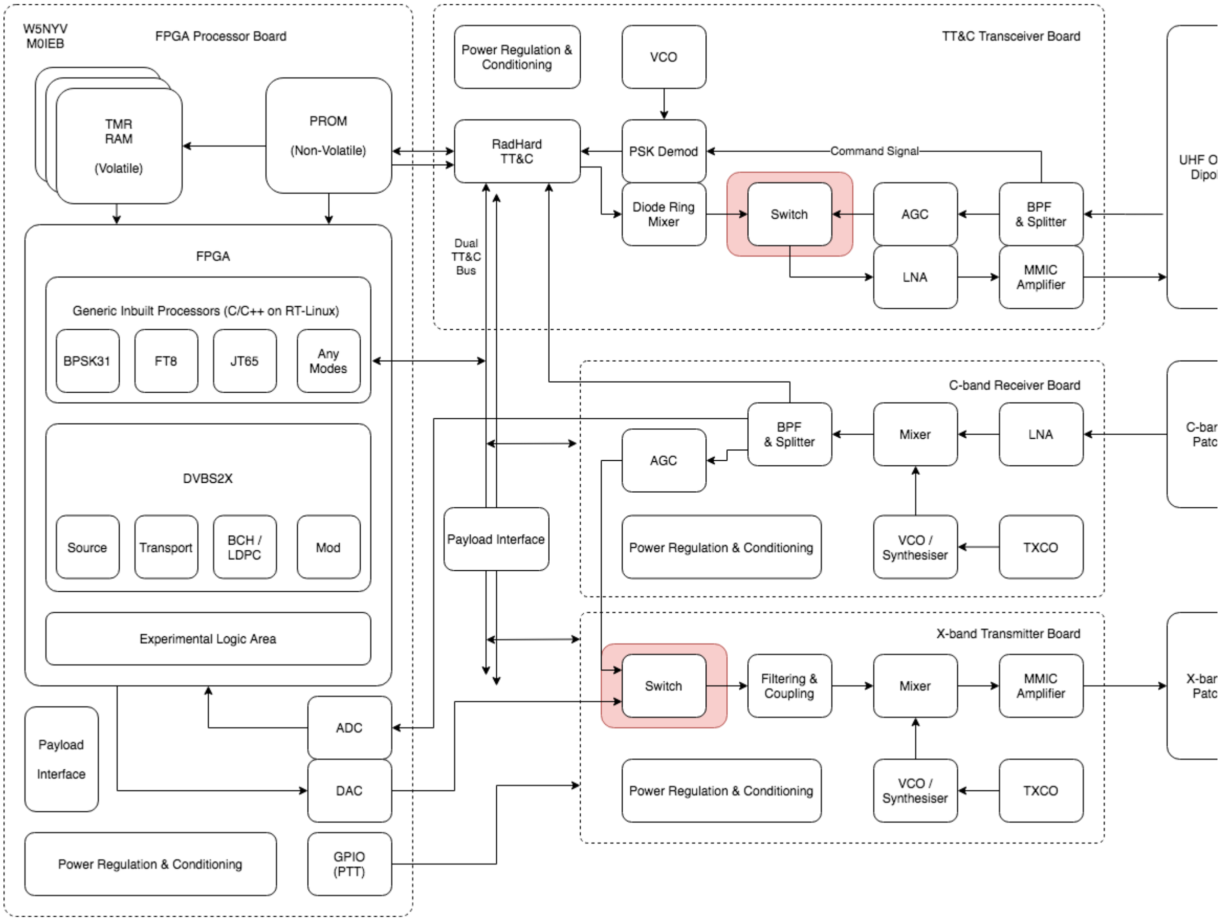
FPGA Processor Board
TT&C Transceiver Board
C-band Receiver Board
X-band Transmitter Board

Antennas are external to the boards and are also considered part of system architecture.

The interfaces between the boards are chosen to allow multiple options for the boards. This means that early designs could have something like a PLUTO for the FPGA Processor Board. Later designs could have a different SDR or custom hardware. The system architecture accommodates modularity.

Hardware States

The two switches in the system diagram create four possible states of radio frequency hardware operation. Switches highlighted below.



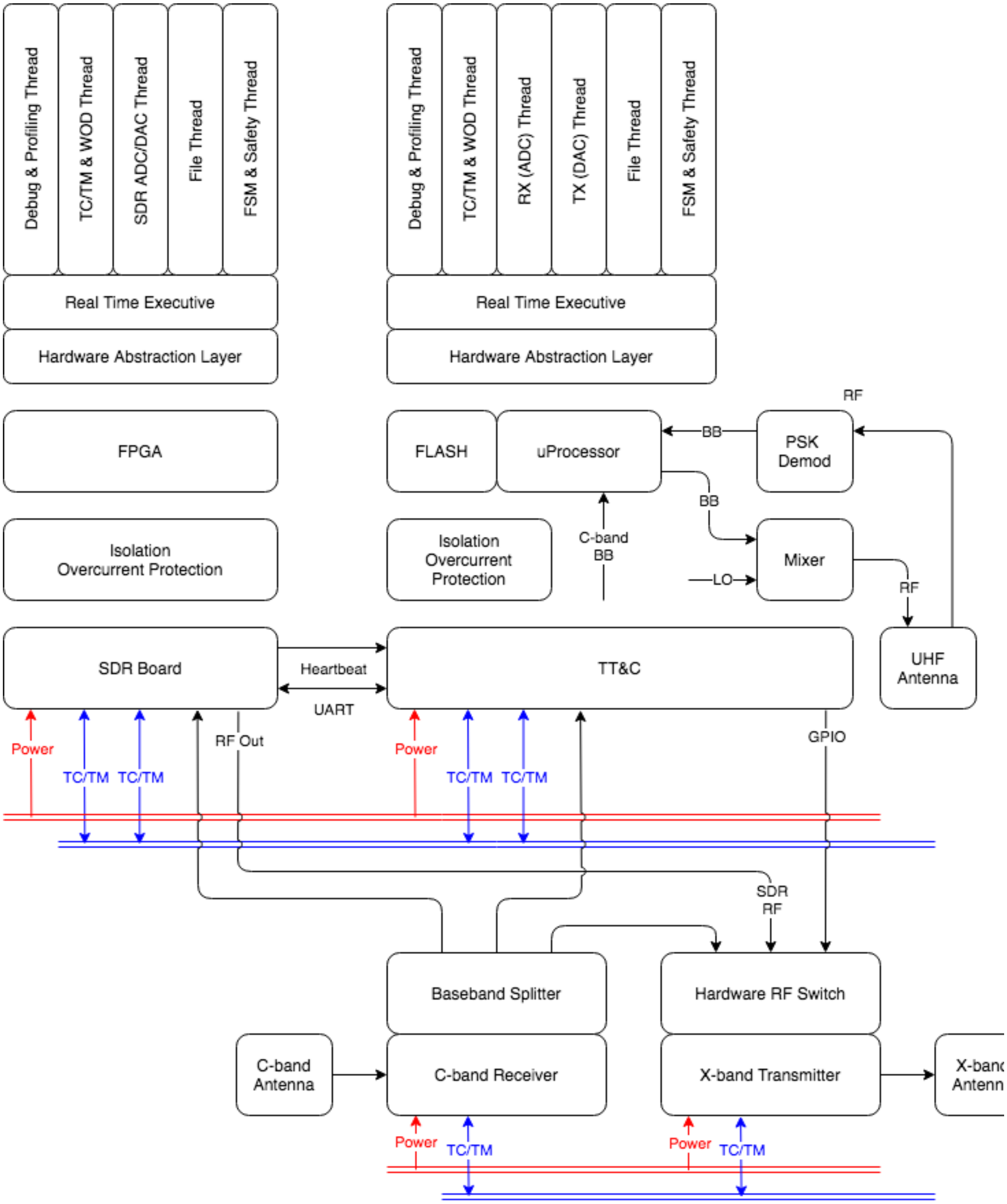
The four RF Hardware States are summarized in the table below. All four states are valid and provide different functionality. Green (darker box background) are default states or define default output.

RF Hardware State	**TT&C Switch**	**Rx C-band**	**Tx X-band Switch**	**SDR**
Transponder	0 = UHF Transponder	ON	0 = X-band Transponder	Not used.
Digital Mode	0 = UHF Transponder	ON	1 = SDR input	Active. Default Output of file-based operations.
Transponder + UHF Telemetry	1 = TT&C Active	ON	0 = X-band Transponder	Not used.
Digital Mode + UHF Telemetry	1 = TT&C Active	ON	1 = SDR input	Active. Default Output of file-based operations.

File-based SDR Operations

SDR has a default directory that is pre-loaded with `control.txt` and IQ files. The IQ files are test modes. They are selected to provide test coverage for space and ground operations. The file `control.txt` is an ASCII file edited by TT&C command. Addition, deletion, enable, disable, and loop are the anticipated commands. This provides a very low bandwidth method to control potentially very high bandwidth SDR output in a useful manner. The same file-driven approach is used in the Telemetry View. AI: add link or reference to that section.

The digital interfaces are summarized below.



User Operating Modes

The user operating modes, description, and control interfaces required are summarized in the table below.

User Operating Mode	**Description**	**Control Interfaces Required**
Analog Transponder	Uplink signals appear in analog downlink.	GPIO controlled.
Digital Transponder	Uplink signals appear in digital downlink. Supports all operating modes. Open loop control. No digital optimization. No Access control.	GPIO controlled
Digital Transponder + Telemetry	Uplink signals appear in digital downlink. Operators must be able to hear themselves in the downlink in order to transmit through the communications resource.	GPIO controlled + Acquisition Control. Acquisition Control consumes and produces Derived and Acquisition Telemetry.
Digital Transponder + Telemetry + Access Control	Uplink signals appear in digital downlink. Operators must be able to hear themselves in the downlink in order to use the communications resources. Operators may be authenticated and/or authorized to transmit through the communications resource.	GPIO controlled + Acquisition Control + Access Control. Acquisition Control consumes and produces Derived and Acquisition Telemetry. Access Control consumes and produces data related to the authentication and authorization protocols desired by the operator.

System Controller

Control of Coding and Modulation

The decisions for which codes rates and which modulation schemes to use per frame depends on the signal to noise ratios reported by the ground stations. More noise on a downlink path means more coding and/or simpler modulation must be used in order to successfully decode downlink frames. This decreases the data throughput.

A low noise (clearly heard) downlink signal means less coding and/or higher order modulation schemes can be used. This increases the data throughput.

By varying the coding and modulation with respect to the signal to noise ratio, the maximum possible bandwidth can be maintained across a wide variety of conditions. A subset of coding and modulation pairs can be selected by the payload. In other words, the payload need not support every code rate and every modulation scheme. For example, the higher-order modulation schemes are intended for high-throughput satellites, and would not be useful for any of the currently anticipated missions. Ground stations are expected to support the entire standard, as they are intended to be re-used for all future systems, terrestrial operations, and to support laboratory and education work.

A finite state machine controls the transition between the code and modulation pairs. A hysteresis curve is used to transition between coding and modulation pairs in order to avoid oscillation between adjacent pairs.

Controlled Response to FPGA Board Failure

A loss of an FPGA board in most systems of this type, where received signals are digitized, processed, combined, and transmitted, often results in the loss of the entire communications resource.

A goal of the system architecture is to mitigate a failure of this type. Therefore, if the FPGA board fails, then received signals are routed directly from the receive board to the transmitter board. This requires additional hardware. If a “heartbeat” is not detected, then switches are activated and RF signals are routed entirely around the FPGA board.

The loss of the coding gain from the FPGA board is non-trivial. However, a controlled response to FPGA board failure preserves the communications resource and increases reliability.

A finite state machine controls the response to a loss of the FPGA Processor Board. The two input signals are hardware heartbeat and software heartbeat.

Acquisition Control

In the Digital Transponder + Telemetry and Digital Transponder + Telemetry + Access Control User Operating Modes, station acquisition is granted if the station wishing to use the payload can hear itself in the downlink. A station receives an occupied channel list from the telemetry view (see System Views, next section). As the station is not yet on the list of stations accepted by the payload, the first transmission is treated as a probe and the candidate station is assigned an unoccupied channel. This probe is not transmitted in the general downlink, but appears in reserved frames.

If the station sees its own ID show up in the list of current stations in the occupied channel list, then it reports success reception back to the payload. All subsequent frames become part of the primary data downlink.

Acquisition Control consumes and produces Derived and Acquisition Telemetry.

Receiving is unaffected by Acquisition Control. Any station can receive without being assigned an uplink channel.

Deliberately assigned communication access provides substantial benefits for the cost of additional complexity.

Blindly transmitting into the payload should not, outside of Analog

Transponder and Digital Transponder User Operating Modes, result in that signal being part of the downlink. This feedback loop does not eliminate all potential problems with interference and misconfiguration, but this control structure gives a substantial benefit for a reasonable cost and advances the amateur radio state of the art.

Access Control

Authorization and authentication protocols are well-studied and can be added to Acquisition Control. These functions allow the communications resource to be reserved or controlled through mechanisms such as allowed and block listing.

Acquisition Control consumes and produces Derived and Acquisition Telemetry.

Access Control consumes and produces data related to the authentication and authorization protocols desired by the operator.

Hardware State Control

Telemetry collection drives the state. Finite state machines and other control interfaces drive the decisions on system state based on the information collected by the telemetry.

Examples of conditions that may drive state changes:

- If something is too hot.
- If SNR is too low.
- If voltage is too high.

AI: The following may be better written as tables. The above sections have a rhythm.

Summary of Control Interfaces

Transmit ON/OFF is a push-to-talk GPIO line for safety and regulatory requirements.

SDR UART has a command line and file transfer interface.

Summary of Finite State Machines

A Finite State Machine in the TT&C manages Switches and GPIO

SDR Heartbeats are GPIO ON/OFF pulses. If stopped, there is an error. A finite state machine manages attempts to reset.

A finite state machine controls the selection of modulation and coding for adaptive coding and modulation.

System Views

A System View defines the displays or interfaces. Views connect the user to the controller and allow the user to see updates from the system model. There are also internal views that connect subsystems and parts to each other. The content here is intended to support an interface control document (ICD).

User View Definition

The minimal guaranteed operating mode is delivered by the analog transponder hardware state. Operators may access the communications resource with legacy communications modes. The mode transmitted in the uplink appears in the downlink signal.

For digital modes, the user's view of the data is determined by the application layer software functions in the ground stations of the operators involved in the communication. Functions in the application layer are supported and enabled by the content of the fields in the downlink packets and information contained in telemetry.

User View Displays

A Pan-Adapter Display of the uplink spectrum accommodates both digital and analog uplinks. A Pan-Adapter display centers the frequency spectrum of interest.

A Multi-Media Display centers the processed data in a User View.

For example, an operator may choose to transmit video from a Raspberry Pi and text from a keyboard, with no audio. They may be receiving AAC encoded audio with still images from a GoPro updated every 20 seconds. They may be joined by someone that just wants to drop off a PDF.

The Multi-Media Display view is HTML5. This allows the data received by the ground station to be accessed with nothing more than a web browser.

Advanced experimenters can take the GSE stream directly and modify existing view software, or write their own view software.

Telemetry View Definition

System architecture must accommodate multiple types of telemetry. These types are Sensed, Derived, and Acquisition.

Sensed Telemetry are values directly sensed and reported. These are values vital for stationkeeping and payload health. These include but are not limited to voltage, current, and temperature.

Derived Telemetry are values calculated from Sensed values. They are in general not directly measured by a sensor. They are calculated using Sensed values by any of a wide variety of techniques ranging from basic math to machine learning.

Acquisition Telemetry is information in the telemetry view that is used for station acquisition, authorization, and authentication. Acquisition Telemetry enables two of the three digital communication states.

The telemetry view is a general purpose telemetry software engine that is driven by data from a file or a table. The engine reads from a file which tells the engine what to do. For example, the AO-51 telemetry program used a telemetry flat file that contained all of the telemetry decoding equations. When the telemetry flat file was edited, the telemetry program then used the new equations. File-driven telemetry control for AREx is expanded to include what text to display and what fields and images to use.

This approach allows operators, students, and experimenters to edit the config file and modify the telemetry display to customize it. This approach is the same approach used to control the file-driven default output for the SDR.

For digital modes, power is a critical value. Values include the expected amount of power received, the expected amount of power transmitted, the noise levels, the measured power received, and the measured power transmitted. These values allow us to derive a very important value called signal to noise ratio (SNR).

SNRs are vital for digital functions such as the station acquisition process and are a necessary input to adaptive coding and modulation. SNR, the list of successfully transmitting and receiving stations (occupied channels), and the list of stations requesting access are the primary inputs to the acquisition algorithm.

The Acquisition Telemetry view can provide the information required for the station acquisition process. The telemetry view would present the identification of all stations that are currently assigned a channel. This is a view that the ground station would receive. If the station successfully hears itself in the downlink after a transmit probe, then it reports success back to the payload. The station is then moved to the main downlink channel and can use the communications resources offered by the payload.

This information can also be interleaved as downlink packets within DVB-S2/X. This approach is considered in-band signaling, in contrast to handling station acquisition through a separate telemetry channel. Either or both approach is valid.

Sensed Telemetry	**Derived Telemetry**	**Acquisition Telemetry**
Voltage	Power	Occupied Channel List
Current		Requesting Stations List
Temperature		

Authors and Contact Information

Dr. Chris Bridges draws upon his experience with multiple ESA spacecraft designs and deployments AMSAT-UK activity. He has extensive academic credentials covering a broad range of modern digital communications theory and practice. Contact Dr. Bridges at c.p.bridges@surrey.ac.uk.

Thomas Parry has experience in the design and manufacture of power and communication systems for ESA, NASA and commercial small satellites. He currently works in mixed-signal IC design for Space and other applications. He is a keen proponent of open-source and community led initiatives. Contact Mr. Parry at yrrapt@gmail.com.

Michelle Thompson is an information theorist specializing in communications theory and design. She is CEO of Open Research Institute, an IEEE Distinguished Visitor, and has contributed to a wide variety of amateur and commercial satellite systems since 1996. Contact Ms. Thompson at w5nyv@arrl.net.

Appendix A

Adaptive Coding and Modulation for Phase 4 Ground

Michelle Thompson W5NYV
Phase 4 Ground Lead

Phase 4 Ground provides digital radio solutions for any payload that complies with the Phase 4 Ground Air Interface document. These projects currently include but are not limited to Phase 4B Payload, Cube Quest Challenge (CQC), Phase 3E, and terrestrial Groundsats.

An Introduction to Coding and Modulation

In analog wireless communications, continuously varying signals are sent from transmitter to receiver. Voice, for example, is directly encoded in an analog transmission by a proportional relationship between baseband and carrier. The changes in audio that make speech intelligible to the ear are proportional to changes in either the frequency (FM), amplitude (AM), or phase (PM) of a transmitted carrier signal.

In digital wireless communications, data such as voice is represented by the digital symbols 1 and 0. Coding is the process of removing unnecessary redundancy in a signal and adding the right type of redundancy. Removing unnecessary redundancy is compression. Adding useful redundancy is channel coding. The type of channel coding we're most interested in is forward error correction coding. This is a way of coding the data where we can recover corrupted parts of the signal.

When we talk about **code rate**, we are talking about the ratio of how many bits go in to the forward error correction coder, or **encoder**, over how many go out. A rate 2/3 code takes in two bits and produces three. The extra bit is produced with mathematics especially designed to make the signal more durable as it travels from transmitter to receiver. The more bits you add, the smaller the ratio. Rates up to 1/9 are common. For a rate 1/9 code, for every bit that goes into the encoder, nine come out. As you'd expect, the more coding, the more durable the transmitted bits are against noise and interference. However, there's a cost. If you compare two signals that are transmitted at the same rate, the one with more extra bits to protect it needs more time to get through. The data rate is lower. It takes longer to transmit the same amount of data.

After the data is channel coded, the resulting bits are transmitted. The simplest type of digital waveform has two distinct states. One state corresponds to a 1, and the other state corresponds to a 0. Each of these ready-to-transmit-values is called a **symbol**. When we send one bit at a time, we have two symbols to choose from. An example of this type of modulation is Binary Phase Shift Keying (BPSK). The **modulation order** is the number of symbols we have to choose from. For BPSK it's two.

This simple BPSK modulation scheme can be dramatically improved. Sending one bit at a time is a great start, but we can do a lot better. If we use four distinct states in our transmitted waveform, then we can send binary data two bits at a time. We now have four symbols instead of two. An example of this type of modulation is Quadrature Phase Shift Keying (QPSK). The modulation order has doubled to four.

How about 8? 16? 32? Yes, to all, and more, all the way up to 256, 512, and even 1024! Sending 1024 bits in a single sample sounds amazing. So, why don't we just send 1024 bits in a single sample all the time?

Engineering is all about trade-offs, and there's another one right here in front of us. The higher the modulation order the more power required. This means that the signal carrier power for transmitting two bits at a time must be twice that of transmitting one bit at a time, assuming that we are transmitting at the same **symbol rate**. We pay for the doubling in information capacity by having to provide double the power. As long as you have enough power, you can use more powerful modulations. If you have too much noise or not enough power, then you have to drop down to a lower modulation order.

Coding and Modulation Techniques in DVB

Traditional communications design assigns a fixed **mod**ulation and forward error correction **cod**ing (MODCOD) to a link. The MODCOD is selected to provide reliable communications under worst case conditions. For example, a microwave link that points down off a mountain is often designed to be good enough to work through rain fade and summer foliage. During clear conditions in the fall with no leaves, plenty of excess link margin is available, but a fixed system designed to work through summer thunderstorms cannot take advantage of this margin. In the Digital Video Broadcasting (DVB) world, this technique is called Constant Coding and Modulation (CCM). Phase 4 Ground uses many DVB protocols and techniques due to their high quality and widespread use in industry. Adapting these protocols to amateur radio is part of our mission.

Since it makes sense to adjust our link to better match observed conditions, one can design a system that uses a variety of MODCODs. An operator can then observe the link and then adjust the MODCOD to take advantage of better conditions. This technique is called Variable Coding and Modulation (VCM). VCM requires intervention of some sort to accomplish. In general, there is no feedback path from the receiver to the transmitter and a human is involved. But what if there was a feedback path from the receiver to the transmitter?

Adaptive Coding and Modulation (ACM) is a technique where the modulation and forward error correction are automatically changed in response to link conditions. As the link improves, higher order modulations and less coding allows increased throughput. Throughput can increase to take better advantage of available link margin. Challenging link conditions are responded to by lower order modulation and more coding. The throughput will decrease, but the link is maintained. The adaptation is enabled by establishing the set of MODCODs to be used, listing the metrics that control the decision to change MODCODs, and defining the algorithm that produces the decision. These three ingredients make up ACM.

With a CCM systems, severe fades can cause total loss of the link and zero throughput. VCM can address some of the challenges of severe fades, but ACM automatically turns fade margin directly into capacity. Maximizing throughput is highest with ACM.

Adaptive Coding and Modulation in Phase 4 Ground

The first challenge to an amateur-radio-centric version of ACM is that all existing implementations of ACM are proprietary. ACM is used in landline modems, 802.11, terrestrial microwave communications, and satellite links. When you are making money with subscribers, leaving margin on the table is not ideal. More efficient links mean more capacity, and more capacity means more subscribers, and more subscribers means more profit.

Most commercial ACM links generally only connect amongst themselves. There is no reason to create and maintain an open standard. Therefore, outside of the limited advice given in the implementation guidelines for DVB and a few white papers from a few companies, there is no open standard for ACM that we can simply adopt. For Phase 4 Ground we have to develop our own implementation of ACM, document it fully, and adjust it as we learn more in the field.

This is a great opportunity for amateur radio. Documenting the engineering trade-offs made in an advanced digital wireless system provides enormous educational opportunity for a wide variety of people, from interested amateurs to engineering students to satellite startups to people interested in machine learning and cognitive radio. Providing a working open-source implementation of ACM that other amateur projects can consider for adoption is a valuable engineering service.

The particular radio problem that has to be solved for space payloads is relatively straightforward. The geostationary and lunar and beyond radio environments are well-characterized. The available modulation schemes and coding rates are drawn from an established set described in the DVB standards (freely available from https://www.dvb.org). Advice from commercial and academic sources exist.

The particular radio problem that has to be solved for terrestrial Groundsats is also relatively straightforward. Groundsats are terrestrial versions of space-based payloads. They provide all the functions of an orbiting platform, but are on the ground. The control loop for terrestrial ACM has to be able to respond faster than for space. This is still well-characterized and advice exists from commercial and academic sources.

DVB allows an extreme resolution of MODCODs. Each and every frame can have a different MODCOD. This enables a link to respond very rapidly. For receiving transmissions from space, rapidly changing links are not the norm. The primary challenge is weather and rain fade or dishes not quite pointed right. For terrestrial links, changes in link quality can be more rapid, especially if the station is mobile. Terrestrial links have multipath, obstacles, noise, signal interference, and can also have rain fade and pointing problems.

There is a simple equation for ACM. In DVB, and for ACM in particular, the symbol rate is fixed. This greatly simplifies the communications system design. After a symbol rate is determined, a set of MODCODs is selected. The bit rate expression is therefore

Bit rate = symbol rate * modulation order * code rate

There are a lot of MODCODs to choose from in DVB. For space projects, the DVB-S family is the standard to reference. For terrestrial, we look to DVB-T. S stands for Space, and T stands for Terrestrial (think “television”).

Phase 4 Ground uses DVB-S2X and DVB-T2. The 2 in DVB-S2X and DVB-T2 stands for second generation. Second generation DVB-T2 and DVB-S2 is backwards compatible (with some effort) to the first generation DVB-S and DVB-T. Second generation standards provide substantial improvements over first generation.

DVB-S2X is an extension to DVB-S2 that provides additional MODCODS and some additional mechanisms. Of compelling interest to us is the additional MODCODs at the lower end of the spectrum that provide enhanced very low signal to noise (VL-SNR) operation. For CQC, VL-SNR operation will provide needed support. For Phase 4B Payload, VL-SNR allows for reasonably sized dishes and opens up the possibility of patch arrays.

A large advantage gained in choosing DVB standards is that the receiver is explicitly told, frame by frame, exactly what MODCOD has been chosen. The receiver does not have to do anything extra to use ACM. The complexity of ACM is in the transmitter.

The second challenge to an amateur-radio-centric version of ACM is that ACM assumes exactly one intended receiver. If the transmission is a QST or CQ, or intended for a roundtable talk group, or is merely open to monitoring by silent listeners, modifications to the standard ACM scheme will be needed.

Maximizing The Bit Rate

There is a very important distinction between analog and digital systems and how to interpret the guidance for best operating practices as set out in part 97.

In analog communications in amateur radio, there is a principle of conservation of power. The least amount of power should be used to ensure reliable communications in normal operations.

Part 97 : Sec. 97.313 Transmitter power standards

> (a) An amateur station must use the minimum transmitter power necessary to carry out the desired communications.

Obviously, emergencies may require a different practice. In digital communications, the spirit of this guidance is best achieved with maximizing the bit rate, or throughput. Maximum bit rate ensures that the communications are achieved in the most efficient manner by providing maximum capacity. If this means transmitting at a higher power than is necessary to simply maintain a communications link, then so be it. It's better to transmit for 450 milliseconds and then almost immediately allow someone else to then use the channel than to transmit for 450 seconds on minimum power using maximum coding and the lowest modulation scheme before relinquishing that particular channel. We equate bit rate with power and assert that this complies with the spirit of part 97.

We want to maximize throughput. This means maximizing the bit rate. In order to get to maximum bit rate, the professional advice is to start out with a stable link and work upwards. Here's an excerpt from Work Microwave's website.

> Start conservatively, approach the optimum: When setting up a link it is wise to start with very conservative settings to have a stable link running in the first place. Even if the "first shot" has not the desired bandwidth efficiency, an incremental approach will be the best way to optimize the link once it is up and stable. Due to numerous parameters and conditions affecting the Es/N0, the best settings will be reached by trial and error and can hardly be predicted beforehand.
>
> "ACM Dos and Don'ts." Work Microwave, 13 Mar. 2016, https://work-microwave.com/acm-dos-donts/

The Es/N0 value is a big clue. It's a critical metric for ACM. It stands for energy per symbol divided by the noise power spectral density. We already know what symbols are. A symbol is the distinct states of the modulator. The simplest one transmits 0 and 1. Two symbols are able to be transmitted so the modulation order is 2. Next most complex is 00, 01, 11, and 10. Four symbols are able to be transmitted so the modulation order is four. Next most complex is 000, 001, 010, 011, 100, 101, 110, 111. Eight symbols are available to be transmitted so the modulation order is eight. An example of this type of modulation scheme is 8PSK.

Energy Per Bit

Es/N0 is commonly used in the analysis of digital modulation schemes, but we're going to dig deeper and look at at a quantity called Eb/N0. This is the energy per bit divided by the noise power spectral density. Eb/N0 is the normalized signal to noise ratio of our link and this value is what drives the adaptation decisions in ACM. Think of Eb/N0 as the signal-to-noise (SNR) per bit. The energy per symbol and the energy per bit are related by the following expression.

Es/N0 = Eb/N0 * Log_2(modulation order)

So for the modulations that we listed above, we have the following relationships.

Es/N0 = Eb/N0 * Log_2(2) *two symbols to choose from*

Es/N0 = Eb/N0 * Log_2(4) *four symbols to choose from*

Es/N0 = Eb/N0 * Log_2(8) *eight symbols to choose from*

This gives us

For modulation order 2: Es/N0 = Eb/N0

The energy required to transmit a symbol of 0 or 1 is the same as required to transmit 0 or 1 bits. Makes sense!

For modulation order 4: Es/N0 = Eb/N0 * 2

The energy required to transmit a symbol of 00, 01, 10, or 11 is twice as much as required to transmit a 0 or 1. Still makes sense.

For modulation order 8: Es/N0 = Eb/N0 * 3

The energy required to transmit a symbol of 000, 001, 010, 011, 100, 101, 110, 111 is three times as much as required to transmit a 0 or 1. We are seeing the pattern.

In ACM, we have to be able to decide when we can afford to move on up to the higher order modulation schemes, which allows us to transmit more bits at once. If all the power we have available to us amounts to about as much power as required to transmit one bit, then we are stuck transmitting one bit at a time in BPSK. If our metrics tell us that we have more than three times the power required for a single bit available to us, then we can transmit a symbol that stands for three bits at once. We can go with 8PSK.

Within the modulation schemes are sets of coding rates. We've seen how spending power can increase the bit rate. How does coding fit in?

Coding Gain

There are two major types of coding. **Source coding** removes unnecessary redundancy so that source data can be more efficiently stored and handled. For example, digital music and video is source coded for compression. Otherwise the directly sampled files would be enormous.

Channel coding puts back in the right type of redundancy to make the transmitted signal resilient. Forward error correction puts in additional bits that allow for both the detection and correction of errors. Better than magic!

In DVB-S2X, the forward error correcting code is called LDPC-BCH. It's an advanced **concatenated block code**. Block code means that groups of bits are gathered up and then mathematically modified with extra bits. There are other types of codes that operate on continuous streams of bits. Those types of codes operate bit-by-bit as long as there are bits in the pipeline. Each block stands alone and is decoded separately. Concatenated means that two different codes are used. The reason these two different codes are used together in DVB-S2X is because using them together cancels out weaknesses. Taken together they make a very high-performance code.

Coding gain is the measure of the difference between the Eb/N0 levels of an uncoded system when compared to a coded system, when both systems are required to provide the same bit error rate. We have the same signal energy available in either case. Coded signals allow us to correct errors, which allows us to transmit at less power.

What can with do with this extra gain? In ACM we can put it right to work in maintaining target bit error rate performance. If we know what Eb/N0 we need, and we know which codes consume that much Eb/N0 to maintain a particular performance level, then we are able to select the code that maximizes bit rate while minimizing bit error rate.

We do this by measuring Eb/N0 at the receiver. This tells us how strong the signal is. Eb/N0 is reported to the ACM controller, and the right modulation and coding is selected for that receiver. In commercial satellite, the ACM controller is centralized and is usually on the ground. For Phase 4B Payload and for Groundsats, it's planned that the controller will be onboard the satellite.

Changing the modulation is the coarse-grain control knob in ACM. Changing the code rate is the fine-grained control knob in ACM.

Putting Metrics, MODCODs, and Algorithms Together

For ACM to work, the modulator must send which MODCOD is being used at the start of each frame. The receiver must be able to handle an arbitrary change in MODCOD without any advance knowledge. The receiver must be able to work fast enough to process the packet or frame without corrupting or dropping it. This puts a lot of pressure on the receiver. This pressure can be alleviated in several ways. One example is using standardized mechanisms in DVB such as time slicing. See Wally Ritchie's paper "Using DVB-S2X and Annex M to implement low-cost Phase 4B Earth Station Terminals" for ideas on time slicing.

Another requirement is that the receiver needs to be able to measure or calculate an estimate of the link quality (Eb/N0) and then communicate this estimate to the payload. The payload must be able to process this reported metric and then adapt the data rate and change the MODCOD sent to the receiver. This maximizes the data rate, complies with the spirit of part 97, and is sparkling with efficiency and style.

Reacting to changes in channel quality makes sense. But can there be additional improvement? Yes, there can! There's a large body of research that shows how throughput and bit error rate performance changes when using linear prediction to estimate the future state of the channel based on past measurements.

There are practical limits to how quickly an ACM system can respond. In general, about 1dB per second is achievable. If something happens and the demodulator comes unlocked, then it's a good idea to go back to the lowest MODCOD. This way, you're starting over with the highest probability of re-connecting and then working your way back up to maximum bit rate.

Assuming that the receiver has acquired the satellite and done all necessary chores to receive the downlink, and assuming the receiver has the necessary authentication, and assuming the receiver can successfully determine which channels are free for transmission to the payload, the receiver now needs to determine what MODCODs it is capable of receiving.

The dish might not be pointed correctly. The receiver might be a bit noisy. The local oscillator might not be rubidium quality. There might be some atmospheric conditions that attenuate the signal. Someone could have dented the dish. A connector could be loose. Some of these factors change very slowly over time, and some of them change more quickly. All of these factors affect receive capability and all of them can be automatically accommodated with ACM.

The standard model of ACM has the receiver monitor and report its Eb/N0 to the controller. In our case, the controller can be in the payload. When Eb/N0 falls below a setpoint, the receiving station is sent a lower MODCOD. The setpoints are configured to provide a minimum level of performance. When going to a lower MODCOD, throughput is reduced but the link is maintained. Eb/N0 reports can be part of the frame structure.

Digital communications performance can be defined by maximum allowable bit error rate. DVB is designed to provide very low error rates. The standard of performance for DVB is called quasi-error-free. DVB allows one uncorrected error per hour of video broadcast viewing. This is a very high standard that works out to a bit error rate of about $1*10^{-10}$ to $1*10^{-11}$.

When you establish the values for Eb/N0 that you're going to allow, they have to be made based on what bit error rate you can tolerate. Quasi-error-free bit error rate in DVB is many orders of magnitude lower than, say, the maximum bit error rate for GSM ($1*10^{-3}$) and lower than the data-centric maximum bit error rate for 3G data ($1*10^{-6}$). Voice is more forgiving than data which is more forgiving than digital video broadcasting. Selecting a baseline bit error rate of $1*10^{-6}$ is not out of line.

Once you have a bit error rate that you want to keep below, then you calculate a table of Eb/N0 values that would cause you to move to a better MODCOD. "Better" could mean higher or lower depending on whether you were doing great or having trouble with the link.

Anyone that's ever worked with set points knows that there's a potential for oscillating when the measured value is very close to the set point. A common approach with ACM is to put in 0.3dB or more of hysteresis. Going up requires a bit more SNR than coming down. This doesn't just prevent oscillating between two MODCODs but can also help maintain demodulation lock. You don't want your radio to work any harder than it has to.

We want the operator to see as much information about the metrics and the link as they desire. Our goal is to provide quality presentations of signal-to-noise ratios, states of lock, channel occupancy, system status, Usersynchronous log visualizations, symbol rate, modulation constellation, data rate, bit error rate, and more. Metrics such as these and more are presented by an application that can be run or not, depending on the preferences of the operator. Some systems provide a bit error rate tester as an application. This can help debug situations of synchronization loss, unexpected bursts of bit errors, or other problems. If the operator doesn't want to see any of this, then they don't have to. It should "just work" without intervention, and provide clear error or failure messages if anything goes wrong.

When a higher MODCOD is selected, the available data rate is increased. This usually isn't a problem. When a lower MODCOD is selected, the available data rate is decreased. This can be a problem. Congestion control must be considered and implemented to avoid losing packets or frames. Buffers and FIFOs to the rescue!

Is maximizing the bit rate enough? What about latency? While ACM considered in the abstract doesn't minimize or maximize latency, the use of DVB-S2X can offer some relief over DVB-S2. Latency will be one of the biggest challenges to operator experience on the Phase 4B Payload. It is impossible to go faster than the speed of light, and the round-trip delay of at least 240mS is substantial. There are things that we can do to mitigate latency such as reducing buffer size and using shorter frame lengths. Providing voice memo as an alternative to real-time voice is another.

Proposed Adaptive Coding and Modulation Scheme

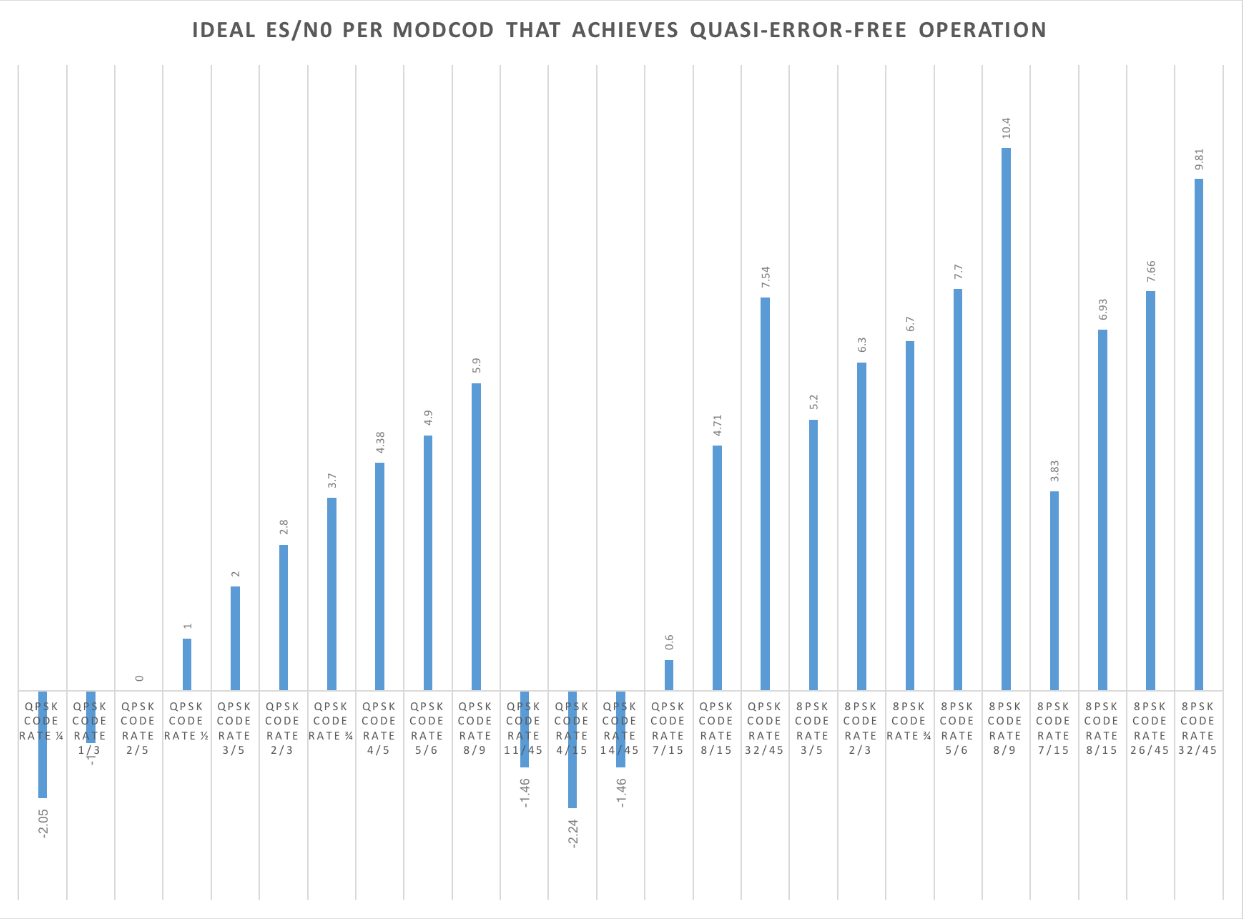
Here's the current proposal for MODCODs, metrics, and algorithm for ACM for Phase 4 Ground. This is an open design that is going into prototyping and testing. The expectation is that this proposal will be reviewed, refined, and retuned to maximize bitrate and avoid commonly encountered challenges. Some challenges are anticipated and have been mentioned above. Others we will certainly discover along the way.

There are choices of frame size in DVB-S2 and DVB-S2X. The CCSDS (Consultative Committee for Space Data Systems) RF Modulation and Channel Coding Workshop, among other individuals and groups, recommends the short frame size for near space-earth transmissions. A selection of the short frame size MODCODs that we believe will work best for Phase 4B Payload are presented in the table below. Short frame size is 16200 bits. Frame size and the presence or absence of pilot signals is communicated in the TYPE field of the physical layer header. Each MODCOD has an identification code. The decimal value of that code, which goes into the PLS field of the physical layer header, is the first column. Ideal Es/N0 is ideal energy per symbol divided by the noise power spectral density in dB in order to achieve a frame error rate of 10^{-5}. This is quasi-error-free operation with no impairments. In other words, very ideal!

PLS Code	**MODCOD Name**	**Rate**	**Ideal Es/N0**
1	QPSK code rate ¼	1/4	-2.05
2	QPSK code rate 1/3	1/3	-1.00
3	QPSK code rate 2/5	2/5	0
4	QPSK code rate ½	½	1
5	QPSK code rate 3/5	3/5	2
6	QPSK code rate 2/3	2/3	2.8
7	QPSK code rate ¾	¾	3.7
8	QPSK code rate 4/5	4/5	4.38
9	QPSK code rate 5/6	5/6	4.9
10	QPSK code rate 8/9	8/9	5.9
216	QPSK code rate 11/45	11/45	-1.46
218	QPSK code rate 4/15	4/15	-2.24
220	QPSK code rate 14/45	14/45	-1.46
222	QPSK code rate 7/15	7/15	0.60
224	QPSK code rate 8/15	8/15	4.71
226	QPSK code rate 32/45	32/45	7.54
12	8PSK code rate 3/5	3/5	5.2
13	8PSK code rate 2/3	2/3	6.3
14	8PSK code rate ¾	¾	6.7
15	8PSK code rate 5/6	5/6	7.7

16	8PSK code rate 8/9	8/9	10.4
228	8PSK code rate 7/15	7/15	3.83
230	8PSK code rate 8/15	8/15	6.93
232	8PSK code rate 26/45	26/45	7.66
234	8PSK code rate 32/45	32/45	9.81

When we look at a chart of these MODCODs, we can see the effect of modulation and coding. We get about 12dB of range just using QPSK and 8PSK. We haven't yet listed the VL-SNR codes that can bring the Es/N0 down to -10dB. They require some additional care and work to implement.



We need to select enough different MODCODs to give the performance we want, but not so many that we have a situation where the algorithm is flailing about making unnecessary changes. The starter list of MODCODs is the following. This gives a MODCOD at about every 2-3dB.

QPSK 4/15 (identification number 218) -2.24
QPSK 2/5 (identification number 3) 0

QPSK 2/3 (identification number 6) 2.8
QPSK 4/5 (identification number 8) 4.38
8PSK 5/6 (identification number 15) 7.7
8PSK 8/9 (identification number 16) 10.4

All measurements have error. There are multiple sources of error and noise. The set of target Es/N0 (or Eb/N0) numbers need to be far enough apart to where link performance instead of noise is the main trigger of an ACM decision.

If three MODCODs turn out to be the best match, then it means we use three MODCODs. If we can use more, then we use more.

Once the MODCODs are selected, hysteresis is applied, and the metrics are monitored, then the choice of which MODCOD to apply to which frame can be usefully made.

While the underlying mechanism is straightforward, there are many problems to solve. Flow control and what type of quality of service needs to be decided. The DVB implementation guidelines give a great start for ACM and describe ways to set up Generic Stream Encapsulation (GSE) to help implement ACM.

Appendix B

Authentication and Authorization in a Multiple-Access User Amateur Radio Satellite Service Digital Payload

Michelle Thompson W5NYV

Paul Williamson KB5MU

Definitions

Authentication is the process of confirming a licensee's declared identity.

Authorization is the process of allowing particular identities access to particular resources.

Satellite is any equipment that serves as the payload for any Phase 4 Ground system. This includes but is not limited to an orbiting satellite payload, satellite simulator (Groundsat), or terrestrial hub.

Phase 4 Radio or ground station is equipment that complies with the Phase 4 Air Interface.

Black List is a list of stations that are positively not authorized to transmit through the satellite. There are probably at least two levels of blacklist: stations that have merely failed to authenticate, and stations which are permanently banned regardless of their authentication status.

White List is a list of stations that are positively authorized to transmit through the satellite.

Ground Control Station is a station that can command the satellite.

Misuse is communications that are illegal or damaging to the communications system.

Introduction

If the satellite has no state, then every frame has to speak for itself. Therefore, authentication and authorization are affected.

Logbook of the World (LoTW) is an QSO confirmation (QSL and awards verification) application from ARRL, built upon the OpenSSL family of cryptographic functions. In order to prevent QSL fraud, LoTW issues certificates to verified licensees that are used to cryptographically sign QSO records, certifying them as originating with the station licensee. These certificates could be used to sign other records for our purposes.

https://lotw.arrl.org

Within the United States, successful LoTW account establishment confirms that the address on file at the Federal Communications Commission (FCC) for an amateur radio licensee can receive and respond to information received at that postal address. That serves as the basis for licensee authentication.

The address given to the FCC comes from the address that that license applicant submits on their form 605. The FCC does not validate this address. The FCC relies upon the volunteer examiner coordinator (VEC) to review the form 605 and ensure that the applicant has provided a mailing address in the United States. The applicant provides the address. The VEC is not required to independently confirm this mailing address. The applicant's name on the form 605 must match the name on accepted forms of identification.

This is the limit of authentication provided to, and therefore by, the licensing authority. In the US, this is the FCC.

There are two fundamental questions.

Does it make sense to exceed this level of authentication?

What sort of authentication and authorization is required of Phase 4 Ground radio traffic?

The default security implementation of Phase 4 Ground is the process described within this document. The Phase 4 Ground process must defer to a process selected by the satellite, if that security process selected by the satellite needs to claim priority. If there is a different process selected by the satellite, then a full specification must be provided to Phase 4 Ground as part of the Air Interface documentation. This serves as a "we can't read your mind" clause.

Phase 4B Satellite State

Phase 4 Ground considers a process of authentication and authorization in this document. This process involves state in the satellite and in the Phase 4 Radio.

The state in the satellite requires the following things.

1. Memory for a database.
2. The ability to read from this database.
3. The ability to write to this database.
4. The ability to demand and process authentication from the Phase 4 Radio.
5. The ability to do some processing using data received from uplink communication frames and information from the database.

The state in the Radio requires the following things.

1. The ability to generate a token.
2. Memory to store the token.
3. Memory to store private key.
4. Memory to store a certificate.
5. Ability to sign a message containing the generated token with the private key.
6. Ability to send the signed message plus the certificate to the satellite when it is demanded by the satellite.

Authentication Process Overview

The process is as follows.

Stations that comply with the air interface can transmit through the satellite. The station is configured with an amateur radio callsign, plus a Secondary Station Identifier (SSID) of TBD bits to allow a particular licensee to operate multiple simultaneous stations, and with the ARRL-issued private key and certificate for that callsign. In addition, the station must generate a token of TBD bits chosen at random. Every transmitted uplink frame contains in its header the callsign, SSID, and token.

The callsign and SSID will be retransmitted in the downlink frame header, but the token is never transmitted on the downlink. Since it is fairly difficult to intercept an uplink transmission, this makes it difficult for an impostor to hear another station's authenticated callsign:SSID and start using it.

The satellite stores the token, the claimed call sign, the claimed SSID associated with the call sign, and a time stamp. This is a tuple that forms the rows of a database.

(satellite time stamp : callsign : SSID : token)

When each uplink frame is received by the satellite, it decides whether to accept it for retransmission on the downlink, or discard it. It may also choose to initiate an authentication transaction with the ground station. Alternately, or in addition, a Ground Control Station may initiate an authentication transaction with any or all active station(s). Unless and until an authentication transaction with a given station has been attempted AND FAILED, the satellite must accept its frames for retransmission (unless that station has already been blacklisted for another reason).

This achieves two important goals. Communication is unimpeded, and the loss of Ground Control Stations can be well-tolerated. When Ground Control Stations are not required for normal communications, system durability and reliability is greatly increased.

When authentication is not frame-by-frame, or required to initiate the process of accessing and transmitting through the satellite, efficiency and performance are greatly increased. In particular, the very first message from a station is not delayed for formalities or blocked entirely; in an emergency situation this could be essential.

Mitigation of Misuse

The risk of bad actors is recognized, but management of bad actors is achieved through Black Listing known bad actors after complaint or system statistics or some other method reveals that the problem is Misuse.

Phase 4 Ground recommends that Misuse must be assumed to be misunderstanding or misconfiguration until proven otherwise. If Misuse is deliberate and there are irreconcilable differences, then Black Listing is the mechanism for resolving the Misuse.

The satellite may be configured to automatically initiate an authentication with each new station it receives, and/or periodically at some interval, or it may rely entirely on Ground Control Stations to request authentications. The satellite controls the rate at which authentication transactions can take place, so it can never be overloaded by the required calculations.

When there is a cause for authentication, a request for authentication can be made by a Ground Control Station. Requests for authentication can be made to all rows with expired time stamps (time stamp < some value : * : * : *), a particular row of the database (* : call sign: SSID : *), all rows that have a particular call sign (* : call sign : * : *), all call signs (* : * : * : *), or some other combination.

Regardless of how the authentication is initiated, the transaction begins with a message from the satellite addressed to one particular (call sign : SSID).

Upon receiving a request for authentication, the station addressed would generate and transmit a response message with the usual header (including callsign, SSID, and token). The token could be the same one it was already using, or it could be a new one (which it would then use for future transmissions). The payload of the message is the cryptographic signature of the message header, including the certificate associated with the callsign. With this information (and its pre-programmed knowledge of the ARRL's root certificate) the satellite can verify the signature.

If the authentication response is not received in a timely manner, after TBD number of attempts, the station is blacklisted.

If the authentication response is received, the satellite checks that the certificate is valid by comparing with the root certificate (LoTW root certificate is through GoDaddy). If the certificate is not valid, the station is blacklisted.

If the certificate is valid, then the signature is checked. If the signature is not valid, the station is blacklisted.

If the signature checks out, then the row corresponding to the station in the satellite security database is updated.

The result of each authentication transaction is reported on the downlink, where it is visible to the ground stations and to Ground Control Stations.

AI: work out the rules for changing tokens. If two (or more!) stations are trying to use the same callsign:ssid with different tokens, the one(s) that can successfully authenticate should be allowed, and the other one(s) should not. But we also need to prevent a bad guy from defeating the system by simply choosing a new token for each transmission. And yet we don't want to stop a station that had to reboot from starting over with a new token.

AI: work out a way for a station that is blacklisted for authentication failure to request a do-over, after the operator corrects the configuration.

Authorization Process Overview

The authentication procedures outlined above just verify the identity of stations trying to use the satellite. Separate authorization procedures are used if not all licensed amateurs are welcome to use the satellite.

For example, an agency that has an MOU with the satellite, and has authority over the operations of a Ground Control Station, could impose an authorization policy. A combination of blacklist and whitelist techniques could be used.

Examples:

"All stations that participated in Red Cross Drills within the past 9 months"
"All stations that support a throughput of greater than 500kbps"
"All stations that have registered to participate in the weekend contest"

In cases where a list of authorized stations is available, the satellite could be configured with a whitelist and instructed to limit all other stations to short message transmissions only. Phase 4 Ground recommends that stations never be entirely banned from access, since (1) banned stations may have critical emergency traffic that should not be blocked, and (2) banned stations may need to message the authorities to request and justify being unbanned.

Black Lists and White Lists are managed by Ground Control Stations. Ground Control Stations can also choose to implement potentially complex policies on the ground, by monitoring the downlink for any unauthorized activity and taking action to stop it. The controlling interest of the Ground Control Station sets Black List and White List policies.

Phase 4 Ground recommends that Blacklists should time out whenever possible. Blacklists should reset upon satellite reset. Blacklisting should be quite temporary by default. If the Ground Control Stations all go offline in a disaster, it is better to open up the system than to leave it locked down in a way that will eventually cause problems.

AI: define uplink and downlink messages relating to authorization. Downlink messages should give non-authorized stations notice so they don't keep trying to transmit, and should ideally give them an explanation so they know why they are not authorized. Uplink messages might explicitly request authorization in certain cases. One use case we've heard is that a Served Agency distributes a secret code that can be used to gain access; that would be implemented through some messaging.

AI: there may be more levels of authorization than simply authorized and not authorized. Some stations may be allowed only short text messages, while others may be allowed to use realtime voice streams, and yet others may be allowed to grab as much bandwidth as they can manage.

Threat Assessments

The primary threat to Phase 4B is people outside the community using the spacecraft. If it's easy to gain illicit access then capacity is drained. Using the satellite without authorization needs to be hard enough that it is rendered rare, while also not making normal legitimate operation a miserable experience.

The community is defined as licensed ham radio operators that are cooperative.

Cooperative means that the operators stop transmitting if asked, are disciplined, do not produce intentional harmful interference, rectify unintentional harmful interference, and defer to emergency traffic.

If the identification of primary threat is accurate, then LoTW cert checking is a solution. LoTW excludes non-hams.

We assume that the large majority of licensed hams are cooperative. We then assume that the list of uncooperative hams, whether they became uncooperative, or have always been uncooperative, is short and can be effectively managed through Black Listing.

Stations can be added to a Black List by several methods including but not limited to observation, complaint, or statistical detection. Positive confirmation with a human in the review loop is ideal.

The threat can be refined. The following categories are either licensed operators that became uncooperative or unlicensed operators that are by definition uncooperative.

Imposters

An operator that takes on the identity of a legitimate operator that has authentication and authorization is called an Imposter.

The required information to become an Imposter must not be present in the downlink. However, it is available through eavesdropping on an uplink. If an Imposter is able to eavesdrop on an uplink, they will be able to obtain the call sign, SSID, and token of the legitimate operator, and form frames that appear to come from a legitimate user already in the database, and therefore would be passed as legitimate traffic by the satellite. Another potential Imposter method is to overwrite the payload part of the frame with a new payload. This requires precise timing and greater transmit power than the legitimate station.

An Imposter can be eliminated by several methods.

1. Frames could be serialized, with duplicate or out-of-sequence frames discarded. This adds some overhead. This method can be circumvented by deserializing or overwriting.
2. "Signed transmission mode": Forcing the signature, by private key held within the Phase 4 radio, of every frame. This adds substantial overhead. This can be circumvented by stealing the private key.

How would these methods be implemented?

If a Phase 4 radio is able to see payloads in the downlink that it did not send, then it could send a message to the Ground Control Station. The Ground Control Station could then put the Phase 4 Radio in "signed transmission mode". This requires something like a traffic summary which would have to have enough information for either an algorithm or an operator to notice extra traffic.

This could be an extension of presence awareness. This could include looking for too large of a variance in Eb/No.

Adaptive Coding and Modulation can potentially hide traffic from potential Imposters. As a side effect, ACM could make it substantially more difficult to steal the identity of a legitimate operator.

An ongoing eavesdropper can defeat anything aside from a cryptographic solution. The cost of preventing ongoing eavesdropping leading to Imposters gaining satellite access must be balanced against the probability of this type of threat.

Jammers

Jammers are stations that produce intentional interference with the goal of denying access to a communications resource. This could range from loud signals aimed blindly at the satellite up to and including correctly formatted frames that are designed to interfere with the normal operations of the satellite.

If the satellite simply doesn't demodulate signals that do not comply with the air interface, then some jammers are eliminated simply due to the fact that the satellite doesn't demodulate waveforms that it is not programmed to receive. However, by aiming loud signals at the satellite the jammer can increase the noise level at the satellite, and therefore the signal-to-noise ratio for one (or more) channels. Since the satellite is FDMA, and the channels (satellite sub-band) are well-known, and the occupied channel list will appear in the downlink, the channel (or all channels) can be targeted for jamming. Causing the satellite ADC to saturate may prevent the satellite from receiving. Detecting the presence and the source of this type of jammer needs to be discussed. Attempting to hide which channels are occupied is insufficient considering that a jammer can simply occupy the entire uplink bandwidth.

A more sophisticated jammer transmits a properly modulated waveform. If the satellite drops traffic that does not contain a token, does not contain a payload, or does not contain a call sign, SSID pair, then more potential jammers are eliminated. The traffic, while demodulated, will not appear in the downlink. The traffic does consume satellite resources.

Signals that comply with the air interface, contain a plausible call sign, SSID, a payload, and a token, but are designed to consume authorization and authorization resources, are also jammers.

Potential attacks include cuasing the satellite, by either following either onboard rules or upon (subverted) direction from the Ground Control Station, to perform functions in excess of processing resources.

Traffic that uses a different token and/or SSID and/or call sign every frame forces a large amount of database activity and processing.

The Ground Control Station could be compromised or tricked into directing the satellite to re-authenticate or re-authorize in excess of processing resources. In this case, the traffic is not at fault, but instead the mechanisms that are in place at the controlling agencies are subverted or abused.

Attacks designed to cause the satellite run out of space in the database are a type of buffer overrun attack. The database is of finite size.

Attacks designed to cause the satellite to process itself to death are a type of denial of service attack. There is a finite amount of processing available to check tokens. There is a finite amount of processing available to provide authentication.

There are several ways to address these potential threats. Some of these methods involve keeping additional state in the satellite. Therefore, the cost of the additional state and/or additional processing must be balanced against the probability of this type of attack occurring.

1. Limit the raw rate of new call sign SSIDs that are added to the database to a reasonable number (TBD) that ensures processing resources are not exceeded. Denial of service attacks are successful in computer timeframes, but not successful if the forced activity happens in human timeframes. In cases where the rate limit is exceeded due to either a large number of legitimate users attempting to transmit, or a large number of attacks, this technique negatively affects the ability of a subscribed legitimate operator to immediately participate in communications on the satellite, upon at least their first transmission and possibly all subsequent ones. This also means that new legitimate users will be dumped on the floor along with the jammers.

2. Develop an algorithm for properly removing rows with implausibly fast-changing information from the operator database. This attempts to clean out the database at the same rate that attackers are adding to it. Properly designed, the legitimate operators that attempt to transmit are unaffected. This requires enough processing to detect patterns of transmissions with rapidly changing data. There could be andom changes that run through all possible combinations (all the bits of an SSID, for example) or oscillating back and forth between two states in order to trick the satellite into doing a lot of work (for example, to force maintenance of two rows of the database by switching back and forth between two different tokens every frame).