Decision Support in Space Situational Awareness

Richard I. Abbot and Timothy P. Wallace

To maintain the space catalog, the sensors of Air Force Space Command routinely track over 10,000 orbiting space objects. Because of the limited number of sensors, however, we cannot maintain persistent surveillance on these objects. This article describes algorithms and systems developed by Lincoln Laboratory to provide commercial and military analysts with better space situational awareness and decision support as they address problems in the space arena. The first problem is collision avoidance in the increasingly crowded geosynchronous earth orbit (GEO) belt, where there is continuous potential for on-orbit collisions between active satellites and debris, dead satellites, or other active payloads. This case is known as cooperative monitoring, since the owners of the satellites of concern share their operating data. Another problem is noncooperative GEO satellite monitoring, in which space analysts have no information about the satellite station keeping and maneuver plans. In this case, space surveillance data provide the only method to determine orbital status. This article summarizes GEO satellite orbits and their control, and describes a cooperative monitoring system for assisting satellite operators in maintaining safe spacing to nearby objects. We also address the noncooperative GEO monitoring problem by using Bayesian networks to combine signature and metric information from space surveillance sensors, which allows us to detect satellite status changes and produce automated alerts.

Space surveillance is the mission concerned with collecting and maintaining knowledge of all man-made objects orbiting the earth. The United States is the preeminent authority on space surveillance and maintains what is known as the *space catalog* of these objects through a global network of radar and optical sensors called the Air Force Space Surveillance Network. This space catalog contains unique identification numbers for each object and an orbital ephemeris that can be used to predict to some degree of accuracy where each object will be in the future.

However, because of the large number of resident space objects (over 10,000) and the limited number of sensors available to track these objects, it is impossible to maintain persistent surveillance on all objects, and therefore there is inherent uncertainty and latency in the catalog. Nevertheless, commercial and military analysts must make important decisions daily with this limited information. Decision support technology and algorithms developed by Lincoln Laboratory allow the analysts to do this work efficiently.

Through cooperation with Air Force Space Command, Lincoln Laboratory has developed an automated warning system that provides selected commercial operators of geostationary communications satellites with daily prediction warnings and supporting information for potential satellite encounters. This system, described in this article, has proven to be a key part of the satellite operator's decision-making process. In this case, the warning system provides the operator with potential encounter information several weeks in advance, and the operator uses this information to plan upcoming orbital maneuvers, or even perform a dedicated collision avoidance maneuver. The net result is that the satellite operator is now cognizant of nearby space threats and makes more informed decisions, which potentially prolongs satellite lifetime and revenue.

Space surveillance analysts, on the other hand, do not control satellites and must determine changes in satellite orbits from Space Surveillance Network sensor data only. Deep-space satellites that occupy the geostationary belt present the biggest challenge, due to the small number of available deep-space tracking sensors. A satellite that maneuvers in this orbital regime without detection may become lost, which will require the analyst to devote additional time and resources to find the satellite, at the expense of sensor resources devoted to the rest of the catalog.

In order to help the operator monitor these deep-space maneuvers, Lincoln Laboratory has developed a decision-support system based on Bayesian Belief networks. This system ingests daily surveillance data from deep-space radars and telescopes, and automatically assesses the orbital state of each geosynchronous earth orbit (GEO) satellite. In the event of any changes, this system can alert the user through either an e-mail notification or through a visual alerting system. This article describes the principal components of this decision system, explains the various types of operator displays, and shows results for selected scenarios.

Satellite Orbits and Propagation

The concept of a geostationary satellite orbit is believed to have originated with the Russian theorist Constantine Tsiolkovsky, who wrote articles on space travel at the turn of the nineteenth century. The idea that a satellite could be placed at a stationary location over the earth so that it could be used for communications is widely credited to Arthur C. Clark, who worked on many details including orbit characteristics, frequency needs, and the use of solar illumination for power.

In this article we mention both geosynchronous and geostationary satellites. A geosynchronous satellite has an angular velocity matching that of the earth, which theoretically requires a near-circular elliptical orbit with a semi-major axis of 42,164.2 km. Figure 1 summarizes important orbital parameters. A geostationary satellite remains over a given location on the earth's surface. A geosynchronous orbit does not necessarily make a satellite geostationary. If the orbit is slightly inclined to the equator, during the course of a day a satellite's latitude will increase and decrease through zero degrees, tracing a small figure eight over the surface of the earth. Also, if the geosynchronous orbit is not circular, the satellite will on average rotate at the same rate as the earth, but when it is at perigee (the closest point to the earth on its orbit) it will move faster and at apogee (the farthest point) slower. This change in velocity will add a slant to the small figure eight shape. Therefore, without a zero inclination and eccentricity, the geosynchronous satellite will not be geostationary.

The first geosynchronous satellite was Syncom 2, which NASA launched into orbit in July 1963. It was geosynchronous in that it had the same angular velocity as the earth, but it was not stationary over one location. The first truly geostationary satellite was Syncom 3, which NASA launched in August 1964. This satellite finally fulfilled Clark's prediction nearly twenty years earlier. Today a narrow belt of geosynchronous satellites orbit the earth near the required earth distance of 42164.2 km. About half of these are currently active; the rest are no longer functioning.

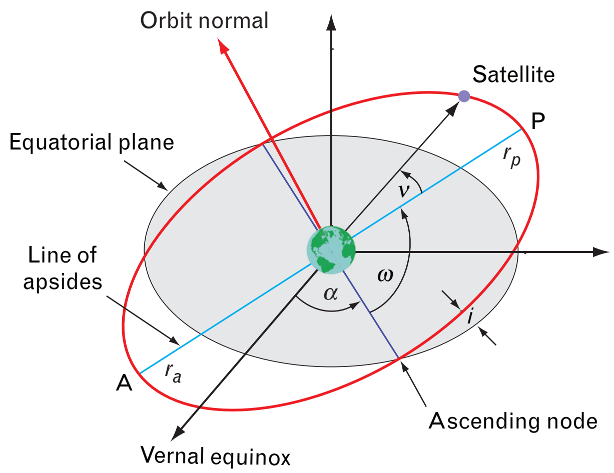


FIGURE 1. Parameters for an artificial satellite in orbit around the earth. The orbital ellipse (shown in red) is described by its semi-major axis *a* and eccentricity *e*. Other parameters are: *i* is the inclination of the orbital plane to the equatorial plane of the earth, A and P are the apogee and perigee of the orbit (furthest and closest points on the orbit to the earth), α is the right ascension of the ascending node, measured from the vernal equinox to the intersection of the north ascending orbit with the equatorial plane of the earth, ω is the argument of perigee measured from the ascending node to the perigee, v is the true anomaly measured from the perigee to the instantaneous satellite location, and r_p and r_a are the perigee and apogee distances given by $a(1 - e)$ and $a(1 + e)$. The line of apsides connects the perigee and apogee.

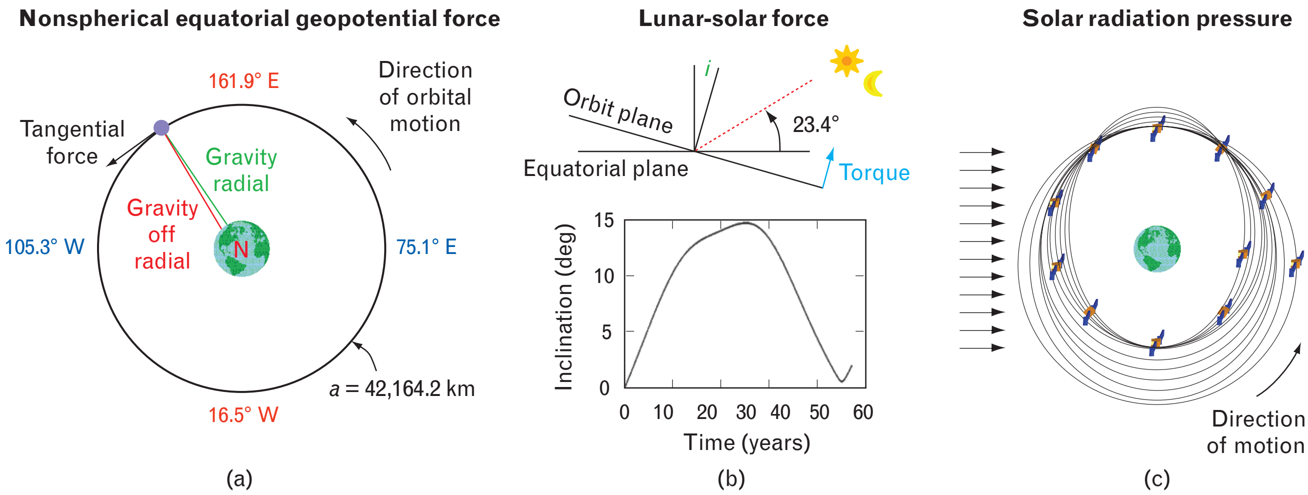


FIGURE 2. The three natural forces affecting the orbit position of a geostationary satellite. (a) The ellipticity of the earth's equator produces tangential forces that cause a drift in longitude. (b) The torques of the sun and moon cause a long-term evolution of the inclination from 0° to 15° and back in a fifty-four-year cycle. (c) The solar radiation pressure causes an annual periodic change in the eccentricity. These natural forces all require counteracting maneuvers by the satellite operator to keep the satellite geostationary.

The realm of geostationary satellites is a bustling belt-like region of space. Satellites are regularly launched into this belt, older satellites are retired, and others have prematurely died and are left to drift through the active satellite population. More recently, aging satellites are guided into graveyard orbits until human intervention can dispose of them. Satellite owners are continually jockeying for advantageous positions in the belt, and thus moving constantly through the region of the active population. Other satellites share common regions of space in clusters, or even in nearly the same locations. All this activity requires vigilance as the region becomes more and more populated. We need to understand the intentions of all these space objects to avoid collisions or communication interference. This has required more accurate satellite tracking plus improved orbit modeling methods and quick and accurate decision making.

A geostationary satellite position is inherently unstable. Even though a satellite operator can maneuver a satellite to a geostationary position, natural forces acting on the satellite will quickly change this position. Figure 2 illustrates these forces. The earth is not a perfect sphere, and the flattening due to its rotation is well known. There is also an ellipticity along the earth's equator. The difference between the largest and smallest radius of the equator does not exceed 192 m, but this differential can have a significant effect on a geostationary satellite, giving it a tangential acceleration.

Mathematically, the nonsymmetric gravity field potential is developed in terms of spherical harmonics (typically Legendre functions). The zonal terms of this expansion are rotationally symmetric and quantify the rotational flattening of the earth. The unsymmetrical mass distribution inside the earth is quantified by the tesseral terms of the expansion. The dominant two tesserals give a longitude dependence that is approximately sinusoidal with four nodes. At these nodes, the acceleration is zero, and therefore a satellite will stay at the node if it was stopped there at rest. Two of these equilibrium points are stable because a small deviation from the node's longitude point will cause the satellite to drift back to the node and oscillate about it. The other two equilibrium points are unstable, because a satellite will drift away from the node given any deviation in longitude. We can think of the stable points as gravity wells and the unstable points as hills. The stable points are at 75.1° E longitude, which is the deeper of the two and is associated with Asia and Africa, and at 105.3° W longitude (over Denver), which is shallower and is associated with North and South America. The higher of the unstable geopotential hills is in the western Pacific at 161.9° E longitude, and the lower peak is at 11.5° W longitude in the eastern Atlantic [1]. An interesting aspect of a satellite left to drift near the western Pacific peak is that it will move down the peak and have enough energy to climb the eastern Atlantic peak and to the other side of the Pacific peak, visiting both geopotential wells in the process.

Another natural force acting on a geostationary satellite is gravitational attraction of the sun and the moon, which do not lie in the equatorial plane of the geostationary orbit. The out-of-plane force of the sun is at its maximum at midsummer and midwinter, and zero in spring and fall. A similar attraction occurs for the moon during its monthly cycle, with the acceleration at its maximum twice per lunar period and passing through zero in between. The lunar and solar perturbations are predominantly out of plane, and thus cause a change in the inclination that has both a periodic and secular nature. This inclination increases to 15° in a period of twenty-seven years and then returns to 0° in the next twenty-seven years.

The third important force on geostationary satellites is caused by electromagnetic solar radiation pressure (SRP). This force has become more significant as the satellites have become larger in size and show more effective area to the sun. The SRP force is always normal to the satellite, which is oriented toward the sun for solar power. Integrated over one half of the orbit, a small velocity increment is gained, which tends to raise the altitude (or apogee) at the opposite point. Over the other half of the orbit a small delta velocity opposes the orbit velocity, which tends to lower the altitude (or perigee). Thus, during the year as the earth moves around the sun, the eccentricity increases and decreases with a magnitude on the order of 0.0005.

Figure 3 shows the changes in semi-major axis, eccentricity, and longitude due to natural forces on a GEO satellite that started to drift near the western Pacific peak of the earth's gravity field potential. The drifts and oscillations caused by natural forces require action on the part of the satellite owner to counteract. This action is discussed in a later section of this article. These forces are also important because they lead to orbits that can potentially be threatening to other satellites, if left without counteraction.

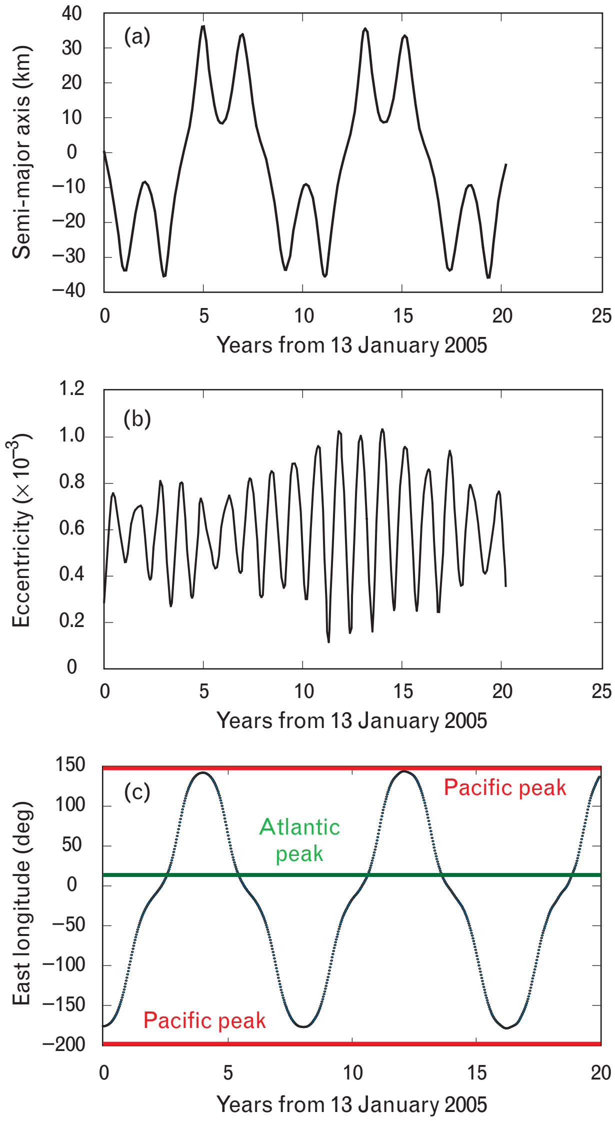


FIGURE 3. The evolution under natural forces of the semi-major axis, eccentricity, and longitude of a GEO satellite that started to drift at 173° E longitude, near the western Pacific peak of the earth's gravity field potential. One orbital position is plotted per day. (a) Over an eight-year period, the semi-major axis varies from –35 km to +35 km from the geosynchronous radius. (b) The eccentricity varies yearly. (c) Over an eight-year period the longitude moves east over the Atlantic peak on to the other side of the Pacific peak until it turns at 150° E and then moves westward back to the initial longitude.

Orbit Determination and Maintenance

There are four primary components in the determination and maintenance of a satellite orbit: (1) tracking data, (2) force models, (3) an estimation theory that ties these components together to continually update the orbit state vector and propagate it into the future, (4) and error analysis. The tracking data are discussed in the following section. The estimation theory can be either least squares or a sequential method. Each has its strengths and weaknesses. The least-squares method is perhaps better suited to batch processing of orbits for many satellites. The sequential estimation approach seems to provide a more realistic estimation of the orbit

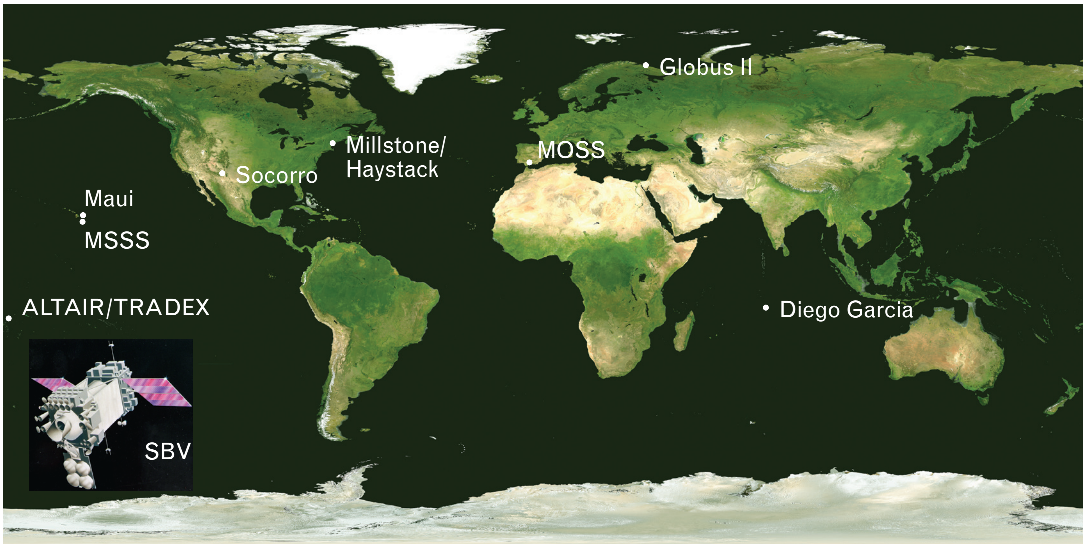


FIGURE 4. The Air Force Space Surveillance Network sensors that track geostationary satellites. The network consists of three radar sites (Millstone/Haystack in Westford, Massachusetts; ALTAIR/TRADEX on Kwajalein Atoll in the Pacific Ocean; and Globus II in Norway) and three Deep Stare Ground-Based Electro-Optical Deep Space Surveillance (Deep Stare GEODSS) optical sites (Maui, Hawaii; Diego Garcia in the Indian Ocean; and Socorro, New Mexico). There are also two other contributing ground-based optical sites; the transportable optical site located at Moron, Spain, and the optical sensors at the Maui Space Surveillance System (MSSS) complex. The Space-Based Visible (SBV) orbiting optical satellite has also been a contributing sensor to the Space Surveillance Network since 1996.

error or covariance because it is better suited to the input of *a priori* error models.

Space Surveillance Network

The Space Surveillance Network, illustrated in Figure 4, consists of a mixture of ground-based radar and optical telescopes. It also includes the Space-Based Visible (SBV) optical telescope situated in a polar orbit at 850 km altitude. The metric measurements from network sensors are used with an orbit determination process to constantly update the state vectors for all earth-orbiting satellites. Additionally, some of the radar cross section and optical signature information can be used for satellite correlation and status change identification. The fusion of the Space Surveillance Network metric, radar cross section, and brightness information reveals much information about each satellite's orbit and state.

The ground-based radar systems in the Space Surveillance Network can provide range, azimuth, and elevation observables, while some can also observe range-rate or Doppler shift of the transmitted radar signal. Optical systems in the Space Surveillance Network provide precise directional information about a satellite with respect to the sensor location. The directional information is either an azimuth-elevation pair or a right ascension–declination pair of observations. The radar and optical metrics are both useful for initial orbit and refined orbit determination. The satellite brightness is also collected and has been found to provide useful information on satellite status.

When metric observations from both radar and optical sensors are fused in orbit determination, each type contributes its unique observables to the process. Orbit determination depends on having an observable system, i.e., a system in which the measurements contribute information to determine all state parameters uniquely. If any of the state parameters are not observable, orbit determination uncertainty increases. Radar range and range-rate measurements are typically precise. When these measurements are fused with precision optical angular measurements, a fully observable system is realized, thus allowing high-precision orbit determination.

To understand how the radar and optical measurements contribute to the observability of a satellite orbital state, it is useful to describe the position and velocity of an orbit in terms of radial, along-track, and cross-track directions. The radial component of a satellite orbit describes the instantaneous position of a satellite along the vector from the earth's center. The along-track component is the position of a satellite with respect to its instantaneous velocity vector. The radial and along-track components form a plane that contains the satellite's orbital ellipse. The cross-track component is normal to both the radial and along-track components and serves to orient the orbital plane. The cross-track component also lies along the instantaneous angular momentum vector.

Radar range and range-rate measurements provide observability of the radial and radial-rate components of a given satellite's orbit. Because the radial measurements add observability of the semi-major axis of the orbital ellipse, the along-track component of the orbit is also well observed, since it depends on the semi-major axis. Optical measurements provide observability of the along-track and cross-track components of a satellite's orbit. Because the along-track component is observed well, the semi-major axis and consequently the radial component of the satellite's orbit are also observed well. Together, radar and optical measurements complement each other by providing overlapping observability of the radial and along-track components. Cross-track observability is provided primarily by the optical measurements, although radar measurements can provide additional cross-track determination if the stations are well distributed around the globe in higher and lower latitudes.

The Space Surveillance Network has been tracking satellites since 1957. The first satellite tracked that year was the first satellite ever launched, the Soviet Union's Sputnik I. Since that time, the network of radar and optical systems has grown, and more than 25,000 satellites have been tracked since the network's inception. Currently, more than 10,000 satellites are maintained in the Space Surveillance Network catalog, and approximately a thousand of the currently tracked satellites are active. The rest consist of debris, launch-related rockets, and unused or failed satellites. The geosynchronous belt contains many valuable satellite assets in geosynchronous or geostationary orbit; about 380 active satellites reside along with more than 750 inactive satellites, rocket bodies, and debris.

Deep-Space Orbit Control

Over its lifetime, the geostationary satellite undergoes a significant amount of orbital activity. After launch it is first inserted into a geosynchronous orbit, followed by station acquisition. Then it undergoes years of station keeping against the drift of the natural forces. From time to time it will have station shifts as the operator decides to move it to a different position over the earth. Invariably it may find itself in a cluster of other satellites in the same vicinity or collocated with another satellite in the same control area. Finally, if it survives failure and is near depletion of station-keeping propellant, it is retired to a graveyard orbit, where it can exist without being a threat to the active population. All of this activity involves thrusting or maneuvering of the satellite, and a resultant change in the predicted knowledge of the satellite's trajectory.

After launch, a geosynchronous satellite is put into a low-earth circular parking orbit. It next undergoes a transfer orbit that has the perigee of the orbit (closest point on the elliptical orbit to the earth) at the parking altitude and the apogee (farthest point) at the geosynchronous altitude. This is a high-eccentricity orbit (e is about 0.73), which allows the satellite to glimpse the geosynchronous belt at the farthest point of the satellite's orbit. Maneuvers are next required to circularize the orbit at the geosynchronous altitude. Also, because the parking orbit has a non-zero inclination while the geosynchronous orbit inclination is near zero, a plane change is required.

The process of geosynchronous orbit insertion requires maneuvers at the satellite apogee. These maneuvers place the satellite in a near-geosynchronous orbit that has a slow drift in longitude. Also, the inclination and eccentricity of this orbit are not yet the desired values. A minimum of three in-plane and one out-of-plane maneuvers are necessary to achieve the required near-zero eccentricity and inclination [2]. The first two burns set one apse at geostationary height and set up the desired drift rate. The third moves the other apse to a geostationary height to achieve the circular orbit. The in-plane maneuvers are done at the orbital apses. The out-of-plane maneuver is performed at the intersection of the drift and required geosynchronous orbit. After these maneuvers are completed, the satellite is moved to a testing location.

When a satellite acquires a geosynchronous orbit and testing is finished, it next needs to be placed in its desired longitude. This step requires an east-west drift initialization maneuver that can be made in either direction, depending on the final destination. For the satellite to drift east, the orbit must be lowered with a retrograde burn. At the lower altitude, the satellite has a shorter orbital period and gets ahead of the earth's eastward rotation, and hence moves east. For the satellite to drift west, the orbit must be raised with a posigrade burn. At the higher altitude, the satellite has a greater orbital period and falls behind the earth's rotation, and hence moves west. Finally, braking burns stop the satellite at its desired location.

Each geostationary satellite is assigned a longitude slot in which it must be kept. The primary limitation in spacing satellites along the geostationary belt is that the limited allocated frequencies must not result in interference between satellites on uplink or downlink. Also, natural forces cause the satellites to move, and it is necessary to ensure that the satellites do not collide. Finally, the satellite must remain within a small distance of its ideal location to ensure that it remains within the ground-antenna beamwidth without tracking; otherwise more complicated antennas would be required. The longitude slots are assigned by the International Telecommunications Union (ITU) with coordination by regional agencies, e.g., the Federal Communications Commission (FCC) in the United States. For commercial satellites the slots range from ±0.050° to ±0.1°. Some satellites (e.g., meteorological, some mobile phone systems, and military communication) often have larger longitude control boxes, since they have a wider coverage beam or use a tracking antenna.

As we discussed earlier, the primary orbital parameters of concern that change due to natural forces are longitude, inclination, and eccentricity. The longitude drift must be counteracted or the satellite will quickly move out of its slot. Inclination must be maintained or the satellite will describe an increasing figure eight and require antenna tracking. Generally, bounds of inclination of ±0.1° are maintained, although if control in inclination is not as critical (because of wide coverage beams or tracking antennas) inclination can drift for some time. Eccentricity must also be maintained. The maintenance of a geostationary satellite in its assigned slot is called station keeping. The strict limits of longitude and inclination (latitude) define a dead zone for the satellite. Two types of maneuvers are done for this station keeping, in the east or west (EW) direction and in the north or south (NS) direction. The satellite must carry enough fuel to perform these maneuvers and maintain its position over its expected lifetime, which can be from ten to twenty years.

Corrections to satellite motion caused by the earth's slightly elliptical equator and SRP require thrusting in the transverse or EW direction. The strategy of these EW maneuvers is to change the longitude drift and to decrease the eccentricity, both in a combined manner. For longitude control, the satellite is allowed to drift toward one longitude limit, and then enough of an impulse is applied in the opposite direction so that the satellite is pushed to the opposite limit, where the natural forces will make it turn and drift back. This maneuver can be done with a single tangential thrust, which also can be timed to correct the eccentricity drift due to the SRP. An east thrust near apogee or a west thrust near perigee decreases the eccentricity. The single station burn does not permit the choice of a new longitude drift rate and eccentricity independently, because the two are coupled (e.g., a tangential thrust of 1 m/sec results in a change in longitude drift rate of –0.352°/day and a mean change in eccentricity of 0.000065) [1]. The two-burn maneuver is commonly used to correct for longitude drift and eccentricity drift, where the two maneuver thrusts are separated by half an orbit. If change in longitude is most important, thrusts must be in the same direction. If change in eccentricity is most important, then east and west thrusts are applied alternatively half an orbital period apart [1].

The NS station keeping is done by changing the orbital plane to maintain correct inclination against the forces of lunar-solar perturbations. This procedure consumes much more fuel than drift corrections; roughly 95% of the satellite's fuel is required to maintain inclination through NS station-keeping maneuvers. Generally, time periods for inclination maneuvers vary from five to fifteen days.

When inclination control is not so stringent (and when a ground antenna can continuously track), the operator can let the satellite drift to save fuel. For example, a 3° inclination bound can be maintained for about 7.5 years if the right ascension of the ascending node starts at 270° [3]. If the maximum possible inclination is only 0.5°, then at least one maneuver is required per year.

For an NS maneuver, any misalignment of the thrust direction away from nominal produces a thrust component in the EW direction (a coupling). This misalignment has to be corrected in the EW station-keeping maneuver, and requires appropriate scheduling of the NS maneuver in the EW maneuver cycle. In all cases, the operators usually give themselves some room for error, knowing that there could be a problem with the performance of the next maneuver.

In theory, we should be able to predict when an operator should be doing a maneuver, by using the orbital mechanics described above and knowledge of the station-keeping bounds. In practice, however, the time when a satellite can undergo a station-keeping maneuver depends on how well the operator knows the true position of the satellite, or how well the last maneuver performed and how much coupling there was, or how much wiggle room the operator likes to maintain for the satellite in the box, or on equipment or man-power availability, or even on the personal schedule of the operator.

There is a small effect on longitude drift that must be considered for orbit control. Satellites must maintain attitude control for proper orientation to the earth. One method of doing this is with momentum wheels, which utilize gyroscopic stiffness to provide three-axis stabilization. These momentum wheels absorb external torque disturbances by a gradual spin-up or spin-down. For the momentum wheels to function properly, the stored momentum of the wheels must be kept within allowable limits. When the limits are exceeded, a momentum-wheel adjustment is required, which involves a thruster firing of suitable magnitude and orientation. The change in velocity values involved are small (< 0.01 m/sec) but can still produce a noticeable drift of the satellite. They can also be used to advantage to provide a small contribution to the EW station keeping.

To maintain the orbit for the satellite and to know when a station-keeping maneuver is required, the operator collects tracking data. These tracking data may be obtained on an ongoing basis (e.g., once per hour), or densely for a limited period following a maneuver in order to check performance of the maneuver and derive a new orbit. The tracking consists of measurements of range to the satellite and possibly angular measures of azimuth and elevation. The range measurements can be time delays of a signal sent and returned by the satellite through a transponder, or they can obtained by using satellite beacons. Usually two ranging stations are involved and are given the largest separation, or baseline, as possible. The range data are precise to a few meters but can be poorly calibrated and have large bias errors. The angle measurements generally have errors of tens of millidegrees and are marginally useful. The consequence of poorly calibrated range data can be severe. Large biases in these data will shift the satellite in longitude, and to a lesser extent in inclination. This error can lead to a satellite being out of its allocated station-keeping box, thus impinging on the transmissions of a neighbor and possibly leading to a collision.

A number of geostationary satellites require station-keeping strategies that are subject to additional constraints. The ring-shaped region of the geosynchronous belt has just one dimension—longitude—to allocate different spacecraft. With increasing demand for geostationary satellite services over certain regions of the world, many GEO satellites today exist in clusters. A cluster consists of satellites in neighboring deadbands plus those which are collocated or which share common deadband regions. The cluster can provide connected or individual satellite services from a number of satellites. A well-known example of a collocated cluster is the Astra cluster at 19.2° E ±0.10 in longitude with six objects, which are kept separated by eccentricity and inclination. Two satellites may also be collocated for a short time as one replaces another. From the surveillance perspective, a cluster is defined as two or more satellites that can come close enough that tracking sensors can mistag them (i.e., the tracking of one is assigned to another in that cluster). Currently, there are nearly sixty clusters with satellites within 0.6° of each other in longitude.

Satellites existing in clusters can be owned by a single operator or by a number of operators and agencies. The single operator of collocated satellites for some configurations must keep the satellites within the beamwidth of a fixed ground station antenna, and must satisfy the above station-keeping requirements and also keep the satellites sufficiently separated to avoid collisions among themselves. When different operators have satellites in a cluster, the operators have to pay strict attention to their own station keeping to avoid interference or a possible collision. It is in the best interest of the different operators to share orbit information, which is routinely done in practice.

There are various approaches to collocation of GEO satellites [1]. The first approach is when different operators are involved and the risk of a collision is ignored (the probability of collision is considered insignificant by the operators). Signal interference can of course be monitored by each operator. In the second approach, the satellites are flown independently, but a safe separation distance is agreed upon and checked before and after maneuvers. A third approach maintains collocation by separation in longitude, eccentricity, or eccentricity and inclination in combination. This approach can still involve different operators who are either exchanging information or assuming that they are keeping to individual allocated orbital regions. The final approach utilizes separation by longitude, eccentricity, or eccentricity and inclination but with offsets so that station keeping for all satellites is done on a predefined schedule. With the same station keeping they all move in their control area in the same manner. Clearly, routine proximity checks should be made for all of these methods. Figure 5 illustrates both longitude station keeping and a collocation of two satellites.

A satellite can also be associated with a cluster if, for example, it is one with a larger longitude control region. Such satellites pass through the longitude boxes of other satellites during their station-keeping cycle. Generally, there is no coordinated effort by operators during these longitude crossings, although proximity analysis must be maintained by the surveillance community.

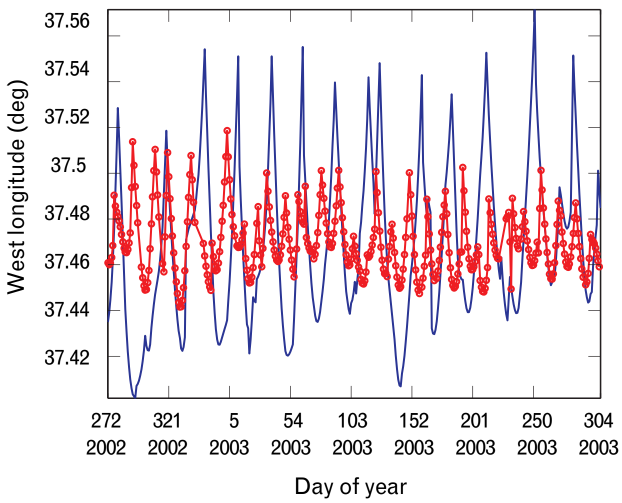


FIGURE 5. An example of longitude station keeping and collocating two satellites. These two satellites shared the same longitude slot for three and a half years. This figure illustrates slightly more than a year of this collocation. One satellite was Telstar 11 (red), which had a longitude box size of ±0.05°, and the other was Satcom C1 (blue), with a longitude box size of ±0.1°. Flying these two satellites at the same longitude location forced operators from different companies to develop a strategy to keep the satellites separated. We played a role in monitoring this collocation and occasionally suggested avoidance strategies to keep the satellites at safe distances. Satcom C1 has since been retired by being boosted into a safe super-synchronous orbit above the geostationary radius.

From time to time a geostationary satellite operator performs a relocation. This move could be done if a more productive longitude slot becomes available, to switch an older satellite with a newer and more capable satellite over a given service area, or to move an older satellite closer to a stable point to conserve fuel and lengthen its lifetime. The rate at which this relocation is accomplished depends on how much fuel and time the operator wishes to allocate. The relocating satellite crosses other active satellites during this move and is more exposed to the dead population. Therefore, monitoring is required to avoid a possible collision.

As a satellite nears the end of its life, the decision must be made of how to dispose of it so that it will not be a threat to the active population. Before 1977, satellites were left to die in place and allowed to drift under the natural forces. Recommendation of a systematic removal of satellites from the geosynchronous belt was made in 1977, when four satellites (three Intelsat satellites and one from the Soviet Union [4, 5]) were disposed by putting them in regions not used by active satellites. Today, the ITU Radio Communication Assembly recommends that a retired geostationary satellite must be sufficiently boosted above its geostationary orbit so that it cannot interfere with existing operational satellites that are within 200 km above the GEO altitude that incorporates both the station-keeping zone and the relocation corridor [6]. The re-orbit, which requires the operator to have a good assessment of the remaining fuel on the satellite, is usually done with a series of thrusts. The last thrusts circularize the orbit and deplete all remaining fuel.

The active geostationary satellite population worldwide is maintained by many commercial operators and government agencies. Their satellite control activity is governed by regulations and recommendations, but for the most part many of these operators and agencies perform their work in various levels of isolation. Most of them generally keep specific information about their satellite operations to themselves, and they are not always completely aware of the geostationary satellite situation around them. The surveillance community that attempts to maintain the orbital catalog for the geostationary satellite population does not have information readily available about all the specific activity of this population, and therefore must determine this information by continuously collecting tracking data for it. In this process, the surveillance community must detect the maneuvers and then quickly determine a new and accurate orbit. Otherwise, a satellite may be temporarily lost to the catalog and require a search to find it again. Also, the post-maneuver trajectory may be on a collision trajectory with another satellite; this possibility must be quickly assessed and a response must be formulated, as discussed in the next section.

Cooperative Geosynchronous Monitoring

Geostationary objects have been launched into orbit for over forty years. Prior to 1977, when their station-keeping fuel was depleted, they could no longer be controlled and were simply allowed to drift. With the 1977 recommendation to re-orbit the geostationary satellites to at least a few hundred kilometers from the geostationary orbit, many were moved to orbits where they could be less threatening to the active satellite population. This re-orbiting, of course, not only depended on the actual height above or below the geostationary orbit but also on the eccentricity, since the perigee and apogee heights could still allow the drifter to reach the geostationary ring.

Satellites also suffer catastrophic failure. Strong solar activity is a major cause of such failure and ultimate loss of communication and control. High-speed solar wind streams give rise to a large flux of charged particles that reach the earth within hours. Many get trapped at geosynchronous altitude, where they form a highly energetic plasma for a short time. Exposed satellite surfaces can build up electrostatic charge, which can lead to an electrical discharge and induced current in electronic systems. Today, operators do make an effort after a failure to remove their own satellites from the active geostationary ring if they can manage sufficient control.

Currently, the number of controlled satellites is over 380. The total number of drifting uncontrolled geostationary satellites (with drift rate of 0.9 to 1.1 rev/day or with semi-major axes of 40,465 km to 42,488 km, respectively, and eccentricity less than 0.1) is near 750. Approximately 150 of these drifters are in a librating orbit and thus cannot cross the entire active population. Of these librators, about 36 oscillate in the geopotential well centered at 105.3° W with periods of 2.5 to 6 years, about 90 oscillate in the other geopotential well centered at 75.1° E with periods of 2.5 to 5.5 years, and about 15 oscillate about the unstable points passing through both wells and with periods from 8 to 10 years [7]. The remaining uncontrolled satellites are circulators far enough from the geostationary orbit not to be captured in oscil-

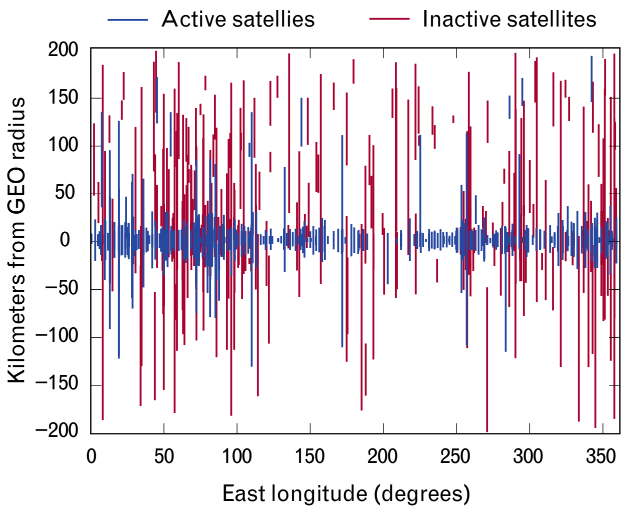


FIGURE 6. A snapshot on a given day of the radial distances from the earth (determined by the perigee and apogee) versus longitude of all active and inactive geostationary satellites within 200 km of the geostationary radius. The active satellites (shown in blue) stay nearly at the same longitude, while the inactive satellites (shown in red) drift in longitude at a rate that depends on how far they are above or below the geostationary radius. An animation of these data would show how the inactive population drifts by the active satellites and thus potentially could be a threat if they have common radial distances.

lation. Their eccentricity is large enough, however, that their perigee or apogee can cross the active geostationary population. They drift around the earth with periods proportional to their semi-major axis. Figure 6 shows a one-day snapshot of the geostationary belt, illustrating the potential threat of the uncontrolled inactive satellite population to the controlled active population, based on common radial distances from the earth.

Figure 7 summarizes the total number of encounters between all active satellites and all inactive satellites during one year. The peak of this distribution depends primarily on the variance of the radial distribution of the drifter population [8]. The question is invariably asked about the probability that a collision will occur in the geostationary ring. This ongoing problem was first studied as early as the 1980s [5]. In our definition of a collision we include the possibility that two solar panels would hit, since this event would have a severe and possibly critical impact on the operation of the geostationary satellite. Relative velocities for a drifter in a 7° inclined orbit are about 370 m/sec. Different methods have been used to estimate the probability of such a collision, and they basically give the same result. If we assume a colli-

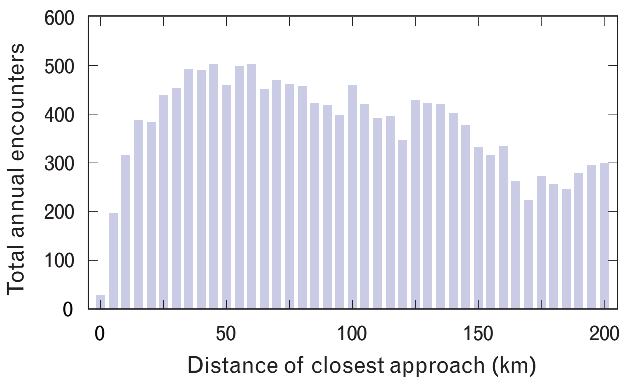


FIGURE 7. Histogram of the number of encounters of all active satellites with the inactive satellites for a year. The peak depends on the radial distribution of the drifter population. This histogram will stay nearly the same in shape but will scale as both populations increase.

sion radius of 50 m (i.e., a cross-sectional area of about 8000 m^2), today's satellite population would yield a collision rate on the order of 1.0×10^{-3} per year, or about one collision every thousand years [9]. This rate has increased by a factor of ten in the last decade. An order-of-magnitude calculation like this one does not, however, consider that there are longitude regions where satellites are crowded. This calculation also assumes that active satellites cannot collide with each other because they are maintained at their assigned position.

A collision and subsequent loss of a geostationary satellite would have an enormous impact. Besides the monetary loss of the satellite, valuable communications would be disrupted or possibly lost completely over the affected area until the satellite could be replaced. A collision would also leave a debris population that would make that longitude region of space unusable until means were available to clear it.

We became actively involved with helping to prevent a possible collision of geostationary satellites in early 1997, when Telstar 401 failed on orbit because of a geomagnetic storm, and because there was no opportunity to boost the satellite away from the active geostationary ring [10]. As Telstar 401 failed at 97° W longitude, its long-term evolution has it oscillating to 113° W longitude and back over a 2.5 year period. Unfortunately, this oscillation causes it to pass through a dense population of geostationary satellites serving the Americas. Figure 8 shows the first cycle of Telstar 401's drift through the geopotential well centered at 105° W longitude. The first crossing came with Galaxy IV in June 1997. The estimated separation distance was less than one kilometer, so we suggested an avoidance maneuver for Galaxy IV. An avoidance maneuver is an additional unscheduled maneuver, which fortunately can be designed as best as possible to also achieve some station-keeping gain. This maneuver resulted in a new predicted crossing distance of six kilometers. Some type of avoidance maneuver strategy was implemented on eight of the fourteen crossings that occurred with Telstar 401 that year. These strategies included extra maneuvers that were unscheduled or existing maneuvers that were modified to increase separation distance.

In 1997 Lincoln Laboratory joined a cooperative research and development agreement (CRDA) with four commercial companies, all of which had many assets at risk because of the Telstar 401 drift. This CRDA permits us to work with the operators to monitor their satellites for the threat of collision, and they in turn sponsor Laboratory research on related topics. Besides monitoring the Telstar 401 crossings, the work of this CRDA initially concentrated on (1) further study of orbit accuracy of geostationary satellites as a function of tracking type—radar and optical—and tracking density, (2) understanding the risk to the active population of the entire drifting population, (3) monitoring the calibration of CRDA partner range data and utilizing it in the orbit estimation, and (4) understanding how to model the station-keeping maneuvers from the different operators.

As we studied the overall threat of the inactive drifters to the active population, and as more satellites failed on orbit, we found it necessary to build an automated geosynchronous monitoring and warning system (GMWS). Figure 9 illustrates the components of this system. The GMWS performs the following steps. It first maintains a list of current CRDA partner satellites that need to be protected from collision. It also forms a threat list utilizing the most recent historical orbit information for all the inactive drifters, and determines those which can cross the active geosynchronous belt ring. This threat list can be supplemented with active satellites that can be a significant threat from time to time. The Space Surveillance Network tracking data are combined with the CRDA partner ranging data into an orbit-determination process to update the orbit state for all the satellites. This process also incorporates station-keeping maneuver information that is requested from each CRDA partner for two weeks in advance.

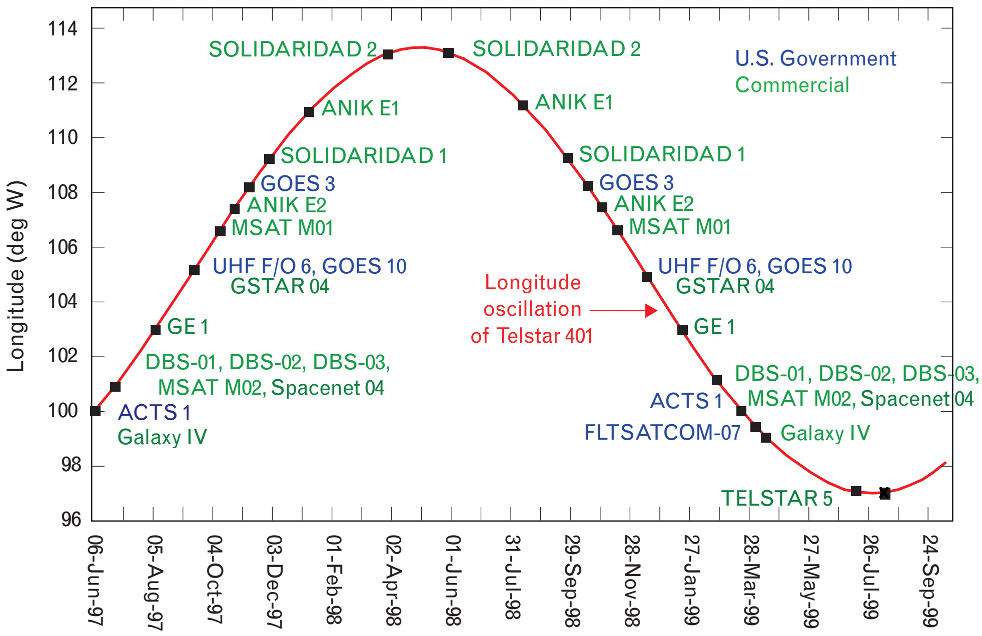


FIGURE 8. After Telstar 401 failed in January 1997, it began to drift in the geopotential well centered at 105° W longitude. This figure shows oscillation of Telstar 401 and the commercial and U.S. government satellites that were crossed during a two-year period, beginning in June 1997 with Galaxy IV (since retired). The first crossing with Galaxy IV was estimated to have a crossing distance less than one kilometer. We suggested an avoidance maneuver to increase the separation to six kilometers. Telstar 5 (now known as Intelsat Americas 5) replaced Telstar 401, which illustrates how the population changes with each Telstar 401 oscillation cycle as new satellites are launched, some are relocated, and others are boosted to graveyard altitudes.

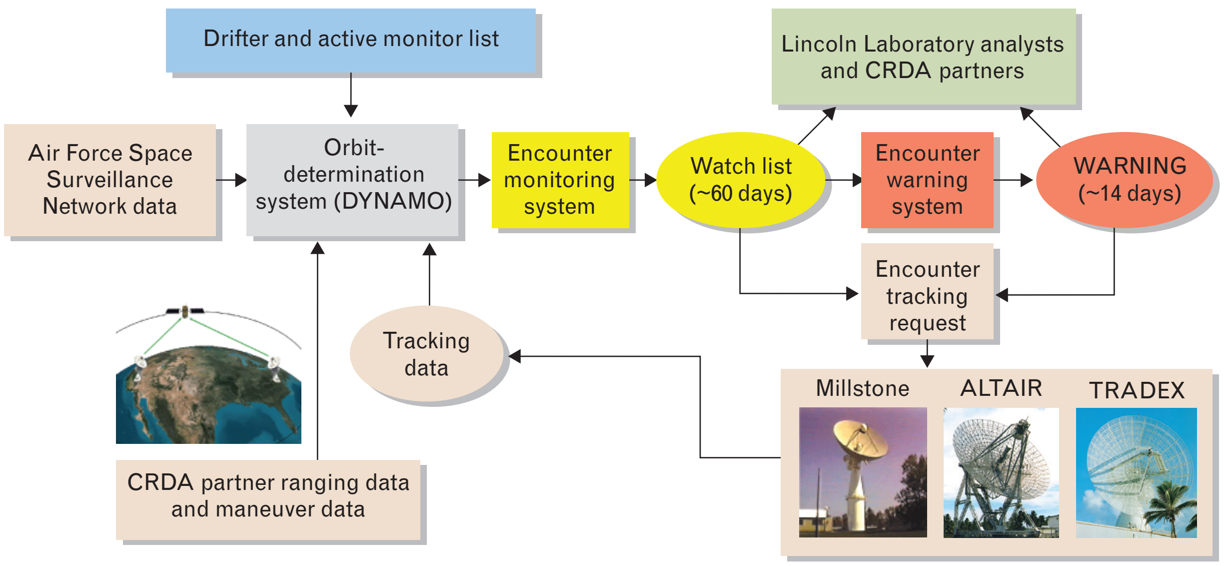


FIGURE 9. Geosynchronous monitoring and warning system (GMWS). This automated system, which monitors active cooperative research and development agreement (CRDA) partner satellites against potentially threatening inactive drifting satellites or other active satellites, computes high-precision orbits for all GEO satellites by using a Lincoln Laboratory orbit-determination system known as DYNAMO. This system fuses Space Surveillance Network data with commercial tracking data, which usually are collected from two widely spaced ground stations, and determines a sixty-day watch list of potential close encounters, and a two-week warning list of close encounters that may require some precautionary action.

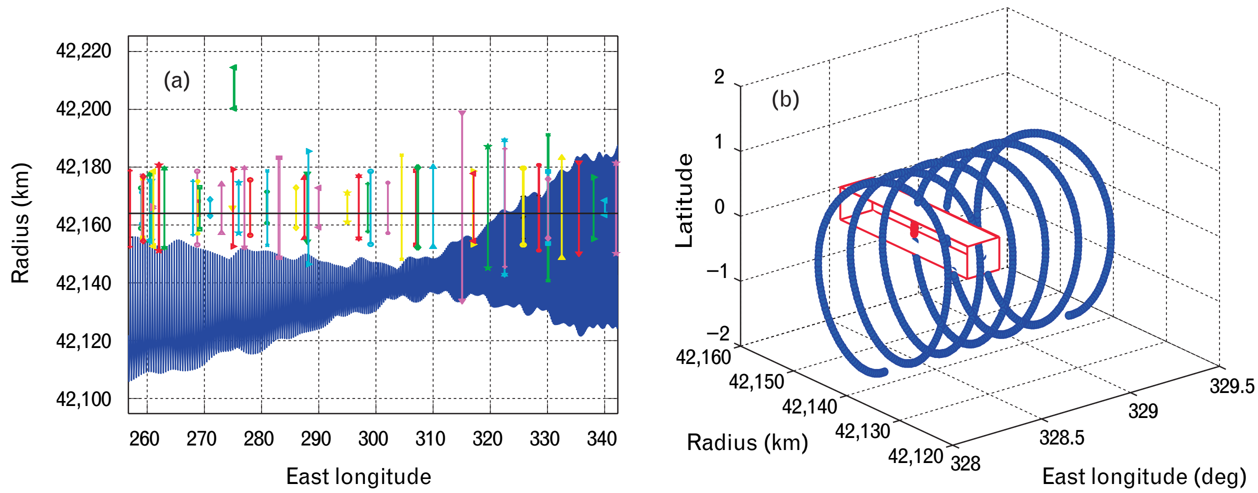


FIGURE 10. Left: the radial distance versus longitude of an inactive drifting GEO satellite in 2006, shown at four-hour spacing for its propagated orbit. Each colored line represents an active satellite with its longitude and radial range computed from its perigee and apogee. The orbital evolution characterized earlier in Figure 2 has the semi-major axis lower than the GEO radius during the start of this period. As a consequence, the orbital radial distances were below the GEO radius, and the drifter could not cross through the active station-keeping boxes, making a close encounter impossible. In the summer, the satellite's eccentricity decreased, keeping the radial spread smaller. Toward the end of 2006, the semi-major axis and eccentricity increased, and the drifter crossed radially into other active satellite boxes. The determining factor as to how close it will get to the active satellite in its box depends on where the drifter crosses through the active box as it passes through the active's orbital plane. Right: a three-dimensional view of a single box crossing. This figure shows the projected crossing of the drifter through an active satellite's station-keeping box in September 2006.

With updated orbits for all objects, the next step is to determine which satellites will actually be close to each other. To do this we look ahead sixty days or longer. Two satellites can come close only if they occupy the same volume of space at the same time. For the inclined drifters this closest crossing would occur near the intersection of the two orbital planes. We can make the specified volume the size of the station-keeping box, although we typically make this volume conservatively larger, with dimensions of 250 km in longitude, radius, and latitude. Satellite pairs that pass this criterion are then put onto a *watch* list, where 'watch' indicates that a very close crossing could occur. The watch can be visualized in various ways. Figure 10(a) shows the longitude of a drifting GEO satellite from 15 February to 10 December 2006. It also shows when the drifter can have a radial distance that can cross an active station-keeping box and become a threat. After September 2006, the drifter met the radial distance criterion for intersecting several active satellite station-keeping boxes, but for only three active satellites did it cross through the plane of their boxes at the proper radial distance. Figure 10(b) shows a three-dimensional view of one of these three crossings in September 2006. The crossing distance for this encounter was 50 km and of no concern. As we discuss later, the orbit accuracy even sixty days ahead is generally on the order of two kilometers, so the point of crossing through the box is well determined. This information allows the operators to visualize where their active satellites should not be. It can also be considered if long-term maneuver planning is being done.

Thirty days from a close crossing, we assess the accuracy of the orbit and the amount of tracking data available for the drifter. With the throughput of the upgraded Deep Stare Ground-Based Electro-Optical Deep Space Surveillance (Deep Stare GEODSS), tracking is usually sufficient. If it is not sufficient, then extra tracking is requested from the Millstone radar in Massachusetts and the Reagan Test Site ALTAIR or TRADEX radars in the Pacific Ocean, if they have coverage and are available. Generally, a track of five separate radar measurements is adequate. Finally, within two weeks of a close crossing, the operators are usually doing station-keeping maneuvers, and a stronger alert with more urgency—called a *warning*—is given at that point, and some precautions may be required.

The next question is how close do crossings have to be to be of concern. We have arrived at certain guide-lines based on orbit accuracy assessment; this topic is discussed in more detail below. A crossing distance of ten kilometers is notable, but generally of no concern. A crossing on the order of six kilometers is generally considered safe, but this information is reviewed by satellite operators and our analysts. In that review process we examine how recent maneuvers for the active satellite performed relative to predictions, and we check again on all maneuvers that are scheduled before the upcoming close crossing. Also, we review the quantity of tracking information for the drifter, as well as the modeling of the SRP, and we make an assessment of the accuracy of the orbit.

Generally, when a crossing separation is within four kilometers and consists of a certain geometry, we suggest an avoidance maneuver strategy. The idea is to avoid having the operator perform a maneuver to take the satellite off the station-keeping plan and then another maneuver to bring the satellite back. Ideally, the operator checks if some advance or delay of an upcoming maneuver can be done without penalty. A change in the maneuver time by minutes or by advancing or delaying it by a day is the most common strategy, since it does provide adequate separation (again with the six-kilometer or greater goal), especially if the encounter is still a few days off. Another strategy, which requires some fuel, is to change the eccentricity of the orbit to increase the radial separation before the crossing, and then change it back after the crossing. The operators have relatively rigid constraints on their station keeping, but they always seem to find a strategy that accomplishes their goals and gains a satisfactory increase in separation.

When we feel an avoidance maneuver strategy should be considered for satellite crossings, we provide the operators with the drifter satellite orbit. Compatibility of our orbits with their orbit determination system is one of the first things checked when we begin to work with CRDA partners. The drifter orbit lets them validate the encounter and plan a strategy to increase separation. Ultimately, they make the final decision about the safety of their satellites. We, however, model the suggested maneuver strategy in our orbit determination to check its effect on the separation of the satellites. Examples of close crossings where avoidance maneuver strategies were performed are presented later.

The most difficult aspect of this monitoring is the proper modeling of a maneuver near an encounter, and the validation that the maneuver resulted in the expected performance. This validation is not difficult for the primary component of the maneuver (either NS or EW) but it can be difficult for the coupled components that depend on the satellite attitude and hence the direction the thrusters fire in. For an EW maneuver, there can be coupling in the radial direction (causing a change in the orbital eccentricity), and for an NS maneuver in both the EW and radial directions. Often these coupling components are not given in advance, but their determination can be critical for close crossings of certain geometries. The orbit-determination process estimates the relevant maneuver components as soon as it gets enough tracking, and these values can be compared with the operator's estimates. Generally this comparison can be done within a day after the maneuver, given the partner ranging data, or with nominally two tracks of optical and/or radar measurements.

For close crossings, the decision to modify a maneuver or specifically perform an avoidance maneuver depends on the orbit accuracy and the encounter geometry. Orbit accuracy was the first issue addressed after the Telstar 401 failure and drift [11]. A drifting geostationary satellite orbit can actually be well determined without much tracking data. With radar-only tracking, we can achieve 0.5 to 2 km (1 σ) accuracy over the period of the tracking data used to determine the orbit. If we examine the orbit error in terms of the components in the along-track or velocity direction, the cross-track or out-of-plane direction, and the radial direction, the error is on the order of 0.3 to 0.5 km, 0.5 to 1.5 km, and 0.05 to 0.1 km, respectively. From the discussion on sensors, we know the angle measurements are the worst for a radar, which degrades the orientation knowledge of the orbit plane and leads to larger cross-track component in the error budget. If the SRP force is sufficiently well modeled, the error usually remains at this acceptable level for many weeks as the orbit is propagated into the future.

With optical data (before the Deep Stare upgrades), and with at least ten tracks of five to ten measurements per track, the error components were roughly 1 km, 0.2 km, and 0.2 km, respectively. Here the cross-track component is better determined. A mixture from both sensors (even with nominally two radar tracks) is very complementary, and total errors can be achieved on the order of 0.5 km, with 0.2 to 0.4 km, 0.1 to 0.2 km, and 0.025 to 0.05 km, respectively, by component. With current Deep Stare GEODSS and radar data, or Deep Stare data alone, 0.2 km accuracies (1σ) are typically achievable, and even better results have been demonstrated.

For the active satellite, these accuracies can be reached, but it depends on how accurate the maneuver information is or how well it can be estimated in the orbit solution. With the addition of calibrated CRDA-partner two-station ranging, orbit accuracy on the order of 50 m is achievable. The limiting factor controlling improved geostationary accuracy seems to be the simplified modeling of the SRP by a simple surface, whereas the geostationary satellite is much more complicated. The momentum-wheel adjustment thrusts also complicate the orbit modeling at this level, if information about them is not available.

With the ability to determine orbits to 0.2 km or better, the obvious question is why do we consider a separation of six kilometers safe and request a avoidance maneuver strategy when the separation is less than four kilometers. Basically, this distance provides a comfortable margin of safety, especially since the cost of a collision is extremely high (as we've already mentioned). If the crossing separation is mostly in the radial component, we may be comfortable permitting a closer crossing to occur. Currently, with sixty CRDA partner satellites, we have to make a decision about five times per month regarding crossings within four kilometers.

A complication arises when an operator is using a satellite that has a xenon ion propulsion system (XIPS) for station keeping. The XIPS thrusting technology is attractive because it is ten times more efficient than conventional liquid fuel systems. Therefore, bigger payloads and longer lifetimes can be achieved at lower cost. There are different strategies for using these systems, but thrusting is generally done a few times per day for intervals up to a few hours. The impulsive maneuver a few times per month is no longer needed when XIPS is used, although a traditional bi-propellant fuel is still often used in conjunction with the XIPS. This frequent thrusting makes obtaining a maneuver-free orbit impossible, and accurate estimates of separation distances of crossing satellites are difficult to determine if the XIPS maneuver information is not available. Also, the XIPS was meant to be autonomous (but with operator intervention), in which case significant advanced planning is required if an avoidance maneuver strategy is required. When the XIPS maneuver information is supplied, it is possible to predict a close crossing as accurately as with the traditional means of station keeping. Therefore, when a drifter is predicted to pass through a XIPS satellite station-keeping box, we need to have the XIPS schedule with corrections as they occur, or we provide the operators with the drifter orbit and have them do the analysis.

We now discuss two operational examples of encounters that were predicted to be very close (i.e., less than three kilometers), and how the separation distance was increased. The first example illustrates an avoidance strategy that delayed the start time of a scheduled station-keeping maneuver. On 2 October 2005, the GMWS predicted a 1.5 km crossing distance between the drifter ASC01 and the active Intelsat Americas 5. The operator had a scheduled EW maneuver for 1 October, which if delayed by one day until after the close crossing would increase the separation distance to 17.5 km, as illustrated in Figure 11. The operator was able to do this maneuver delay without the satellite being out of its station-keeping box. The crossing was later reviewed with post-encounter tracking and orbit determination. This review validated the crossing at a slightly greater distance of 17.9 km.

A second example illustrates another type of avoidance strategy, in which a small eccentricity change is made for the active satellite orbit in such a way as to increase the radial separation with the drifter during the closest crossing. The encounter involved the drifting Telstar 401 with the active satellite MSAT 01 on 16 May 2006. A west maneuver had been performed for MSAT 01 seven days prior to the predicted encounter. After we estimated the radial coupling of that maneuver we found the crossing distance to be 2.4 km, with 0 km in the cross-track component, 0.2 km in the radial component, and the rest in the along-track direction. There were no scheduled maneuvers before the encounter, so the operator decided to schedule an avoidance maneuver that involved a two-maneuver change of eccentricity. The operator made this avoidance maneuver one of the yearly sets of required eccentricity control maneuvers, which therefore resulted in no additional fuel cost for the life of the satellite. The first eccentricity maneuver resulted in a 4.6 km total separation, but more importantly the radial separation was increased to 2.4 km, which was considered safe, given the errors in that component as discussed above.

The GMWS also has components to monitor the active versus active population. These components can

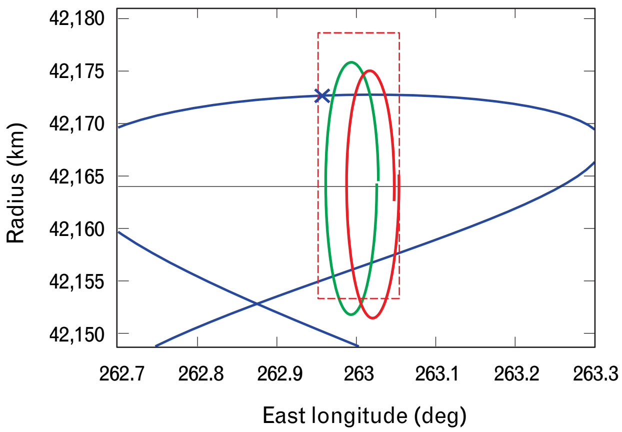


FIGURE 11. A two-dimensional view of the effects of avoiding a close crossing, estimated to be less than 1.5 km. Given the value of the active satellite (Intelsat Americas 5), and the risk of losing it, an avoidance strategy was performed by delaying by one day an EW station-keeping maneuver. The dashed box is the active satellite station-keeping region in the radial and longitudinal projections. The encountering drifter is ASC01 (launched in 1985); its trajectory is shown in blue. ASC01 is a high-inclination drifter (nearly 9°) passing through the active box with a relative velocity of 0.5 km/sec. The blue x along the ASC01 trajectory represents where ASC01 passed through the plane of the active satellite. The green trajectory inside the station-keeping region is the trajectory predicted for the active satellite during the day of the encounter, but before the maneuver was changed. With a change of maneuver, the active satellite trajectory became the red path and the separation distance was increased to 17.5 km.

monitor an active satellite with all the other active satellites in all phases of its lifetime from geostationary insertion to the retiring re-orbit. The GMWS also contains a prototype system to monitor the infringement of a neighboring satellite on the station-keeping box of a specified active satellite. There is a substantial challenge in monitoring an active population for which the maneuvers are known with the remainder of the active population for which the maneuvers are not known. This problem requires maneuver detection, which can be done in a few different ways (one method is discussed in the next section).

With the metric tracking data of the current optical and radar sensors of the Space Surveillance Network, the detection of a maneuver is not inherently difficult. It is necessary, though, for a satellite to be tracked after the maneuver. This tracking may not occur immediately afterward, however, and there is some chance—if the maneuver was large enough—that the satellite could be temporarily lost and a search could be required. The radar angle measurements are not the most accurate of the four radar measurements; they cannot detect a maneuver as capably and quickly as the range and range rate measurements. The high-precision angular measurements and increased throughput of modern optical data show the greatest promise. In simulations, maneuvers on the order of 1 m/sec (typical of relocation burns) are detectable in as few as fifteen minutes [12]. Typical EW station-keeping maneuvers on the order of 0.1 to 0.01 m/sec can be detected within twelve to twenty-four hours.

The quick post-maneuver recovery of the orbit accuracy to its pre-maneuver level with routine nominal tracking, or perhaps with extra tasked tracking, is also a challenging problem. Detecting the maneuver and recovering an accurate orbit are both required to make quick decisions if a collision trajectory is a possibility. The most promising development has been with a sequential estimation filter for orbit determination [12]. With the ability to yield a realistic covariance for the orbit, it is easier to establish confidence that new observations that do not match the orbit within the covariance imply that a maneuver has occurred. Once the maneuver has been detected, there are three possible approaches to estimating a new post-maneuver orbit. One approach simply disregards the pre-maneuver orbit and uses new tracking data as they are available to compute an initial orbit for the satellite. The second approach forces the filter to accept the post-maneuver tracking that indicated the maneuver and that otherwise would be rejected as not fitting the orbit within the covariance. Mathematically, the orbit covariance is opened up to accept the new data, while the filter still retains memory of the pre-maneuver orbit. The third method involves the utilization of both the pre-maneuver and post-maneuver orbits from the IOD to determine the approximate maneuver time and delta velocity where the two trajectories best intersect.

Which of the three methods to be implemented depends on how quickly and accurately a post-maneuver decision has to be made with regard to the satellite's new orbit. The third method shows the best promise for producing the quickest and most accurate post-maneuver orbit, but it involves more steps, and hence makes automation more difficult. A system that can automatically determine all active satellite maneuvers with confidence and quickly determine accurate orbits with minimal amount of tracking is under development.

We finish this section with a summary of the encounter monitoring work. The GMWS has been operational since April 2001. Since our initial work with Telstar 401, we have monitored well over 1250 crossings within ten kilometers and with about 136 unique drifters involved. For sixty CRDA partner satellites, this monitoring now finds on the order of 250 crossings per year that are less than ten kilometers. Since we began our monitoring work with Telstar 401 in 1997, we have recommended about 65 avoidance strategies. We now average about eleven strategies per year for the sixty active CRDA partner satellites currently being monitored. We now recommend—about two-to-three times per year—a specific unscheduled maneuver to increase the separation distance to a safer level. Most of the time, however, we rely on our accuracy assessment and slight adjustments of scheduled maneuvers.

Noncooperative GEO Monitoring

As described at the beginning of this article, Air Force Space Command performs the task of tracking and cataloging all space objects. This task is accomplished mainly by processing metric observations of space objects, including range, azimuth, and elevation from radar sensors or azimuth and elevation from optical sensors. Metric observations are used to build orbital element sets that characterize the orbits of space objects.

The GEO belt is a particular challenge because its great range renders many objects untrackable by some sensors, particularly radars. Also, certain regions of the GEO belt are quite crowded with satellites, so the possibility exists that an observation of one satellite may be incorrectly tagged as an observation of another nearby satellite. If a major malfunction occurs, or the satellite operator underestimates the depletion rate of maneuvering fuel, a GEO satellite may fail in place, causing it to drift through the GEO belt, as described in the section "Satellite Orbits and Propagation."

As tracking sensors observe GEO satellites to obtain metric observations, the sensors also collect signature information in the form of photometric measurements from optical sensors and radar cross section measurements from radar sensors. This signature information adds to the already voluminous metric data stream, and for that reason has traditionally not been used to assist in the construction of element sets for catalog maintenance and status monitoring. We have observed that this signature information has been used by radar operators at the Lincoln Laboratory Space Surveillance Complex as well as by Laboratory analysts to distinguish one satellite from another or to help determine the status of a satellite. This process suggests that signature information might be routinely useful if automated processing could be developed such that the work load on the analyst does not increase significantly.

We have developed a system that processes signature information on GEO objects along with element sets from Air Force Space Command, and then performs automated information fusion. The majority of our signatures are from the Space-Based Visible sensor [13], but we have also processed signature information collected during tracking operations of the Millstone L-band radar and the Haystack X-band radar.

The case of a GEO satellite failure on orbit can be used as an example of how signature and metric information is fused. The satellite will start to drift out of its slot, but it will most likely be days or even weeks before that drift is evident. Signature information might indicate a loss of attitude control much sooner, providing a tip that drifting behavior is imminent. Loss of attitude control also means that the satellite status has changed, and the satellite is no longer able to perform its mission.

Several processing steps are involved in fusing signature and metric data to obtain more reliable satellite status information. First, relevant information is extracted from the signatures and element sets. We have developed several new algorithms to assist in this process, and the performance of these algorithms is critical to overall system performance. The extracted information becomes evidence that is combined in dynamic Bayesian belief networks, which compute an assessed status for the satellites over their entire life history.

Bayesian network nodes have discrete states, and maintain their belief in each of those states as a belief vector. The GEO Bayesian networks have several different kinds of nodes. The signature data nodes are the simplest and have the states *nominal* (NOM) and *anomalous* (ANOM). The stability data nodes are similar and have the states *stable* (STAB) and *unstable* (UNST). The metric data nodes are more complicated and have four states—*in-slot*, *moving*, *drifting*, and *graveyard* (NOM, MOV, DRFT, and GYRD). Evidence of these three types is accumulated and fed to the satellite status nodes, which have six states—*nominal in-slot*, *anomalous in-slot*, *nominal moving*, *anomalous moving*, *drifting*, and *graveyard* (NOM, ANOM, NMOV, AMOV, DRFT, and GYRD). Some of the state abbreviations are the same, although they may mean slightly different things. There should be no confusion as long as the type of node is kept in mind.

It is a challenge to display the six-state output of the multiple satellite status nodes. We have chosen to do this by constructing a history in which the Bayesian state of maximum probability determines the color of the function. The belief in nominal states is plotted normally, and the belief in anomalous or end-of-life states is inverted, i.e., subtracted from the number one. Figure 12 shows an example in which GEO weather satellite GOES 8 is in one orbital slot, moves to another slot around July 2003, and is finally put into a graveyard orbit at its end of life. The purple curve shows the longitude of the ascending node (LAN) of the element sets for this object, showing constant longitude value while the satellite is in a slot and a decrease in longitude as the satellite was moved. When the graveyard state is entered, the longitude decreases much more rapidly.

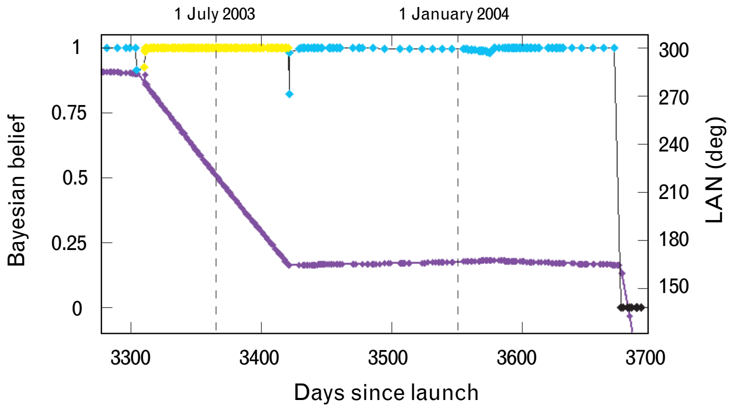


FIGURE 12. Bayesian history for geosynchronous satellite GOES 8 during the period from early 2003 into mid-2004. This satellite starts in one orbital slot, moves to another, and is then put into a graveyard orbit. Blue represents the nominal in-slot state, yellow represents the nominal stable but moving state, and black represents a graveyard state. The purple curve shows a rapid decrease in longitude of the ascending node (LAN) during the graveyard state, which is a typical end-of-life scenario as a satellite is transferred to a graveyard orbit.

The Bayesian network reflects this history, with blue representing the NOM state, yellow the NMOV state, and black the GYRD state. Note that the GYRD state is shown at a belief value of 0.0. Since this is an anomalous state, the belief in it is actually 1.0 but it is inverted to make it easier to detect visually and to make the curve dip when the state of the satellite is uncertain or anomalous. This is the nominal end-of-life behavior of a GEO satellite.

Figure 13 shows the another typical end-of-life case of a GEO satellite. Here the satellite is in the NOM state (blue) but then suddenly becomes ANOM (red) as a major failure occurs. The operator is unable to boost the satellite into a graveyard orbit, so the satellite drifts out of its slot, as shown by the purple LAN curve and the DRFT state (brown) in the Bayesian plot. Both ANOM and DRFT are anomalous states and so are inverted.

Because this system is automatically fusing information from incoming data, significant state changes are detected around the clock. Alerts can be sent to notify users of significant status changes, thus obviating frequent checking of the Bayesian network graphs. This process can cue further investigation of the reason for the alert, further sensor tasking, or any other appropriate action.

Information Extraction from Element Sets

The GEO Bayesian network described in this article combines diverse evidence into an overall assessment of satellite status at different times over the lifetime of the satellite. The goal of element-set processing is to provide metric evidence for this Bayesian network. There is an inherent mismatch between an element set, created from selected metric observations over some significant time span, and the requirements of the GEO Bayesian network, in which evidence is ordinarily time-stamped, reflecting a relatively short collection interval. To bridge this gap we need to process the element sets to obtain derived evidence that fits the Bayesian network requirements.

Another potential problem is the lack of evidence independence. Consecutive element sets for an object frequently are based on almost the same input data, and sometimes the exact input data, i.e., the last element set, can be propagated forward to obtain the current one. Ideally, the evidence input to the Bayesian network would be independent; practically, it is possible to work with some correlation. Unfortunately, however, we do not know which metric observations were used in the calculation of any given element set. We do not even know the time span represented by those observations.

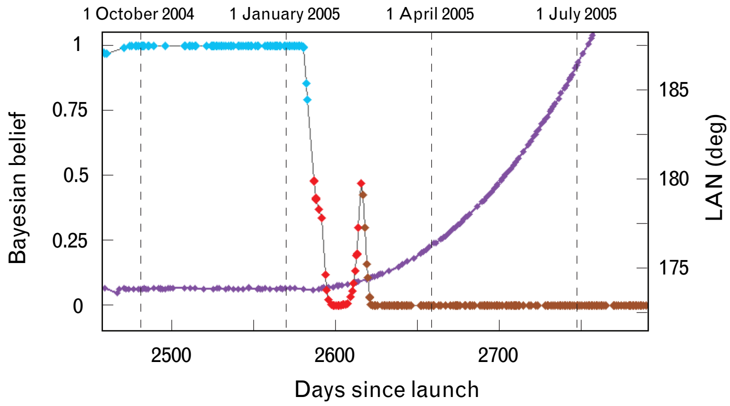


FIGURE 13. Bayesian history for the failure of a GEO satellite in 2005. Loss of attitude control results in a change from the nominal in-slot state (blue) to an anomalous state (red) when a failure occurs. Cessation of station keeping eventually causes the satellite to drift out of its slot, as shown by the brown drift state as well as the purple LAN curve. In this scenario, the red anomalous state warned of impending drift before that drift actually occurred.

Often the epoch date and time of the element set are the same as the most recently processed observation, but not always. We need to keep these issues in mind as we determine the time spacing of our derived evidence.

An element set describes the inertial motion of a satellite, and can be propagated into the future to predict its future location. To obtain evidence relevant to satellite status from element sets, we look for thrusts, which always represent evidence relevant to status. A correct assessment that a satellite thrusted not only tells us about the orbital status but also presents evidence that it was stable at the time of the thrust, since otherwise boosts are not attempted. However, thrusts may occur infrequently in some GEO belt locations, especially when highly accurate station keeping is not required. In these cases, we may need to output intermediate evidence based on the LAN. Also, if a satellite is abandoned, it will start to drift; of course, there will be no thrust event corresponding to this status change. Thus drift evidence must sometimes be output on the basis of LAN variation and unrelated to any boost.

Table 1. Element-Set-Derived Evidence

Evidence type	*Epoch*	*Evidence value*
station keeping	date/time	walk rate
“to slot” thrust	date/time	walk rate
“from slot” thrust	date/time	walk rate
other thrust	date/time	walk rate
moving	date/time	walk rate
drifting	date/time	confidence
in slot	date/time	confidence

Table 1 shows the type of evidence that is developed by preprocessing the element sets for a GEO object. Seven different types of evidence are used, as appropriate, each consisting of a triplet as shown. Some evidence types (those which involve a thrust, plus moving) supply a new walk rate, which is the average daily change in LAN, to the GEO Bayesian network. The others supply a confidence that the specified state exists, given the elapsed time and the LAN. Evidence of this form is all the GEO Bayesian network requires to perform its processing. The next section provides an outline of how this evidence is developed.

Element-Set Filtering

Not all element sets are created equal. The assumptions upon which element-set calculation is based require inertial motion, i.e., no thrusts of any kind, just natural forces. When a satellite maneuvers in some way, and pre-thrust and post-thrust observations are combined, the resulting element set is corrupt and should not be used. Even when there is no nearby maneuver, sometimes a few bad observations can corrupt an element set. These specifics support the general rule that we must perform some preprocessing whenever real data is processed, since we can be certain that occasional anomalies will surface.

Figure 14 shows a typical GEO satellite history of walk rate. The gradual slope down represents drift and the sharp jump up shows the station-keeping thrust to remain in the orbital slot. Note that, especially around the thrust times, outlier walk rates exist that come from corrupted element sets. It might appear that we could filter this function to obtain a better idea of walk rate, which is probably true, but that choice is not the best idea. Instead we need to mark outliers as bad and avoid using them.

The first step is to estimate some global properties of this satellite. Operational GEO satellites spend most

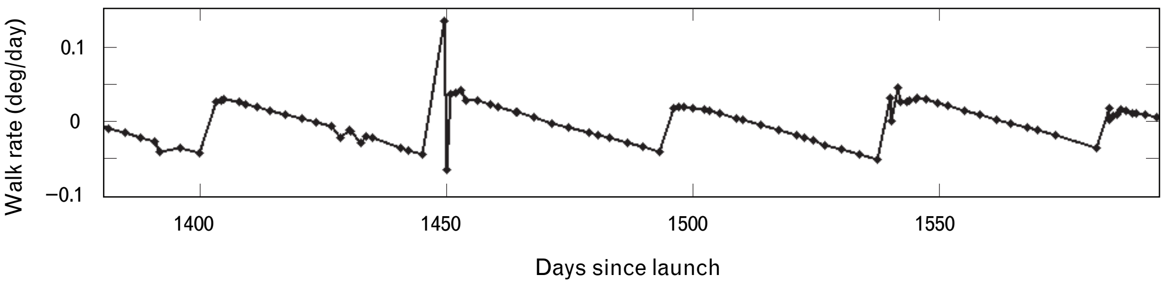


FIGURE 14. Typical GEO satellite history of walk rate. A perfectly positioned GEO satellite has a zero walk rate, maintaining a constant longitude with time. Satellites not located at an equilibrium point have an element-set-derived walk-rate history similar to this plot, in which the constant decrease in walk rate results from natural forces. The down slopes represent positional drift over time and the sharp periodic upward slopes represent a station-keeping thrust maneuver to maintain the satellite position within the orbital slot. The large jumps and glitches reveal corrupted element sets, which must be filtered out to obtain an effective monitoring system.

of their time in one slot or another, so we should be able to estimate the kind of station keeping they typically do by looking at the median absolute walk rates. Some satellites perform station keeping to a small fraction of a degree, say 0.1°, while others range over a full degree or more. Similarly, some satellites are stationed near the equilibrium points of the GEO belt, and experience little acceleration, so that the daily change in walk rate is small, while for others it is larger. Briefly, we estimate the typical change in absolute walk rate, the expected range of walk rates, and the expected absolute difference in walk rate from point to point. This series of steps enables us to identify probable outliers and effectively remove them.

Although this approach is generally simple and effective, there are several cases in which it fails. First, sometimes a satellite is moving or drifting; the algorithm above works best when the satellite is almost stationary. There are times in which a short period of motion to a new slot generates a group of element sets, or elsets, that may look like an outlier. We detect these special cases with special code. For example, LAN is basically the integral of walk rate, so an outlier in walk rate may be a bad elset or it could be a short intense thrust that moves the satellite to a new slot, perhaps close by the old one. A short deviation from normal walk rates thus can be validated by a longitude change. In the case of LAN, it is pretty hard to imagine a short excursion to a different longitude followed by a return to the same longitude. We use this fact to detect bad LAN values and obtain a validated LAN function suitable for detecting short slot moves.

Thrust Detection Algorithm

After accomplishing this important preprocessing, we turn our attention to thrust detection. We define four types of thrusts: *normal station keeping*, *from slot*, *to slot*, and *other*. Before we identify the type of thrust, however, we need to detect thrusts.

We know that small absolute drift rates, say less than 0.15° per day, are typical of in-slot behavior. Rapid changes in drift rate, or values outside this range, suggest that a thrust may have occurred. However, drifters will exceed this value periodically without any thrust required. Element-set propagation shows us if the change in longitude observed is consistent with drift.

With this approach we can develop a candidate list of elsets straddling apparent thrusts, but unfortunately we sometimes have multiple candidate thrusts that represent only a single actual thrust. This confusion occurs because elsets tend to be adversely affected around thrust times. Even with effective bad elset detection, some adjacent thrusts need to be merged.

The logic for the merger is fairly simple. First, *other* types of thrusts should not be merged because frequent thrusts are expected when a satellite is moving to orbit. For the remaining cases, if the time span between the elsets straddling the candidate thrusts is less than two days, we merge them. Alternatively, if the three elsets straddling the two thrusts are consecutive and the total elapsed time is less than seven days, we merge them.

Further culling has been found to be necessary. Since in most cases GEO station-keeping walk rates resemble the graph of Figure 14, we can determine the typical sign of walk-rate change between thrusts to figure out if there is increasing or decreasing walk rate at this point in the GEO belt. In most cases this rate can be determined analytically, but in some cases unusual station keeping near the equilibrium points might confuse us. So we go with what the data are telling us. A legitimate thrust must be in the reverse direction to the drift, forming the sawtooth of Figure 14. If an apparent station-keeping thrust is not consistent with this expectation, we may have a spurious thrust.

This incorrect direction is not enough evidence by itself to delete a thrust; we rely on propagation for that. We propagate both straddling elsets forward to a time two or more weeks ahead of the older elset. If their longitude is close, and a fraction of the total longitude change, then the thrust is determined to be spurious and deleted.

Now that we have a list of thrusts, we need to convert them to evidence suitable for input to the Bayesian network. What we have is a pair of elsets that are believed to straddle a thrust of the specified type. What we would like is to identify the time of the thrust itself. Depending on the elset spacing, that additional information may or may not be highly helpful in evidence combination, but we would like to obtain it if possible.

In theory, this information should be relatively easy to obtain if a thrust can be modeled as an impulse, or a very short burn. By propagating the two elsets toward each other in time, we can measure the distance between the two satellite positions at each time. A minimum distance should reveal the approximate time and place of the thrust. Atypical thrusts or elsets that are bad in some way or poorly selected will be problematic for this approach.

In practice, we observe that this procedure seems to work well less than half the time, in the sense of yielding a well-defined minimum at a time within the spanning elsets. In the other cases, the minimum often occurs outside the period of the spanning elsets or the minimum is just not very small. Initially, we restrict the propagation times to the interval between elsets; as a consequence many failures occur at one of the endpoint times. This complication results in an error that depends on the spacing between elsets, but is usually small enough that the problem is not critical. Another issue with this approach is that the computation time can be significant when elsets separated by significant time are propagated.

Element-Set Evidence Generation

We now have a list of thrusts, along with their estimated times and types. We need to convert these to individual pieces of evidence to be consumed by the GEO Bayesian network. Table 1 summarizes the different types of evidence that we need to produce.

The thrusts themselves could be mapped easily to some of the evidence types, but between thrusts we may need to output some intermediate evidence, for a variety of reasons. Among these reasons are the following. (1) A short move from one slot to another can be accomplished by allowing a satellite to drift to the new slot. There is no associated thrust but we want to identify this status change. (2) Thrusts may be separated by significant time, such as one or two months or more, so we may need more timely evidence of the metric state. (3) A satellite may have started drifting since its last station-keeping thrust.

To perform these functions, we maintain an estimate of slot width when we examine station keeping for this satellite. If the satellite's position exceeds this slot-width limit by more than 30%, evidence is output of drifting. The associated confidence depends on how much the LAN exceeds the limit. If the satellite stays within the slot, evidence is output of "in slot" on roughly a weekly basis.

If the satellite is in a drifting or moving state, evidence is output periodically, including either a confidence if drifting or a walk rate if moving. A simple test for drifting is to see how much the walk rate varies. A moving state or graveyard orbit usually has a constant walk rate, while a drifting state shows positive and negative rates, as well as rates that go through zero. Of course at the start, drifting does not show all these rates.

Deriving Bayesian Network Evidence Vectors

We have now converted element-set information into logical evidence that contains the important information relating to satellite status, is associated with discrete times rather than time intervals, and occurs at an appropriate rate to feed data to the Bayesian network. In order to process that evidence it must be converted into vectors suitable for input to the Bayesian network. Two Bayesian network nodes are pertinent: a metric evidence node with four states—nominal in-slot (NOM), moving (MOV), drifting (DRFT), and graveyard (GYRD) and a stability evidence node with two states—stable and unstable. Each metric evidence input is converted into a four-vector representing belief in these four states. Some of the metric evidence inputs are also converted into a two-vector for the stability node.

We first outline the derivation of the metric four-vector. Given station-keeping or to-slot evidence, the probability of NOM is set to be a thousand times greater than MOV, DRFT, or GYRD. Given in-slot evidence, which comes with a confidence, this confidence is used for NOM, the number one minus this confidence is used for DRFT, and the other states are assigned probabilities of 0.001. Drift evidence is handled similarly to in-slot evidence, except that the confidence provided is assigned to DRFT and the number one minus this confidence is assigned to NOM. It is not necessary that our input evidence numbers sum to 1.0; it is the ratios of these values that matter, not the absolute values.

The most complexity arises from the cases in which the evidence supplied is of any of the types "from slot," "other," or "moving." Here we are given a walk rate, and from that we determine appropriate confidences for the NOM, DRFT, and GYRD states. The same algorithm is used for each of these evidence types, and is based on nominal walk rates.

Metric evidence associated with a thrust also must be converted to a two-vector for the stability node, which assumes the states "stable" and "unstable." The satellite may or may not be nominal in terms of configuration and operation, but it will not execute a thrust unless the operators are confident the satellite is stable. If the evidence indicates a normal station-keeping maneuver, then we use 0.999 for the stability confidence and 0.001 for the unstable confidence. Those numbers are 0.99 and 0.01 for "to slot" and "from slot" evidence, and 0.95 and 0.05 for "other" thrusts. Normal station keeping is strong evidence that everything is fine, but we are a little less confident that everything is nominal immediately after a thrust of a different type.

These numeric values may seem heuristic, but the real power of a Bayesian network comes from the network structure and the many conditional probabilities within the Bayesian network. As long as the input evidence numbers are reasonable, we obtain good performance. High-precision probability estimates are not required.

Photometric Signature Processing

We obtain information about satellite status from both radar and photometric signatures because signature information alone can sometimes distinguish a nominal satellite from an anomalous satellite. We would like to extract this information in order to combine it with our other sources of information and arrive at the best estimate of satellite status.

There are two main approaches to the problem. The most direct approach is to predict the signature and then compare the received signature to the prediction. If we are able to do this comparison, then any signature can be assessed, regardless of the conditions under which it was collected. We simply run a satellite simulation that is appropriate to the type of signature as well as the collection geometry. This procedure sounds straightforward, but in practice there are several major difficulties with this approach. The most obvious is that we may not know the exact configuration, the exact materials, and the exact modes of operation of the satellite when it is operating nominally. The use of guesses or approximations in this situation will result in an inaccurate prediction.

A second problem concerns the algorithms used for predicting signatures. The most accurate known methods of prediction are often computationally demanding. Ray tracing, for example, is a good method of predicting photometric signatures, but including enough rays, enough bounces, and an accurate enough model of the interaction of light with the various materials is very computationally expensive. Similarly, the method of moments is an effective method to predict radar cross section, but in the common case of satellites much larger than the radar wavelength the computation is prohibitive. So approximate high-frequency methods are generally used in this case.

These issues have motivated us to develop a second method of signature assessment based on signature comparison. The concept is that a historic database of signatures is stored, and new signatures are compared to selected historic signatures. The resulting distances are processed to estimate a probability that the signature is consistent with a nominal satellite of that type. This method can also be used as a classifier to determine which of several types of satellite best matches the new signature, or to test certain other hypotheses.

Existing classifiers might at first look seem applicable to our problem, such as the nearest-neighbor method or artificial neural networks. However, some of our signature vectors are of different dimensions, or consist of as little as a single number, rendering such methods in-

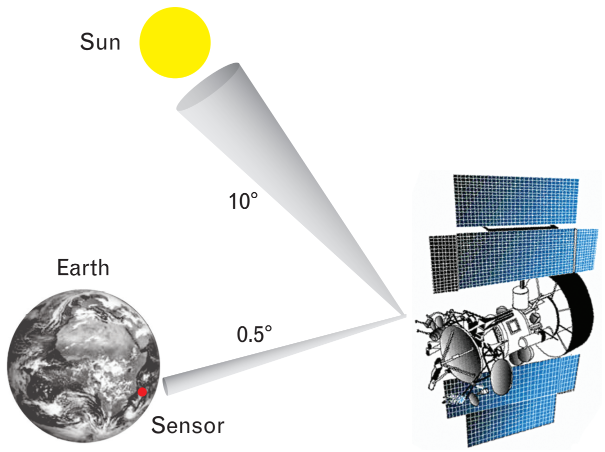


FIGURE 15. Photometric data parameters. Two signatures should show comparable photometric intensity if they are taken on similar satellites in similar configurations, and both the illumination angles and observation angles are similar. By using a coordinate system centered at the satellite being observed, and assuming some continuity of photometric intensity, we can define two small cones based on sun and sensor azimuth and elevation within which signatures may be compared.

applicable, ineffective, or both. These limitations have motivated the development of the hypothesis-testing method described below.

The signature comparison part of the method is data dependent, after which data-independent algorithms process the calculated distances to obtain the final assessment. The first step is to consider the phenomenology of the signature and identify all parameters that significantly affect the signature. The purpose of this step is to enable us to specify which historic signatures are expected to be comparable to each new signature. The second step is to define a distance measure reflecting the similarity of the selected historic signatures to the new signature. After the numeric distances are defined, the rest of the algorithm is common to all data types.

Photometric Signature Comparison

Most of the main GEO belt sensors are photometric, since few radars can track objects at that range. Even those which can perform GEO tracking have difficulty searching large volumes of the belt, a task better suited to photometric sensors using the sun for illumination. Let us consider the situation from the point of view of a target satellite. Two angles—called the sun angles—determine the position of the sun. Two other angles—called the sensor angles—determine the position of the sensor. We define two small cones within which the angles are considered close enough for associated signatures to match in some sense, as shown in Figure 15. We call the set of historic signatures contained within this region the *cohort* of the new signature.

Almost all GEO satellites are either three-axis stable with solar panels that rotate once a day to remain pointed at the sun, or spinners with solar panels tiling a spinning cylindrical drum. Most are communications satellites that can reorient their communications antennas, but rarely do. It follows that if a satellite is operating nominally, then two signatures taken at different times but with the same four angles should be very similar. The configuration of the three-axis stable object might be thought to present a problem, but the solar-panel position is a generally a function of the sun azimuth, which is one of the four angles. If the sun azimuth matches, then the expected solar-panel position is the same, so we have the required conditions in which the configuration, illumination, and viewing geometry are all the same.

Now, as just mentioned, we define two small cones in parameter space in which the observing conditions are close enough to produce similar signatures. The argument is that in the four-parameter space we have variation that is continuous or at least relatively smooth over a small region. It is difficult to plot four-dimensional data by using two- or three-dimensional displays, but Figure 16 shows a graph of signature brightnesses in which the two sensor angles are limited to a fixed range and the sun angles are allowed to vary. In this graph, each colored symbol represents a Space-Based Visible satellite signature, and the signature brightness is encoded in both color and size. Although Space-Based Visible detections have multiple associated brightness values, the displayed values represent an average over a short interval of time such as twenty seconds. The small purple symbols are the dimmest, and the large red symbols are the brightest. The small ellipse defines the cohort region of the signature number 703307, which is the only signature in the graph represented by a diamond. The slice of sensor-angle space selected for this graph is slightly larger than the slice that would be in the cohort of 703307 to enable the analyst to look around the corner, so to speak, and see if any helpful data just missed being included in the cohort. Those historic signatures actually in the cohort of 703307 are represented by circles. All other signatures not belonging to the cohort

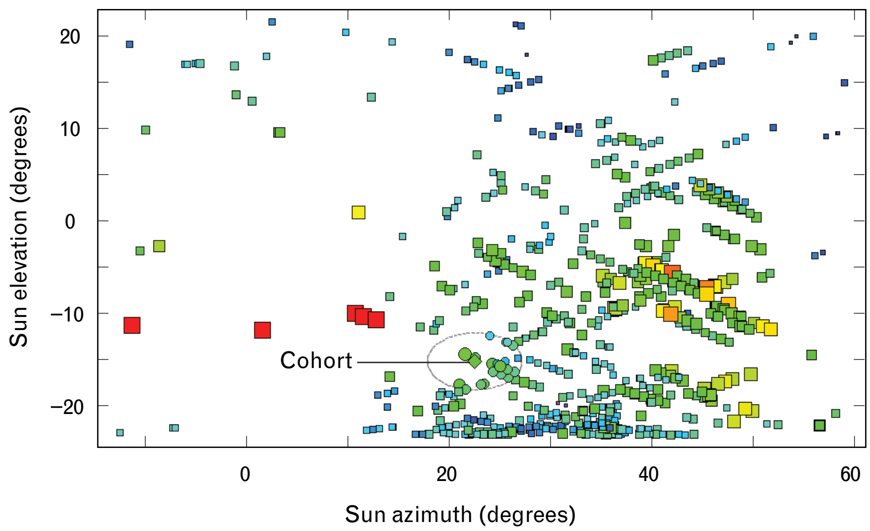


FIGURE 16. Visualizing signatures in parameter space. A photometric signature is assessed by comparing it to historic signatures from nominal satellites of the same type. Signatures chosen for comparison (the cohort) must inhabit the same region of a four-dimensional parameter space. In this example a fixed slice is taken in sensor azimuth and sensor elevation and the signatures are graphed as functions of sun azimuth and sun elevation. Brightness is encoded in both size and color. The small ellipse contains the cohort of its central signature, plotted as a diamond.

are represented by squares. There seems to be reasonable continuity in brightness in this parameter space; nearby signatures are quite likely to be similar in brightness.

After having defined the cohort, the next step is to define some distance measure that reflects the difference between signatures. The obvious choice is simply the absolute value of the difference in average magnitude between two signatures. This simple distance measure has proven effective.

A few isolated exceptions look unusually bright or dim; such outliers are present to some extent in most real data. Several unusually bright squares are somewhat isolated but do suggest more rapid variation in their immediate vicinity. Also, the orange squares at right center appear to show larger variation than normal. Both of these regions show specular effects, in which some flat structure on the satellite is reflecting the sun like a mirror. It is necessary to handle these specular regions differently, since their extremely rapid variation in brightness is inconsistent with the assumption that we can determine a small region of parameter space within which signature brightness is essentially constant.

Because many GEO satellites are three-axis stable types with large solar panels, we expect that speculars from such panels are very bright, when visible. We define our coordinate system for GEO satellites to have its origin at the center of mass of the satellite and to have x pointing toward the earth center, y pointing in the opposite direction to the velocity vector (west), and z pointing north. Our four angles are defined with respect to this coordinate system. When the sun vector is projected onto the x-y plane, the sun azimuth θ_{sun} is the number of degrees counterclockwise from the x-axis of the projection. The sun elevation ϕ_{sun} is the number of degrees the projection needs to be rotated counterclockwise to point to the original sun vector. The exact same definition holds for the sensor azimuth θ_{sensor} and elevation ϕ_{sensor} (these are similar to ordinary spherical coordinate angles, except that the typical coordinate system ϕ must be subtracted from 90° to obtain our elevations).

The specular condition occurs when the angles of incidence equal the angles of reflection. If we ignore details of the bidirectional reflectance function, we can create a graph with axes representing deviation from the solar-panel specular condition in both azimuth and elevation. Assuming that the panel tracks the sun in azimuth but not in elevation, which is common for GEO satellites, our coordinates become $x_{dev} = \theta_{sun} - \theta_{sensor}$ and $y_{dev} = \phi_{sun} - \phi_{sensor}$. In this coordinate system, we can produce a histogram of the available data for nominal satellites of this type, as shown at the left in Figure 17.

In this two-dimensional histogram, white bins with a dash contain no data. The colors in the other bins encode the average brightness by using a color scheme similar to that of Figure 16. The large bold numbers indicate that three or more samples were used to compute the average. Large normal numbers indicate that only two samples were used, and the small numbers indicate that only a single sample was available. These magnitudes are not the original magnitudes actually measured by the sensor, but rather the magnitudes that would have been observed at a range of a thousand kilometers. This magnitude normalization is analogous to the radar community's radar cross section, in which the measurement serves as a measure of the target's size and shape without regard to measurement range.

It seems that the specular region is approximately elliptical, and extends roughly 12° in azimuth by 8° in elevation. There also appears to be an offset θ_{off} of several degrees, which can be explained by solar panels that are not exactly tracking the sun. This mode of operation can occur if the panels' output is more than required, as with a new panel designed for a long lifetime with expected future degradation. If we avoid pointing the solar panels directly at the sun we can reduce panel temperatures as well as limit possible battery overcharging problems.

In this specular region our argument is that the signature brightness can change rapidly, so comparisons to nearby signatures may be ineffective in assessing satellite status. This graph also suggests yet another coordinate system, a normalized one in which zero is at the center of the ellipse and the value of one is at the edge of the ellipse. To accomplish this normalization, we map the ellipse to the unit circle with the semi-major axis a (about 6°) and the semi-minor axis b (about 4°), and we define the normalized specular distance as

$$d = \sqrt{\frac{x_{dev} - \theta_{off}}{a} + \frac{y_{dev}}{b}}\ .$$

The right side of Figure 17 shows a graph of Space-Based Visible signatures in which photometric magnitude is plotted as a function of the normalized specular distance d. The blue squares represent signatures collected on active satellites, with periodic station keeping to keep themselves in their slot. These satellites are presumably operating normally, with solar panels tracking the sun. The red squares are signatures collected on old, dead, drifting satellites, presumably not even stable, or at least not tracking the sun. The y-axis scale represents the number of magnitudes brighter than the median magnitude for this class. Note that the red squares are pretty close to the median, while the blue squares follow roughly the region between the lines shown. There are very few outliers in each case until we exceed the bounds of the original ellipse (i.e., when $d > 1$).

To assess signatures in this region, we simply measure their deviation from the expected region between the lines. This measure is a simple model of expected photometric brightness in the specular region, and defines a region of parameter space in which signature comparison is not used. The result is an effective method of assessing specular signatures, which can be seen by noting the separation between the blue and red squares.

Single Distance Statistics

We now return to the question of the non-specular signatures. We have a new signature and selected nominal historic signatures (the cohort), which are expected to be comparable to the new signature by virtue of their close similarity in the four angles defining the parameter space. For each cohort signature a distance can be computed, which in this case is just the absolute difference in photometric magnitude. We first consider the statis-

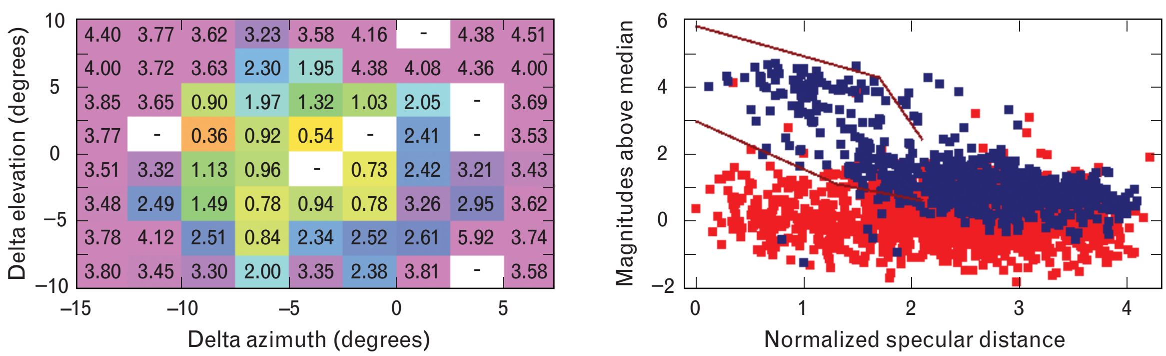


FIGURE 17. Solar-panel specular analysis. The assumption of photometric continuity is violated in specular regions, such as where large rotating solar panels reflect the sun directly to the sensor. (a) We graph the brightness of nominal signatures as a function of offset from this specular condition in both azimuth and elevation. (b) Changing coordinates so that both angles are combined into a distance from the specular condition enables us to graph magnitude of signatures from both nominal (blue) and failed (red) objects. The specular condition is observed in almost all of the nominal signatures but in few of the anomalous signatures, as expected.

tics of a single distance, and then we examine methods of combining the evidence from the multiple available distances.

Given this single distance, the question is whether the satellite is nominal (hypothesis H_1) or anomalous in some way (H_0). Bayesian hypothesis testing theory is applicable [14] by using the probability density functions (PDF) $p(R|H_i)$, where R is the measurement—in this case a distance. Given *a priori* probabilities and a cost function, a weighted ratio of these PDFs produces a likelihood ratio, defining a method for choosing between the two hypotheses. In this two-hypothesis case, we assume equal *a priori* probabilities, and a trivial cost function weighting any type of error equally. Then the solution reduces to a simple ratio of the PDFs.

What remains is to estimate the PDFs, one of which represents the expected distances when two nominal signatures are compared (H_1), while the other represents the expected distances when one anomalous signature is compared to one nominal signature. We base these estimates on histograms of sample distances. To obtain the sample distances we need an *a priori* status for the satellites in question over the total data collection time span. Generally this status determination is easy, since nominal GEO satellites generally stay in their slots, while anomalous satellites generally either drift out of their slots or possibly are retired to graveyard orbits if sufficient control is retained by the operators.

There are many exceptions, of course. A satellite might lose attitude control while in a slot, and then regain it before drifting out of the slot. This situation would result in anomalous data that appear to belong metrically to a nominal interval. Moving satellites represent a bigger problem. A satellite moving east or slowly moving west is probably moving to a new slot, and may appear nominal during the move. A satellite moving west rapidly may be in the same situation, or it may have been boosted into a graveyard orbit and abandoned. The latter case is easy to detect once the satellite circles the globe.

These exceptions do not represent a big problem for determining our single-distance statistics, however. A sensor like SBV sweeps up large chunks of data over long time periods, and all we need is one set of data in which we are sure a satellite (or satellites) is nominal, and another set in which we are sure it is anomalous. We can ignore ambiguous data. To demonstrate the calculations, we need a GEO satellite class that has plenty of SBV data on both active and inactive satellites. The Russian Gorizont satisfies these requirements and is used in our examples.

Figure 18(a) shows histograms of Gorizont distances for H_0 and H_1. Figure 18(b) is an estimated single-distance log-likelihood ratio that is the logarithm of the ratio of the histogram values. This log-likelihood ratio gives us an assessment based on a single distance. In the case in which only a single distance is available, this single measure represents our assessment, but the case in which many distances are available is more important as well as more common. For convenience we normally

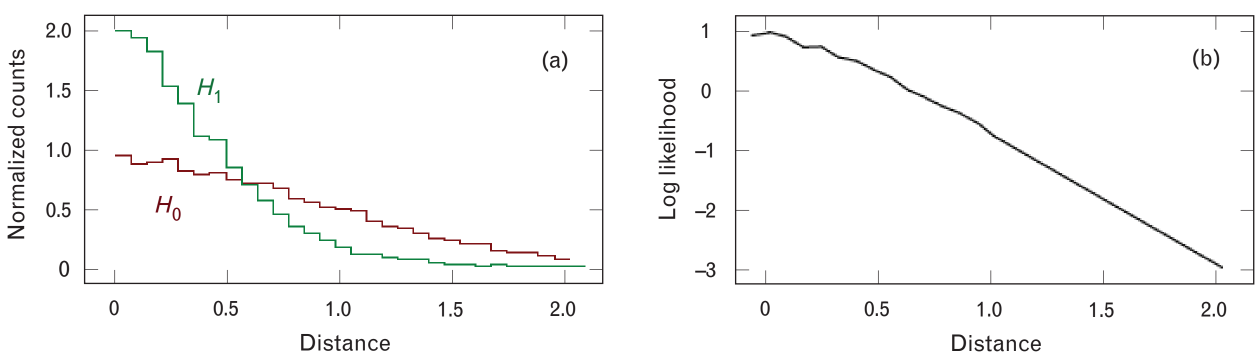


FIGURE 18. Single-distance histograms and log-likelihood ratio. (a) If we compare nominal signatures to nominal historic signatures we get the green histogram curve H_1, which is normalized to approximate a probability density function. Similarly, if we compare anomalous signatures to nominal signatures we get the red normalized histogram curve H_0. (b) The log-likelihood ratio estimate of these histograms can be used to determine the probability that a single unknown signature belongs to the nominal class. It is convenient to use the log-likelihood ratio, since exponential falloff is represented by a straight line, which may be reasonably extrapolated beyond the limits of the existing data.

work with log likelihoods, where ratios become differences, products become sums, and straight lines represent exponential change.

The histograms in Figure 18(a) are normalized to have unity area, since they are being used to approximate PDFs. The log-likelihood ratio function is just the log of the ratio of the histogram curves for distances between about zero and 1.0. At higher distances the data become sparse and the estimates fluctuate, so we have fit a line through the actual points to give a smoother estimate. The line, which is a good fit to the data, represents exponential drop-off because we are in log space.

Note that the log-likelihood ratio λ in Figure 18(b) crosses zero at a distance of just over 0.5. This means that distances less than about 0.5 are evidence for a satellite being nominal, while distances greater than this value are evidence of an anomaly. The exact crossover point depends on the size of the cohort region. A large cohort region means we are comparing more data taken under slightly different conditions, so a nominal satellite will show larger distances on the average. In the limit of very small cohort regions, the brightness repeatability of the actual sensor and the accuracy of the sensor will be the limiting factors.

Sometimes the SBV sensor scans down the GEO belt, collecting data on overlapping fields of view. In this case, we sometimes have two collections on a satellite separated by a short time period, typically around twenty seconds. If the satellite is stable, the observed magnitudes are generally quite similar. If the satellite is tumbling, the magnitudes may be more different. We would like to be able to use this evidence for monitoring purposes, because sometimes no other data are available for comparison, but there is a real problem with doing that.

Recall that we never compare a new signature to a signature from an anomalous satellite. It would be difficult to determine the status of the new signature when comparing to a relatively random quantity like the brightness of an anomalous satellite. But if two signatures are only twenty seconds apart, for example, then we are really trying to compare either two nominal signatures or two anomalous signatures. The solution is to estimate another log-likelihood function λ_p, defining H_1 to mean both signatures are from a stable satellite, and H_0 to mean both from an unstable satellite. When we do that, we obtain a λ_p similar to the λ of Figure 18, but crossing zero around 0.25 instead of near 0.5.

One subtle point is that we are really producing evidence for a slightly different hypothesis with these pair statistics. Instead of evidence for nominal or anomalous, we are really providing evidence for stable or unstable. An anomalous object that happens to be very slowly tumbling, or is stable in an anomalous attitude or configuration, will appear to be stable. The majority of anomalous GEO satellites suffer some loss of stability, but this loss is not guaranteed. The option to omit the pair distances from our assessments facilitates the detection of unusual anomalies, but generally the pair processing improves system performance.

Multiple-Distance Combination

If the distances developed in the previous section were independent, then the individual log likelihoods could simply be added to obtain the overall log likelihoods of each hypothesis, and the differences of these summed log likelihoods would be our desired assessments. It is easy to see that this independence is not the case, since every distance depends on the same new signature. So in reality there exist a number of correlated distances, each of which represents evidence that a given signature matches a signature in the database of nominal signatures. The question is whether to accept H_1, object nominal, or H_0, anomalous.

Consider the following two-step process for determining an overall log likelihood, given n multiple correlated distances [15]. First we compute the n individual single-distance log likelihoods and sum them as if they represented independent statistics. The resulting quantity S_n will overestimate the total log likelihood, but will be corrected by multiplying by a factor K_n less than unity and depending on n. The overall log likelihood Λ_n becomes $\Lambda_n = K_n S_n$. With this approach, all the distances are used and their relative contributions are equally weighted. Varying n can be handled, as long as K_n is computed for each n.

In the single-distance case, we use a non-parametric approach to estimate the likelihood functions. No *a priori* functional forms are used or even hypothesized. Taking this approach further, we can fix n, and then analyze the statistic S_n in the same way the individual distance statistics are analyzed.

We need to obtain a large number of S_n samples representing the two hypotheses. To do this, we use the same intra-database distances discussed above in the estimation of the single-distance likelihoods. However, we now take them in groups of n so that we obtain multiple S_n estimates. For example, if a certain signature is compared to twelve other signatures in the database generating twelve distances, for $n = 5$ the first five are used to make one S_5 sample, the second five are used to make another S_5 sample, and the last two are discarded.

This procedure can be used for the distances representing both H_0 and H_1, whereupon the PDFs are estimated by using histograms, and then a likelihood ratio Λ_n is formed as before. Previously, by doing this estimation for fifteen multiple values of n, we developed an analytic expression for K_n [15]. With all data sets analyzed, we found an approximate linear relationship between $\log n$ and $\log K_n$, $\log K_n = m \log n + b$. This procedure reduces our problem to finding appropriate values for the correlation parameters m and b.

It is possible to laboriously combine distances in groups of n for many values of n, calculating resulting ratios and then estimating the values of m and b that best fit the data. It is important to realize that the merit of the system is not reflected in how well m and b match the estimates derived from the original data set; the original data set is not beyond reproach. The real test is of the quality of the assessments generated by the system. The advantage of an analytic model is best illustrated when we optimize the parameters of that model, as discussed in the next section.

Parameter Optimization

We can take a number of new signatures from active and dead satellites and assess them, obtaining two groups of confidences that each signature represents a nominal satellite. We can then form a histogram from each group, as shown in Figure 19. In Figure 19, the satellite class is the Gorizont class, and the correlation parameters are –0.704 and 0.488. From the histogram it is easy to create a numeric merit related to how much information is extracted when a signature is assessed. For confidences near the middle of the graph, we can use the log likelihoods for the good signatures (active satellites); this measure will be positive if the confidence exceeds 50% and negative otherwise. We can subtract the log likelihoods for the bad signatures (dead satellites), since that will reward low confidences, which we desire in this case.

For large confidences, such as those over 0.99, we invoke the law of diminishing returns and scale back the merit contribution so that it is only slightly better to have a confidence of 0.99999 than 0.99. Finally, we set false-alarm penalties. Clearly, we should not have more than 1% of the bad signatures assessed as nominal with 0.99 confidence or better. Nor should we have more than 9% in the range of 0.9 to 0.99. Similarly, we should not have more than 1% of the good signa-

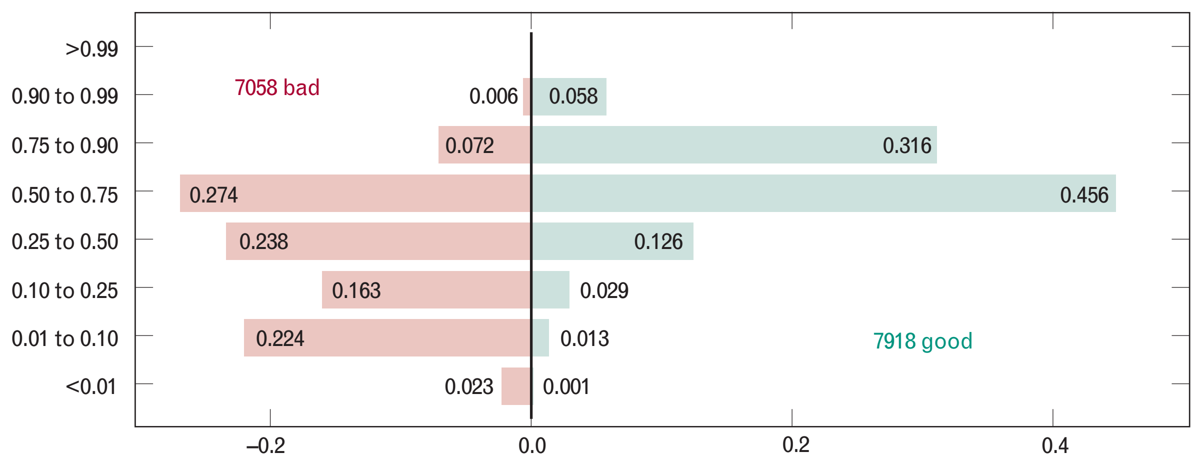


FIGURE 19. Confidence histogram for Type 3 with merit 1.44 (0.73 + 0.71). Given a large collection of signatures of a given type, including both nominal and anomalous satellite states, we can assess each signature and then form a histogram of the resulting assessments, which are confidences that the satellite is nominal. The nominal signatures are shown in green at right and the anomalous signatures in red at left. Good performance is reflected in high green assessments and low red assessments. An overall merit value derived from this graph is useful in optimization of algorithm parameters.

tures assessed as 0.01 or lower. A substantial penalty is assessed whenever these false-alarm limits are exceeded. The header at the top of Figure 19 shows a typical result in which the merit associated with both classes of signatures is reported as well as the total.

By combining these factors into an overall merit we have a number that can be used for optimization. Not only can this number guide the adjustment of the correlation correction coefficients *m* and *b*, but it can be used to optimize almost any parameter. For example, the sizes of the cohort regions can be adjusted iteratively with the merit guiding the search. Large changes in the cohort regions will necessitate a recalculation of the likelihood functions. This recalculation can be incorporated into the iteration if the likelihood function estimation is automated.

Radar Signature Processing

Although photometric signatures are the most plentiful and practical to use for GEO monitoring, signatures from those few radars which are able to track GEO satellites can assist. To add a new source of data we need first to determine the parameters that define the cohort, and second to define a suitable distance that captures signature similarity in a fashion useful for discriminating nominal from anomalous satellites. After that, the procedures for estimating likelihood functions and combining multiple correlated distances are the same as above. We briefly present an example of a radar signature and explain its parameter space and distance definitions.

The Millstone radar is a high-power coherent L-band sensor used mainly for deep-space tracking. It uses circular polarization on transmit, and receives both the principal and orthogonal polarizations (PP and OP). The sample rate is variable, and depends on the coherent processing done by the real-time processor when tracking. Typically we see samples on the order of every few seconds, but they can easily vary from several per second to one per minute. The length of tracks is also variable, from just a few seconds to the better part of an hour.

When we view GEO satellites from a fixed ground radar, the relevant parameter space depends to some extent on the satellite itself. As in the photometric case, we use coordinate systems centered on the satellite to simplify processing. The two angles from the satellite to the radar (azimuth and elevation) are necessary parameters. If the satellite is spin stabilized with a cylindrical drum of solar cells, then those two viewing angles are sufficient to define our parameter space. If the satellite is a three-axis stable type, then the flat solar panels are rotating to follow the sun, so the sun azimuth forms the third angle defining the cohort space. This angle is really a proxy for the configuration of the satellite, which clearly affects the signature.

From basic physics, we expect that the radar cross section is nearly constant for a stable GEO satellite. The relative radar cross-section values of the PP and OP components should depend on the physical configuration and orientation of the satellite. So differences in PP and OP radar cross section are logical components of a distance (i.e., measure of the difference) between two signatures. Millstone operators also have a tradition of using the PP/OP ratio as one discriminant to identify GEO objects.

The left-side graphs in Figure 20 show two signatures from a GEO object, taken on different days but within two hours of the same time of day. The radar cross section is quite consistent, and the OP radar cross section is greater than the PP radar cross section. The right-side graphs show two signatures from another GEO object of a different type. The signatures are taken about six hours apart, and although the PP value is similar, the OP value is different. Since both of these objects are three-axis stable, we would expect different signatures at different times of day. Because the two-hour difference is small in the left-side graph tracks, there isn't much change in the signatures. We expect the signatures to be similar from one day to the next, so the fact that the signatures are taken on different days should not prevent them from matching.

Work with other GEO data suggests the following definition: the provisional cohort of a new signature is any signature whose sensor azimuth and elevation are within a few tenths of a degree of the new signature and whose sun azimuth is within a specified angle such as 15°. The former requirement is likely to be satisfied for any signature in the same longitude slot as the new signature when observed from a fixed ground sensor; the latter requirement translates to a one-hour time-of-day difference.

Now that we have defined the provisional cohort, we need to define the distance measure. We retain the logarithmic units of dBsm for all our radar data, because the smoothing effect generally improves performance. There are several important components to our distance measure, including the difference in mean PP radar cross

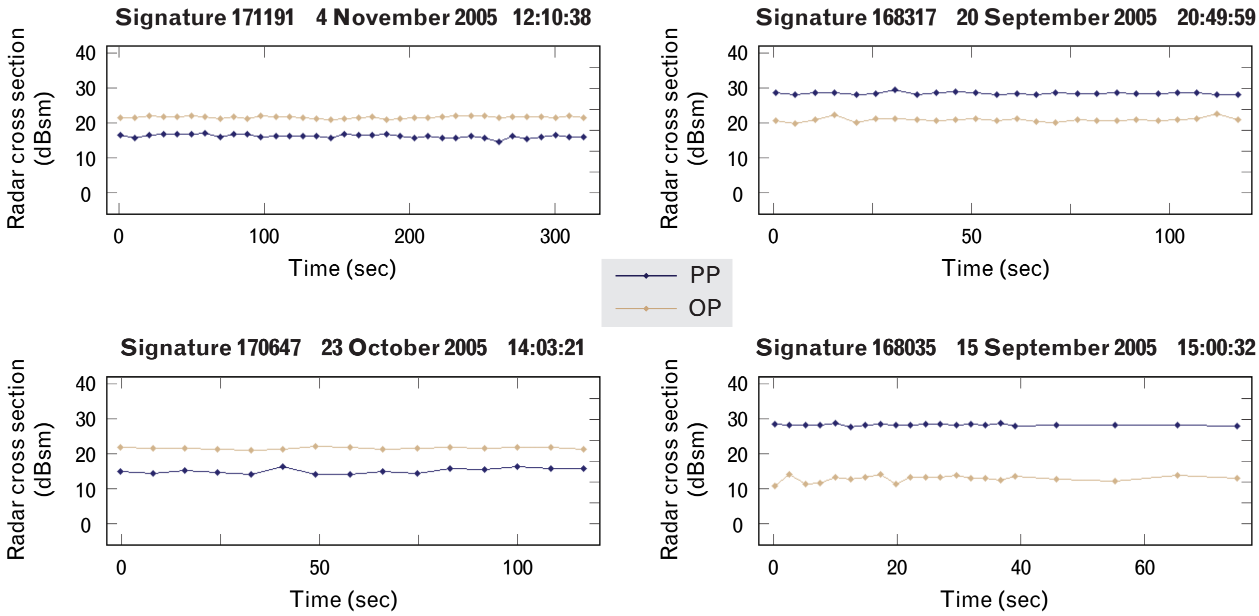


FIGURE 20. Principal polarization (PP) and orthogonal polarization (OP) signatures. Dual polarization radar signatures of GEO satellites, when available, can provide useful information about status. At left are two pairs of signatures from a GEO object taken within two hours, so the solar-panel azimuth difference would be expected to be less than 30°. At right are two pairs of signatures from another GEO object taken six hours apart, where the solar-panel azimuth difference would be 90°. There is much greater similarity in the left pair, which suggests that comparison is feasible as long as the configuration of the satellite is similar in both signatures.

section, the difference in mean OP radar cross section, and the difference in PP/OP ratio. Our basic distance measure is based on a sum of these three differences. However, in some cases one of the channels is weak and may drop out altogether, causing big jumps in our logarithmic scale. In this case, the relevant channel is discounted or even dropped altogether from the distance measure. The scaling is adjusted so that the numeric variation is the same as if all three terms were used.

Also, variation has been found to be useful in determining satellite status. Greater variation is indicative of loss of stability, an anomaly we wish to detect. Under certain conditions, however, solar-panel speculars may cause rapid variation even though the satellite is operating nominally. To prevent specular-related variation from adversely affecting our distance measure we use an adaptive algorithm to estimate the normal variation found in the nominal signatures of the cohort. Only variation significantly in excess of the normal variation results in significant increases in distance.

The resulting distance measure is complicated to write down, since there are conditional calculations and an adaptive variation estimate, but it is straightforward in concept. The resulting merits generally exceed the merits for the SBV photometric signature case. This result is not surprising, since more information is captured in the separate PP and OP time histories than in a single SBV intensity.

Bayesian Network Information Fusion

We now have evidence derived both from element sets and signatures, and we need to combine this evidence to produce an overall status. We perform information fusion by using a Bayesian network, which is described in the sidebar "Bayesian Networks" (page 328). Bayesian networks fit our problem nicely. Some critics of the Bayesian network approach have pointed out that these networks require probabilistic inputs, which human experts have difficulty estimating. Fortunately, we have probabilistic signature assessments that are automatically derived. The Bayesian network approach is not good at estimating continuous parameters, since these need to be quantized into discrete states. But the status monitoring application has natural discrete states of interest.

Figure 21 illustrates the Bayesian network for a GEO satellite. The central blue nodes are the most important in the network. They represent the hypothesis that the satellite status is in any of the six states (NOM, ANOM, NMOV, AMOV, DRFT, and GYRD). As evidence comes into these nodes from two or more directions, the nodes update their belief in the six states. We create a status node at every time in which signature or metric evidence is available, and also at any additional time at which we want to know the status.

Similarly, the red node represents the hypothesis that the signature evidence depicts a satellite in one of the states NOM or ANOM. The purple node represents the hypothesis that the satellite is in the states STAB or UNST, based on elset analysis suggesting that the satellite thrusted at this time, which requires stability. The green nodes represent the hypothesis that the elset evidence depicts a metric state from the set (NOM, MOV, DRFT, and GYRD).

The small Bayesian network in Figure 21 shows the three cases: signature evidence at time t_0, metric evidence from the set {moving, drifting, in slot} (see Table 1) at time t_1, and metric evidence from the set {station keeping, to slot boost, from slot boost, other boost} at time t_2. As each type of evidence is processed, two or three nodes are added to the Bayesian network, and the new evidence is processed. The evidence arrives in arbitrary order, and the Bayesian network is constructed dynamically as this happens.

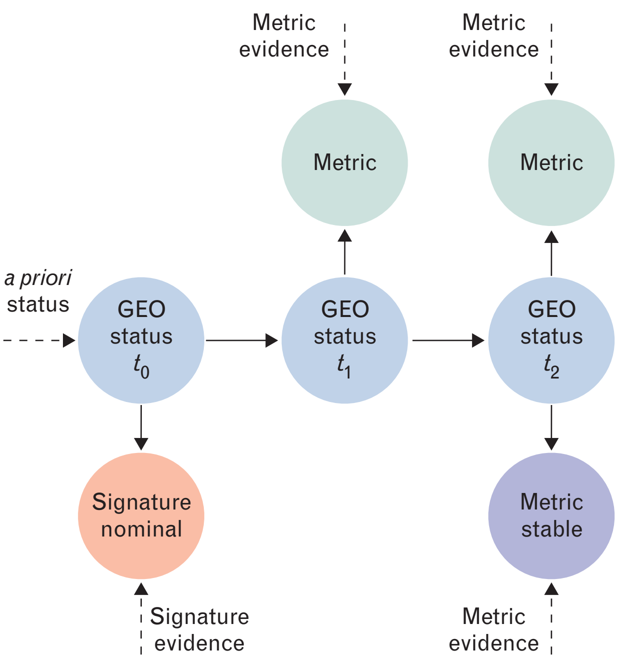


FIGURE 21. The GEO-satellite-monitoring Bayesian network consists of central status nodes that provide the information of most interest, plus peripheral data nodes that receive the incoming evidence. As more evidence is collected, the network grows, and the entire network is updated, providing a continuous assessment of status. The solid arrows show a causal relationship; e.g., the status at time t_0 is a cause of the status at time t_1, as well as a cause of the status of the signature collected at time t_0.

The solid arrows represent a causal relationship between one node and another. For example, we treat the GEO status at time t_0 as a cause of the status at time t_1, since the previous status will be highly correlated with the current status, especially if not much time has elapsed. Similarly, the true status of the satellite at time t_0 causes a signature collected at t_0 to be nominal (or not). Each such arrow has an associated conditional probability matrix relating the states of the child node to the states of the parent. In the case of Figure 21, we have 6 × 6 matrices connecting the status nodes, 6 × 4 matrices connecting the metric nodes to the status nodes, and 6 × 2 matrices relating the red and purple nodes to the status nodes. The information contained in the conditional probability matrices is a key factor in Bayesian network performance.

Conditional Probability Matrices

We can only sketch an outline of the construction of the conditional probability matrices, but that should be adequate to provide a basic understanding of the system. First, consider the GEO status conditional probability matrix between the nodes at times t_0 and t_1. We introduce the concept of a failure rate, which is the probability that a satellite will fail during a given time period, such as the current month. This failure rate clearly depends on the age of the satellite, the type of satellite, and the owner/operator. The rate is weighted by the time interval to obtain a probability of failure p_f. Similarly, we have a recovery rate, which is the rate at which a failed satellite will be recovered by the operator during a given time period, such as per day. This recovery rate depends on the same or similar factors as the failure rate. We also weight the recovery rate by the time interval to get a recovery probability p_r.

We can arrive at an estimate of the probability of moving toward a new slot, p_m, based on the history of similar objects. Similarly, we can arrive at the probability of stopping at the slot p_s based on the historic rates of GEO moves and typical longitude changes. A probability of graveyard boost p_g can also be estimated from the age of a satellite and the history of that class.

BAYESIAN NETWORKS

We define a Bayesian network to be a directed acyclic graph, as illustrated in Figure A, in which each node represents a variable, the directed arcs represent a causal relationship, and the strength of the causal relationship is contained in associated conditional probabilities [1]. When a Bayesian network is constructed to reflect the judgments of experts who determine the causal relationships and conditional probabilities, we call it a Bayesian belief network, or causal network. In such a network, each node is a hypothesis that can assume one of a finite set of states, such as (true, false) or (red, green, blue, unknown). The computed belief is a non-negative vector that sums to unity, where each component represents the probability or belief that the node is in the corresponding state. If a network has n states, and its single parent has m states, then the associated conditional probability matrix is n by m. If there are two parents, with m and l states, then the conditional probability matrix becomes n by $(m \times l)$.

Each directed arc must point to a variable that is "directly caused" by it; "indirect causes" or correlations follow logically from the structure of the network itself, and so come for free. Figure B, as an example, shows the causal relationships between sex with states (male, female), hair length (short, long) and height (≥5′6″, <5′6″) [2]. Hair length is not a direct cause of height; if we know the sex then the two variables can be considered independent. These variables are correlated, though, if sex is unknown, and that fact is automatically accounted for by the network structure.

The acyclic requirement is important; if A causes B and B causes C and C causes A, we have a feedback cycle that cannot be represented in a Bayesian network.

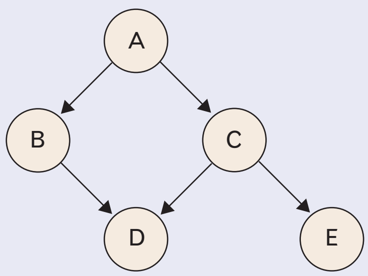


FIGURE A. Directed acyclic graph.

Loops may exist in the network, but their arrows must be converging or diverging at a node.

It can be shown by the chain rule that the complete joint probability distribution function for all the variables in a Bayesian network can be represented as a product of the conditional probability matrices [1]. This result goes hand in hand with the fact that a correctly updated Bayesian network provides the exact probabilities of all variables in the network, given the input evidence. All that is needed is an effective algorithm to update the network.

Updating Procedures

A tree-like Bayesian network that contains no cycles and also contains no loops of any kind may be efficiently updated by using local message passing. We define updating to mean processing a piece of external evidence such that all nodes of the network receive updated beliefs. In this procedure, each node receives prior information π as well as evidence λ. The π is mediated by the conditional probability matrix, and then mul-

Armed with these estimates, we can fill in the matrix a row at a time. The lower-probability events can be done first until there is only one left, which is assigned the remainder.

For example, in the first row, we are investigating transitions out of the NOM state. We expect a failure to take some time to manifest itself as drifting, so we assign 80% of the p_f to a transition to state ANOM and 20% of it to DRFT. Similarly, we expect a moving satellite to be nominal perhaps 90% of the time. So we assign 90% of the p_m to NMOV and 10% to AMOV. The p_g is all given to the transition to GYRD, and the remainder (usually quite large) is given to NOM, the probability of remaining nominal. The other rows follow in similar fashion.

The other conditional probability matrices are data tiplied term by term by the λ vector and normalized to produce the overall belief for that node. The only π needed *a priori* is the probability of the root node; all of the others receive their π vectors from their parents.

The evidence λ passes back through the conditional probability matrix to the parents, enabling them to update their beliefs on the basis of evidence incoming from their children. As each λ vector representing incoming evidence appears, each node (hypothesis) of the entire network is updated to the correct belief, given the totality of the evidence received so far. The complexity of this algorithm is of order n^2, where n is the number of nodes. The computation required by each node depends on the size of the conditional probability matrices, which in turn depends on the number of discrete states of the node and its parent node.

The general directed-acyclic-graph Bayesian network is more problematic. If the network is small, then loops may be broken up by instantiating certain nodes to fixed states. The computational complexity of this approach is exponential in the number of loops, and so the method is not practical for networks with many loops.

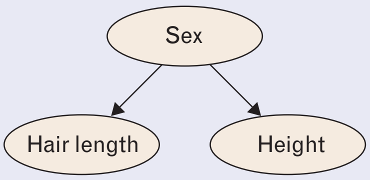


FIGURE B. Directed-acyclic-graph indirect causes.

Stochastic methods have also been proposed, in which certain nodes receive random values and multiple runs are used to estimate the true probabilities. Various methods using this type of Monte-Carlo approach have been proposed.

Coding theory has recently introduced turbo codes, which are actually loopy Bayesian networks that are updated by a method equivalent to the local message passing methods [3]. There is no guarantee in general that ignoring loops will result in a system that will converge to the correct solution, or even converge at all, but the turbo codes are very effective. K. Murphy and Y. Weiss have investigated this issue [4].

J. Pearl describes a useful method called *clustering*, or reducing a general Bayesian network to a smaller network containing nodes with more states [1]. For example, a two-state node and a four-state node can be combined, resulting in an eight-state node containing all the information of the original nodes, yet removing a loop. The resulting loopless network can be updated by the efficient and deterministic local updating algorithm.

Automatic methods can create conditional probability matrices for these product nodes from the smaller conditional probability matrices of the original nodes. Some product states may be illegal or not useful, however, because of a lack of perfect orthogonality between the original nodes. By omitting these nodes we may lose the ability to maintain two smaller conditional probability matrices and automatically compute the larger conditional probability matrix, but gain both a better representation of the problem and improved performance.

References
1. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Mateo, Calif., 1988).
2. F.V. Jensen, *Bayesian Networks and Decision Graphs* (Springer, New York, 2001).
3. R.J. McEliece, D.J.C. MacKay, and J.-F. Cheng, "Turbo Decoding as an Instance of Pearl's 'Belief Propagation' Algorithm," *IEEE J. Select. Areas Commun.* **16** (2), 1998, pp. 140–152.
4. K. Murphy, Y. Weiss, and M. Jordan, "Loopy Belief Propagation for Approximate Inference: An Empirical Study," *Proc. 15th Conf. on Uncertainty in Artificial Intelligence, Stockholm, 30–31 July 1999,* K.B. Lasky and H. Prade, eds. (Morgan Kaufmann, San Francisco, 1999).

matrices that connect our metric and signature data nodes to the status nodes. These represent the strength of the link between our incoming evidence and the actual satellite status. Higher-quality sensors have tighter coupling. The metric nodes do not provide evidence useful for discriminating between nominal and anomalous states like NOM and ANOM, so the conditional probability matrices show equal evidence for these states related to how strongly the evidence favors an in-slot condition.

Some of these conditional probability-matrix entry decisions may seem arbitrary, but high accuracy is not needed here. As remarked above on the precision of the input evidence calculations, the power of the Bayesian network system is the structure of the network and the fact that many pieces of evidence are combined. As long

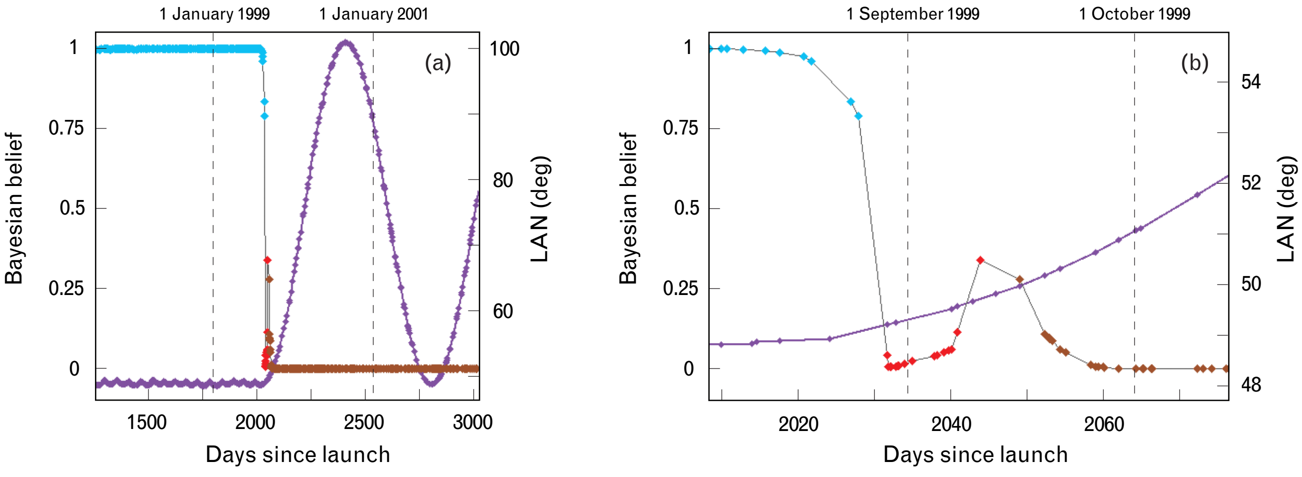


FIGURE 22. GEO satellite failure. (a) The composite history graph gives an overview of a satellite failure in 1999. The LAN curve (purple) shows the expected oscillation about the stable equilibrium point after failure. (b) Zooming in on the time of failure shows points representing individual pieces of evidence that were fused by the Bayesian network to determine status. Figure 23 shows further analysis of this event.

as those many pieces of evidence and probabilities are reasonable, we obtain good performance.

Iterative Status Assessment

Given a good flow of data into the Bayesian network, we can make a good estimate of the status of a satellite throughout its lifetime. This estimate is not only useful in and of itself; it can be useful in improving signature assessment. We need historic signatures from nominal satellites in order to assess new signatures by using signature comparison. New signatures that support the belief that a satellite is nominal may be automatically incorporated into the historic signature database, leading to improved performance in signature comparison.

To see how this comparison would work, imagine that several new signatures are received on a GEO satellite. Initially, these signatures are assessed by comparing them to known good signatures in the historic signature database, but the new signatures are also added to the database, with unknown status. If the resulting Bayes-

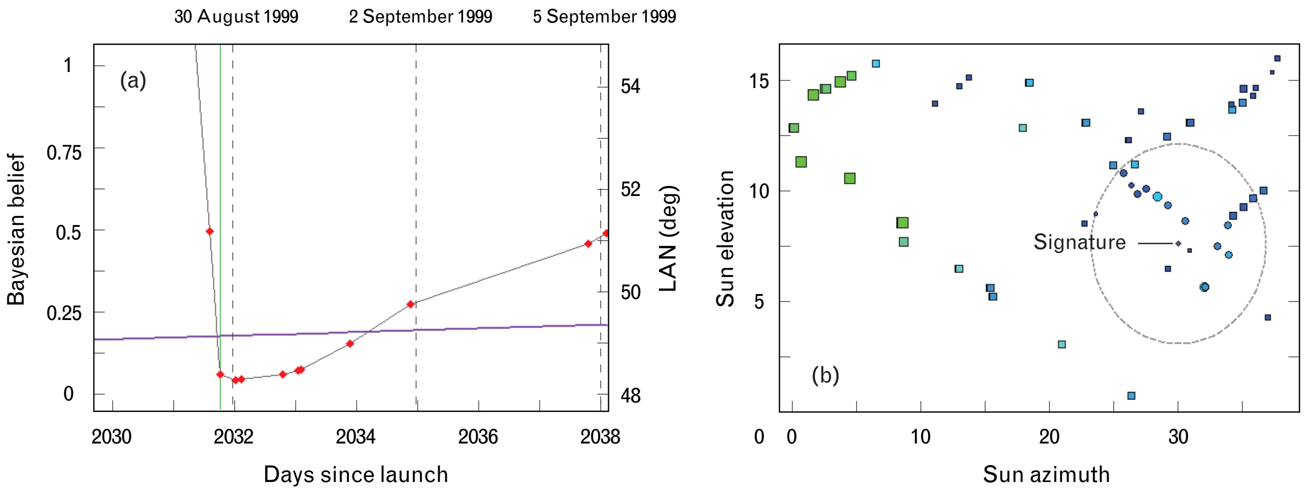


FIGURE 23. Satellite failure supporting evidence. (a) A further zoom into the satellite failure history plot of Figure 22. Moving the mouse over a point causes a balloon to pop up and give the details of that specific piece of evidence, which in this case is the first piece of evidence— an SBV signature with confidence 0.02—strongly supporting an anomalous state. (b) A graph of the cohort of this signature, with nearby signatures. The small purple diamond represents the signature under analysis.

ian network time history of that satellite shows that it is believed to be nominal with high confidence at those times, say over 90%, then those several new signatures could have their status set to nominal, and they would be available to help assess future signatures.

In fact, as more historic signatures become available, the confidence of any signature of the same type within their cohort region will change; thus we can recalculate confidences for that type. Once this recalculation is done, the Bayesian network assessments will change slightly, so the status of all the signatures can be reset to match the latest Bayesian beliefs. This procedure could even iterate several times, although the changes precipitated by adding several signatures to a large historic data set would not normally be significant.

In our discussion of the estimation of signature distance statistics, we pointed out that identifying sets of signatures representing nominal and anomalous satellites was not difficult on the basis of metrics alone. Not every signature needed to be assigned to one of the sets to obtain data sets that were large enough to estimate the target likelihood functions. However, given Bayesian network assessments of the histories of the satellites of a type, we can automatically assign, say, the H_0 signatures to be taken at a time when the satellite status has less than 10% Bayesian network belief of being nominal, and the H_1 signatures to be taken when the belief was over 90%. This type of iteration has proven very useful with several categories of satellites.

Monitoring Example

The left graph in Figure 22 shows the Bayesian history for a GEO satellite that failed in 1999. The purple curve of element-set LAN shows that the object was in a slot near 49° until the time of failure, after which it began to oscillate about the stable equilibrium point at 75.1°. The composite Bayesian history curve shows a short region of red anomalous state followed by a brown drift state. By zooming in on the failure time we can examine individual points on the Bayesian history curve, representing pieces of evidence, as shown in the right graph in Figure 22.

The left graph in Figure 23 shows a further zoom on the failure portion of the Bayesian network history. Moving the mouse over the individual points shows pop-up information about that individual piece of evidence. This is the first piece of evidence strongly supporting a state of anomalous. Inside the yellow balloon

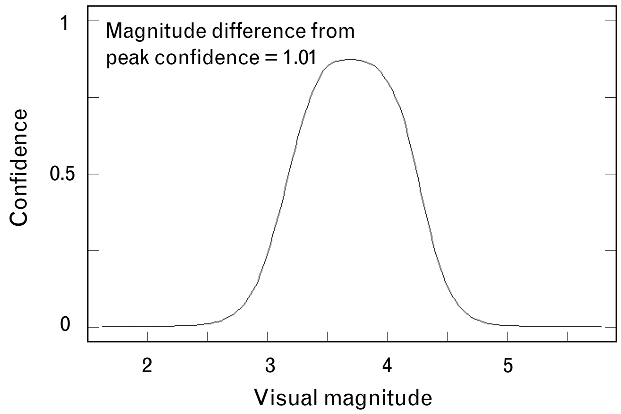


FIGURE 24. Photometric potential confidence curve. Signature 170239 is dimmer than would be expected from its cohort of parameter-space neighbors. This result can be quantified by the potential confidence curve, which shows the confidences that would be obtained if this signature had different magnitude. In this case the magnitude is about 1 magnitude dimmer than expected, resulting in a confidence much lower than the maximum of around 0.7.

we see the Bayesian state, the date, the type of data (SBV in this case), the signature number, and the confidence nominal (which is 0.02). The right graph is a cohort plot showing the signature under analysis as a tiny purple diamond, and the cohort signatures as the blue circles within the ellipse. (Recall that the squares are signatures that are slightly outside the cohort in the other dimensions of the parameter space, which in this case is sensor-angle space.)

The circles are larger than the diamond, which indicates that this signature is dimmer than expected. The question might arise as to how much brighter it would need to be to have a reasonable assessment. That question can be answered by a potential confidence curve, illustrating the range of confidences that any signature could receive by comparison to the historic database in this region of parameter space. Figure 24 shows the potential confidence curve for the signature in Figure 23.

The current system has been available only recently, but we can set back the clock and process historic data as a test of algorithm performance. One key performance metric is the false-alert rate, along with, of course, true-alert rate. When the Bayesian network changes state significantly, e-mail alerts are sent out. False alerts can be caused by a bad element set or a cross-tagged signature, and they waste analyst time. When enough new data are processed, and the event can be looked at from the perspective of both past and future, the Bayesian network

Table 2. Historic Alert Replay Results

1996-06-02	21:22:12	Init nominal	(elset 4979145)
1997-10-17	20:22:23	Nominal moving	(elset 4979401)
1997-10-19	01:30:50	Nominal	(elset 4979402)
1999-08-30	00:57:10	Anomalous	(SBV 170401)
2000-11-22	16:45:48	Nominal	(SBV 335434)
2000-11-25	18:20:16	Drifting	(elset 4979797)

will usually do a good job, as indicated in Figure 22. But in the one-sided situation, one or two bad pieces of data can cause an alert, which is typically followed by another alert indicating a return to the previous state.

Table 2 shows a typical alert history over the last ten years of this same object. It is initialized to NOM, and then experiences a false alert to MOV about sixteen months later, followed immediately by a return to NOM. The next item in the table is the true failure event we have already discussed, in August 1999. Note that there is no alert for DRFT; this is because ANOM and DRFT are both considered failure states, so once alerted to ANOM the transition to DRFT is just a normal progression and the analyst does not need to be informed of it. Of course the Bayesian network display will display DRFT for any analyst who looks at the history of this object.

A little more than a year later, there is a false alert to NOM, followed quickly by a return to DRFT, which is the true state. To understand these false alarms better, we can examine the graphs in Figure 25. In the left graph we see some element-set parameters around the time of the first false alert. The large dip in walk rate caused the system to alert to a move, and the element set being processed when this alert was sent out—as reported in the alert—was indeed this element set. This type of short apparent increase in absolute walk rate can be symptomatic of a short move in the GEO belt, but in this case it is easy to see from the LAN curve that no move has actually occurred. This bad element set is near one of the station-keeping maneuvers, which can be seen as the rapid drops in the sawtooth walk-rate graph. Bad element sets are most probable at such a time because of accidental combinations of pre- and post-maneuver observations, or even the use of cross-tagged observations.

The right graph in Figure 25 shows the element sets around the time of the second false alert. The small glitch in the center of the graph is responsible for the false alerts, in conjunction with some SBV signatures that just happened to have a brightness consistent with a stable payload. There was enough change in the element sets for the system to decide that a boost had occurred. Since the satellite must be stable for an intentional boost to occur, this is evidence against drifting. The alert list shows SBV evidence as triggering the alert, but in this case that evidence is the straw that breaks the camel's back, or the morsel of evidence that flips the Bayesian

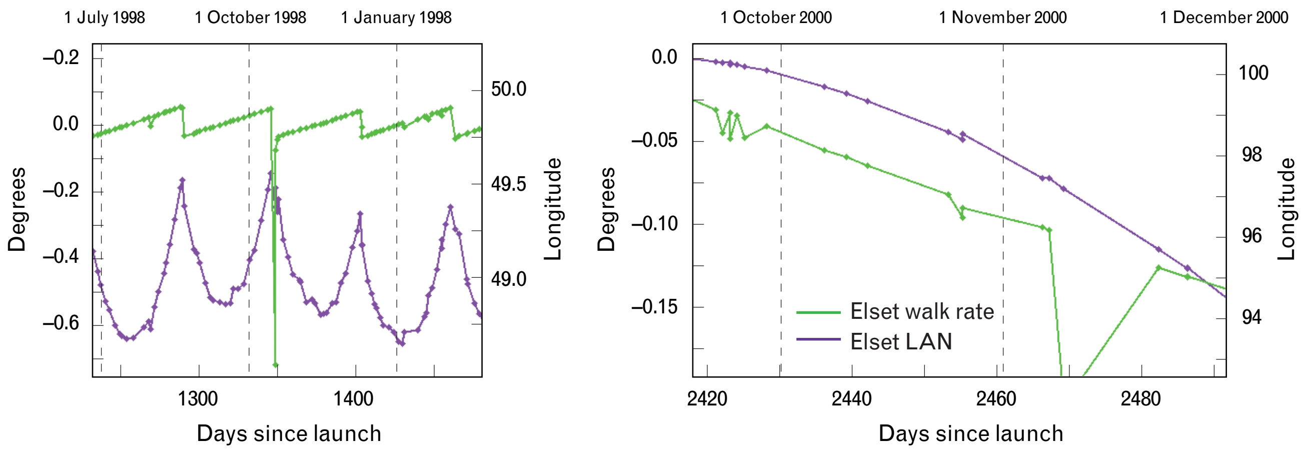


FIGURE 25. Bad evidence causes false alerts. Left: plots of element-set walk rate and LAN during a normal station-keeping period. The large dip in walk rate indicates a bad element set. This bad element set is responsible for generating a false alert, since it looked like the start of a GEO belt move. Right: the same functions during a period of drift. The small glitch in the center of the graph coupled with ambiguous SBV signatures was responsible for another false alert.

network. The more important cause of the alert was the bad element sets.

In this example we obtained one false alert every five years, and we correctly sent a true alert. With this false-alert rate we could monitor well over a thousand GEO satellites without exceeding one false alert per day. So the Bayesian network fusion seems to be effective enough to result in a usable system. We cannot conclude much about the true alert performance from an example with a single alert, but as evidence piles up, very few true alerts are likely to be missed. A better performance metric would be the time delay between actual satellite status change and occurrence of the corresponding alert.

Summary

The GEO belt currently consists of over 380 active satellites. They are controlled by numerous commercial and government operators who are continuously maneuvering them from the time of their initial launch until final retirement. Many reside in crowded regions over the earth, and, as well, they all orbit within a ring of a large population of inactive satellites that were left to drift as they were depleted of station-keeping fuel or failed on orbit.

One major consequence of this crowding is that there is a real and increasing probability that two satellites could collide. The long-lasting impact of such an event would be a debris-filled and possibly unusable region of space as well as the loss of valuable assets and services. The ongoing active satellite maneuvers are generally privy information to the satellite operator, and so continuous monitoring by the space surveillance community is required to prevent losing the satellites for some time after maneuvers occur. A lost active or inactive satellite could pose a collision threat that could not be foreseen.

We have developed a system to monitor a number of active satellites under commercial R&D sponsorship. These satellites are monitored against all threatening inactive satellites as well as the remainder of the active population. The philosophy of this monitoring is to provide a sufficient early warning of a potential collision to provide the operator time to modify routine station-keeping maneuvers (without additional fuel expense), or in some cases to require a specific avoidance maneuver to obtain a satisfactory separation. The decision-making process relies on two things. One is the evaluation of the accuracy estimates for the close separation distances. This accuracy depends in part on the quality, quantity, and timeliness of the metric tracking data. The other depends on the upcoming plans for normal station-keeping maneuvers of the satellite that may be threatened.

We have generally strived to keep the closest separation distances to greater than 6 km, but for cases in which the encounter geometry and error assessment indicate, we have let closer crossings occur (to within 2 km). This separation distance is significantly greater than the estimated errors, but, given the cost and results of a collision, we prefer to be rather conservative with how closely we allow satellites to approach each other. For the future, we need to continue to improve the orbit accuracy and error assessment process so that we can more confidently allow closer crossings to occur.

In the case of noncooperative satellite monitoring when nothing is known *a priori* about satellite status or owner intentions, we must rely solely on measurements from the space surveillance network. Because of the large number of satellites in the GEO belt, it is essential to provide a highly automated system—it is impractical to expect space analysts to give much personal attention to every single satellite. The available space surveillance data contains the essential metric measurements used to maintain the catalog, but it also includes measurements of photometric brightness and radar cross section. We have developed methods based on hypothesis testing to estimate the probability, given a signature measurement, that the corresponding satellite is nominal. Since we may not be able to predict a signature of a nominal satellite, we rely on signature comparison, comparing new signatures to historic signatures from nominal satellites. A data-dependent comparison step produces correlated pattern-recognition distances, whose subsequent processing is the same for all data types.

Individual signatures often contain relatively little information, but if collected in sufficient volume they are useful inputs into our automated status monitoring system. This system uses dynamic Bayesian networks to fuse all available evidence, deriving an assessment of satellite status over the lifetime of the satellite. This lifelong assessment conveniently satisfies the requirements of the signature-comparison-based method of signature assessment, which requires a set of historic signatures collected during periods when the satellite status was nominal.

By constructing a Bayesian network containing status nodes whose states are operationally significant, we guarantee that the Bayesian belief associated with those nodes will provide information useful to space analysts. The Bayesian network itself contains causal information, time evolution information, and state-transition probability information. These relationships are built up on the fly as new evidence arrives and the network is enlarged, but they are based on the same analyst knowledge that enables traditional manual analysis. Each new piece of evidence triggers a Bayesian network update that updates all beliefs in the network, so the historical status nodes are updated as well as the more recent ones.

We continue to refine algorithms and look for new sources of data that could improve the automatic status monitoring system. As new space surveillance sensors come on line, we hope and expect that more accurate and more timely assessments will result, providing higher-quality decision support to space analysts.

Acknowledgments

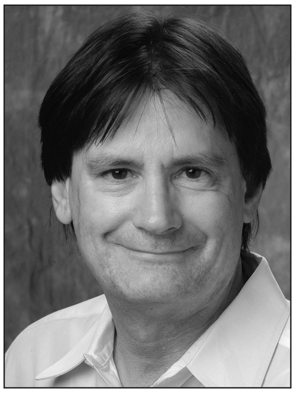
Both authors would like to thank Forrest Hunsberger for motivating and helping us to develop this article.

The first author would like to thank all those who have contributed to the Geosynchronous Encounter Analysis CRDA project over the last ten years. For technical support I would specifically like to mention Lori Thornton, Zach Folcik, Susan Shulman, Miquela Vigil, Elizabeth Evans, Jayant Sharma, Ray Moy, and Eric Phelps, all of whom are still at Lincoln Laboratory, and Bob Clouser, Chad Morris, Brian Pantano, John Conner, and Alana Sette, who left valuable contributions. I would also like to thank John Beusch, Dave Chan, Ramaswamy Sridharan, and Dave Whited, who have contributed to the development, management, and growth of this project. I also would like to thank the commercial satellite companies that have been involved as CRDA partners: SES Americom, Telesat Canada, Intelsat Global Services, SATMEX, Loral Skynet, and PanAmSat. Finally, I owe a great deal of gratitude to Marilyn Lewis, who has always provided helpful information on the GEO belt population, and to the personnel at the Millstone Hill and ALTAIR/TRADEX radars for the valuable and timely tracking data they have provided.

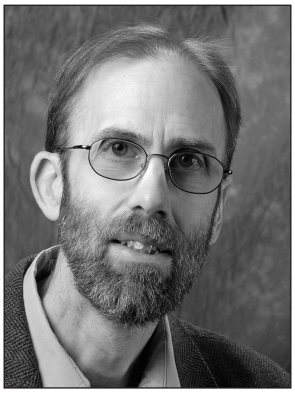
The second author would like to thank the present and former coworkers who made important algorithmic and software contributions, especially Karl Voskuil, David Rysdam, Terry Mallberg, Mona Kosik, Dan Salomon, and Tom Morgan. I would also like to thank Ken Fields for his valuable data analysis and his user interface and algorithmic suggestions, and Dave Chan, Mark Czerwinski, and John Beusch for management support.

REFERENCES

1. E.M. Soop, *Handbook of Geostationary Orbits: Space Technology Library, vol. 3* (Kluwer, Dordrecht, The Netherlands, 1994).
2. J.J. Pocha, *An Introduction to Mission Design for Geostationary Satellites* (Reidel, Dordrecht, The Netherlands, 1987).
3. P. Berlin, *The Geostationary Applications Satellite* (Cambridge University Press, Cambridge, U.K., 1988).
4. L. Perek, “The Scientific and Technical Aspects of the Geostationary Orbit,” *Acta Astron.* **17** (6), 1988, pp. 587–598.
5. M. Hechler, “Collision Probabilities at Geosynchronous Altitudes,” *Adv. Space Res.* **5** (2), 1985, pp. 47–57.
6. “Environmental Protection of the Geostationary Satellite Orbit,” ITU-R Recommendations, 1994 S Series Vol., Fixed Satellite Service, International Telecommunication Union, Geneva, Switzerland, ITU-R S.1003, 1994, pp. 364–367.
7. C. Hernández and R. Jehn, *Classification of Geosynchronous Objects, Issue 8,* European Space Agency, European Space Operations Centre, Darmstadt, Germany, Feb. 2006.
8. C. Toews and E. Phelps, “Statistical Analysis of Close Approach Distances in the Geostationary Environment,” *Proc. AAS/AIAA Astrodynamics Specialist Conf., Big Sky, Mont., 3–7 Aug. 2003*.
9. R. LeClair and R. Sridharan, “Probability of Collision in the Geostationary Orbit,” *Proc. Third European Conf. on Space Debris* **1**, *Darmstadt, Ger., 19–21 Mar. 2001,* pp. 463–470.
10. R.I. Abbot, R. Clouser, E.W. Evans, J. Sharma, R. Sridharan, and L.E. Thornton, “Close Geostationary Satellite Encounter Analysis: 1997–2001,” *Spaceflight Mechanics 2002, Advances in the Astronautical Sciences, vol. 112, pt. 1, 2002* (Univelt, San Diego, 2002), pp. 219–237.
11. R.I. Abbot and J. Sharma, “Determination of Accurate Orbits for Close Encounters between Geosynchronous Satellites,” *Proc. 1999 Space Control Conf., 13–15 Apr. 1999,* Project Report STK-254, Lincoln Laboratory, Lexington, Mass., pp. 71–83.
12. B.S. Aaron, “Geosynchronous Satellite Maneuver Detection and Orbit Recovery Using Ground Based Optical Tracking,” Master’s Thesis, Dept. of Aerospace Engineering, MIT, Cambridge, Mass., June 2006.
13. G.H. Stokes, C. von Braun, R. Sridharan, D. Harrison, and J. Sharma, “The Space-Based Visible Program,” *Linc. Lab. J.* **11** (2), 1998, pp. 202–238.
14. H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Pt. 1: Detection, Estimation, and Linear Modulation Theory* (Wiley, New York, 1968).
15. T.P. Wallace, “An All-Neighbor Classification Rule Based on Correlated Distance Combination,” Technical Report 1030, Lincoln Laboratory, Lexington, Mass., 5 Nov. 1996.



RICHARD I. ABBOT
is a staff member in the Space Situational Awareness group. His research interests include celestial mechanics, astrodynamics, geodesy, and astrometry. He received B.A. degrees in astronomy and mathematics from the University of South Florida, and a Ph.D. degree in astronomy from the University of Texas at Austin. As a graduate student, he focused on dynamical resonances in the solar system, although he also worked on NASA-funded projects for lunar and satellite laser ranging and astrometry of the outer-planet satellites for the Voyager fly-bys. As a researcher in the Earth, Atmosphere, and Planetary Sciences department at MIT he was involved in developing early geodetic applications with the Global Positioning System. Since coming to Lincoln Laboratory in 1989, he has worked on numerous projects regarding orbit determination for artificial satellites. He is a member of the American Astronomical Society.



TIMOTHY P. WALLACE
is a staff member in the Space Situational Awareness group. He received a B.S.E.E. degree in electrical engineering from Cleveland State University, and M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University. Since joining Lincoln Laboratory he has worked in a variety of areas, including radar imaging, radar systems analysis, and optical and radar space surveillance. Recently he has concentrated on problems of satellite monitoring and identification, and he has developed algorithms for fusing photometric and radar signature information with metric observations. His research interests include signal processing, pattern recognition, and information fusion.