COGNITIVE COMMUNICATIONS FOR NASA SPACE SYSTEMS

David Chelmins*, Janette Briones, Joseph Downey, Gilbert Clark, Adam Gannon

NASA Glenn Research Center, 21000 Brookpark Road, Cleveland, OH 44135, USA
*dchelmins@nasa.gov

Abstract

The growing complexity of spacecraft constellations, communication relay offerings, and mission architectures drives the need for the development of autonomous communication systems. NASA has traditionally launched single spacecraft missions that are served by the Space Communication and Navigation (SCaN) program. Operations on SCaN networks are typically scheduled weeks in advance, and often each asset serves a single user spacecraft at a time. Recent movement towards swarm missions could make the current approach unsustainable. Additionally, the integration of commercial communication service providers will substantially increase the data transfer options available to new missions.

NASA science missions have found benefit in launching swarms of spacecraft, allowing coordinated simultaneous observations from different perspectives. Inter-spacecraft communication (mesh networking) is an enabler for this architecture, as are CubeSats that allow cost-effective provisioning of distributed mission assets. As more complex swarm missions launch, one challenge is coordinating communication within the swarm and choosing the appropriate mechanism for telemetry, tracking, control, and data services to and from Earth.

Cognitive communications research conducted by SCaN aims to mitigate the increasing communication complexity for mission users by increasing the autonomy of links, networks, and service scheduling. By considering automation techniques including recent advances in artificial intelligence and machine learning, cognitive algorithms and related approaches enable increased mission science return, improved resource utilization for service provider networks, and resiliency in unpredictable or unplanned environments.

The Cognitive Communications Project at the NASA Glenn Research Center develops applications of data-driven, non-deterministic methods to improve the autonomy of space communication. The project emphasizes development of decentralized space networks with artificial intelligence agents optimizing communication link throughput, data routing, and system-wide asset management. This paper discusses the objectives, approaches, and opportunities of the research to address growing needs of the space communications community.

1 Introduction

Mission needs typically spur innovation in the field of space communication. NASA’s communication architecture is no exception: from the early days of the Apollo lunar missions, to the remarkable construction and operation of the International Space Station, to the more recent InSight mission to Mars leveraging a small spacecraft communications relay (MarCO) [1]. In each case, communications system engineers match the mission needs to an appropriate communications architecture and capability, delivering the correct data throughput, with the correct latency, with the required reliability, at an acceptable cost.

Traditionally, NASA’s space communication infrastructure has been government-owned or contracted. Early dedicated networks including the Spacecraft Tracking and Data Acquisition Network and Manned Space Flight Network consisted of ground stations that provided limited but acceptable data communications capability for uncrewed and crewed spacecraft, respectively, in the 1960s. NASA added geostationary relay satellites to the infrastructure beginning in the 1980s, providing continuous coverage of low Earth orbit through the Tracking and Data Relay Satellite System (TDRSS) [2]. Today, NASA operates three communication networks: the Space Network (SN), the Deep Space Network (DSN), and the Near Earth Network (NEN).

The modern space communications marketplace continues to evolve. While commercial satellite relays and ground stations have existed since the earliest days of spaceflight, only recently has NASA considered the regular use of these systems to meet its mission needs [3]. Multiple companies have proposed new mega-constellations of satellites in low Earth orbit, providing high-rate and low-latency communication capability [4]. NASA is developing cognitive communications technology to reduce the burden of bridging its legacy government-owned/operated communications systems with the use of commercial systems.

The growing number of operational spacecraft requires more sophisticated techniques to cooperatively share spectrum as well as mitigate intra- and inter-network interference when necessary. Regular use of increasingly high frequencies including Ka-band (26.5 – 40.0 GHz) and beyond necessitate more intelligent techniques to adapt to time-varying atmospheric conditions [5], [6]. More complex multi-hop
network topologies envisioned around other bodies such as the Moon and Mars provide a challenge for optimal in-space data routing [7]. Software-defined radio provides a platform to address many of these challenges. The Space Communication and Navigation (SCaN) program office sponsors the Cognitive Communications project at the NASA Glenn Research Center. The purpose of the project’s research is to leverage the flexibility and adaptability of software-defined radio within the context of autonomy, providing an overall benefit to the mission without increasing the operations cost.

2 Defining Cognition

Merriam-Webster defines “cognitive” as relating to, being, or involving conscious intellectual activity such as thinking, reasoning, or remembering [8]. While many of the best examples of cognitive systems are biological, a significant effort in artificial intelligence (AI) has been exerted to develop artificial systems that exhibit abilities resembling thinking, goal-oriented reasoning, and remembering. A cognitive radio, as defined by seminal works in the field, is a “brain empowered” wireless device capable of reacting to and learning from its environment to enhance communications [9], [10]. Thus, cognitive radios should be aware of the operational environment, capable of adapting their operational parameters, and able to improve from past actions to enhance future performance [11]. The principles of cognitive radio are broadly applicable across the protocol stack to cognitive networking and even entire cognitive communication systems [12], [13].

NASA’s Cognitive Communications project adapts the formal definition of “cognitive” as follows: any system, or part of a system, that is able to mitigate obstacles, respond to and learn from its environment, and achieve beneficial goals to the completion of its primary mission. Such a cognitive system can perform these activities with minimal to no human interaction. Finally, the cognitive system must have the ability to adapt to changing conditions by producing reasonable outcomes in scenarios that extend beyond the pre-programmed knowledge of its original inception. The Cognitive Communications project defines a cognitive engine (CE) as a decision-making algorithm that enables part of a cognitive system. Multiple CEs can apply to various levels of the communications protocol stack, from a single radio frequency (RF) or optical link to complex distributed applications.

A system designer could implement each CE in many different ways utilizing different decision-making methods, including those based on machine learning (ML), so long as these methods align to the goals of the overall cognitive system. In general, CEs must rely on multiple inputs and process data in different ways to come up with a usable solution. One CE design approach aggregates various ML and deterministic algorithms into a single framework, evaluates performance in parallel, eliminates poor-performing algorithms, and aggregates the remaining algorithms to deliver an optimal solution for the particular problem at hand. This approach requires environmental feedback to optimize toward particular objectives. The decision process (Fig. 1) provides a general approach on how CEs interact with the outside world and behave intelligently in that environment, to maximize their objectives.

3 Focus Areas

The Cognitive Communications project performs research in four distinct but intertwined areas:

- Links – concerning point-to-point connections between two devices
- Networks – concerning multiple devices routing information among multiple links
- Systems – concerning the interaction among devices and supporting ground- and space-based infrastructure
- Enabling Technology – concerning the on-board processing, sensing, and adaptation capability of a device that allows it to participate in cognitive links, networks, or systems

These areas broadly resemble the Open Systems Interconnection (OSI) model [14] with some notable differences. The highly directional links common to space communications necessitate an approach to medium access control, typically a function of the data link layer, different from that of terrestrial wireless where transmissions from user equipment are generally omnidirectional. System infrastructure (ground stations or relay satellites) featuring directional antennas are generally incapable of supporting many simultaneous users.

CEs applied broadly across all levels of the protocol stack will determine link optimization, network routing, and system management. While each of these focus areas can mature independently, the end goal is to transition towards an overall cognitive system-of-systems, optimized across all OSI layers. The spacecraft itself and the communication provider networks must perform joint cross-layer, distributed decision-
making that conforms to the mission objectives and network capabilities. In the next several sections, this paper will discuss each focus area, its potential optimizations, and benefits in the context of a cognitive space communications system.

4 Cognitive Links

A space communication link is a wireless connection between two radios over distance with at least one of the radios in space. Currently, NASA missions determine the exact communication system configuration of the radio prior to launch. In the case of an RF system, the mission designer typically allocates a single frequency and bandwidth to the radio, applies for the corresponding spectrum license, and negotiates service with a communications service provider (through space-based relays or ground stations). The traditional approach is robust and proven, yet inflexible to real-time changes. The communication system can fail to communicate under several plausible scenarios including:

- The receiver encounters interference resulting in loss of lock
- Mission hardware degrades, reducing the transmit power or increasing receiver noise figure
- The communications service provider cannot schedule a sufficient number of contacts

In each of these cases, the only remedy in a traditional, inflexible communications approach is to keep transmitting and hope the result improves over time. Using a software-defined radio, a mission operations team can program the system to adapt to predictable or gradual failures. However, most real-time issues (especially those that are unlikely or unpredictable) can result in loss of mission data. Cognitive link capabilities include technologies, algorithms, and protocols applicable to the physical and data link layers. The prime benefit of a cognitive link approach is on-board, autonomous mitigation of real-time issues. A second, significant benefit is the ability to improve performance and efficiency of the communication link.

4.1 Radio Frequency Interference Mitigation

One example of a cognitive link capability is RF interference mitigation, which automatically senses and avoids spectrum interference by changing frequency, bandwidth, data rate, and antenna pointing. An automated approach was developed [15] that significantly reduces data loss from RF interference while increasing throughput. Fig. 2 shows the RF interference mitigation concept, where the space-based transceiver is located on the International Space Station. Previous space-based testing using NASA’s ScAn Testbed has shown that RF interference is most likely encountered on the ground. That is, transmitters local to the ground station are more likely to introduce interference in a link than (1) a space- or ground-based transceiver in the spacecraft or (2) a space-based transceiver pointing at the ground station. The authors in [8] describe a cognitive engine that optimizes:

- Throughput (maximized) – the number of data bits transferred per second
- Occupied bandwidth (minimized) – occupied spectrum
- Transmit power (minimized) – amount of communications power on the spacecraft

![Fig. 2 RF interference mitigation concept](image)

4.2 Radio Link Optimization

Prior work has shown adaptive coding and modulation (ACM) successfully optimizing throughput over a communication link with varying margin [16]. Typically, ACM uses fixed signal-to-noise ratio thresholds based on theoretical characterization. A new approach is to implement a cognitive engine that decides when to change modulation, coding, and transmission power based on observed channel conditions and mission platform constraints. Results have demonstrated that a neural network-based reinforcement learning algorithm performing multi-objective optimization is feasible for satellite communication [17]. The authors of the referenced work developed a cognitive engine that tested multiple radio settings so that the system could learn how to adapt to achieve multiple goals for satellite communication in a dynamically changing environment.

4.3 Automatic Receiver Configuration

In an effort to reduce operator burden when switching between communication relay providers, self-configuration of a software-defined radio may be possible by sensing the inbound signal to perform signal recognition. This technique relies on signal processing mechanisms to recognize signal waveform parameters such as modulation scheme [18], coding, and data rate. Using these parameters facilitates system self-configuration and link acquisition even in the presence of noise or weak signals. For deep space systems that have a significant round-trip time (RTT) and multiple possible waveform configurations, such a method could save one RTT or more.

4.4 Deep Learning Communication Links

One intriguing development in the cognitive links area is the creation of learned communication systems, which use deep learning (AI) techniques to minimize end-to-end message
reconstruction. This method offers potential improvements to traditional systems in three key areas: jointly optimizing modulation and encoding, utilizing the non-linear function approximation capabilities of a neural network to account for power amplifier distortions, and a relaxation on the assumption that system noise follows a Gaussian distribution. In [19], [20], the authors introduce an autoencoder model (see Fig. 3) that performs physical layer optimization. This work was extended in [21] where a generative adversarial network (GAN) was used to learn an arbitrary channel model that includes non-linearities and memory effects.

Although this approach shows promise, a real-time adaptation will impose a significant computational burden for today’s space processors. Additionally, due to the amorphous nature of the self-learned system, new protocols must be implemented to account for symbol timing acquisition and transmitter/receiver coordination during training. Recently, this area has received significant attention in the research community and several potential architectures have been proposed. In [22], the need for a feedback path during training was removed and a live implementation was demonstrated using hardware radios.

5.1 Delay Tolerant Networking

NASA has a demonstrated need for networks that are tolerant to delay and disruption, known as Delay-Tolerant Networking (DTN). DTN can refer to a network that is affected by disruption or delay, an architecture [23], a set of tools, a protocol [24], [25], or a specific implementation of a protocol [26]. For the sake of clarity, this paper will disambiguate its use of DTN as it relates to the various ways in which cognition may be added to different elements of the DTN portfolio. Generally, DTN is an architecture. An implementation of DTN is a set of protocols and techniques that realizes a network that is tolerant to delay and disruption. One key aspect of DTN is the capability to customize an implementation to match its environment: if an environment is not anticipated to carry links with long delays, for example, protocols such as LTP [27] may not be necessary.

5.2 Intelligence in the DTN Architecture

One aspect of DTN study is the identification, maintenance, and assignment of endpoints and their mapping to Minimum Reception Groups (MRGs). The DTN architecture allows an entire constellation of spacecraft (for instance) to act as a single, shared DTN endpoint in a larger network. To date, however, there generally has been a 1:1 correspondence between physical machines with specific endpoints. In an intelligent system, this does not need to be the case.

An intelligent DTN application can provide flexibility: rather than each node in a constellation acting as an individual endpoint, the entire constellation may serve as a single endpoint (with a shared logical bundle protocol agent). The constellation can then send and receive DTN data in a coherent, locally coordinated fashion, reducing the number of times that bundles must be encoded and decoded. This reduces unnecessary overhead, allowing implementation of DTN in non-traditional locations, such as between individual nodes in a constellation that may not individually suffer from large degrees of delay or disruption.

An intelligent approach to realizing a delay-tolerant network is to monitor a constellation for failures and/or new available assets, and to adjust the MRG accordingly. The system might also predict and move data to specific elements of an MRG that would be most likely to need that data for future events. Another potential advantage is the freedom to manage data movement between nodes in a MRG; this capability (Fig. 4) allows selection of a prime data downlink based on expected availability, and dynamic reassembly of data fragments within the context of a distributed MRG. This area of study is referred to as Drop Data Anywhere (DDA).

An additional area relates to the means by which data might be prepared and processed: this applies not only at the data’s origin but also, through the application of virtualization, at each incremental hop in data’s movement toward its destination. A cognitive routing scheme could ensure that the
data passes through incremental nodes that have the correct processing ability.

Routing between multiple delay-tolerant networks is a common challenge. While a number of approaches have been explored to date, there is still more work needed on the integration of physical and link scheduling with network routing decisions. For example, there is active work in the area of applying Spiking Neural Networks (SNNs) to the area of information transfer through delayed and/or disrupted networks [28]. The referenced author has considered using Software-Defined Networking (SDN) protocols to propagate routing information through a network.

5.3 Cognition in the DTN Protocols

Cognition in the DTN protocol suite would blend global (architectural) goals with local ones. For instance, specific protocols might be necessary to support the maintenance and establishment of an MRG. While these protocols would not be cognitive in and of themselves, the protocols would serve as a means to realize cognition in the larger network. One such protocol example is the ability to dynamically discover, enumerate, and control radio capabilities, allowing the system to understand its environment and communicate its decisions to others. Work in this area can build on existing protocols that are well-known [29]. It can also build on newer entries into the protocol arena such as [30], which is an approach to the management and monitoring of a delayed and/or disrupted network: while not suited for a well-connected network, this protocol is useful for situations where coherent networks need to exchange information or commands through hops that are somehow constrained.

The other approach to protocol-level cognition is through optimization of protocol parameters themselves. If, for example, a system knows what the expected link round trip time and link rate will be, it can optimize elements of its own protocols to use that information without having to rediscover that information for itself. Work here has explored modeling existing protocols [31] to predict the impact of a specific protocol-level decision.

5.4 Legacy, Infrastructure, and Bootstrapping Intelligence

One major challenge associated with DTN involves its infusion into existing infrastructure. Traditionally, NASA’s space network infrastructure has focused on information transfer from an asset in space to the ground. Expanding such a network requires increasing the number of available dedicated antennas or number of unique spectrum bands available to support missions. The network will expand until the supply of either frequency or antennas is exhausted.

Recent decreases in launch costs, an evolution in mission design to move toward distributed (e.g. multi-spacecraft) missions, and a general increase in demand for space-faring missions have all begun to place a severe strain on the fixed pool of resources available. This has led to applied research at the intersection of resource allocation and DTN. Automated scheduling techniques to reserve network resources (e.g., antennas, bandwidth) are one mitigation. To implement automated scheduling, DTN can find an optimal network routing solution, and a machine-to-machine interface can schedule time on the corresponding nodes. This creates a strong sense of vertical integration between different elements.

Additionally, MRGs can fuse multiple short physical events (i.e., antenna time) into one larger logical event. This allows missions to create a long logical event without necessitating the implementation of point solutions to support such a use-case. Given that smaller scheduling blocks allow for improved flexibility when scheduling, this offers benefits to both the mission and the service provider: missions can get more aggregate service time, and providers can increase the duty cycle of the resources they have available to them.

5.4 Virtualization in Future Cognitive Networks

Virtualization is an attractive mission-level solution to support flexible on-board processing and routing capabilities in a cognitive network. When implemented correctly, the overhead associated with virtualization can be minimal, while offering benefits to reliability, load balancing, and the ability to scale-up / scale-down.

One candidate for virtualization in a space environment involves the Core Flight Executive (cFE) framework [32]. The common development libraries, framework, and environment has allowed for code re-use across a diverse set of missions. Work is ongoing in the area of building cFE applications as Real-Time Executive for Multiprocessor Systems (RTEMS) virtual machines. This would allow an intelligent system to migrate elements of its own execution to any compatible virtualization environment in real-time. It also offers the network a capability to treat members less like an individual, and more like a distributed piece of a coherent whole, realizing a cloud-compatible approach to service management in space environments.

Notably, both intelligent systems (e.g. inference and on-line learning capabilities) as well as network functions themselves (e.g. DTN) can be realized as discrete sets of coherent functions, each of which can execute on a participating member of a cloud-centric service architecture in a space setting.

6 Cognitive Systems

The Cognitive Systems focus area aims to optimize performance across entire space communication systems, improving the interaction between mission spacecraft and service provider infrastructure. With increased system autonomy, the mission spacecraft can negotiate access to communication services based on its current data transfer needs. This architecture of automated resource allocation from spacecraft-initiated requests, called User-Initiated Service (UIS), aims to provide more responsive access to high-performance space communications [33], [34], [35].
6.1 User-Initiated Service
In current practice, network operators manage access to the highest performance services based on requests from each mission’s operations staff. Mission operators must anticipate spacecraft commanding and data transfer needs potentially weeks in advance. However, an increasing number of spacecraft have event-driven service requirements with demand that is difficult to predict. Current practice limits the network’s ability to negotiate schedules rapidly. The ideal network infrastructure incorporates automation and cognitive techniques to allocate resources exactly meeting each user’s immediate demand. In the UIS concept, the mission spacecraft itself originates a request for communications or navigation service based on current needs. For example, a spacecraft recording scientific data from a transient astronomical event can automatically send a request for data downlink service when its onboard storage is nearly full. In this new paradigm, access to high-performance service can be requested using machine-to-machine communications over low-rate, high-availability control channels.

The use of control channels to request resources from a wireless network is seen during the base station association process of terrestrial mobile networks or the Proximity-1 hailing channel used by Mars spacecraft [35], [36]. Examples of existing control channels include the multiple-access capability of NASA’s TDRSS and various commercial services that provide low-rate, on-demand messaging capabilities. Fig. 5 shows an overview of the process. A request for service is received over the control channel by a central Event Manager that is aware of both spacecraft and communications network capability. The Event Manager contracts on behalf of the requesting spacecraft for communication service with any network capable of establishing a link with the mission spacecraft. The Event Manager then sends the necessary waveform configuration, ground station or space relay location, and contact time back to the requesting spacecraft via the same control channel. At the start of the negotiated access window, the spacecraft exchanges data with its mission operations center over the contracted high-rate data channel.

Fig. 5 Overview of the User-Initiated Service process

An on-orbit experiment using NASA’s ScAn Testbed demonstrated the concept with TDRSS [37]. The low-rate S-band multiple access service was used as a control channel to send requests to schedule high-rate Ka-band single access service for data transfer. Though both services were provided by the TDRSS, the high-performance Ka-band link is capable of supporting data rates 5,000 times greater than those of the S-band multiple access system [38]. For missions that do not have space-to-space communications capability, the feasibility of establishing a control channel between an Earth-orbiting spacecraft and an omnidirectional antenna co-located at a NASA ground station site was evaluated in [39]. Future work will demonstrate the use of a low-rate commercial relay service as an additional control channel.

6.2 System-Wide Intelligence
There has been a growth of commercial ground station and relay satellite networks in recent years, offering more potential links to a spacecraft than ever before. By adding scheduling interfaces for commercial networks, the Event Manager can schedule service with one of several networks on behalf of the spacecraft. This capability expands the capacity of communications service available to the spacecraft. Such a multi-provider framework offers a heterogeneous mix of links representing a trade-off between different characteristics. The Event Manager must choose the optimal link.

System optimization could take place across many factors: availability, cost, latency, data volume, contact time, etc. Machine learning techniques have recently shown great promise in handling these types of multi-objective optimization problems including in the space communication environment [17]. Using these and similar techniques, the Event Manager could select the optimal link to meet the UIS request while balancing network load across multiple assets. An added benefit is that machine-learning techniques can detect service provider anomalies by examining mission performance across multiple providers or provider performance across multiple missions. Time-varying characteristics such as optimization for weather conditions affecting a given ground station could be considered. Furthermore, a cognitive engine learns from past allocations to improve mission communications performance over time.

7 Enabling Technology
7.1 Reconfigurable Hardware
The hardware necessary to implement cognitive communications capabilities on-board spacecraft typically mimics the hardware that enables artificial intelligence and machine learning on the ground. Radiation-hardened space processors tend to be about two generations behind terrestrial processors. Examples of enabling cognitive processing capabilities for space include:

- General purpose Graphics Processing Units (GPUs) and multi-core processors
- Neuromorphic processors
- Field-Programmable Gate Arrays (FPGAs)

In each case, these technologies must have low size, weight and power (SWaP) for integration into the spacecraft communication system, with tight coupling to the functions of the spacecraft software defined radio. Additionally,
processor radiation tolerance is necessary for long-term survivability in the space environment, although software techniques (e.g., triple mode redundancy, regular system resets) can mask radiation effects to some extent.

Neuromorphic processors implement a non-von Neumann computing architecture that utilizes analog and digital electronic circuits to mimic the neuro-biological architectures present in the nervous system. Neuromorphic systems exhibit increased energy efficiency, execution speed, and robustness over traditional computing architectures, and can provide pattern recognition capabilities for SWaP-constrained applications [40].

7.2 Cognitive Processing Challenges
Enhanced onboard processing of science data products reduces the amount of data transferred to a mission operations center, therefore reducing cost and demand on the network. High fidelity science instruments (e.g. synthetic aperture radar) are capable of generating volumes of data that far surpass a spacecraft communication system’s capability, requiring compression or preprocessing [41]. The machine learning algorithms discussed throughout this paper may also require significant computation during the training process.

8 Conclusion
The objectives of cognitive links, networks, systems, and enabling hardware discussed in this paper aim to provide increased autonomy and reliability for NASA’s communication architecture. This will require increased on-board processing in the space environment, eventually resembling a terrestrial cloud computing architecture. Instead of simply providing point-to-point links, the future architecture for space communication must include communication, processing and storage. In such a service-oriented architecture with distributed cognition, all of the optimizations and concepts discussed in this paper become possible.

Machine learning and related automation technologies are a new thrust in space communication. Implemented correctly, these technologies have the potential to make communications networks more efficient and resilient in the challenging space environment as they have done on the ground. As the envisioned NASA space communications network evolves into a cognitive system-of-systems, this will enable improved science return, resource efficiency, and reliability for both missions and the communication networks providers.

9 References


[29] IEEE, Station and Media Access Control Connectivity Discovery,’ IEEE Standard 802.1AB-REV.


