JOURNAL OF SPACECRAFT AND ROCKETS
Vol. 37, No. 6, November–December 2000

Generalized Characteristics of Communication, Sensing, and Navigation Satellite Systems

Graeme B. Shaw,* D. W. Miller,† and D. E. Hastings‡
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

A generalized framework has been developed for classifying space system architectures and their characteristics for any mission in communications, sensing, or navigation. For each of these applications, the overall mission objective is to transfer information between remote locations and to do so effectively and economically. Consequently, satellite systems can be represented as information transfer networks, serving origin–destination information markets. Their capabilities are characterized by four quality-of-service parameters that relate to the isolation, rate, integrity, and availability of the information transferred. This generalized approach has led to significant conclusions in a variety of applications and has allowed a comprehensive characterization of satellite systems in general. Based on the arguments presented, deduced from organized, qualitative analysis, it would appear that the potential offered by distributed satellite systems is very great.

Nomenclature

A	=	aperture area, m^2
a	=	acceleration, m/s^2
D	=	aperture extent, m
D_c	=	cluster (sparse) aperture extent, m
D_s	=	satellite (filled) aperture extent, m
E_s	=	energy per symbol, J
$E[\,]$	=	expected value
g	=	gravitational field strength
I_{sp}	=	specific impulse, s
$N_{i,o}$	=	noise power (i = input, o = output), W
N_0	=	noise power density, W/Hz
n	=	frequency of reference orbit, rad/s
n_r	=	number of receive channels
n_s	=	number of satellites in a cluster
n_t	=	number of coherent transmitters
R_c	=	cluster radius (from reference orbit), m
$S_{i,o}$	=	signal power (i = input, o = output), W
t	=	time, s
$t_{\mathrm{sym},1}$	=	dwell time (symbiotic cluster, single satellite), s
ΔV	=	velocity increment, m/s
λ	=	wavelength, m

Introduction

THE primary goal of this research is to develop a consistent methodology for quantifiable analysis of any satellite system designed for communications, navigation, or sensing applications. The emphasis of this introductory paper is to introduce the concepts that are needed to construct this generalized analysis. This involves the identification of the characteristics that are general to these types of satellite systems and also the definition of a framework for classifying space system architectures.

Distributed Satellite Systems

Recently, increases in the available processing power, improvements in navigation, and advances in the manufacturing process have all made the concept of a distributed satellite system feasible. The term distributed satellite system is used to refer to a system of many satellites designed to operate in a coordinated way to perform some specific function. This definition encompasses a wide range of possible applications in the commercial, civilian, and military sectors. The advantages offered by such systems can mean improvements in performance, cost, and survivability compared to the traditional single-satellite deployments. The term distributed satellite system can have two different meanings.

1) Distributed satellite system can refer to a system of many satellites that are distributed in space to satisfy a global (nonlocal) demand. Broad coverage requirements necessitate a separation of the satellite resources. At any time, the system supports only single-fold coverage of a target region. The local demand of each region is served by the single satellite in view. Here, the term distribution refers to the system being made up of many satellites that work together to satisfy a global demand.

2) Distributed satellite system can also refer to a system of satellites that gives multifold coverage of target regions. The system, therefore, has more satellites than the minimum necessary to satisfy coverage requirements. A subset of satellites that are instantaneously in view of a common target can be grouped as a cluster. The satellites in the cluster operate together to satisfy the local demand within their field of view. Note that the cluster may be formed by a group of formation-flying satellites, or from any subset of satellites that instantaneously share a common field of regard. The cluster size and orientation may change in time, as a result of orbital dynamics or commanded actions. In any case, the number of satellites in the cluster is equal to the level of multifold coverage. In this context, distribution refers to there being several satellites that work together to satisfy a local demand. The entire system satisfies the global demand.

The most important characteristic of all distributed systems, common to both of the preceding concepts, is that more than one satellite is used to satisfy the overall (global) demand. This is the basic distinction between a distributed and a singularly deployed system. Within the classification of a distributed system, the main difference between the two concepts described lies in the way that the local demand is served. Specifically, the distinction is the number of satellites used to satisfy this local demand: the cluster size n_s thus characterizes the level of distribution, with larger cluster sizes corresponding to higher levels of distribution. The lowest level of distribution, with a cluster size of one, corresponds to the first meaning of distribution described.

Satellite Systems as Information Transfer Networks

Most current satellite applications provide some kind of service in communications, sensing, or navigation. The common thread linking these applications is that the satellite system must essentially perform the task of collection and dissemination of information. Data that contains pertinent information are gathered by the satellite,

Received 12 May 1999; revision received 20 April 2000; accepted for publication 5 May 2000.

*Research Fellow, Department of Aeronautics and Astronautics. Member AIAA.

†Assistant Professor, Department of Aeronautics and Astronautics. Member AIAA.

‡Professor, Department of Aeronautics and Astronautics. Fellow AIAA.

either from other components of the system (on the ground, in the air, or in space) or from the environment (local or remote). Some interpretation or processing of the data may be performed, and then the satellite disseminates the information to other system components. The generalization made is that all satellite systems are basically information transfer systems and that ensuring information flow through the system is the overall mission objective. This is easily understood for communication and remote sensing systems. Perhaps more surprising is that navigation systems such as the NAVSTAR Global Positioning System (GPS) are also information disseminators. The GPS satellites use the information uploaded from the control segment to construct a signal that is transmitted to the ground. GPS receivers can use the information in the signal, including not only the navigation message contained therein, but also the phase of the signal itself, to determine a navigation solution. As with communications and remote sensing, the performance of the system relies on the flow of information through the satellite network.

Whereas the format and routing of the information being transferred may be different for different applications, the physics characteristic of information transfer systems are, of course, invariant. This common thread linking all systems (navigation, surveillance, communications, and imaging) establishes a context for a generalized analysis and is particularly useful in the study of distributed systems.

To generalize, most satellite systems can be represented as information processing networks, with nodes for each satellite, subsystem, or ground station. The satellite network connects a set of source nodes to a remote set of sink nodes and, in doing so, addresses a demand for the transfer of information between them. Figure 1 graphically represents a simplified version of such a network for a communication system consisting of three satellites and two gateway ground stations. The system transfers data between users distributed throughout its coverage area, using several spot beams, which are the input and output interfaces for the satellite nodes. The satellites can also route information through ground stations. Even in this simple example, there are many possible routes for information to travel through the network. Some paths involve only satellites nodes, whereas others involve both satellites and ground stations.

Accepting the abstraction of satellite systems to information networks allows satellite system analysis to be treated with the well-developed mathematics of network flow and information theory. The principles of network flow apply to the overall routing and flow of information, whereas the transmission of information over each individual link is governed by the rules of information theory.

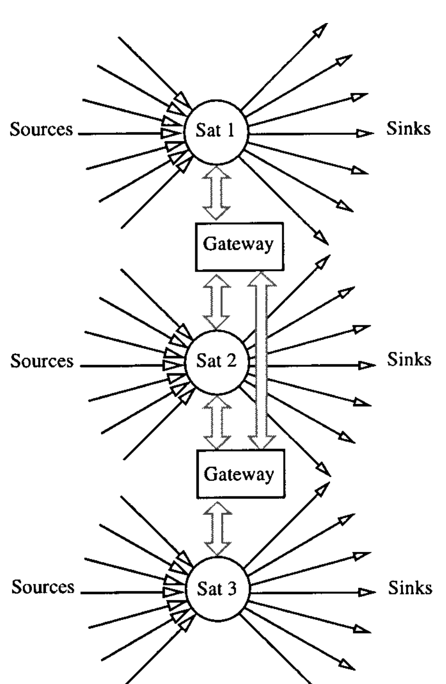

Fig. 1 Network representation of a simple communication system.

The information symbol is the atomic piece of information demanded by the end users. The symbol for communication systems is either a single bit or a collection of bits. For imaging systems, the symbol is an image of a particular scene. This compound symbol has many component pixels, each of which has some value, defined by a sequence of data bits. The symbol is the image and not the individual pixel because pixels on their own carry little or no information and are of no use to the end user, who demands images. For a navigation system, the symbol is a user navigation solution. The NAVSTAR GPS system is an interesting example because it addresses this demand without transferring user navigation solutions through its satellites; they only relay their position and time to the users. With this information from at least four satellites, the user terminal can calculate the navigation solution, assembling the information symbol from several constituent parts.

To be a contributing element to the system, each satellite node must receive information from some other node, be it a source, a ground station, or another satellite. Once this information has been received, the satellite may perform some processing and reduction before relaying the information to the next node in the network. This destination node may likewise be an end user, a ground station, or another satellite in the system. Although some data reduction may be done, information must flow through the satellites. Because of this continuity constraint, every satellite must be able to communicate with at least one other node in the network. For all satellites, the energy conversion system, for example, solar arrays, must provide the energy for the transmission of this information. For active systems, the satellite must also provide the energy needed to receive the information in the first place. These are systems such as radar and lidar that illuminate a target and detect the return. The satellites must transmit a signal with enough energy to make the round-trip journey to the source and back. The source adds the information to the signal, but returns only a fraction of the incident energy, depending on its cross section. Note that under this definition, and contrary to intuition, communications satellites are passive because they only relay received information to a destination node.

The volume of demand served by the system is limited by the market (demographics, capture, and exhaustion) and by the system capabilities. For information networks, the quantity, quality, and availability of the information arriving at the sinks are fair measures of the system's capabilities and represent the quality of service delivered to the users. Four quality-of-service parameters can be defined to measure system capabilities:

1) Isolation characterizes the system's ability to isolate and identify the information signals from different sources within the field of regard. The isolation capabilities of a system determine the level of cross-source interference that is admitted. Multiple access schemes for communication systems are methods of signal isolation. Analogously, the resolution of an imaging system allows spatially separated sources to be isolated.

2) Information rate measures the rate at which the system transfers information between the sources and the sinks. This is most familiarly associated with the data rate for communication systems. The revisit rate is the corresponding parameter for imaging systems. Information must be sampled at a rate that matches the dynamics of the source or end user. For example, a high-speed cruise missile must be tracked with a high sampling rate. Similarly, a GPS receiver on a high-dynamic aircraft must receive information from the satellites at a rate that is sufficient to allow navigation solutions to be updated very quickly.

3) Integrity characterizes the probability of making an error in the interpretation of an information symbol based on noisy observations. For communications, the integrity is measured by the bit error rate. The integrity of a surveillance radar system is a combination of the probability of a missed detection and the probability of a false alarm because each constitutes an error.

4) Availability is the instantaneous probability that information symbols are being transferred through the network between known and identified (isolated) origin–destination (O–D) pairs at a given rate and integrity. Availability is a measure of the mean and variance of the other capability parameters, but is not a statement about component reliabilities. At any instant, the network is defined only by

its operational components, and so all networks are assumed to be instantaneously failure-free. Should a component fail, the network changes by the removal of that component. Generally, the capabilities of the new network will be different from those of the previous network.

Basically, the rate and integrity correspond to the quantity and quality of the information exchanged between a single O–D pair, the isolation measures the ability to serve multiple O–D pairs without interference, and the availability measures how well the system does all this, at any particular instant. These quality-of-service parameters measure the capabilities of satellite systems over all likely operating conditions. The actual operating point is set to match the demands of the market that the system is to serve. This demand is represented by a set of functional requirements, specific to an individual information transfer. The requirements specify minimum acceptable values for each of the quality of service variables.

Because the availability implicitly includes a reference to the other characteristics, the requirements simply enforce that, for a specified level of isolation, rate, and integrity, the availability of service exceeds some minimum acceptable value. Architectures that support capabilities exceeding the requirements of the market are viable candidates for the mission.

The degree to which a system is able to satisfy the demands of a market is a critical consideration for system analysis. In fact, the probability of satisfying the system requirements that correspond to the market is the correct measure of system performance. This is sensitive to component reliabilities because failures can degrade the system such that the new capabilities violate requirements. Architectures that can tolerate component failures without significant degradations in the capabilities are good candidates for the mission.

Satellite System Classifications

The information network representation of satellite systems and the definition of the four capability parameters supports a generalized classification of satellite systems, distributed or singularly deployed. Categorizing the different architectures and identifying those issues and problems characteristic of each class allows immediate architectural decisions to be made for any given mission. Classifications are, therefore, necessary that allow system identification and highlight the most important system characteristics.

Distribution

Earlier, it was pointed out that the level of distribution exhibited by a system is defined by the cluster size. Although the cluster size is the primary form of system categorization for distributed systems, additional classifications beyond this are necessary. Specifying a cluster size of 10, for example, says nothing about the way that the satellites coordinate to satisfy the local demand. The first type of classification is, therefore, based on the level of coordination exhibited by the system elements, and is related to the network architecture. The following two sections refer to Fig. 2.

Collaborative

Each separate satellite operates independently and is able to isolate signals satisfactorily. Although an individual satellite addresses a given source (or sink), other satellites (or sensors) may be needed for connectivity across the network. The cluster size can be as low as unity, but may be more if multiple satellites are needed to satisfy rate, integrity, or availability requirements for the size of the market that is addressed. The defining characteristic is that the network architecture consists of uncoupled parallel paths from the set of sources to the set of sinks. Most communication satellite systems are collaborative, because each satellite can support local point-to-point communications, although in some cases they rely on the constellation for connectivity across the network. Examples of collaborative remote sensing systems are the commercial distributed imaging systems such as SPOT or ORBVIEW.[1] These systems feature constellations of several satellites, each capable of recording images with good resolution. The size of the constellation determines the coverage and revisit time of the system. Traditional singular deployments are by definition collaborative.

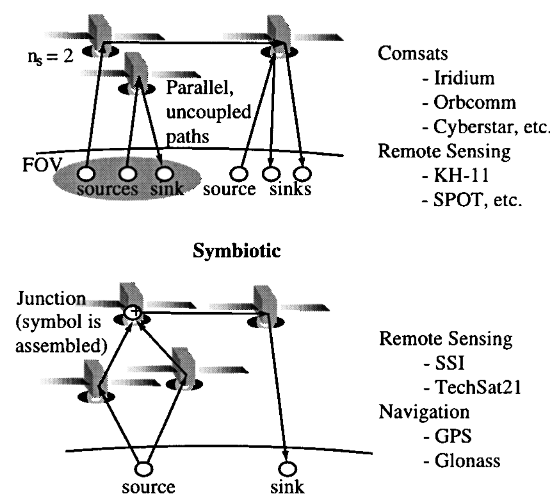

Fig. 2 Classes of distribution for satellite systems.

Symbiotic

The separate satellites cannot operate alone, exhibiting a symbiotic relationship with the others in the system. No single satellite can sufficiently isolate the signals or transfer information symbols from the sources to sinks. Only by the coordinated operation of several elements can the system perform the design function. The cluster size of symbiotic systems must be greater than unity. The defining characteristic is that the network architecture features junctions of paths from separate satellites where the information symbols are assembled before delivery to the sinks. An example of a symbiotic system is the proposed separated spacecraft interferometer (SSI).[2] Here, the signals from two small apertures are combined and interfered to obtain very high resolution images. GPS is also symbiotic because the signals from several satellites are used to assemble a navigation solution within the user receiver.

Architectural Homogeneity

The second level of classification specifies the level of homogeneity exhibited by the system architecture.

Local Cluster

Some proposed systems involve a local grouping of satellites that are in close proximity. The clusters can be made up of formation-flying satellites or can even involve the physical tethering of satellites. If there is only a single cluster in the system, such as with the SSI, the architecture is simply termed as a local cluster.

Constellations

These are systems that feature a large number of similar satellites in inertial orbits, each with their own unique set of orbital parameters. Walker Delta patterns or Molniya orbits support these types of constellations. Systems such as GPS and Iridium are characterized as being constellations. Cluster sizes greater than unity can be formed if the constellation supports multiple coverage of target regions.

Clustellations

A system may involve more than one local cluster. Each cluster orbits as a group, and several clusters can be placed in separate orbits. An architecture that utilizes several local clusters is classified as a clustellation because it features a constellation of clusters.

Augmentations

An augmented system has a hybrid architecture featuring primary and adjunct dissimilar components that perform different subsets of the mission. The system is designed such that the combined capabilities of the different components satisfies the overall mission objective. An example of an augmented system would be the combined use of different platforms or sensors to perform active and passive

Table 1 Satellite system classifications

Class	Local cluster	Constellation	Clustellation	Augmentation
Collaborative	EchoStar	LEO Comsats SPOT, OrbImage, etc.	Skybridge	SBIRS high and low
Symbiotic	SSI	GPS	TechSat21	Sat+UAV bistatic radar

surveillance. Within this analysis framework, the space-based infra red systems, SBIRS Low and SBIRS High, are collectively classified as augmented. Another example of an augmented system is the proposed concept of using both uncreated aerial vehicles and space assets for tactical reconnaissance of a battlefield.

Operational

A third level of classification groups systems according to their operational characteristics. This type of classification is the most abstract. The list shown here is by no means exhaustive and covers only some examples of the operational classifications.

Active or Passive

Remote sensing may be active or passive, with marked differences in capability and cost. This is primarily due to the additional power requirements needed to overcome the two-way attenuation losses associated with active systems.

Track, Search, or Imaging

Tracking targets using staring sensors involves different scaling parameters than searching for targets with scanning sensors. The detailed imaging of a static scene differs from either tracking or searching. These differences are all related to the extent over which the ground must be illuminated or viewed.

Distributed or Concentrated Market

The market addressed may involve multiple sources or sinks, distributed over a wide area, or could involve small numbers of sources or sinks concentrated in specific locations. Conventional communication satellites (Intelsat) serve concentrated sources and sinks. Weather satellites are examples of systems that address distributed sources, and concentrated sinks, whereas DirecTV broadcast satellites serve concentrated sources and distributed sinks. The mobile communication systems (Iridium, Globalstar, ICO) are characterized by a distributed market of both sources and sinks. As with track vs search, the difference between a concentrated and distributed market lies in the amount of ground that must be illuminated or viewed.

Table 1 gives some examples of existing or proposed systems for each class that has been introduced. All satellite systems for missions in communications, sensing, or navigation can be similarly classified using these different categories. If any trends in the capabilities, performance, and cost can be found within and between classes, quick decisions can be made in choosing an architecture for a particular mission.

Characteristics and Advantages of Distributed Architectures

Stated very simply, for a distributed architecture to make sense, it must offer either reduced lifetime cost or improved capabilities compared to traditional singular deployments. A system architecture that offers improvements in any of the four generalized capability parameters should be given serious consideration during the system design.

The system lifetime cost accounts for the total resource expenditure required to build, launch, and operate the satellite system over the expected lifetime. This includes the baseline cost of developing, constructing, launching, and operating the components of the system and also the expected costs of failure. These additional costs arise from the finite probability of failures occurring that could compromise the mission. Should such failures occur, economic resources must be expended to compensate for the failure. One example would be the cost to build and launch a replacement, whereas another is the lost revenue associated with a reduced capability. The options to lower the expected cost of failure are to reduce the impact of any failures that do occur, or to lower the component failure rates such that these failures are less likely.

As a result, all of the reasons supporting the use of distribution relate in some way to improving the capability characteristics or to reducing the baseline or failure compensation costs. The following sections detail these relationships, highlight the general trends observed within and between the different classes of systems, and hint toward the type of applications for which distributed architectures are best suited.

Signal Isolation Improvements

In general, different signals can be isolated by exploiting differences in their frequency content, by sampling at times that match the source characteristics, or by isolating spatially separated sources using a high-resolution detector. By definition, each satellite in a collaborative system independently satisfies the isolation requirements of the mission. Distribution, therefore, makes no difference to the isolation capabilities of a collaborative system.

Isolation capabilities can be improved with a symbiotic architecture. The reason is straightforward; by separating resources spatially over a large area, the geometry of the signal collection is different for each detector. Combining the received signals can assist isolation of the different sources due to field of view changes, different times of flight, or different frequencies or phases of the received signals. Larger spatial separation of the apertures means that the phase difference between signals arriving at different detectors is increased, further separating the sources. This is best demonstrated with an example.

Example: Isolation Improvements and Spacecraft Arrays

The advent of economical, fast integrated-circuit technology has recently surpassed the previously restrictive data processing requirements of forming large sparse and synthetic apertures in space. Many people have now started to claim that their use offers potential benefits by reducing the mass and cost of remote sensing systems for high-resolution imaging.

The angular resolution of any aperture scales with the overall aperture dimension, expressed in wavelengths λ. That is, $\theta \approx \lambda / D$.

An array is an aperture excited only at discrete points or localized areas. The array consists of small radiators or collectors called elements. An array with regularly spaced elements is called a periodic array. To avoid grating lobes in the far-field radiation pattern, the elemental spacing of a periodic array should be less than one-half of the wavelength.

The concept of the spacecraft array involves forming a large thinned aperture from a set of satellites, each acting as a single radiator element. Because the spacing between satellites is very much greater than the characteristic wavelengths, grating lobes can be avoided only by positioning the satellites to avoid periodicities. This can be done by a random placement of satellites or by arranging them such that their relative separations are prime.[3]

The resolution of sparse arrays can be very much larger than an equivalent filled aperture. This arises from the enlarged overall array dimension resulting from splitting and separating the aperture into elements. Consider the case of an imaging system capable of 1-m resolution at a wavelength of 0.5 μm (green-visible). A geostationary satellite would require diffraction-limited optics 18 m across. Similar resolution for lower frequencies (X band, etc.) requires even greater aperture sizes. This is clearly impractical for filled aperture systems. A filled aperture must be supported over its entire extent, leading to heavy structures. Even if mass can be kept low through the use of advanced materials, impressive deployment techniques would be required to stow such an antenna within the launch shroud.

A sparse aperture can be made very large indeed, the only requirement being that the signal at each aperture be known, with measured and preserved phase. Widely separated elements connected through light tethers or booms could easily extend over length scales of 10–100 m (Ref. 4). For even larger baselines, a sparse array of separated spacecraft allows resolutions in the submilliarcsec range.

Rate and Integrity Improvements

For many applications, the requirements for a high information rate or high integrity drives the designer toward very large apertures and high-power payloads. The probability of correctly detecting information symbols in the presence of noise is a function of the energy in each information symbol. Collecting or transmitting symbols at a high rate with a low probability of error, therefore, requires high-power signals. This in turn leads to high-power transmitters or large apertures to collect more power or to concentrate the power radiated.

Distributed systems can offer large improvements in both rate and integrity compared to singular deployments by reducing the impacts of noise and interference. The interfering noise can arise from several sources, including 1) thermal noise from resistive heating of electrical components in the receiver, 2) noisy radiation sources in the field of view (FOV) of the instrument, 3) jamming from unfriendly systems, 4) interaction with the transmission medium (rain, bulk scatterers), and 5) background clutter.

The ergodic property of thermal noise means that integrating over multiple detectors gives the same processing gain as integrating the signal plus noise from a single detector over time. The advantage is that there is no penalty in rate. Also, the interference from noisy radiating bodies in the FOV of one satellite may not be an issue for a second satellite due to the differing viewing angle of the scene. Jamming interference may also be satellite specific; an enemy can easily disrupt a single satellite but would struggle to jam an entire group of satellites that may be spatially separated. In general, the level of improvement in rate and integrity that is offered by distributed architectures varies across the different system classes.

Rate and Integrity Improvements for Collaborative Systems

For a given level of integrity, a collaborative system can achieve higher rates by summing the capabilities of several satellites that individually operate at modest rates. This is equivalent to division of the top-level task into smaller, more manageable tasks that can be allocated among the elemental components of the architecture. The responsibilities of each satellite in the collaborative system reduce linearly with the number of satellites in the cluster. Each satellite can allocate more of its resources to each source, satisfying higher rate requirements. Increasing the number of satellites in the cluster yields linear increases in the achievable rate of information flow from each source. The limit is reached when each satellite is dedicated to a single user. The maximum rate for that user would then be the maximum supportable rate of the satellite.

Equivalently, at a given rate, the level of integrity increases with the number of satellites in a collaborative cluster. The energy per symbol E_s increases with the number of satellites, as a result of the increased dwell time allowed by the task division. If each satellite coherently integrates the received signal, linear increases in the dwell time result in linear increases in the energy per symbol to noise density ratio (E_s/N_0). The integrity will then improve almost exponentially because the error probability can be approximated as an exponential function of E_s/N_0.

In actuality, the integrity improvements will not be quite this good because this result assumes the detection is limited only by stationary white noise. Linear integration gains are achieved only with coherent integration of a signal plus random noise. This is the case if the dominant interference source is thermal receiver noise. Unfortunately, any correlated interference experiences the same linear gains during the integration process. This is a critical consideration for active systems where clutter returns from the ground are not at all suppressed by time integration. For this reason, collaborative distribution cannot improve the clutter-limited performance of radar or lidar systems over that achievable with singular deployments.

Rate and Integrity Improvements for Symbiotic Systems

Unlike collaborative systems, a symbiotic architecture does not, in general, give a simple linear improvement in rate capabilities with increases in the numbers of satellites. In fact, the relationships between the number of satellites and the achievable rate and integrity are different for passive systems and active systems.

Consider first passive symbiotic clusters that form sparse receiver arrays. The signal power to noise power ratio (SNR) behavior of sparse arrays is identical to a filled aperture of the same physical collecting area. To show this, consider a cluster of n_s satellites, each with aperture area A. Assume the array is illuminated by a distant source. Each satellite measures the radiation field, and the signals from the different satellites are then combined to deliver a single waveform to a detector. In the most general case, the input signal power varies across the array. When unit impedance is assumed throughout, the average signal power is

$$E[S_i] = E\left[\frac{1}{n_s}\sum^{n_s} s_{ij}^2\right] \tag{1}$$

where the subscript i refers to the input side of the array and s_{ij} is proportional to the envelope of the rf signal voltage for the jth satellite. Unless the array is so large that the signal strength varies across the array due to path length differences, the signal strength across the array will be constant. In this case, $s_{ij} = s_{ik} = s_i$, and the signal power per satellite is $S_i = s_i^2$. Under the assumption that all signals are cophased by a bank of phase shifters, the output signal voltage after integration is $n_s s_i$ and the output signal power of the array $S_0 = n_s^2 S_i$ because all of the signals add in phase.

If the dominant noise source is thermal noise, then the noise input at each of the satellite apertures will be independent, with zero mean. The average input noise power will be

$$E[N_i] = E\left[\frac{1}{n_s}\sum^{n_s} n_{ij}^2\right] \tag{2}$$

The noise does not add in phase because it is uncorrelated, and so the output noise voltage after combining is given by

$$n_0 = \sum^{n_s} n_{ij} \tag{3}$$

and its square is the output noise power

$$N_0 = \left(\sum^{n_s} n_{ij}\right)^2 \tag{4}$$

$$= \sum^{n_s} n_{ij}^2 + \sum_j^{n_s}\sum_k^{n_s} n_{ij}n_{ik} \tag{5}$$

Because the noise sources are independent with zero mean, the second term is zero, leaving the average noise power to be $E[N_0] = n_s E[N_i]$. The output SNR is, therefore, equal to

$$\mathrm{SNR} = S_0/E[N_0] = n_s S_i/E[N_i] = n_s(\mathrm{SNR})_1 \tag{6}$$

where $(\mathrm{SNR})_1$ is the signal to noise power ratio for a single satellite of the cluster. The improvement in SNR compared to a single satellite is, therefore, n_s, the number of satellites in the array. The same SNR is achievable with a filled aperture of area $n_s A$ receiving the same signal and with the same average thermal noise temperature. This of course makes sense because the same amount of energy is collected over the same collection area in both cases.

Even larger benefits in SNR can be obtained with active symbiotic systems. Recall that active systems are defined to be systems that have to provide the power for the signal to make the round-trip journey to the information source. The active system may have several transmitting satellites that illuminate the source. If the transmitters radiate coherently, the power incident on the information source is increased quadratically because the signal amplitudes add. Alternatively, if the transmitters radiate independently, the power at the source sums linearly. The incident power is then reflected back to be collected by the cluster of satellites, with the receive characteristics described. The resulting SNR improvement for the symbiotic system compared to a single satellite is given by

$$(\mathrm{SNR})_{\mathrm{sym}} = n_t^2 n_r (\mathrm{SNR})_1 \tag{7}$$

Note that n_r can be greater than n_s; if each of the n_s satellites transmit incoherent but uniquely identifiable signals and each satellite receives all n_s transmissions, a total of $n_r = n_s^2$ different signals can be coherently integrated. This is the operating principle behind Techsat21, the U.S. Air Force's (USAF) most recently proposed space-based radar concept.

The integrity is a function of E_s/N_0, given by multiplying the SNR by a dwell time corresponding to the duration over which the signal is integrated. For a tracking mission, the symbiotic cluster must cycle through all of the targets one at a time, so there is no difference in the dwell time compared to the single satellite case. For a search mission, however, there is a penalty paid for coherence. The beamwidth scales with the overall synthetic aperture dimension as opposed to the physical aperture size of each satellite. For a given area coverage rate, the symbiotic cluster must scan its smaller beam more quickly:

$$t_{\mathrm{sym}} = (D_s/D_c)t_1 \tag{8}$$

For coherence only on receive, multiple receive beams can be formed simultaneously to fill the satellite FOV. The dwell time then scales the same as that of a single satellite.

The resulting E_s/N_0 relationships for both the search and track missions are summarized in Table 2. To simplify the results, it has been assumed that the symbiotic cluster has n_s satellites and can operate in three different modes, a passive mode in which all n_s satellites are used to form a coherent receive array, an active mode in which each satellite independently transmits (incoherent) and all the satellites coherently receive all of the signals ($n_r = n_s^2$), and a coherent transmit and receive mode ($n_t = n_r = n_s$).

Interestingly, collaborative and symbiotic clusters both achieve linear improvements for passive missions but for quite different reasons. The collaborative system gains benefit from task division, increasing the allowable dwell time, whereas the symbiotic cluster achieves the same linear improvement by increasing the SNR.

Notice that the symbiotic system with coherence on transmit and receive is not well suited to the search mission unless many satellites can be deployed over a reasonably short extent, such that $n_s^3 > (D_c/D_s)$. The largest and most realizable benefit from distribution for the search mission can be gained with several independent transmitters and coherent integration of all of the received signals. This is the reason supporting the development of the Techsat21 space-based radar program.

Of course, in the presence of a heavy clutter background, the detection is not noise-limited and the results change somewhat. However, this is where symbiotic clusters can really help. As a direct result of the smaller beamwidths that are characteristic of symbiotic systems, the clutter rejection of the system is greatly improved compared to single satellites or collaborative systems. Consequently, the improvements in the E_s/N_0 seen in Table 2 are conservative estimates of the benefits offered by symbiotic clusters.

Table 2 Factor of improvement in the energy per symbol to noise density ratio for distributed clusters compared to singular deployments

	Collaborative		Symbiotic		
Mission	Passive	Active	Passive	Active coherent receive	Active coherent transmit/ receive
Search	n_s	n_s	n_s	n_s^2	$(D_s/D_c)n_s^3$
Track	n_s	n_s	n_s	n_s^2	n_s^3

The tempting conclusion to draw from this is that symbiotic clusters are beneficial for missions requiring high rates and integrity. Unfortunately, there is a crucial factor that has been omitted so far. The data processing requirements placed on symbiotic systems are extremely restrictive and are on the frontier of what can be achieved with today's technology. This issue will be discussed in a later section.

Availability Improvements

If carefully designed, a distributed architecture can often lead to improved capabilities by increasing the availability of system operations. Losses of availability can arise due to increased numbers of users accessing the limited resources of the system, signal attenuation from blockage or weather, or from statistical fluctuations due to noise and clutter. More commonly, a loss of availability can also be attributed to poor viewing geometries or poor coverage statistics. For example, reconnaissance satellites may have to image scenes over two or more continents, relaying the data to multiple downlink stations across the world. There will be times during the orbits of these satellites when they are not passing over important targets. The system is unavailable at these times because images of the targets cannot be recorded. The revisit time of the satellites effectively specifies the maximum availability that is built into the system. Of course, very high availabilities can only be achieved by constellations giving continuous coverage over the target regions.

Note that the availability of a system is related to the variance of the supportable isolation, rate, and integrity and, as such, is sensitive to worst-case scenarios. Distributed architectures can lead to increased availability through better coverage of the demand, or reducing the variability of the capabilities.

Improved Visibility and Coverage Geometry

There are some instances when distribution and multifold coverage improve the availability by reducing the variability of the system capabilities. By making the behavior of the system more predictable, the probability of operating within acceptable bounds is increased. The capabilities of the system are particularly sensitive to coverage variations, and it is here that distribution can lead to improvements. The multifold coverage that is characteristic of distributed architectures supports consistent capability in two ways.

1) The first way is by reducing the variance of the visibility, defined as the number of satellites in view from a ground station. Generally, the visibility is a function of both space and time. The number of satellites in view from a location changes in time and is usually different at other locations. The capabilities of a satellite system are frequently dependent on the visibility. Large variations in visibility can, therefore, cause large fluctuations in the isolation, rate, or integrity. The designer faces the choice of sizing the system for the worse-case coverage, or accepting losses of availability at times when the visibility is below average. Increasing the number of satellites in the constellation not only increases the visibility, but also reduces the variance. According to the central limit theorem, as the number of satellites is increased, the minimum visibility converges toward the average value. This assists the designer, improving the availability of systems based on average coverage characteristics.

2) The second way in which the multifold coverage supports consistent capability is by reducing the impact of the variability in the capabilities of individual satellites in collaborative systems. The geometry of the coverage over target regions can have a large impact on the sensitivity of the system. Frequently, the isolation, rate, or integrity that can be supported by a single satellite can be spatially and time varying, depending on the viewing angle, the transmission path, and the detector characteristics. Favorable coverage geometries minimize the impact of these variations, ensuring that the combined operation of the collaborative configuration achieves consistent levels of capability.

These two concepts are most easily understood with the help of a simple example.

Example: Visibility and Geometry (Distributed Space-Based Radar[5])

Consider a collaborative space-based radar system consisting of a cluster of satellites in common view of a theater of interest. The system satisfies requirements on the rate (target update) and integrity (probability of detection or false alarm) by summing the capabilities of several small radars that independently search the same target area. The cumulative rate is, therefore, directly proportional to the number of satellites in view of the target area, as indicated in Table 2. Variations in the visibility translate directly into variations in the achievable rate or integrity. This can result in a loss of availability if the visibility drops below that necessary to support the requirements. The availability can be improved if the system is designed to use an even greater number of smaller satellites to satisfy the detection requirement. As the number of satellites increases, the spatial and temporal variations in the visibility are reduced. The minimum visibility approaches the average value, and the achievable detection rate changes over a much smaller range.

Furthermore, larger configurations of satellites result in more favorable coverage geometries. The multifold coverage leads to a wide distribution of viewing angles surrounding the target. This is particularly important for slow-moving targets. The radar return from slow-moving targets is difficult to distinguish from the ground clutter. Normally the different velocities of the target and the ground relative to the radar give rise to different Doppler shifts that separate the target and clutter in frequency, allowing detection. The return from slow-moving targets is often buried in the clutter because of the low relative velocities. A viewing angle parallel to the target's velocity maximizes the Doppler shift between the target and the ground in the frequency spectrum, increasing the signal isolation and improving the probability of detection. Because the target's velocity vector is unknown a priori, receivers must be placed at all possible viewing angles to ensure detection. With receivers located at all angles around the target, the distributed space-based radar concept increases the probability of detecting slow-moving targets by being less sensitive to the orientation of target velocity vector. This effectively increases the availability by reducing the probability of failing the detection requirement.

Reducing the Baseline Cost

Initial deployment costs for a given satellite constellation include costs associated with development, production, and launch of the system's original complement of satellites. Additional expenditures beyond the initial deployment costs are necessary to maintain the constellation over a given time period. These costs include the production, storage, and launch costs associated with the on-orbit or on-ground spares and also all of the operations costs. The baseline system cost is the sum of the initial deployment costs and the maintenance costs. Baseline costs are typically very high. For distributed satellite systems to be considered viable, they must be at least competitive in cost as compared to traditional systems.

Conventionally, system cost estimates can be made using basic parametric models such as the USAF unmanned spacecraft cost model (USCM), or the small satellite cost model (SSCM).[6] These models consist of a set of cost estimating relationships (CERs) for each subsystem. The total cost of the system is the sum of the subsystem costs. The CERs allow cost to be estimated as a function of the important characteristics, such as power and aperture. Frequently, they are expressed as a power law, regressed from historical data.

Care must be taken in applying the SSCM to distributed systems. Although each satellite in a distributed system may be small, the SSCM was derived assuming single-string designs and modest program budgets. This is clearly unsuitable for a distributed system of perhaps 1000 satellites, with a total system cost of several billion dollars. Unfortunately, the use of USCM generally leads to high costs for distributed systems. This is due to two factors:

1) In partitioning the mission and allocating tasks among separate components, the total hardware resources required on orbit are often increased. Among other things, this is a result of having to add redundancy to overcome serial reliability problems. Consider the case of single satellite satisfying a demand with reliability of 0.9 over the mission lifetime. To achieve the same overall reliability with two collaborative satellites of half the size, an additional redundant satellite is also necessary. In this example, the total resources on orbit for the distributed system is 50% more than for the single deployment. Because the CERs base cost on characteristic resource, the result of this increase of total hardware is an increase in cost.

2) Typically, the USCM power laws in the CERs are nonlinear, with an exponent less than unity. There is a higher marginal price per kilogram of mass, or per square meter of aperture for smaller systems. As a result, it is more expensive to divide a large system (especially aperture or power) into smaller components.

It would appear then that distributed satellite systems are characteristically more expensive than singular deployments. However, there are additional factors that can sway the balance in favor of distribution.

First of all, there is a question as to the validity of using the USCM for estimating the cost of modern distributed satellite systems. The basic problem here is that the model is based on regression from historical data of past military satellite programs. As such, the CERs of the USCM may not reflect modern trends or practices. The programs from which the model was derived were not subject to the same budget constraints as modern systems. Stated simply, past military satellite programs were expensive because they were allowed to be. An additional point is that conventional cost models, being based on historical data, reflect an industry that was crippled by a conservatism and a reliance on risk avoidance. The high baseline cost of space systems was perhaps the largest reason for the conservatism. The enormous initial expenditure, added to the characteristically high risk, led to a reliance on tried and tested practices and established technologies. Unfortunately this doctrine was self-supporting, being usually more costly than modern alternatives and, thus, serving only to refuel the conservatism.

There are, however, some indications that things are changing. The advent of small satellite technology has heralded a new era of satellite engineering that minimizes costs by risk management rather than risk avoidance.[7,8] A willingness to accept some risk can lower the cost of satellite programs, enabling more missions to be flown and allowing new technology and innovative techniques to be implemented.[6,8,9] The use of commercial off-the-shelf technology can lead to substantial cost savings in development and operations (legacy systems often require specially trained operators). By accepting high risk and implementing strategies to manage failures, small satellites have been successfully designed, built and operated at a fraction of the cost of traditional systems.[10] Should distributed satellite systems really proliferate the market, they will achieve low costs by lowering the requirements on individual satellite reliability, taking advantage of the redundancy built into the architecture.

The changes in the space industry have not been restricted to the small satellite arena. The commercial satellite industry is just now beginning to realize the benefits of modernized design practices, moving away from the concept of the handcrafted satellite and instead adopting the production-line approach to satellite design and construction. Standardized bus designs with modular interfaces to many different payloads reduce the development time and simplify assembly and test. Whereas the CERs of the USCM assume a single-string design, favorable economies of scale can result from bulk manufacture. The production of a larger number of small units allows quicker movement down the learning curve. This is made possible by economies of scale in manufacture and by modifying the way that satellites are built and assembled. Such practices are poorly represented by existing cost models.

The cost of launching a satellite system can make up a significant portion of the baseline costs. This is especially true of distributed satellite systems featuring many small satellites. Typically, launch costs do not scale linearly with mass. The price per kilogram is higher for lower-mass payloads. Unless bulk-rate contractual agreements can be made with launch providers, learning curve discounts do not apply to launch costs. This would suggest that the launch costs of distributed systems is greatly increased compared to traditional singular deployments. However, although each satellite in a distributed system may be small, when considered as a whole, the entire system can be huge. Economies of scale support the larger launch vehicles, and so, subject to volume and orbit constraints, it is cheaper to deploy the initial constellation using large launch vehicles. An entire orbital plane of satellites could be deployed on a single launch, giving the added benefit of distinct performance plateaus. The initial launch costs of distributed systems, therefore, scales more like that of large satellites and should be priced based on the total constellation mass rather than on the individual satellite mass.

Collaborative distributed systems also offer the possibility of being able to ramp up the investment gradually, to match the development of the market. Only those satellites needed to satisfy the early market are initially deployed. If and when additional demand develops, the constellation can be augmented. The cost of constructing and launching these additional satellites is incurred later in the system lifetime. Because of the time value of money, the delayed expenditure can result in significant cost savings.

Each of these factors help to offset the apparently high costs suggested by conventional parametric cost modeling. Consequently, the baseline cost associated with a distributed satellite system may actually be smaller than for a comparable large-satellite design. This is not always true because of extreme sensitivity to the application.

If the proposed microsatellite systems become a reality, the current costing paradigm will change completely. Cost models that scale with unit cost, modified only by a learning curve, are not really applicable to microfabrication or batch processing techniques. The microfabrication of solid-state components involves huge production runs, and so the cost is reasonably insensitive to the actual number produced, being dominated by startup costs. An interesting caveat to be considered here is the increased component reliability resulting from mass manufacture. As a result of the manufacturing process, mass-manufactured products have a very low variability in production standards and, therefore, have a characteristically high reliability.

Reducing the Failure Compensation Cost

In addition to the baseline costs, expenditure is necessary to compensate for any failures that cause a violation of requirements during the lifetime of the system. Expected failure compensation costs can be minimized by lowering the probability of failures, by reducing the impact of failures so that compensatory action is not needed, or by reducing the cost of that action. Clearly the expected failure compensation costs are closely related to the overall reliability of the system. System reliability can be improved by deploying redundancy, or by improving the quality of the components. Both of these options add to the cost of the system. Generally, a larger initial expenditure in improving the system reliability leads to smaller compensation costs.

Note that distribution can improve reliability only if there is redundancy in the design. A distributed architecture with total resources that can only just satisfy the demand is a serial system and is subject to serial failure modes. Under these conditions a failure in any component will lead to a failure of the system. The system reliability would be the product of the reliabilities of the components, which would decrease geometrically with the number of serial components. Only by adding redundancy can a distributed architecture take advantage of parallel reliability. System failure of a redundant architecture occurs only if all parallel paths fail.

In general, most architectures will require some redundancy to satisfy mission reliability requirements throughout the expected lifetime. Frequently the cost associated with this redundancy is less for a distributed architecture than it is for traditional systems. This reliability cost accounts for the production, storage, and launch of on-orbit or on-ground spares necessary to maintain operations. For a distributed system these spares often represent only small fractions of the initial deployment.

For collaborative systems, the system degradation is linear with the number of satellite failures. When the number of satellites drops below that needed to satisfy the size of the market, either replacements must be launched, or the system will incur opportunity costs corresponding to the part of market that is not served. Deployed redundancy simply provides initial capability over and above that necessary to satisfy the market. In the absence of any other compensatory action, the system capabilities will continuously degrade toward the minimum acceptable levels. If enough redundancy is deployed, this point will not be reached within the system's designed lifetime.

Redundancy in symbiotic systems has a different role. Individual satellite failures do not have a linear relationship with the degradation of the overall system. In fact, a small number of satellite failures may have no noticeable impact on the system capabilities. If, however, the number of satellites in the cluster falls below some safe limit, the cluster will simply not operate at all. For example, the users of GPS can obtain navigation solutions provided at least four satellites are in view. Usually, many more than this minimum number are visible from any ground location, but should failures occur such that this is not the case, a navigation solution cannot be obtained at all. This can happen in some ground locations if as few as two satellites fail from the existing constellation of 24 satellites. The consequences of failure, in terms of the opportunity cost, are, therefore, very much greater for symbiotic systems.

For certain satellite missions, a distributed architecture may also lower lifetime costs by reducing the cost of any failure compensation that is necessary. A recent design study at the Massachusetts Institute of Technology (MIT)[11] showed that distributed systems appear to yield the greatest cost savings under two conditions:

1) The first condition is met if the components being distributed make up a large fraction of the system cost. It is prudent to distribute the highest cost components among many satellites. Do not carry all of your eggs in one basket.

2) The second condition is met if the components being distributed drive the replacement schedule of the spacecraft.

These savings manifest themselves in a number of ways. First of all, for the distributed architecture, the replacements represent only a small fraction of the initial deployment, whereas in a traditional design, the entire space segment must be replaced after a failure. Also, the replacements, on average, occur later, thus realizing larger savings from discounting to constant year dollars. The potential savings over traditional singular deployments are demonstrated very well in the following example, taken from the MIT design study.

Example: Replacement Costs (Polar Orbiting Weather Satellites[11])

Instruments aboard polar orbiting weather satellites, such as the proposed National Polar Orbiting Earth Sensing Satellite (NPOESS) system, are classified as either primary or secondary. Because the primary instruments provide critical environmental data records, failure of a primary instrument necessitates replacement. A secondary instrument is one whose failure may be tolerated without replacement. If an orbital plane's complement of sensors are all located on a single satellite, failure of any primary sensor will require redeployment of all of the plane's sensors. By distributing the primary instruments intelligently across a cluster of several smaller spacecraft, it may be possible to reduce the cost of the system over its lifetime because the plane's entire complement of sensors are not redeployed after every failure.

Consider the following three configurations illustrated in Fig. 3. The blocks labeled as A, B, and C represent three primary instruments required in a given orbital plane.

The total costs over a 10-year mission life were calculated for each of the three cluster configurations. As shown in Fig. 4, the costs over the 10-year period are broken down into three categories; namely initial deployment, required spares, and expected replacements. Initial deployment includes the development, production, and launch

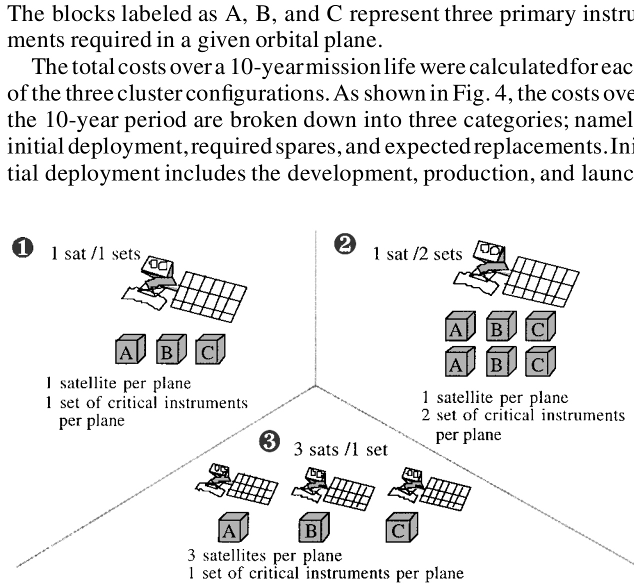

Fig. 3 Satellite and sensor configurations.

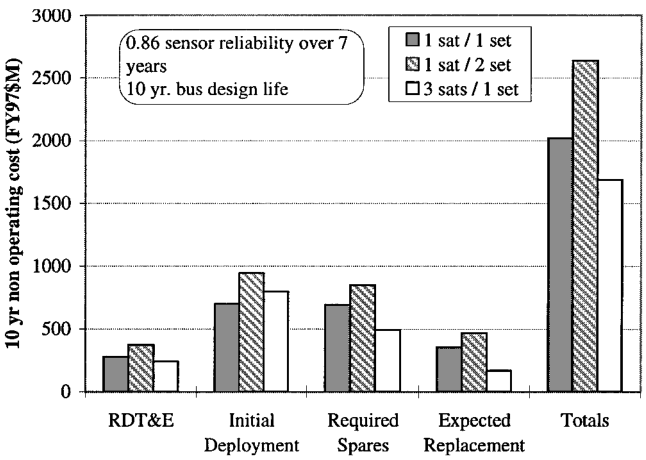

Fig. 4 Total system costs over the 10-year mission life of a polar orbiting weather satellite system.

costs for each orbital plane's original complement of spacecraft. The number of required bus, payload, and launch vehicle spares were derived from a Monte Carlo simulation of the mission, assuming reasonable component reliabilities.

Figure 4 shows that the initial deployment cost is least expensive for the {1sat/1set} configuration. Adding a redundant sensor to the single-satellite configuration greatly increases initial deployment cost in terms of larger bus size, additional instruments, and more expensive launch vehicles. The {3sat/1set} configuration, although being launched on a less expensive vehicle, is slightly more expensive than the {1sat/1set} configuration due to the duplication of bus subsystems and some sensors on each of the three smaller satellites.

Distributing the primary instruments among three satellites significantly increases the reliability of each individual satellite. Higher satellite reliability and lower replacement launch costs result in the {3sat/1set} configuration having the lowest expected replacement cost. Once again, the slight increase in reliability gained from adding redundant primary instruments for the {1sat/2set} configuration is outweighed by the higher bus, payload, and, launch costs.

To summarize, distribution within a satellite mission may reduce the replacement costs over the lifetime of a mission. A modular system benefits not only because a smaller replacement component has to be constructed but also because there are huge savings in the deployment of the replacement. These savings are the greatest when the component(s) being distributed make up a large fraction of the system cost and drive the replacement schedule.

Issues and Problems

There are some factors that are critical to the design of a distributed architecture that were irrelevant to the design of traditional systems. Depending on the application, these issues may be minor hurdles, or could be so prohibitive that the adoption of a distributed architecture is unsuitable or impossible.

Modularity vs Complexity

An important issue to be addressed is the level to which a system should be distributed. How much can the system be divided into smaller components and still offer the benefits discussed earlier? The central issue here is the trade between the advantages of modularization and the cost of complexity.

Modularization

In distributing the functionality of a system among separate satellites, the system is essentially being transformed into a modular information processing network. The satellites, subsystems, and ground stations make up individual modules of the system, each with well-defined interfaces (inputs and outputs) and a finite set of actions. Such systems are analogous to the distributed inter- and intranet computing networks and, as such, are subject to similar mathematics.

One beneficial aspect of modularization comes from an improved fault tolerance. System reliability is by nature hierarchical in that the correct functioning of the total system depends on the reliability of each of the subsystems and modules of which the system is composed. Early reliability studies in distributed computing[12] showed that the overall system reliability was increased both by applying protective redundancy at as low a level in the system hierarchy as was economically and technically feasible and by the functional separation of subsystems into modules with well-defined interfaces at which malfunctions can be readily detected and contained. Clearly, subdividing the system into low-level redundant modules leads to a multiplication of hardware resources and associated costs. However, the impact of improved reliability and lower replacement costs over the lifetime of the system can outweigh these extra initial costs.

These arguments suggest that a system should be separated into modules that are as small as possible. However, there are some distinct disadvantages of low-level modularization that must be considered. The most important of these are the costs and low reliability associated with very complex systems.

Complexity

The complexity of a system is well understood to drive the development costs and can significantly impact system reliability. In many cases, complexity leads to poor reliability as a direct result of the increased difficulty of system analyses; failure modes were missed or unappreciated during the design process. For a system with a high degree of modularity, these problems can offset all of the benefits discussed earlier.

Although each satellite in a distributed system might be less complex, being smaller and having lower functionality, the overall complexity of the system is greatly increased. The actual level of complexity exhibited by a system is difficult to quantify. Generally, however, it is accepted that the complexity is directly related to the number of active interfaces between the components of the system. Although the actual number of interfaces in any system is architecture specific, it is certainly true that a distributed system of many satellites has more interfaces than a single satellite design. Network connectivity constraints mean that the number of interfaces can increase geometrically with the number of satellites in a symbiotic architecture. This is an upper bound; collaborative systems exhibit linear increases in interfaces with satellites. The complexity of a distributed system is, therefore, very sensitive to the number and connectivity of the separate modules.

The impact of this additional complexity is difficult to evaluate, especially without a formal definition of how complexity is measured. Recent studies at MIT[13,14] would, however, suggest that complexity can cause significant increases in development and qualification time, increases in cost, and losses of system availability. For these reasons, the level of modularization must be carefully chosen. Only with thorough system analysis and efficient program management can the impacts of complexity be minimized.

Clusters and Constellation Management[15,16]

Earlier it was argued that by combining the capabilities of many individual elements, systems of small or microsatellites can be used for high-rate or resolution applications. For these applications, the relative positions and dynamics of the satellites in the cluster are a critical factor in the design.

Equations of Motion for Local Clusters

Recall that in local clusters a group of satellites fly in formation, and their relative positions are controlled within specified tolerances. The relative motion of the satellites can be predicted by linearized perturbations of the equations of motion about a reference orbit. The linearization is valid if the cluster diameter is small compared to the radius of the reference orbit. For circular reference orbits, the linearized set of equations are known as the Hill's equations[17]:

$$\ddot{x} - 2\dot{y} - 3n^2x = a_x, \qquad \ddot{y} + 2\dot{x} = a_y, \qquad \ddot{z} + n^2z = a_z \quad (9)$$

The acceleration terms on the right-hand side represent all noncentral force effects (drag, thrust, gravity perturbations, etc.). The right-handed local coordinate frame has x pointing up and y aligned with the velocity direction of the reference orbit. These equations can be used to estimate the propulsive requirements placed on satellites constrained to orbit in clusters. The different cluster configurations are characterized by different degrees of freedom in x, y, and z in Eq. (9).

There are, in fact, many ways to create local clusters. Two will be considered here. The first is to fly the satellites in rigid formation, maintaining their relative positions and orientation. This involves constant values of x, y, and z for each satellite because the position is fixed relative to the reference orbit. The second option, which proves to be usually more realizable, is to allow the cluster configuration to rotate, maintaining only the relative intersatellite separations. This option has x, y, and z that are constrained to follow circular trajectories around the reference orbit. The propulsion requirements for each of these options are described hereafter.

Rigid Clusters

Because the reference orbit is assumed to be a circular Keplerian orbit, rigid clusters must feature some satellites in non-Keplerian orbits. These non-Keplerian orbits are characterized by either a focus that is not located at the Earth's center of mass, or orbital velocities that do not provide the proper centrifugal acceleration to offset gravity at that altitude. The Earth's gravity will act to move these satellites into Keplerian orbits, giving rise to tidal accelerations exerted on the satellites that are a function of the cluster baseline and orbit altitude. To maintain relative position within the cluster, these accelerations must be counteracted by thrusting.

The required amount of thrusting to maintain the cluster for a single orbit can be estimated from Eq. (9), by setting all time-varying terms to zero and integrating over a single orbital period ($=2\pi/n$). The result is that

$$\Delta V \text{ (per orbit)} = 2\pi n\sqrt{9x^2 + z^2} \tag{10}$$

If the cluster radius is $R_c = \sqrt{(x^2 + y^2 + z^2)}$, this result suggests that the ΔV requirements for rigid clusters scale to first order as $10nR_c$ ms^{-1} per orbit. At low-Earth-orbit (LEO) altitudes, $n \approx 0.001$ rad/s, and so $\Delta V \approx 0.01R_c$ ms^{-1} per orbit.

For a particular propulsion specific impulse I_{sp} and propellant mass fraction f_p, the lifetime as a function of the ΔV per orbit is[16]

$$\text{Lifetime (years)} = -\left(\frac{\text{years}}{\text{orbit}}\right)\left(\frac{\text{orbit}}{\Delta V}\right) I_{sp}\, g\, \ell n(1 - f_p) \tag{11}$$

For reasonable propellant mass fractions of around 10%, even propulsion systems with high specific impulse (2600 s) cannot maintain a 100-m cluster in LEO for more than six months. This makes the implementation of rigid clusters extremely unlikely.

Circular (Dynamic) Clusters

An alternative to holding the clusters rigidly is to allow the satellites to rotate in circles around each other, in a plane defined by a normal aligned in the viewing direction, such that their relative separations are preserved. In this case, x, y, and z are only constrained to lie on a circle. The period of rotation of the circle is the same as the orbital period, so that the satellites have the same natural frequency as the reference orbit. In the plane of the cluster, the general motions of the satellites are described by

$$x' = 0, \qquad y' = R_c \cos(nt + \alpha), \qquad z' = R_c \sin(nt + \alpha) \tag{12}$$

where α is a phase angle. These equations can be transformed into the Hill[17] frame by standard rotational transformation through angles ϕ and ψ corresponding to the azimuthal and elevation angles of the line of sight to the nadir (negative x) direction. This gives the constraints on x, y, and z for satellites in the Hill frame. Sedwick et al.[16] integrate Eq. (9) with these constraints for all values of ϕ and ψ over a single orbit, and conclude that, at worst, the ΔV per orbit scales as $3nR_c$. From the same arguments as were made for rigid clusters, this leads to lifetimes of at most 18 months for LEO clusters of 100-m diam.

However, there are some angles (side looking at 30 deg off nadir), at which there is no ΔV required to maintain the circular configuration. These represent free-orbit solutions to the problem. If this off-axis viewing angle can be tolerated, the propulsion requirements are reduced to only that needed to overcome perturbations. Over reasonable cluster sizes, these perturbations exert negligible differential forces to distort the cluster and only act to perturb the cluster as a rigid body.

The results presented in this section would suggest that maintaining clusters is prohibitively difficult if the cluster is required to move only as a rigid body. This is unfortunately the requirement placed on optical interferometers, which need differential paths to be preserved very accurately such that the same wavefront is measured at the different apertures in real time. However, if a sparse array is to be formed at radio frequencies, there is a possibility for the signals from different satellites to be combined during postprocessing after digitization. Time delay can be easily introduced during the interfering process, and the distance of any given satellite from the source is no longer an issue. This relaxes a degree of freedom in Eq. (9) because the satellites are no longer bound to move in a plane. No results have yet been presented, but it is suggested that great propellant savings will be realized from allowing this behavior.[16]

Spacecraft Arrays and Coherence

There are many technical difficulties involved with the design and construction of symbiotic clustellations that form spacecraft arrays, mainly due to the requirement for signal coherence between widely separated apertures. This is especially true for systems intended for Earth observation. Interferometric techniques are not well suited to Earth observation from orbit because the Earth forms an extended source, unlike the astronomical sources that lie embedded in a cold cosmic background. This forces a need for very high SNRs and high sampling densities,[18] leading to designs featuring a large number of satellites. For high-resolution imaging applications, requiring either long integration times or high SNRs, the situation is made worse by forward motion of LEO satellites limiting the time over the target. This forces more simultaneous measurements to be made to reach the required SNR and, therefore, requires even greater numbers of elements. Furthermore, although there may be no grating lobes to consider, the thinned and random array will exhibit large average sidelobe levels. For randomly distributed arrays, the ratio of the power in the main lobe to the average sidelobe power is approximately equal to the number of elements in the array.[19] For many detection applications, the maximum sidelobe power should be much lower (more than 10 dB lower) than the main beam power. Using this measure gives bounds on the order of 10 for the minimum number of satellites that must be used to form the sparse array.

The formation of sparse apertures using a large number of satellites is complicated by data processing, presenting a barrier to the adoption of sparse array technology. The most generic problems, common to both active and passive clusters, involve using spacecraft arrays as receivers. The signals from each element of a receiver must be combined coherently. The data processing requirement scales quadratically with the number of elements, and the equipment becomes very costly as the aperture size grows. The actual exchange of signals between receivers and combiners also poses a difficult challenge. For an optical interferometer the exchange is done simply by routing the analogue signals from the pair of collectors to a common combiner, constraining the optical paths to be equal in each case. It is difficult to adopt the same strategy for arrays with many elements. Because the satellites are remote from each other, there is no easy way of simultaneously combining the signals from all of the different elements in an analogue form. For these arrays, the combining is easier done during postprocessing after digitizing the signals while preserving the phase.

This effectively limits the applicability of spacecraft arrays for passive sensing. A passive receiving spacecraft array must record information over a reasonably long period of time to integrate the SNR. This then necessitates enormous storage capacity onboard each satellite, because all of the phase information must be preserved. Sampling the carrier wave at the Nyquist limit with 8-bit quantization would result in storage rates of 96 Gbits/s for an X-band detector. Even with high-speed buffers, the required storage capacity after only a few seconds of integration time is prohibitive. Of course, the receiver can filter and mix the input signal down to a lower intermediate frequency before the A/D conversion, greatly reducing the load on the data processing. This results in no loss of information provided the information bandwidth is known to be small compared to the carrier frequency. In general, the bandwidth of the information collected may be as high as the receiver bandwidth. Sometimes, however, the nature of the target is such that the information content is known to be bandlimited over a reasonable range (kilohertz to megahertz). In these cases, digitization and storage can still be problematic, but at least manageable. An active symbiotic system may benefit here because the characteristics of the transmitted signal are known, and the information content is limited only to the changes observed in the received signal.

There are also technical challenges associated with spacecraft arrays as active transmitters for tracking applications. These systems track targets with a narrow beam, optimizing the signal to noise from the target while nulling the clutter and noise. Correct phasing of the array at the desired angle and range to illuminate the target must be performed in real time. Returns from the target are used by a feedback controller to vary the phase at each element to steer the beam. To do this, each array element must have accurate information about the relative positions of all other array elements. Continuous communication between satellites is needed. Furthermore, the time constant for the detection (including signal reception, combining, processing, and phasing of the transmissions) must match the dynamics of the target. For small local clusters, the slow dynamics of the array may allow this to be carried out if the processing capability exists. For more dynamic clusters this would be a very tricky task.

Summary

A generalized framework has been developed for classifying space system architectures and their characteristics. This framework is applicable for any mission in communications, sensing, or navigation. The generalization is possible because, for each of these applications, the overall mission objective is to transfer information between remote locations and to do so effectively and economically. Summarizing, 1) satellite systems are information transfer systems; 2) information transfer systems serve O–D markets; 3) satellites and ground stations are nodes in a network; and 4) the capabilities of the system are characterized by the isolation, rate, integrity, and availability parameters.

Distributed architectures are enabling for small satellite designs because they expand their useful range of applications to include high rate and resolution sensing and communications. The capabilities of many small satellites are combined to satisfy mission requirements.

A distributed architecture makes sense if it can offer reduced cost or improved capabilities. Distribution can offer improvements in isolation, rate, integrity, and availability. The improvements are not all-encompassing, and in many cases are application specific. Nevertheless, it appears that adopting a distributed architecture can result in substantial gains compared to traditional deployments. Some of the more important advantages that distribution may offer are 1) improved isolation corresponding to the large baselines that are possible with widely separated antennas on separate spacecraft within a cluster; 2) higher net rate of information transfer, achieved by combining the capacities of several satellites to satisfy the local and global demand; 3) improved availability through a reduced variance in the coverage of target regions (this reduces the need to overdesign and provides more opportunities for a favorable viewing geometry); 4) staged deployment on an as-needed or as-afforded basis; 5) progressive technology insertion and modular upgradeability, reducing the impact of technology-freeze dates; 6) improved reliability through redundancy and path diversity; and 7) lower failure compensation costs due to the separation of important system components among many satellites (only those components that break need replacement).

There are some problems, specific to distributed systems of small satellites, that must be solved before the potential of distributed architectures can be fully exploited. The most notable of these problems are 1) an increase of system complexity, leading to long development time and high costs; 2) difficulty of maintaining signal coherence among the apertures of separated spacecraft arrays, especially when the target is highly dynamic; and 3) the need for autonomous operations (if autonomy is not implemented, operations costs will dominate, and for symbiotic systems human intervention may not be sufficiently timely).

The resolution of these issues, and the proliferation of microtechnology, could lead toward a drastic change in the satellite industry. It seems clear that distribution offers a viable and attractive alternative for some missions. More analysis is warranted to completely answer the question of where and when distribution is best applied, but the potential prospects of large cost savings and improvements in performance are difficult to ignore. It is, therefore, likely that increasing numbers of distributed satellite systems will be developed in both the commercial and military sectors.

References

[1] *The Satellite Remote Sensing Industry*, KPMG Peat Marwick LLP, 1996.

[2] Jilla, C., and Miller, D., "A Reliability Model for the Design and Optimization of Separated Spacecraft Interferometer Arrays," *11th Annual AIAA/USU Conference on Small Satellites*, AIAA, Reston, VA, 1997.

[3] Greenaway, A. H., "Prospects for Alternative Approaches to Adaptive Optics," *Adaptive Optics for Astronomy*, edited by D. M. Alloin and J. M. Mariotti, NATO Publishers, 1993, pp. 287–308.

[4] Stephenson, R., Miller, D., and Crawley, E., "Comparative System Trades Between Structurally Connected and Separated Spacecraft Interferometers for the Terrestrial Planet Finder Mission," MIT Space Engineering Research Center, TR SERC 3-98, Massachusetts Inst. of Technology, Cambridge, MA, 1998.

[5] Wickert, D. P., Shaw, G. B., and Hastings, D., "Impact of a Distributed Architecture for a Space-Based Radar," *Journal of Spacecraft and Rockets*, Vol. 35, No. 5, 1998, pp. 703–713.

[6] Bearden, D., "Cost Modeling," *Reducing Space Mission Cost*, Microcosm, 1996.

[7] Sellers, J., and Milton, E., "Technology for Reduced Cost Missions," *Reducing Space Mission Cots*, Microcosm, 1996.

[8] Wertz, J. R., and Larson, W. J. (eds.), *Reducing Space Mission Cost*, Microcosm, 1996.

[9] Parkinson, R., "Introduction and Methodology of Space Cost Engineering," AIAA Short Course, 28–30 April 1993.

[10] Fleeter, R., "Design of Low-Cost Spacecraft," *Space Mission Analysis and Design*, edited by J. R. Wertz and W. J. Larson, 2nd ed., Microcosm, 1993.

[11] "Project Foresight," Space Systems Engineering Final Rept. 16.89, Dept. of Aeronautics and Astronautics, Massachusetts Inst. of Technology, Cambridge, MA, 1997.

[12] "System Reliability and Integrity," Infotech International Ltd., 1978.

[13] Schwarz, R., "A Probabilistic Model of Satellite System Automation on Life Cycle Costs and System Availability," M.S. Thesis, Dept. of Aeronautics and Astronautics, Massachusetts Inst. of Technology, Cambridge, MA, June 1997.

[14] Dutton, E. S., "Effects of Knowledge Reuse on the Spacecraft Development Process," M.S. Thesis, Dept. of Aeronautics and Astronautics, Massachusetts Inst. of Technology, Cambridge, MA, June 1997.

[15] Yashko, G., and Hastings, D. E., "Analysis of Thruster Requirements and Capabilities for Local Satellite Clusters," *10th Annual AIAA/USU Conference on Small Satellites*, AIAA, Reston, VA, 1996.

[16] Sedwick, R., Kong, E., and Miller, D., "Exploiting Orbital Dynamics and Micropropulsion for Aperture Synthesis Using Distributed Satellite Systems: Applications to TechSat21," AIAA Paper 98-5289, Oct. 1998.

[17] Hill, G. W., "Researches in the Lunar Theory," *American Journal of Mathematics*, Vol. 1, No. 1, 1978, pp. 5–26.

[18] Swift, C., and Levine, D., "Terrestrial Sensing with Synthetic Aperture Radiometers," *IEEE MTT-S International Microwave Symposium Digest*, 1991.

[19] Steinberg, B. D., *Principles of Aperture and Array System Design*, Wiley, New York, 1976.

A. C. Tribble
Associate Editor