1

The Foundations of Satellite Networks

1.1. Introduction

Satellite networks are principally characterized by the use of a satellite. They were designed with the aim of achieving a global coverage of the earth and to provide different services, such as voice transmission, short messaging services, global positioning system (GPS) localization, etc. These networks also enable two distant points located anywhere on the earth to communicate. However, building a network always requires that a number of precautions be taken and, more specifically, that a precise set of specifications relating to performance be presented. Theoretically, these specifications are determined from market research that aims to forecast the different uses of the system. It would be very naïve to think, however, that this research could forecast the future needs of users with any degree of certainty.

The market can change very quickly and a property that was overlooked when the network was being designed can become the determining factor in getting one step ahead of the competition. In this event, it is imperative that the network follows the change.

For example, let us consider land mobile networks, which have undergone countless changes in infrastructure. At first operators were attached to the idea of covering the largest possible area in a populated zone. Then, however, the market grew exponentially and operators refocused their efforts on regions that already had a wide coverage for capacity reasons: in certain situations, a station was no longer able to meet the demand from the zone it was covering, a zone that is called a *cell*. An effort was then made to divide the overloaded cells into smaller cells that were associated with less powerful (and therefore less *noisy*) stations, thereby enabling radio channels to be used once again.

The idea of using more, smaller and less powerful cells can also be applied to satellite networks. To increase the capacity of a system, satellites can be *brought closer* to the earth, thus improving acuity for differentiating between ground terminals, as illustrated in Figure 1.1: to cover an equivalent air surface, a satellite that is far from earth must use a much more precise spotbeam than a satellite that is closer to earth.

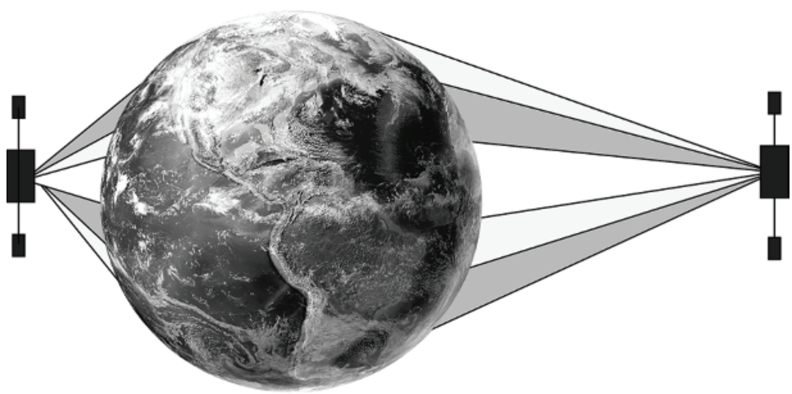


Figure 1.1. *The reuse of frequencies by satellites with varying altitudes*

This chapter will first present the places where a telecommunications satellite can be positioned in the vicinity of the earth. We will then see how several satellites can work together to guarantee a global coverage. We will also tackle the question of the time it takes to pass through a network according to where it is located and, in concrete terms, we will discuss the resulting broad categories of satellite networks. The characteristics of cellular satellite networks will then be presented and the problem of handover in LEO satellite networks will be introduced.

1.2. Satellite orbits

An orbit is a path the satellite follows when there are no perturbations. There are many different types of satellite orbits. Contrary to common opinion, a device's path without propulsion has nothing at all to do with its weight, though it is conditioned by precise rules (described next). Before this, it is useful to go over some basic properties of the ellipse.

1.2.1. *Characteristics of the ellipse*

Let us consider a plane with an orthonormal coordinate system $(O, \vec{i}, \vec{j})$. Given the two real positive constants a and b where $a > b$, the ellipse ξ centered on O of a semi major axis a and a semi minor axis b is the set of points $P(x, y)$ that verify the following equation:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad [1.1]$$

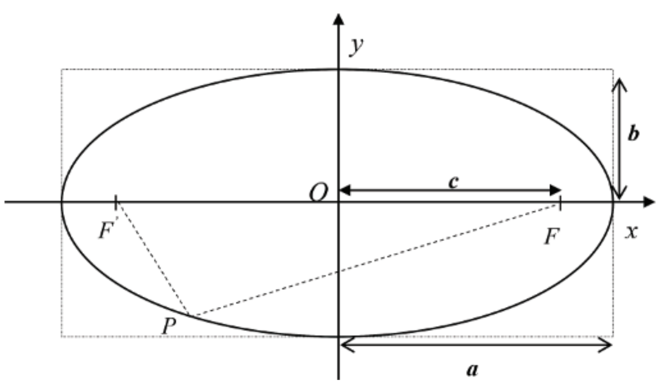


Figure 1.2. *A few properties of the ellipse*

A general parametric form of the ellipse exists. It can be formulated as follows:

$$\varepsilon = \left\{ P(x,y) \quad with \left| \begin{matrix} x = a\cos(k) \\ y = b\sin(k) \end{matrix} \right. \right\} \qquad \text{with} \quad k \in \mathrm{IR} \qquad [1.2]$$

Figure 1.2 shows different parameters. The two foci of the ellipse appear at points *F(c, 0)* and *F′(-c, 0)* with $c = \sqrt{a^2 - b^2}$. The *eccentricity* of the ellipse is thus defined by $e = c / a$. For every point *P* of ξ, the following relation is verified:

$$PF + PF' = 2a$$

1.2.2. *Kepler's laws*

The parameters used today to describe satellite orbits are inspired by the work of Kepler (1571–1630) [SUN 05, MON 05].

These laws describe the way in which planets move around the sun and are typical of the following parameters:

– the *orbit of planets*, that is the path they follow over time;

– the *instantaneous speed* of the path of each orbit by the associated planet;

– the *orbital period* of a planet, which is the total time it takes to complete its orbit;

Kepler's laws can be summarized as follows:

– the orbit of each planet is an ellipse where the sun is one of the foci;

– the area swept out by a line between the sun and the planet is constant over a unit of time;

– the square of an orbital period of a planet is proportional to the cube of the major axis of its orbit.

These laws can, in fact, be applied to any system where a celestial body with a large mass (the sun), which is known as the *primary element*, determines the movement of bodies with small masses in comparison to the former (the planets), which is known as the *secondary elements*. More specifically, they describe the movement of satellites (*secondary elements*) around the earth (the *primary element* in this case).

It is useful to locate these orbits in relation to the movement of the earth around the sun. Seven parameters, often known as *Keplerian parameters*, can be used to characterize the movement of a satellite.

1.2.3. *Orbital parameters for earth satellites*

Let us take *(Oxyz)* as an orthonormal coordinate system, where *O* is the center of the earth, *Oz* its rotation axis and *(Oxy)* its equatorial plane, taking the direction of *(Ox)* as the intersection of the equatorial plane with the ecliptic plane (plane of the rotation of the earth around the sun).

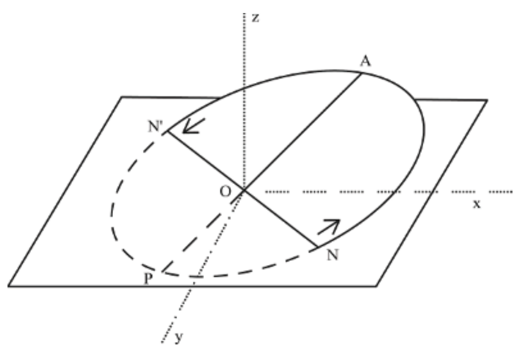


Figure 1.3. *The remarkable points of a satellite in orbit around the earth*

There are several remarkable points for a satellite in orbit (Figure 1.3):

– Apogee *A*, a point where the satellite is furthest from the earth;

– Perigee *P*, a point where the satellite is closest to the earth;

– the nodes *N* and *N`*, the points where the satellite passes through the equatorial plane. In *N*, the satellite passes above the equatorial plane and in *N`*, it passes below.

The movement of the satellite can be described with seven parameters.

Description of the orbit

The orbit is an ellipse given by:

– the semi major axis *a*, half the length of the major axis;

– eccentricity *e*: when $e = 0$, the orbit is described as circular and in the opposite case, it is described as elliptical.

Position of the orbit in relation to the earth

The parameters characterizing the position of the ellipse are:

– *inclination i* defined as the angle between the orbit plane and the equatorial plane, which is also the angle between the normal orbit and the rotation of the earth. By convention, inclination is between 0° and 90° when the satellite is turning in the same direction as the earth and between 90° and 180° when it is turning in the opposite direction;

– the perigee, which is the angle (ON, OP);

– the right ascension of the ascending node (Ω), which is simply the angle (Ox, ON).

Position of the satellite in the orbit

Now, when the orbit is defined, it is useful to specify where the satellite is located on it. To do so, it is necessary to specify an observation date and a place on the orbit:

– *the observation date t* is the moment the satellite is observed;

– *the average anomaly M* is the angle the satellite makes with the perigee. The angle is positioned according to the direction of the orbit (hence the subtle difference between *N* and *N`*).

1.2.4. *Orbital perturbations*

The orbits described previously are perfect. In reality, two types of perturbations should be taken into consideration:

– *gravitational perturbations* only involve gravitational forces, but are caused by elements that are not taken into account by the previous model. Kepler's laws, in fact, consider that each planet has a point of volume equal to nil. In reality, the earth has a volume that is not insignificant, especially when low orbit satellites pass close by. The tide, the moon and the sun are also factors. Given that these forces derive from a potential, they do not reflect a satellite's loss or gain of energy. Nevertheless, the resulting deviation of the Keplerian satellite path is not always desirable;

– *non-gravitational perturbations* notably involve atmospheric drag and the pressure of (direct or indirect) solar radiation or radiation from another source (particularly in infrared).

1.2.5. *Maintaining and surviving an orbit*

The earth has large magnetic fields, which have an impact on its environment. Indeed, certain zones beyond the earth's surface emit radiation that is powerful enough to endanger the electronic components that pass through it.

Two energetic tori called the *Van Allen belts* circle the earth around the equator. These belts emit intense radiation that is very dangerous for the electronic equipment on board a satellite. The inner belt particularly affects altitudes between 1,500 and 5,000 km, while the outer belt affects those between 13,000 and 20,000 km.

Satellites orbit in four main regions:

– the LEO zone between the outer edge of the atmosphere and the inner Van Allen belt at an altitude of between 400 and 1,500 km;

– the *medium earth orbit* (MEO) zone between the two Van Allen belts at an altitude of between 5,000 and 13,000 km;

– the *high earth orbit* (HEO) zone whose apogee is beyond the Van Allen belts but whose elliptical orbit spans one or two of the previous zones;

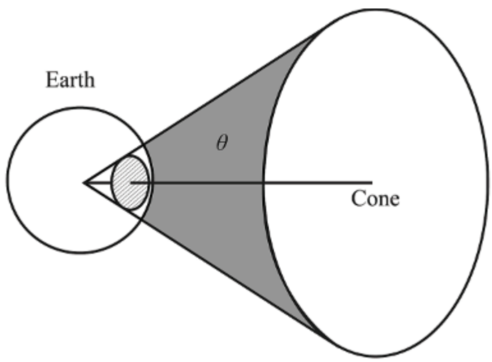
– the *geostationary earth orbit* (GEO) zone, which can be considered as a specific case of HEO for geostationary satellites at an altitude of 35,786 km.

1.3. Time, time variation and coverage

Geostationary satellites are regularly criticized for their transmission time. Transmission time corresponds to the time it takes for electromagnetic waves to be propagated into space with a speed that is well estimated to be equal to that of the speed of light in space (300,000 km/s). It therefore takes 240 ms to go to a geostationary satellite and back. This explains why one of the main aims of low orbit satellites is to reduce the communication times of old systems. However, while geostationary satellite times correspond in an almost uniform manner to the some 35,800 km route separating them from earth, the picture become less uniform when the satellite views earth at a large angle, on the one hand, and when its movement is no longer synchronized with the rotation of the earth, on the other hand.

1.3.1. *Geometric data*

The closest point on the surface of the earth for radio communication with a satellite is located precisely at its vertical. Starting at this point on the surface of the earth and expanding out in concentric circles, reception gradually declines the greater distance from the satellite. In fact, given the lowest altitude at which satellites orbit (i.e. 200 km), the relief of our planet can be considered non-existent. The final coverage zone is therefore described as a disc on the surface of the earth; this disc is known as the *footprint* of the satellite [RES 95]. It is possible to give it a radius; however, a more important parameter is the radius of its solid angle of coverage, as illustrated in Figure 1.4. The solid angle is therefore the cone centered on the center of earth, from the axis passing through the satellite, whose angle of opening is θ.



Dashed region: Intersection of the surface of the Earth and the cone

Figure 1.4. *Solid angle of coverage*

In fact, the coverage zone can be seen as the intersection of a circular cone starting from the center of the earth and from its surface. The knowledge of the angle θ is sufficient to be able to characterize it, if the vertical of the satellite, and thus the axis of the cone, is known. The advantages of this are twofold:

1) All users covered by a satellite with a coverage θ are at the intersection of the surface of the earth and the cone centered on earth with the angle θ whose axis passes through the satellite.

2) All satellites with a coverage θ covering a user are contained in the cone centered on the earth with angle θ whose axis passes through the user.

An essential element used to characterize coverage is the angle of elevation ω, which the user uses to view the satellite. In the event that the receiver is in the middle of an isolated area, and if the landscape is flat, they can pick up signals from the semi-sphere above their head. If plane P is tangential to the surface of the earth at the level of the receiver, the latter receives signals from the semi-space of the border P, which does not include the earth. However, the presence of the relief masks, in the majority of cases, at least one part of this space. In fact, a satellite situated too close to the horizon is changed by any obstacle that could get in its way: mountains, hills, buildings, trees, etc. To deal with this, the horizon line must be *raised* in some way so that the satellites are more *vertical*.

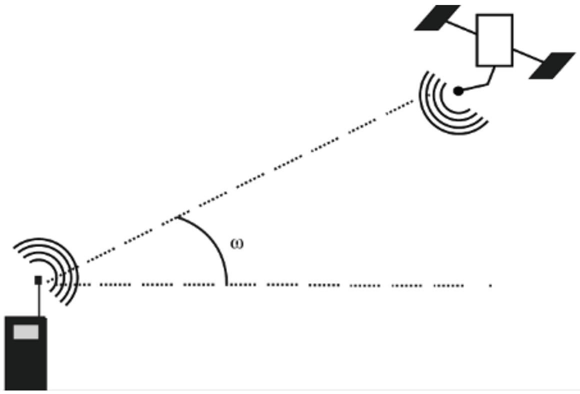


Figure 1.5. *Angle of elevation*

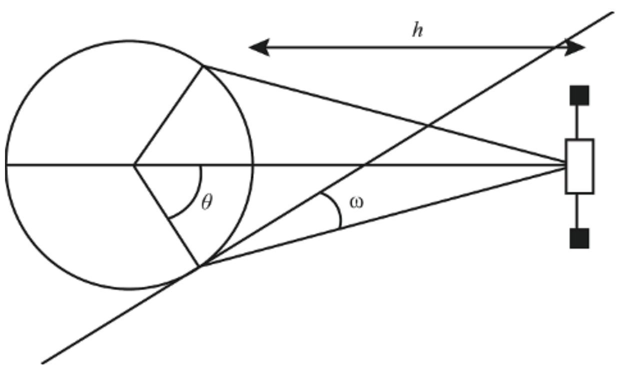


Figure 1.6. *The minimal angle ω*

Figure 1.6 shows how the minimal angle ω is measured. This angle is used to view a satellite.

If the altitude h of a satellite and the angle of elevation ω, which it is hoped is visible, is known, it is easy to determine the angle of coverage θ.

The various parameters are linked by the equation [ALT 99]:

$$h = R_T\left(\cos\theta\left(1+\tan\theta\tan(\theta+\omega)\right)-1\right) \qquad [1.3]$$

where R_T represents the radius of the earth. This equation is also written as:

$$\frac{h+R_T}{R_T} = \frac{\cos\omega}{\cos(\theta+\omega)} \qquad [1.4]$$

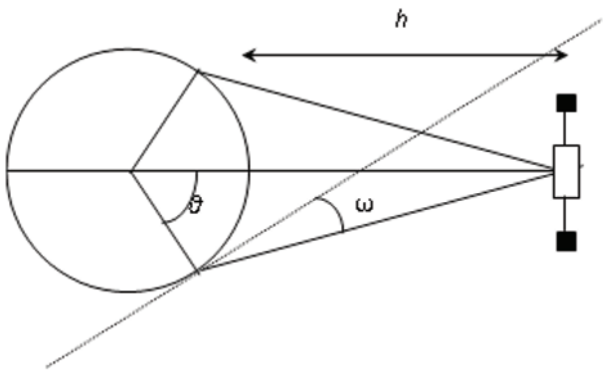


Figure 1.7. *Relation between θ, ω and h*

1.3.2. *Approximation of coverage*

The area covered by a satellite with a coverage θ has a surface area of $2\pi\left(1-\cos\theta\right)R_T^{\ 2}$ [ALT 99, CAP 03, SOR 99]. As the total surface area of the earth is $4\pi R_T^{\ 2}$, at least 2/(1-cosθ) satellites are necessary to completely cover the earth.

It is, however, difficult to reach this level as each satellite covers a disc and the discs must overlap. Even if all satellites were immobile, finding the minimum number of discs required to cover a sphere is a mathematical problem that remains, to this day, unresolved [OCT 88]. More specifically, our problem [SOR 99], in particular when there are a large number of satellites, is that the cells covered by satellites are, in fact, hexagonal. Therefore, the area of a hexagon makes a ratio of about 0.827 with the area contained within the circumscribed circle. It is therefore necessary to add approximately 21% of satellites to the minimum number.

What is more, the coverage must take the movement of the satellite into account. So, when we want to create a permanent service, at least in a given zone, it must be guaranteed by either a geostationary satellite or have a family of satellites at its disposal: in this instance, we talk about *constellations* that ensure the global service. In this case, the movement of the satellites means that the covering satellite is constantly changing, even for an immobile user.

1.3.3. *Time interval between two successive intersatellite transfers*

A satellite transfer occurs when a user with an ongoing call switches from one satellite to another.

If a user comes close to the edge of a zone covered by a satellite, their coverage time by the latter will be very limited. This problem can be resolved by introducing larger recovery zones between neighboring satellites so that users at the outer edge of an area are covered by at least two satellites. The time interval between two intersatellite transfers is equal to the maximum time the same satellite can deal with a user.

1.3.4. *Time and time variation*

Links between users are often made with return channels. More specifically, in telephone communication, a conversation between an individual *A* and an individual *B* comprises two links *A–B* and *B–A*. Therefore, if some time d_1 (respectively, d_2) is required for a user on earth *A* (respectively, *B*) to reach a satellite, the delay caused by the network and perceived by *A* while he or she asks a question to *B* is $2(d_1 + d_2)$ (Figure 1.7).

In general terms, the additional perceived response time is two times the time required to cross the network. We can make a good approximation that this time is proportional to distance; radio waves move, including in the atmosphere, at a speed close to that of light in space. The transaction time via a geostationary satellite is therefore a minimum of $d = T+2(d_1+d_2)$:

$$\frac{4 \times 35\,800}{300\,000} = 0.477 \text{ s}$$

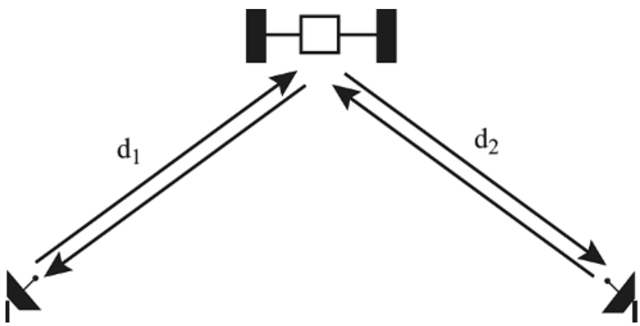


Figure 1.8. *Perceived response time (d) via a satellite with a processing time T*

The behavior of low orbit satellites is different from that of other categories of satellites: their times are very low and vary a great deal, unlike MEO and GEO satellites whose times are quite significant and vary very little. LEOs are therefore unquestionably able to break new boundaries with regards to applications. Nevertheless, some LEOs are not able to cope with the variations in perceived time, even if it means prematurely delaying the information received [WAL 84, BAL 80].

1.4. Orbital paths

The previous section described the criteria required for establishing a communication. These criteria are, however, local and cannot give a satisfactory perspective. So, in addition to establishing a communication, it is also necessary to guarantee a permanent service with a coverage that spans the whole planet [BAL 80]. There are therefore other equally important elements to be taken into account, including the specific restrictions relating to links with earth, networking and even cohabitation with concurrent systems. Some of these elements to be taken into consideration are:

– the *synchronization* of satellite paths with the rotation of the earth. For an observer on the earth, it is desirable for a satellite to stay in the same position (this is the case for geostationary satellites) or for it to periodically find itself at the same place over the course of the day;

– *earth-satellite transmission time* conditions the latency of the network (i.e. the duration of the transmission from start to finish);

– the *angle of penetration* of the earth's atmosphere that corresponds to the angle of elevation;

– the size of the *coverage cells*, the smaller the cells, the more important it is to reuse radio frequencies and the more the maximum number of users increases;

– the *total area covered by a satellite* (which can contain a number of cells), affecting the frequency of intersatellite transfers as well as the number of satellites required to cover a given zone;

– the *number of satellites required*, specifically to guarantee that the whole planet is covered for a service;

– the *required power of transmission*, the power can be used to increase the amount of information transmitted or to improve the portability of reception antenna (mobile telephones).

The characteristics of each of these systems will be analyzed in more detail.

1.4.1. *GEO-type systems*

These systems have a transmission time of 0.27 s, which is very high. They maintain a fixed position above the equator. Their high altitude enables them to cover a large area of the globe. Indeed, just three geostationary satellites are enough to cover almost the entire surface of the earth.

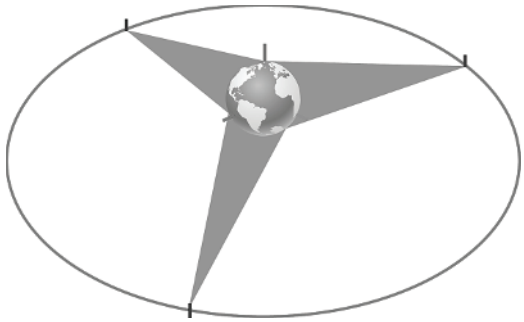


Figure 1.9. *GEO system, constellation with three satellites*

Nevertheless, these systems have significant coverage problems. They are unable, due to simple visibility problems, to cover the poles as well as any location with latitude over 81°. Furthermore, the angle of penetration in the atmosphere, in practice, makes it difficult to establish telecommunication links above 75°. However, even at latitudes between 45° and 75°, the angle of the satellite is small causing problems the moment the relief is not flat. High buildings in urban areas can therefore block coverage.

It would be possible to resolve this defect by using alternating and slightly inclined orbits. Satellites therefore complete one orbit in a figure of eight each day. These are non-geostationary geosynchronous paths.

1.4.2. *Elliptical systems*

In a certain sense, elliptical systems resolve the blocking problems GEOs have in high latitudes. Their orbit inclination of 63.4° enables them to position themselves above the equator. The angular velocity of the satellite on its elliptical orbit is reversely proportional to the square of its distance in relation to the center of the earth (the second of Kepler's laws). Therefore:

– *at its perigee*, the satellite moves very quickly in relation to the earth, even faster than a satellite with a circular orbit situated at the same location at the same time;

– *at its apogee*, the satellite moves very slowly in relation to the earth, slower indeed than a satellite with a circular orbit situated at the same location at the same time.

Satellites with an elliptical orbit are able to better center their footprint over the north and can make more precise adjustments (of coverage, synchronization). They do, however, regularly cross the *Van Allen* belts, considerably reducing their life expectancy.

However, both elliptical and geostationary systems are generally a very great distance from the earth, which has significant consequences on signal times and weakness. Transmission system designers have therefore looked for other solutions.

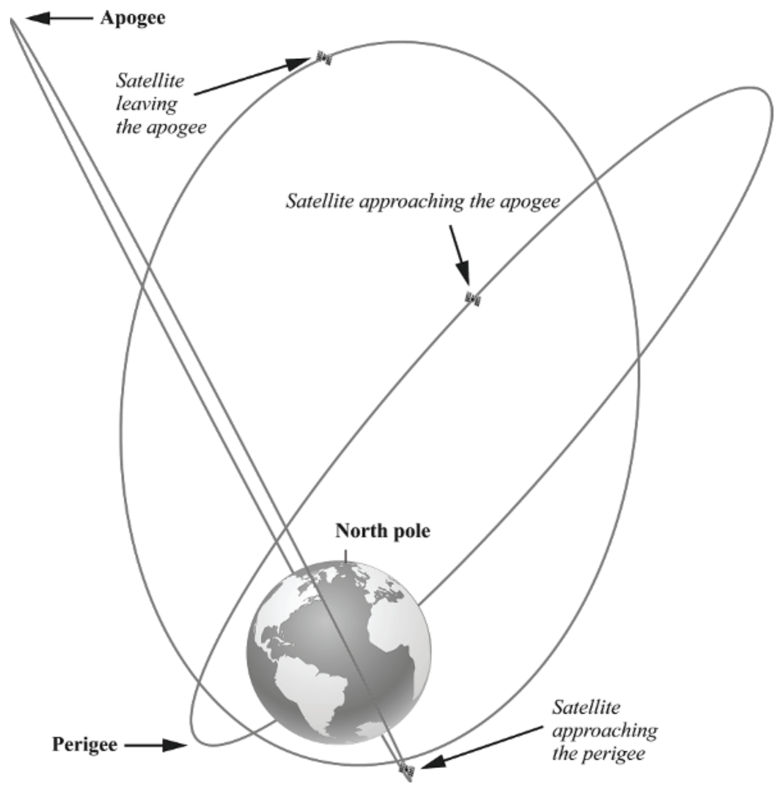


Figure 1.10. *Elliptical system, LOOPUS constellation*

1.4.3. *MEO-type systems*

MEO systems have a lower transmission time than their geostationary counterparts (110–130 ms at an altitude of about 13,000 km). Most MEO constellations have 10 or more satellites to guarantee a global coverage. Some systems only use an orbit with an inclination of zero, in which case coverage is excellent at the equator but rapidly declines as soon as the latitude increases. Other systems have slightly inclined orbits, requiring more satellites, but improving coverage.

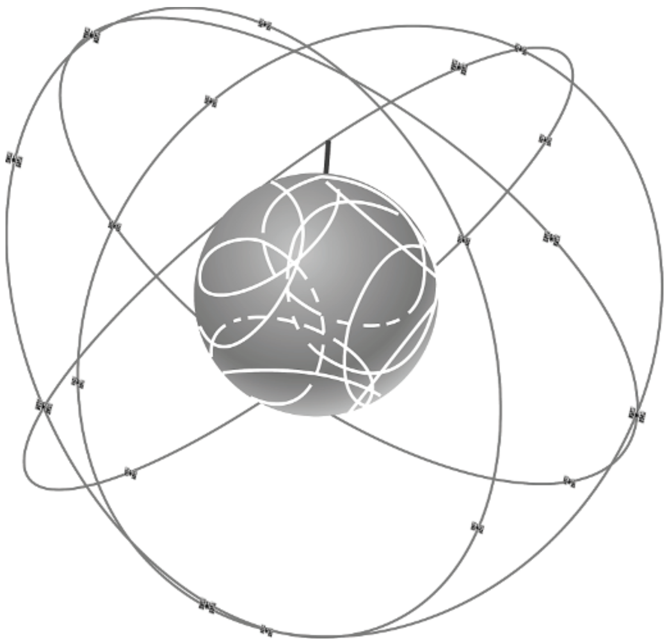


Figure 1.11. *MEO system, Spaceway NGSO constellation*

1.4.4. *LEO-type systems*

For LEO systems, transmission time varies between 20 and 25 ms; the precise time is at least as sensitive to the signals' angle of penetration in the atmosphere as to the precise altitude of the satellite. However, the devices are so close to the earth that a large number (at least 50) is required to cover the planet (due to the coverage reasons described in the previous section). The name *satellite constellation* mostly derives from this compulsory panoply. It is worth noting that to maintain coherent coverage zones, it is necessary that the movements of the satellites between themselves be synchronized very precisely.

Walker [WAL 71, WAL 84] defined two main types of constellations which continue to determine satellite constellation design today (Figure 1.12):

– Inclined constellations (Delta constellation by Walker) [WAL 71]

Inclined constellations have orbits with the same inclination *i* and with right ascensions of the ascending node *Ω*, which are regularly spaced out over a 360° area. In these constellations, both ascending and descending satellites guarantee that a zone is covered. Depending on the inclination *i*, their coverage will be better at the poles or at the equator;

Figure 1.12. *LEO system, Iridium constellation*

– Polar constellations (star constellation by Walker) [WAL 71, WAL 84, BAL 80]

Polar constellations have a series of orbits passing quite close to the poles and are organized in such a way that the ascending satellites (going from south to north) cover half the earth. The orbits all have the same inclination *i* that is close to 90° and have right ascensions of the ascending node *Ω*, which are regularly spaced out over a 180° area. Their coverage is therefore very dense at the pole and weaker at the equator. Additionally, it could be said that on the one side of the earth, all the satellites travel from South to North, whereas on the other side, they travel from North to South. As a result, satellites on the former orbit circle opposite and in reverse direction to those on the latter orbit. This phenomenon, which is known as *seams*, causes problems when communication between the satellites is sought.

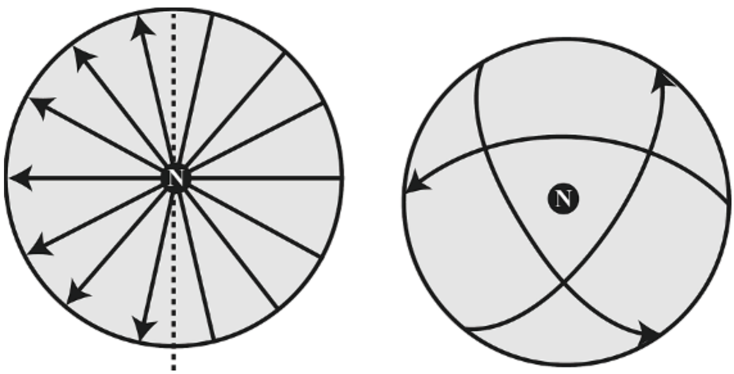


Figure 1.13. *Two types of popular constellations viewed from the pole*

1.5. Characteristics of cellular satellite systems

Cellular satellite systems are designed so that any authorized user, located anywhere on the earth, can have access and directly communicate with another user, either via a mobile phone or landline or via a public network

These systems are satellite constellations that are located on a number of interconnected orbital planes in a *ring* structure circling the earth. Each satellite generates several straight spotbeams that form cells on the surface of the earth. Unlike the cells in land networks that are fixed, these cells move with the satellite and sweep the surface of the earth.

Constellations of cellular satellite systems are designed in such a way that the satellite coverage footprints, which are symmetrically distributed, overlap to guarantee a continued coverage throughout the projected orbital path, as shown in Figure 1.14.

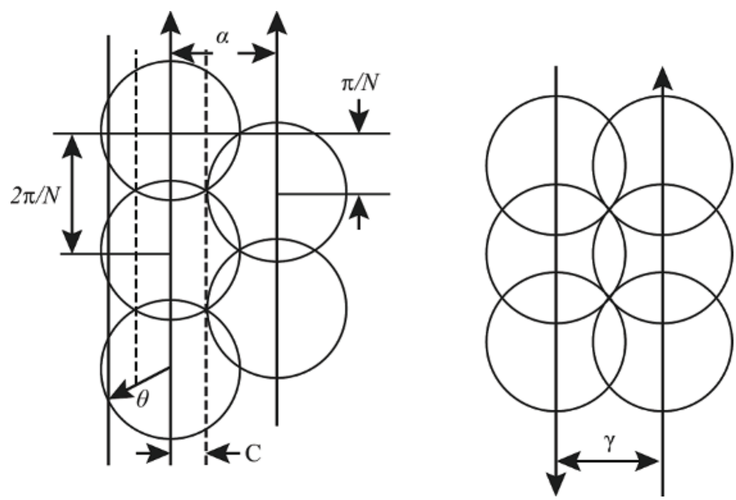


Figure 1.14. *Projection of coverage*

With half street width:

$$C_j = Cos^{-1}\left[\frac{Cos\,\theta}{\mathrm{Cos}\,\pi/\mathrm{N}}\right] \quad [1.5]$$

where N is the number of satellites in the orbit, and θ is the radius of the satellite coverage.

The direction of this constellation is determined by the phase angle ψ separating the satellite planes and the angular separation α_a between the ascending nodes of the successive orbital planes. In every constellation, there are as many orbital planes as coverage interfaces (interfaces separating these planes) and as many satellites as coverage interfaces between the satellite planes.

If a constellation of a orbital plane a is designated for a $(a–1)$ corotational plane, the angle of separation γ between the last and the first planes in the contrarotational sector will be equal to $2C_j = 180 - (a–1)\ \alpha_a$ so that there will be neither gap nor radio silence at any point while the satellites of these planes are moving in opposite directions.

Likewise for the $(a–1)$ corotational planes, the angle of the interplane phase $\psi = \pi/N$, separating each satellite from its neighbors in the adjacent planes, will leave no radio silence or gap in the coverage interfaces between the planes.

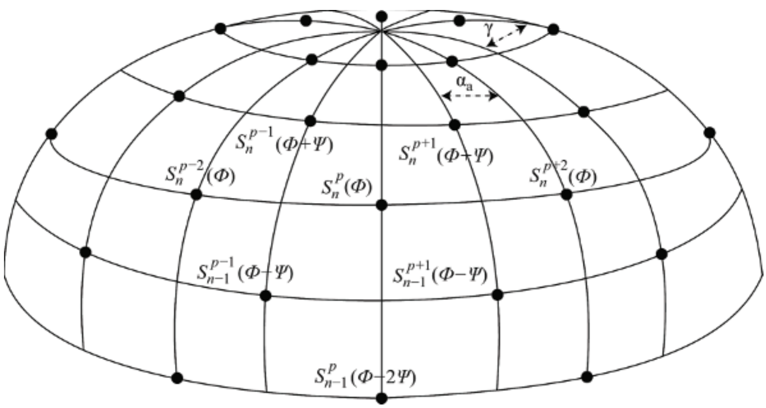


Figure 1.15. *Hemisphere of a constellation*

The common characteristics shared between satellite systems in order to ensure permanent mobile communication services are as follows [ALT 99]:

1) A "plurality" of corotational and contrarotational orbits that depend on the height of the orbit.

2) A number of equidistant satellites in each orbital plane.

3) The difference in the latitude phase between the satellites in the corotational sector is constant and depends on the height of the orbit.

4) The area of the coverage zone overlapped by each satellite with the neighboring satellites is a function of its latitude position.

The differences between the architectures of different satellite constellation systems in the world essentially reside in the inclination and the height of the orbits and, as a result, in the number of orbital planes and satellites in each plane.

1.6. The advantages of LEO systems

Three geostationary satellites located 36,000 km above earth guarantee a global coverage. Though they are perfectly suited to providing broadcasting services (radio, television), they are badly adapted to interactive multimedia services given that a transmission via a geostationary (GEO) satellite introduces a delay of about 270 ms thus making real-time applications difficult.

New low orbit (between 500 and 2,000 km of altitude) constellation systems avoid the ever-growing size of the frequency spectrum that is very sensitive on geostationary orbits. The times introduced by the propagation times, less than 20 ms, are compatible with communication protocols (TCP/IP, etc.), thereby ensuring the transparency of service in comparison to land networks. Moreover, the low altitude of the satellites makes the link budget more favorable and thereby reduces the emission power, the size of the satellites (LEO satellites generally have a mass of 500 kg in comparison to GEO satellites that weight several tons) and the size of the terminals (in particular, the antennae).

LEO constellations do, however, have some drawbacks. Countless satellites are required to guarantee a global coverage (the initial plan for the *Teledesic* system proposed a constellation comprising 840 satellites). LEO satellites, which are said to be scrolling as their rotation speed is higher than that of the earth, are only visible to a user on the earth for a few minutes, implying that the transmission regularly needs to be switched from one satellite to another to guarantee a continuity of service.

1.7. Handover in LEO satellite networks

Handover in LEO satellite networks has important characteristics including: a short propagation time, low power demands for the satellite or the user and a more effective allocation of the frequency spectrum. LEO satellites are considered to be the future of communication and information systems. However, LEO satellites are not stationary in relation to a fixed user on the surface of the earth. In fact, the speed of the satellite footprint on the earth is generally faster than the speed of the rotation of the planet and the user. The visibility period for a satellite in a cell is very brief and thus during one connection a user can be served by many spotbeams and satellites. Consequently, the link will switch from one spotbeam (potentially from a satellite) to another. This transfer is commonly known as handover [PAP 04a].

Consequently, maintaining a continuous communication on a LEO satellite system can require a change between one or several links as well as the Internet Protocol (IP) addresses of the *endpoints* of a communication [PAP 04a, CHO 00]. Handovers from the link layer and the network layer are therefore compulsory in satellite networks.

Managing the mobility of LEO satellites is therefore much more challenging than those of GEO or MEO systems. The mobility of a LEO satellite system is quite similar to that of cellular radio systems, though there are a number of differences. In fact, in the two systems, the relative position between the cells and mobile terminals is constantly changing, requiring the mobile terminals to switch (handover) between adjacent cells. The difference between the mobility of these two systems is as follows: in cellular systems, mobile terminals move through the cells, whereas in LEO systems, it is the cells that move through the mobile terminals. Cells in LEO satellite systems are larger than those in cellular radio systems. Moreover, the speed of a mobile terminal can be ignored in LEO satellite systems, since it is insignificant compared to the rotation speed of LEO satellites. However, unlike cellular systems on the earth where the movement of mobile devices is difficult to forecast, in LEO satellite systems, it is possible to forecast the movement of satellites and thus predicting the next satellite or spotbeam is relatively simple. At any moment, a real picture of the satellite constellation can be obtained, making it easier to select satellites in a communication path between two endpoints and therefore avoid unnecessary handovers. Handovers in satellite networks can be broadly categorized as presented in the following two sections.

1.7.1. *Link-layer handover*

Link-layer handover occurs when a change of one link or more is required between the communication endpoints because of the dynamic connectivity models of LEO satellites.

There are three types of link-layer handovers:

– *Spotbeam handover*: when users cross the border between the adjacent spotbeams of a satellite, an intrasatellite or spotbeam handover occurs. As the coverage zone of a spotbeam is relatively small, handovers between spotbeams are more frequent (every 1–2 min) [CHO 00];

– *Satellite handover*: when the connection established between a satellite and the attachment point of a user is transferred to another satellite, an intersatellite handover occurs;

– *Intersatellite links (ISLs) handover:* this type of handover occurs when the interplane of ISLs is extended due to temporary changes in the distance and the angle of visibility between the satellites in adjacent orbits. The progressive connections using these ISLs are derouted, thereby prompting an ISL handover.

The performance of the link-layer handover can be evaluated using two classic Quality of Service (QoS) criteria [TIA 01]:

– the probability of a blocked call (P_{b1}): the probability that a request for a new call be blocked;

– the probability of a forced termination (P_{b2}): the probability of the failure of the handover procedure causing the loss of an ongoing call.

There is a compromise between P_{b1} and P_{b2} in the different systems involved in the handover. Priority can be given by processing new calls and handover requests differently so as to reduce the ongoing calls being loss due to a failure in the handover procedure [OBR 99].

1.7.2. *Network-layer handover*

When one of the communication endpoints (satellite or user) changes IP address due to a change in the satellite coverage zone or the mobility of the user terminal, a network handover – or high-layer handover – is required to move the existing high-level protocols links (transmission control protocol (TCP), user datagram protocol (UDP), stream control transmission protocol (SCTP), etc.) to the new IP address. This is known as a network handover or high-layer handover [PUR 95]. Three different classifications can be used during the process of transferring a call [PAP 04a]:

– *hard-handover* systems: In these systems, the ongoing link is left before the next link is established;

– *soft-handover* systems: In these systems, the ongoing link is left only when the subsequent link is established;

– *signaling-diversity* systems: This system is similar to the *soft-handover* system, with the only exception that the signal flow of the old and the new links as well as user data are maintained by the old link during the handover procedure [PUR 95].