Managing diversity with handover to provide classes of service in satellite constellation networks

Lloyd Wood, George Pavlou and Barry Evans
Centre for Communication Systems Research, University of Surrey, Guildford, United Kingdom
{L.Wood | G.Pavlou | B.Evans}@surrey.ac.uk

Abstract

The rosette satellite constellation network with intersatellite links (ISLs) presents unique properties, in providing locally separate ascending and descending network surfaces of interconnected satellites with which the ground terminal can communicate. We present a novel approach exploiting this rosette geometry, by use of control of handover and management of satellite diversity, to determine which surface a ground terminal will select for communication. This effectively provides a form of network ingress control. By allocating traffic separate paths through the network with different degrees of delay, different levels of service become available between ground terminals. This allows us to exploit the topology of the rosette constellation network to provide varying classes of service, and thus support for quality of service (QoS).

1.0 Introducing diversity

Diversity is a widely-used term. It simply indicates that a diversity of communication methods is available, i.e. that for a satellite user, there is more than one satellite available for a ground terminal to communicate with at all times. At the physical level, diversity and multiple satellite visibility can be used to combat shadowing by buildings or terrain, to provide redundancy and increase the likelihood of a satellite always being in view of the terminal. The percentage of time that degrees of diversity are available for a number of commercial proposals has been analysed in detail.[1]

Two basic implementations of satellite diversity are *switched diversity* and *combined diversity*.

Switched diversity simply means that the ground terminal has a choice of multiple visible satellites with which it can communicate, and that after selecting a satellite, the terminal establishes a single duplex radio link with that satellite.

Combined diversity is when the ground terminal communicates across multiple satellites simultaneously. This is also referred to as artificially-introduced multipath.[2] This is exploited in CDMA-based systems such as *Globalstar*, which recombine signals passed through more than one satellite at a shared gateway station as a way of combating shadowing and reducing errors. This combined diversity can also be exploited to ensure soft handovers.

Ground-to-space diversity, across the air interface, can be exploited at various layers of the network protocol stack. Physical diversity can be exploited in constellations without ISLs, e.g. *Globalstar*'s use of CDMA and power recombination of signals across multiple satellite transponders. It can be exploited at the data-link layer, via TDMA management as in *ICO*.[3]

A different form of ground-to-space diversity use is planned in *SkyBridge*, which uses dual switched diversity between neighbouring satellites in its two subconstellations to avoid sending transmissions from satellites seen as being in the part of the sky already inhabited by the geostationary arc. This is necessary because *SkyBridge* is based around reusing Ku-band frequencies already in use by geostationary satellites.[4] Since *SkyBridge* does not have ISLs, its use of diversity in this way means only a minor alteration in end-to-end path delay between ground terminal and terrestrial gateway, with no other visible effects from this diversity from a networking viewpoint. This dual diversity could be exploited even when satellites are in another part of the sky, providing a degree of redundancy in the communication infrastructure.

Beyond that, coding diversity and network-layer diversity between a terminal and multiple satellites are not currently proposed for exploitation in commercial constellations, although multihoming of ground terminals – using available satellite diversity to increase redundancy and thus fault tolerance – has been identified as desirable from a networking viewpoint.[5]

Network-layer path diversity for satellites communicating via ISLs is already present in the original 840-active-satellite design and the Boeing 288-active-satellite redesign of *Teledesic*, due to the large number of satellites and their close spacing, and the eight ISL terminals on each satellite.[6] Multipath routing in the ISL mesh can also broadly be considered a form of routing diversity.[7] Use of diversity is fundamental to traffic engineering. Although the *Teledesic* designs have considerable diversity in the space segment, their geometries dictate that they have no diversity or redundancy in the ground-space interface; a ground terminal is simply allocated to a single satellite. If that single satellite is shadowed or has failed, the ground terminal is unable to communicate with the satellite network. There is no use of diversity over ground-to-space links.

2.0 Managing diversity in rosettes with ISLs

Rosette constellations with ISLs offer the opportunity to implement diversity, not just across multiple visible satellites or between satellites within the ISL mesh, but across multiple ascending satellites (moving from south to north) and descending satellites (moving from north to south) that form part of local ascending and descending ISL mesh surfaces visible to ground terminals. These mesh surfaces, though overlapping and part of the same single constellation ISL mesh, are locally separate due to their high relative speeds; to reach the locally-visible ascending mesh from the locally-visible descending mesh, traffic must travel over the highest latitudes using intra-plane ISLs. This choice of surfaces is illustrated in Figure 1.

In the *ns* satellite extensions,[8] we introduced a simple handover procedure. When the current satellite drops below the minimum elevation angle deemed necessary for reliable communications by a ground terminal and handover is required, the terminal looks for the highest visible satellite and establishes a new link to that satellite. We further modified this procedure so that the terminal can find and select between the highest visible ascending and highest visible descending satellites. To investigate the effect of this choice of meshes on constellation traffic, we modified our satellite simulations so that ground terminals could control whether handover would take place to ascending or to descending satellites. A terminal will use one satellite for communication, but will always hand over to a visible ascending (or descending) satellite if it is physically possible. If multiple co-rotating satellites are visible on the selected mesh surface, the highest of those satellites is selected.

This careful use of available diversity offers the opportunity to control end-to-end path delay across the ISL mesh to a considerable extent, depending on whether the two terminals that are communicating have established links with satellites on the same, or different, ascending or descending surfaces. Packets can be injected into the ISL mesh consistently on one ‘mesh surface’ of the rosette chosen by the ground terminal, before normal shortest-path routing, based upon use of a link delay metric, routes them to their destinations across the ISL mesh.

Network traffic between two co-located terminals using neighbouring satellites on the same (ascending or descending) surface will have very short path delay times. Traffic between co-located terminals using counter-rotating satellites on different surfaces will need to travel over intra-plane links across highest latitudes, and will have longer path delay times, shown conceptually in Figure 2.

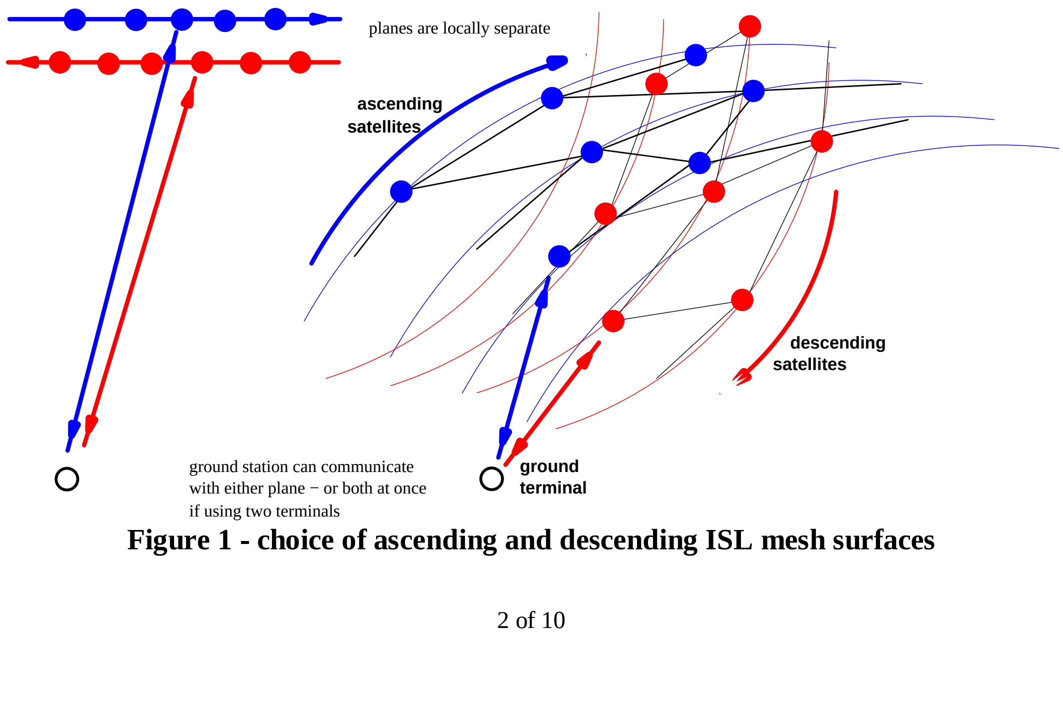


Figure 1 - choice of ascending and descending ISL mesh surfaces

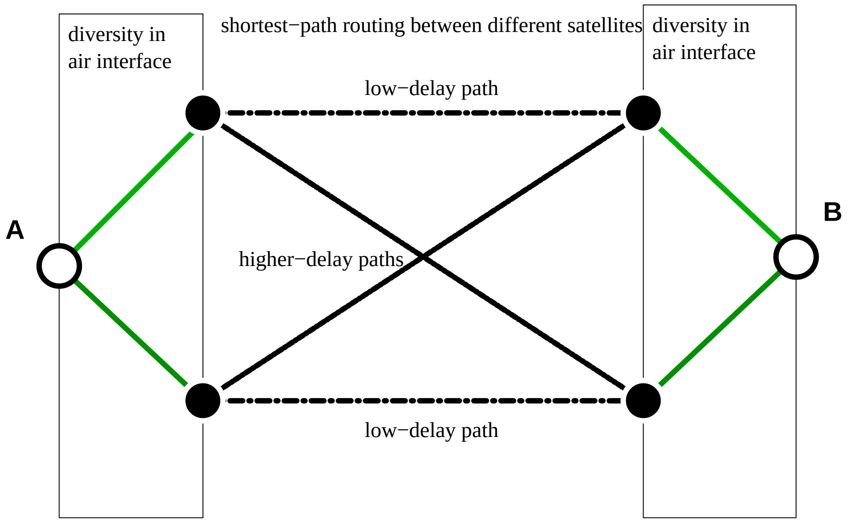


Figure 2 - diversity permits four shortest-path routes across the constellation

3.0 Providing double surface diversity

To examine the effects on the communications between ground terminals of selecting ascending or descending surfaces, it was necessary to use a constellation geometry providing not just a minimum of dual satellite coverage, but a minimum of double *surface* coverage, throughout.

Having more than one satellite visible at all times from areas with coverage is not sufficient; a minimum of one satellite from each of the ascending and descending surfaces must be visible to provide double surface diversity.

Commercial proposals that combine rosette constellation geometries with ISLs are few, but do include Motorola's LEO *M-Star* proposal[9] that predated their *Celestri* proposal, their *Celestri* proposal,[10] and the Hughes *Spaceway NGSO* proposal.[11] However, the geometries of these systems do not provide double surface coverage, and therefore do not allow a terminal to select between ascending and descending satellites at every point in time. We examined the LEO *Celestri* and MEO *Spaceway NGSO* constellations in detail by simulation with *SaVi* and *ns*.

3.1 **Using *Celestri***

The *Celestri* geometry is described as giving single coverage of the Earth at elevation angles above 16° between 60° of latitude.[10] Our simulation based on that description, shown in figure 4, indicates coverage from a single satellite in places, as well as the occasional (but small) gap in coverage; handover decisions between surfaces can be dictated by this low varying coverage.

We took the proposed *Celestri* design as our base, and simply lowered the minimum mask elevation angle of *Celestri* ground terminals from 16 to 10 degrees, as shown in Figure 5. (This presumes that link budgets could be recalculated and equipment dimensioned to meet this new requirement.) We achieved double surface coverage, with the separate layers of ascending and descending satellites covering the Earth entirely between 60° of latitude. We were then able to compare handover strategies: choosing the highest visible satellite versus a controlled choice of highest visible ascending or descending surface, as illustrated for the terminals shown with the *Celestri* satellite network topology in Figure 3. Resulting path delay curves for traffic between the terminals are illustrated.

Figure 6a shows use of uncontrolled seek-highest-satellite handover, which does not discriminate between ascending and descending surfaces, and therefore wanders between them. This shows considerable variation in the path delay between terminals over the course of a day as the Earth rotates beneath the constellation. When the terminals happen to both be using the same ascending or descending surface, where satellites are nearby in the ISL mesh topology, the delay over the path is low (<40ms). However, most of the time the terminals are using satellites on different surfaces, where the satellites are distant from each other in the ISL mesh, and path delay is higher (54-78ms). We then examined control of handover to a chosen surface. The four different path delay curves resulting from use of controlled handover between the same terminals, to ensure that terminals remain on the same ascending or descending surface of satellites, are shown in Figure 6b.

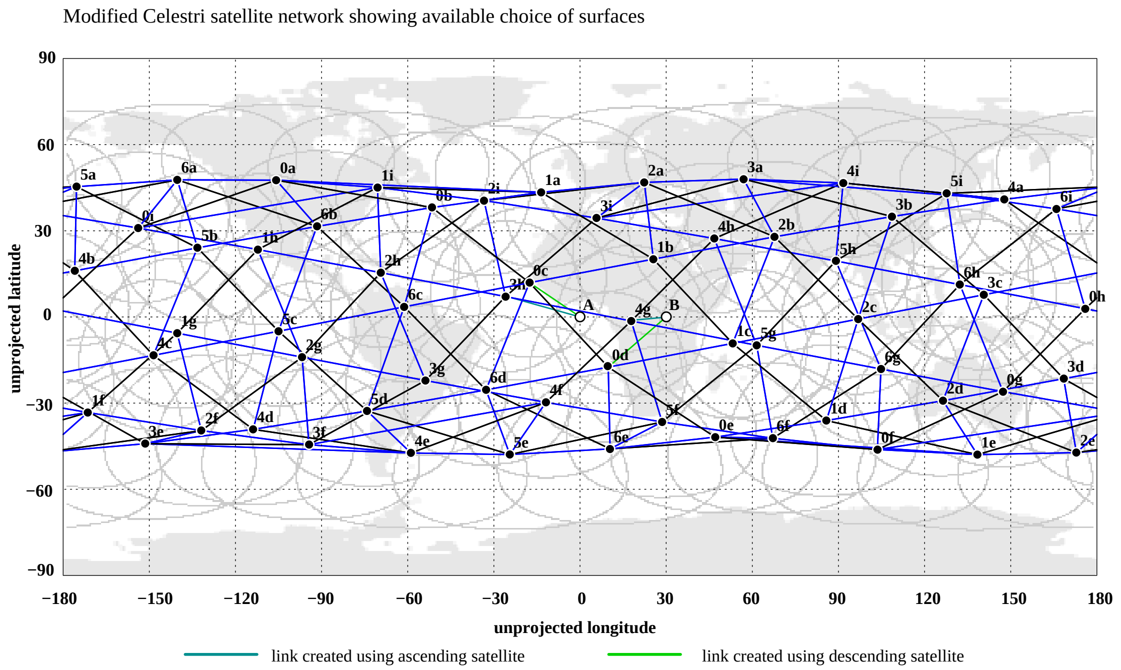


Figure 3 - *Celestri* network (lowered elevation angle) showing handover choices

This controlled handover exhibits lower delays for the shorter paths between ground terminals, where the terminals are sending traffic to satellites near each other in the ISL mesh. Controlled handover also provides a second set of larger delay curves, due to longer paths over highest latitudes in the ISL mesh, where the terminals are sending traffic to satellites that are further apart in the mesh.

The ISL mesh is still using shortest-path routing based on a delay metric; by controlling where the source and destination of traffic connect to the ISL mesh – a form of ingress control – we choose different paths across the ISL mesh for traffic.

By having ground terminals use the same or different surfaces, we get two clear classes of delay between the terminals. By having multihomed gateway stations communicate with both ascending and descending surfaces in the satellite network, these delay classes become accessible to single-link ground terminals that do not themselves exploit diversity simultaneously. Although the single link terminal is only communicating with one surface, it can send traffic to the multihomed gateway using two different routes to two separate downlinks on two separate satellites.

To use both classes of service simultaneously when communicating with other ground stations that do not exploit diversity, the ground station must itself exploit diversity and communicate with both surfaces; this combined switched network-layer diversity requires multihoming by maintaining two communication links using two sets of terminal equipment.

Traffic between ground stations not exploiting diversity will be effectively uncontrolled, arguably itself constituting a third class of service. Terminals exploiting diversity could allocate lowest-priority traffic to either surface to use up available air-interface spotbeam capacity after higher-priority traffic has been allocated; this low-priority traffic will experience uncontrolled, undiscriminating, handover.

We examined the delay properties of this approach in detail. As the *Celestri* geometry provides coverage between 60° of latitude, we examined this range of latitudes at intervals, taking advantage of the symmetry of the constellation around the Equator, as shown in figure 7. We moved one terminal across the whole 360° circle of longitude, and repeated each simulation run for these handover choices:

- terminals A and B both have uncontrolled (seeking-highest-satellite) handover mechanisms.
- terminals A and B prefer ascending satellites.
- terminals A and B prefer descending satellites.
- terminal A prefers ascending satellites, while B prefers descending.
- terminal A prefers descending satellites, while B prefers ascending.

Summary results are presented in figures 8 to 10, in the form of comparisons of average path delays experienced across different latitudes for the various handover scenarios.

The differences in path delay between the first class of service and the second, longer delay class are clearly visible for terminals at the Equator in figure 8. The differences between the classes lessen as the spacing between the terminals increases, until there is no appreciable difference at 180° separation.

The average path delays experienced with the uncontrolled service vary between the sets of delays for the two classes, and are extremely sensitive to minor differences in longitude and to initial handover choices at the start of each simulation.

As the latitude of the communicating terminals increases towards the limits of coverage of the constellation (figures 9-10), the differences in path delays for the service classes drops to nothing. This is because handover is increasingly dictated by available coverage, and the different paths begin to overlap and share common links. Uncontrolled handover continues to oscillate between the two sets of classes.

We did not examine having one terminal control same-surface handover, while handover at the terminal it is communicating with remains uncontrolled and is based around handover to the highest visible satellite.

3.2 Using *Spaceway NGSO*

We also attempted to achieve full double surface coverage for a MEO constellation by modifying the proposed Hughes *Spaceway NGSO* constellation.[11]

By lowering the minimum elevation angle of the constellation's ground terminals from the stated value of 30° (Figure 11) to 25° (Figure 12), we achieved a minimum of dual satellite coverage across the Earth's surface. This, however, turns out not to be full double surface coverage throughout the course of a day; there are gaps in a surface's coverage between individual planes that are visible in path delays when terminals pass from one plane of satellites to another.

This gap in double surface coverage between planes causes terminals to initiate handover to satellites in the counter-rotating plane, from ascending to descending satellites or vice versa, even when handover is controlled for choice of surface. This alters delay significantly, and results in interruptions to the otherwise clearly-distinct delays resulting from different path lengths across the ISL mesh.

The effects of these gaps in full double surface coverage are shown in the left-hand graph in Figure 13a. As the Earth rotates under the orbital planes, one terminal reaches the edge of a plane's coverage and is forced to hand over to the counter-rotating mesh surface, as the current surface that its plane is on can no longer provide coverage. Slightly later, the other terminal is forced to make a similar handover, so that the two terminals again share a surface. This dramatically changes the route and total path delay between terminals, even though controlled handover is being used to ensure handover to a corotating plane whenever possible. Without available diversity, handover is dictated by available single coverage, and resulting routes and path delays are dictated by handover.

The four gaps in double coverage between the four planes of *Spaceway NGSO* are clearly visible in the resulting delay traces shown in Figure 13a.

To compensate for this by making minimal changes to the *Spaceway NGSO* geometry, either:

- the altitude of the satellites can be increased to increase coverage, which is likely to be undesirable for MEO due to the proximity of the Outer Van Allen belt, or
- another orbital plane can be introduced so that the existing planes can be brought closer together and the gaps in double coverage between planes can be closed.

We required closer planes, and the second option, which directly addressed this, was successful. Adding an extra plane of five satellites to the existing *Spaceway NGSO* proposal's geometry of four planes increases its number of active satellites from 20 to 25, but also ensures that handover always takes place between co-rotating planes as the closer planes now overlap sufficiently.

The effect of this change on total path propagation delay between terminals is shown in the right-hand diagram in Figure 13b. Five slight patterns of disruption are now visible in each delay trace at the times when handover occurs between the five sets of neighbouring planes as the distance in-between terminals changes between their handovers to the succeeding planes. The two sets of delay traces are clearly separated, and we have achieved distinct classes of service.

Any such alteration of an existing constellation design will require a reworking of link budgets, equipment requirements, frequency mappings and allocations for overlapping spotbeams and so on.

From a networking and services viewpoint, it would seem sensible to begin with acceptable application latencies and traffic delay bounds as design constraints, and then to work from there to classes of service for traffic, path delays and then to an acceptable constellation geometry and routing approach that meets the delay requirements. Equipment able to meet the needs of this top-down design can then be dimensioned once link budgets are determined for the ground-space interface of the constellation.

However, the commercial proposals that we have examined are described in applications to the US Federal Communications Commission for allocation of radio frequencies. The FCC application is often the first step towards obtaining agreed worldwide allocation of frequencies at a World Radio Congress. (FCC applications are also primarily legal, rather than technical, documents, and this is reflected to some extent in their content.)

Given this, commercial proposals are driven by the need to meet power and interference requirements while working at specific available and allocated frequencies in the ground-space interface. This results in a link-budget-driven bottom-up approach to constellation geometry, the protocol stack, and the resulting constellation network design.

Our observation that these commercial proposals are not optimised for the network traffic they are intended to carry, and that they offer widely-varying delays, is unsurprising; the frequency allocation process simply does not consider these needs as goals.

4.0 Classes of service and reliability

The constellation network is homogenous in terms of ISL design and capacity. We know that the total path delay is roughly proportional to the number of hops (n) that the path traverses. As there is always an uplink and downlink, n is 2 or greater. As the reliability of transmission across each hop is imperfect, we can express the probability of a packet successfully traversing the path as:

$$P_{\text{success}} = (1-d_{\text{uplink}})(1-d_{\text{isl}})^{n-2}(1-d_{\text{downlink}})$$

where d_{link} is an estimation of the probability of the packet being discarded due to errors in transmission along the link or to any congestion at the queuing head of the link.

This introduces a reliability dimension to the use of two sets of paths, allowing different levels of guarantee of delivery. The longer set of paths constituting the second class of service can be expected to be generally slightly less reliable than the shorter, first, class of service; this must be considered in service specifications.

Long-delay and short-delay traffic from ground terminals in different locations will share individual intersatellite and air-interface links. We envisage all traffic being marked for class of service at ingress on entering the network, and being treated differently according to marked class at each node. The path delay and class separation that geometry and handover have provided by design is maintained, and even enhanced, by this treatment even though traffic is mixed. This requires a differentiated-services approach to the QoS model within the constellation.

5.0 Handover and network state

The handover mechanism is also worth examining. Consider a single-homed ground terminal that does not exploit ground-space diversity, and which is undertaking handover to a co-rotating satellite, perhaps as part of a street of coverage so that the satellites are adjacent neighbours in the network mesh. Such a handover will involve moving soft or hard network state from one satellite to another. This movement of state is clearly more straightforward for adjacent satellites across a single link than when carrying out handover to a distant counter-rotating satellite elsewhere in the rosette mesh, particularly when e.g. assignment of new spotbeam capacity and coordinating timings for soft handover are considered. This must be considered for the hard state associated with virtual circuits, or for the soft state associated with routing and addressing.

Full double coverage, where handover never takes place to a counter-rotating satellite, minimises the difficulties and disruptions caused by handover in a rosette satellite constellation network with ISLs. Use of ISLs in the rosette encourages a streets-of-coverage handover approach to constellation geometry.

Conclusions

Use of diversity and control of handover in rosette constellations can create sets of multiple separate and redundant paths of different delays between ground terminals. Control of handover can go a very large way towards control of traffic in the constellation; handover control can provide ingress control for the rosette constellation network.

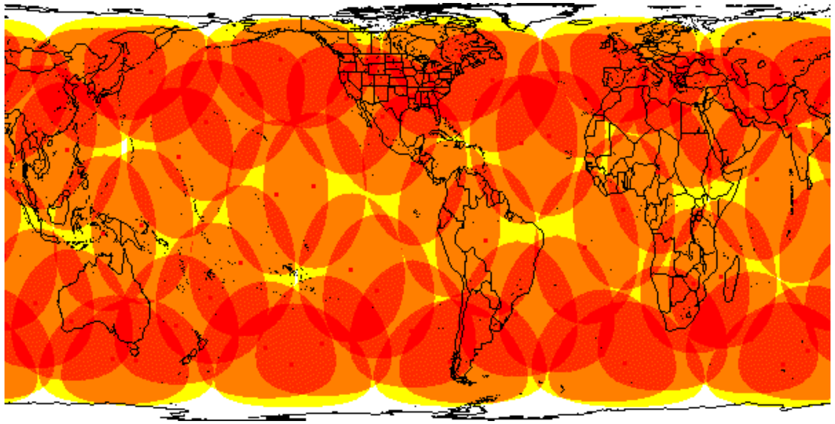
By detailed simulation of rosette constellation networks, we have shown that using controlled handover to manage surface diversity is particularly useful for nearby terminals, with a considerable difference in delays between the choices of available path. This difference in delays decreases as the terminal locations approach the limits of coverage for the constellation, and as terminal separation approaches its maximum of 180° of longitude.

The commercial rosette constellations that we examined were not dimensioned to be able to offer this flexibility of service to traffic. We showed how this flexibility of classes of delay for network traffic could be easily introduced once simple modifications to the geometries of the commercial *Celestri* and *Spaceway NGSO* proposals are carried out.

This novel approach relating handover, diversity, and delay can significantly reduce path delays for traffic across the rosette constellation, as well as allowing sets of delays that permit clear classes of delay and service that support QoS. This approach also minimises the disruption to network traffic and to the movement of network state between satellites that is introduced by handover.

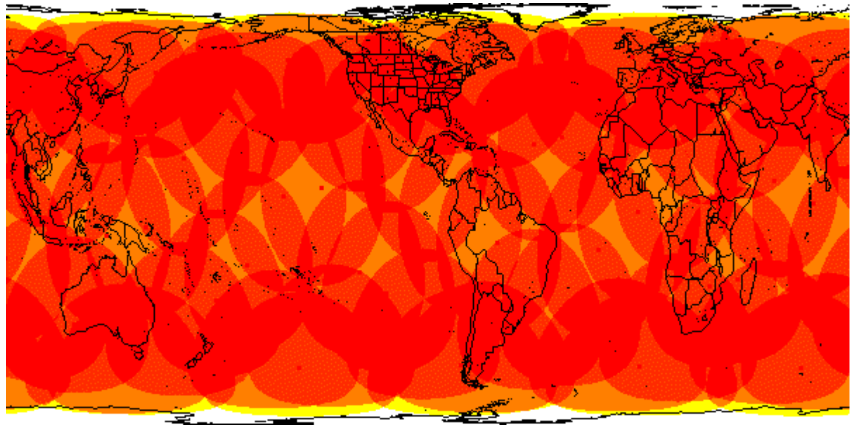
References

1. M. Gkizeli, Payam Taaghol and B. G. Evans, 'Service availability of LEO/MEO satellite systems', *International Journal of Satellite Communications*, Special Issue: LEOs – Little and Big, vol. 17. no. 4, pp. 291-302, July/August 1999.
2. M. Werner, H. Bischl and E. Lutz, 'Mobile User Environment and Satellite Diversity for NGSO S-PCNs', Proceedings of the International Mobile Satellite Conference (IMSC'95), pp. 476-481, Ottawa, Canada, 6-8 June 1995.
3. T. E. Wisløff, 'Dual satellite path diversity in non-geostationary satellite systems', Dr. Ing. Thesis 1996:84, Department of Telematics, Norwegian University of Science and Technology.
4. P. L. Spector, J. H. Olson et al., 'Application of SkyBridge LLC for authority to launch and operate the SkyBridge system', filing with the US Federal Communications Commission, SkyBridge LLC, 28 February 1997.
5. N. Schacham, 'Protocols for multi-satellite networks', Proceedings of MILCOM '88, pp. 501-505, 1988.
6. M. A. Sturza, 'Architecture of the Teledesic satellite system', Proceedings of the International Mobile Satellite Conference, pp. 212-218, 1995.
7. L. Wood, G. Pavlou and B. G. Evans, 'Effects on TCP of routing strategies in satellite constellations', to appear in *IEEE Communications Magazine*, special issue on Satellite-Based Internet Technology and Services, vol. 39 no. 3, March 2001.
8. T. R. Henderson and R. H. Katz, 'Network Simulation for LEO Satellite Networks', Proceedings of the 18th AIAA International Communications Satellite Systems Conference (ICSSC), Oakland, California, 10-14 April, 2000.
9. M. Werner, C. Delucchi and K. Burchard, 'ATM Networking for Future ISL-based LEO satellite constellations', Proceedings of the International Mobile Satellite Conference 1997 (IMSC '97), pp. 295-300, Pasadena, California, 16-18 June 1997.
10. M. D. Kennedy, B. Lambergman et al., 'Application to construct, launch and operate the Celestri multimedia LEO system', filing with the US Federal Communications Commission, Motorola Global Communications, Inc., June 1997.
11. F. A. Taormina et al., 'Application of Hughes Communications, Inc. for authority to launch and operate Spaceway NGSO, an NGSO expansion to the Spaceway global broadband satellite system', filing with the US Federal Communications Commission, Hughes Communications, Inc., 22 December 1997.
12. P. A. Worfolk and R. E. Thurman, '*SaVi* - software for the visualization and analysis of satellite constellations', developed at The Geometry Center, University of Minnesota.



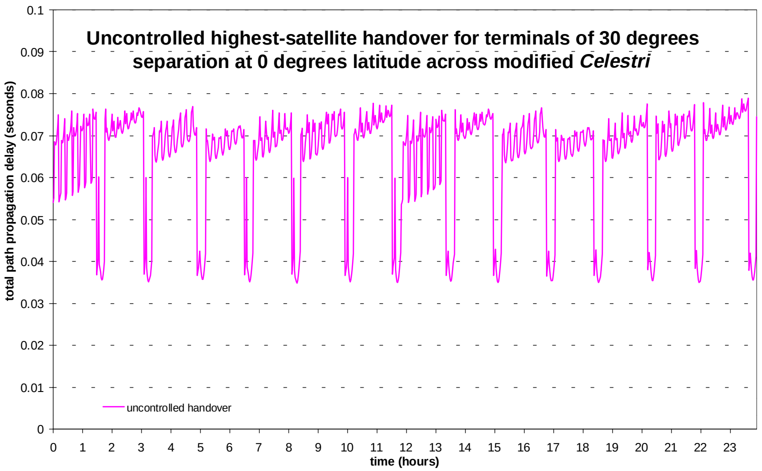
cylindrical projection, using minimum elevation angle of 16°

Figure 4 - *Celestri* LEO constellation, shown in *SaVi*[12]

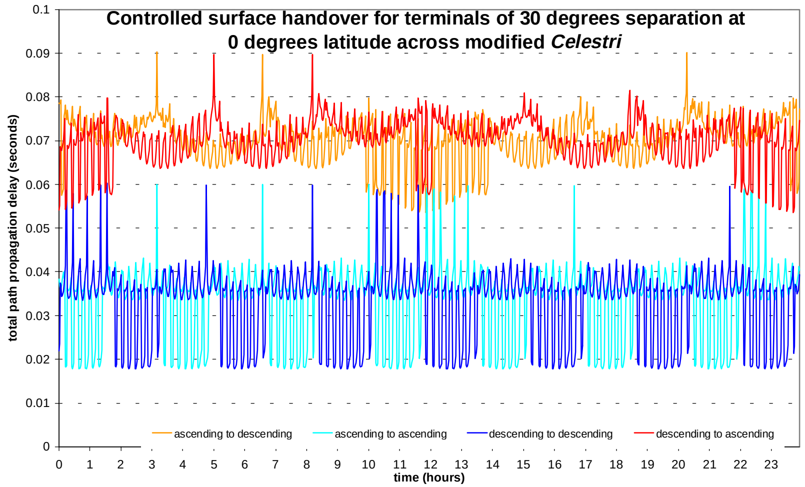


cylindrical projection, using minimum elevation angle of 10°

Figure 5 - modified *Celestri* LEO constellation

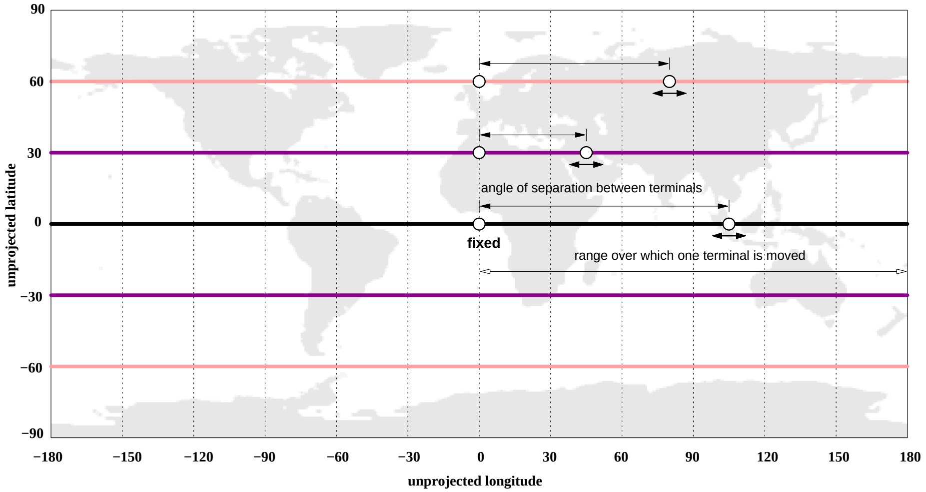


a. Uncontrolled handover choosing the highest visible satellite regardless of surface.



b. Controlled handover. Different surfaces: long delays. Same surface: short delays

Figure 6 - comparison of handover strategies across modified *Celestri* LEO constellation for multihomed terminals A and B in Figure 2



terminal A is fixed. B is moved over from 0° to 180°, and from –180° to 0°

Figure 7 - measurement of path delays at different latitudes between multihomed terminals

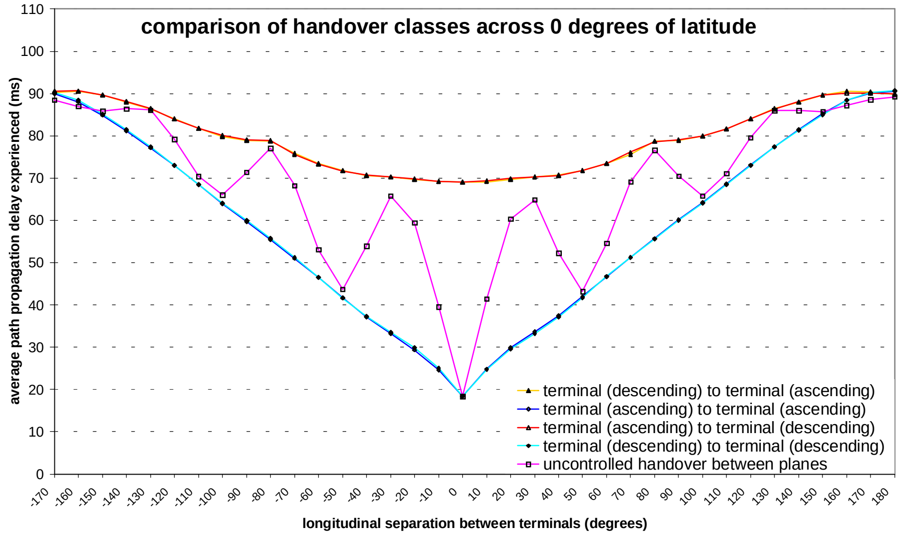


Figure 8 - different mean path delays between terminals on Equator

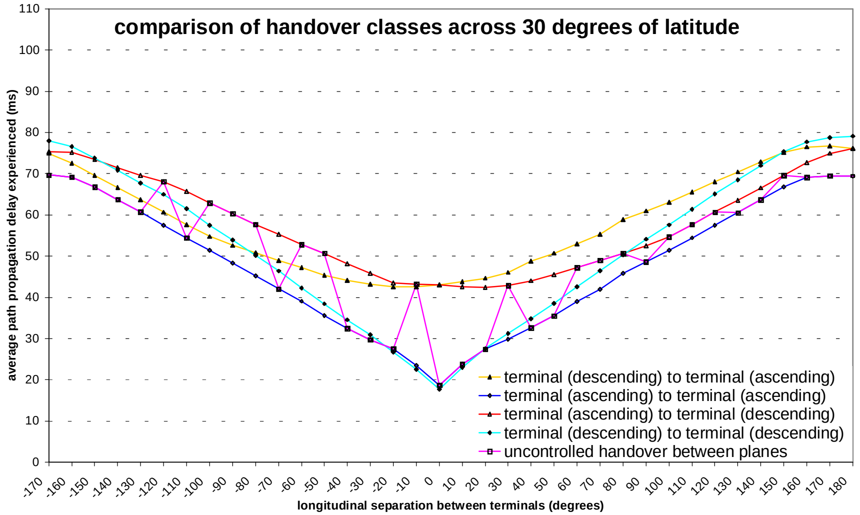


Figure 9 - different mean path delays between terminals at 30° latitude

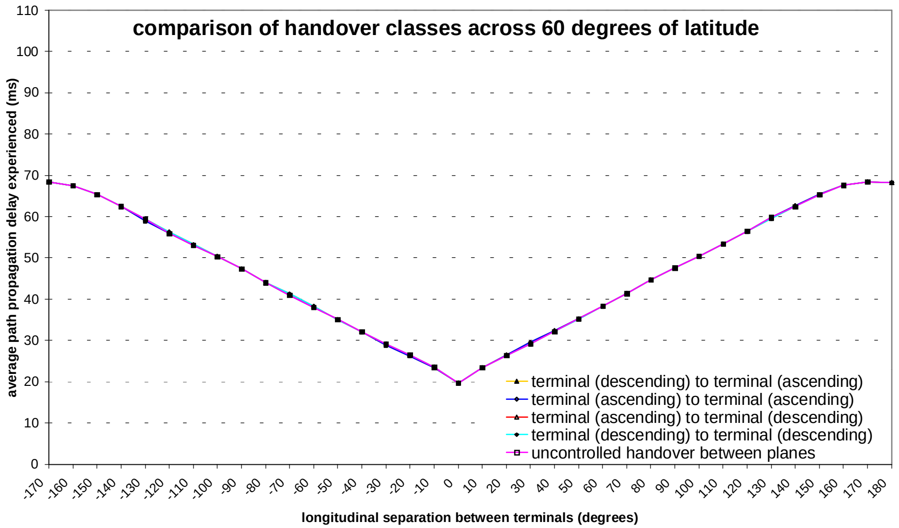
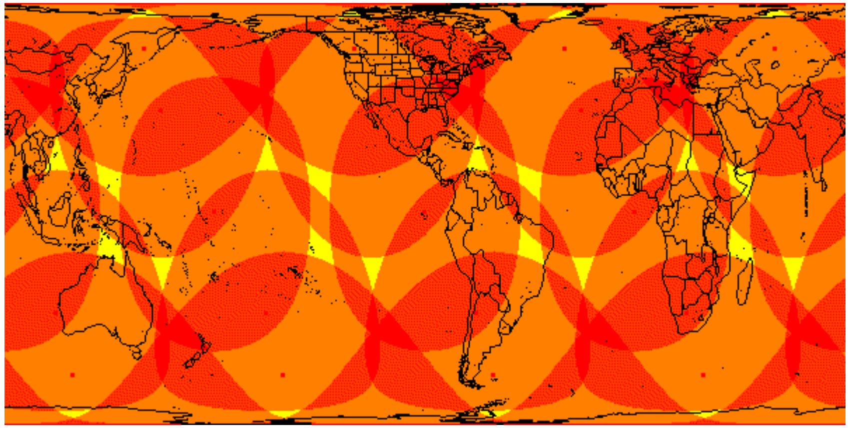
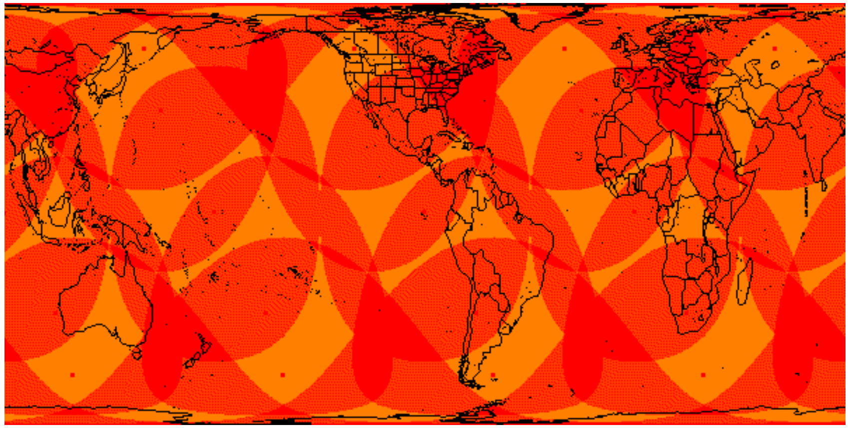


Figure 10 - different mean path delays between terminals at 60° latitude



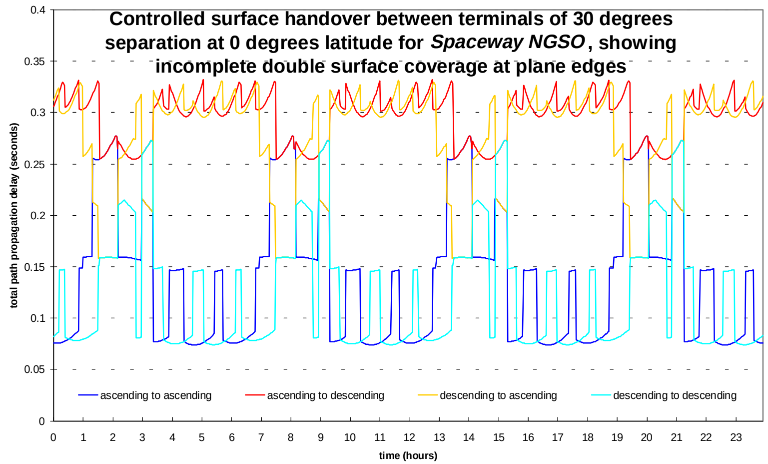
cylindrical projection, using minimum elevation angle of 30°

Figure 11 - *Spaceway NGSO* MEO constellation, shown in *SaVi*[12]

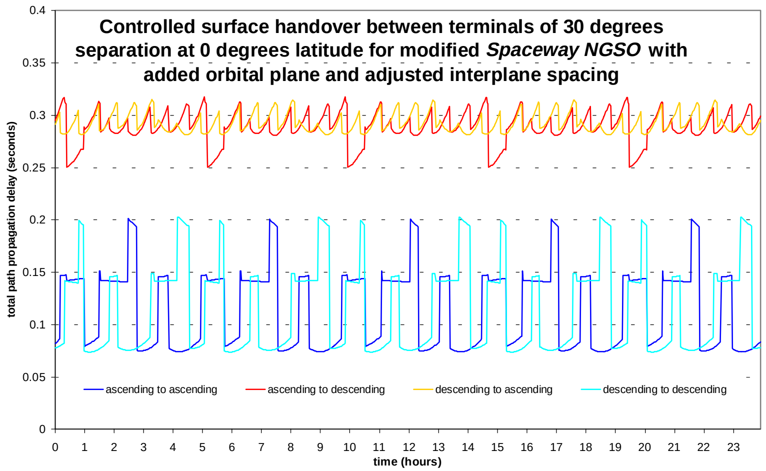


cylindrical projection, using minimum elevation angle of 25°

Figure 12 - adjusted *Spaceway NGSO* with minimum dual coverage



a. Minimum dual coverage, but not double surface coverage. Resulting gaps between planes.



b. Added plane. Different surfaces: long path delays. Same surface: short path delays

Figure 13 - comparison of dual-coverage *Spaceway NGSO* constellation against constellation with added plane showing clear classes of service. Controlled surface handover used between multihomed terminals A and B in both cases.