An Optimization Approach to the Planning of Earth Observing Satellites

Nicola Bianchessi, Vincenzo Piuri, Giovanni Righini, Manuel Roveri

Università degli Studi di Milano, Dipartimento di Tecnologie dell'Informazione
Via Bramante 65, 26013 Crema, Italy
{bianchessi,piuri,righini,roveri}@dti.unimi.it

Giovanni Laneve, Antonio Zigrino

Space Software Italia SpA
Viale del Lavoro 101, 74100 Taranto, Italy
{laneve,zigrino}@ssi.it

Abstract. We describe a complex optimization problem related to the COSMO-SkyMed project for the observation of the Earth surface through a constellation of satellites equipped with Synthetic Aperture Radar instruments. We describe the optimization problem and we present some algorithms we have tested.

1 Introduction

Earth observation satellites are platforms equipped with instruments for optical, radar or infra-red observation, placed on low orbits around the Earth. Platforms and instruments can be controlled in order to answer requests of observation coming from various users. These requests are emitted at any time by the users towards a planning and control center, that is responsible for the management of the mission and the control of the satellites.

In this paper we consider the problem of selecting the tasks corresponding to requests for observation in order to optimize suitable objective functions, while complying with a lot of technical and managerial constraints.

The work has been developed as part of the COSMO-SkyMed project. The constellation under study is made of four satellites equipped with SAR (Synthetic Aperture Radar) instruments. Each satellite repeats the same trace at the end of a cycle of 16 days. The cycle lasts 4 days considering the whole constellation. Each satellite runs 14.8125 orbits per day, that is 137 orbits every 16 days. The four satellites of the constellation are placed at every 90 degrees on their orbital plane around the Earth.

Related Literature. A problem similar to ours has been studied with regard to SPOT5 satellites employing optical technology. Some benchmark instances have been made public by (Bensana, Lemaitre and Verfaillie 1999). Vasquez and Hao presented a tabu search algorithm (Vasquez and Hao 2001) and some upper bounds (Vasquez and Hao 2003) for single-satellite instances of the problem.

The problem has been re-formulated more recently, taking into account also scheduling aspects arising with "agile" satellites (PLEIADES project). Simple instances (single satellite, single orbit) have been the subject of the ROADEF Challenge 2003, promoted by ROADEF, the French operations research society, in cooperation with CNES and ONERA (Verfaillie et al. 2002). Some heuristic algorithms were proposed by (Lemaitre et al. 2002). (Cordeau and Laporte 2003) presented a tabu search algorithm for such simplified problem, where capacity constraints and transmission to ground stations are not taken into account.

(Wolfe and Sorensen 2000) presented three heuristic algorithms of increasing complexity to schedule observations of a single satellite, presenting results on small instances with up to 50 requests.

SAR satellites have been considered by (Harrison, Price and Philpott 1999) who solved very small instances, concerning the schedule of a single satellite with up to 50 requests in a time window of a few minutes.

More recently (Globus et al. 2003) studied genetic algorithms for a problem with one or two satellites on a small time horizon with no memory and transmission constraints.

A richer model, more similar to ours, was considered by (Frank et al. 2001), who developed a greedy stochastic algorithm, without reporting about computational results obtained.

In the remainder we first define all operational constraints of our model and then we illustrate our algorithmic solution.

2 System Description

2.1 System Management

Image acquisition requests can be submitted by users at any time but for the optimization purpose we consider a given set of requests. Each request corresponds to a set of images and each image corresponds to a set of observation opportunities.

The system is intended for use by multiple classes of users with different priorities: therefore the requests can be treated in different ways according to the user class. In particular, some requests can be classified as *high priority*, while the others are *low priority*.

All requests have an associated *deadline* and all high priority requests must be scheduled before their deadline. It is assumed that the feasibility of high priority requests is guaranteed. Low priority requests also have an associated deadline, meaning that if they are satisfied they must be satisfied before that deadline; otherwise they must be disregarded.

We identified at least two significant objective functions for the system: the standard objective is to maximize the value of the requests satisfied; occasionally it may be of interest to minimize the time necessary to acquire images of particular urgency or importance. In this last case we suppose that a set of particular requests is given and they must be scheduled as soon as possible.

When presented with this problem we were also given some limits on the computing resources available: in particular, the computation of a plan on sixteen days can take up to approximately 2.5 hours, while 50 minutes are allotted for computing a four-days plan.

2.2 Image Acquisition and Technical Constraints

Each satellite is built in a compact way around a SAR (Synthetic Aperture Radar) instrument. The SAR instrument can acquire the image of a *swath*, that is a rectangular strip of the ground at the right or left of its ground-track. This implies that the Earth surface is not completely accessible every day by each satellite, but it is over a period of several days, with a number of acquisition opportunities decreasing with the distance from the poles.
Each opportunity is indicated by DTO (*data taken opportunity*) in the remainder. The duration of an image acquisition is proportional to the length of the swath and it is given. Owing to the SAR technology, there is no difference between day and night observations. All DTOs for each satellite are given; the planning and scheduling problem includes deciding which of them must be taken.

Satellite Configuration and Set-Up. There are three parameters that define the configuration of the satellite and the SAR instrument. First, the satellite has two *orientations*, named "right-looking" and "left-looking" and it can rotate from one orientation to the other. Second, the SAR instrument can take images with different *look-angles*: to this purpose it can switch between several different positions. Third, the SAR instrument may work in different *operating modes*, that is it can observe swaths of different width with different resolution and it can consume different amounts of energy for each acquisition.

Therefore set-up operations may be necessary between two consecutive image acquisitions and the set-up times only depend on the two DTOs involved. Moreover the SAR instrument must be off for a minimum given amount of time between any two consecutive acquisitions (two acquisitions are defined as consecutive when no acquisition is executed between them). During this time interval it is not allowed to execute set-up operations either.

Since the changes in orientation, look-angle and operating mode cannot be done simultaneously, the overall set-up time is the sum of four terms $\Delta = \Delta_{rotation} + \Delta_{angle} + \Delta_{\mathrm{mode}} + \Delta_{off}$ corresponding to the four mentioned parameters.

Operational Profiles. There are some operational constraints on the sequence of image acquisitions that the SAR instrument can do. The constraints concern the activity time of the SAR instrument in each operating mode and they represent in a compact way the constraints on the energy consumption of the instrument. The activity time is normalized through a factor depending on the operating mode: for instance one minute in operating mode *A* may be equivalent to 30 seconds in operating mode *B* from the viewpoint of the operational constraints, if operating mode *A* requires a better resolution and therefore a greater amount of data taken per unit of time. There are two kinds of operational constraints: the first bounds the average value of the workload; the second bounds its variance.

Constraint on the average value of the workload. The total (normalized) activity time in each day cannot exceed a given threshold T_{day}. This constraint must be satisfied on each time window one day large.

Constraint on the variance of the workload. The total (normalized) activity time in each orbit cannot exceed a given threshold T_{orbit}; this constraint can be violated provided that the following property holds: in whatever way the temporal axis is subdivided into a set of contiguous time windows, whose width is equal to the duration of an orbit, the total activity time of each satellite cannot exceed T_{orbit} in two or more such time windows if they are not separated by at least one day. Such time windows are referred to as *peak orbits*. As a consequence of this constraint, the total activity time cannot exceed the value $2T_{orbit}$ in any peak orbit. The value of T_{orbit} is given by T_{day} divided by the number of daily orbits.

Remark. Since the constraints on the operational profiles are very conservative, it is allowed to violate them whenever this is necessary to take a DTO related to a high priority request.

Request Splitting. Some images to be acquired can be related to one another because they derive from a unique request, that has been split into several images: this happens when a request concerns an area too large to be completely observed within one swath. When a large target is split into several images there may exists a combinatorial number of ways to combine the possible swaths in order to cover the target area; it is assumed that the choice on which combination to consider has been done a priori, according to rules and considerations related to how easy it is to reconstruct the overall image (for instance, the orbits from which the images are taken must be either all descending or all ascending). Moreover all images related to the same target must be taken with the same look angle.

The splitting of large targets in multiple acquisition requests is managed as follows: as soon as the first partial acquisition of the target is done, all the other images related to the same target are marked as "medium priority" and they are given precedence with respect to low priority ones. This is done in order to make unlikely the case in which a split target is only partially acquired.

2.3 Data Transmission

All data acquired by the radar instrument are stored in memory and must be transmitted to ground stations.

Memory. Each satellite is equipped with two identical memory devices of given capacity. Each time an image is taken, it is stored in one of the two memories. Each image must be segmented in a given number of files. If an image is segmented in more than one file, all files must be stored in the same memory device.

Transmission Channels. A connection between a satellite and a ground station is possible only when the ground image of the satellite position is close enough to one of the ground stations located on the Earth surface. We call *down-link opportunity* (DLO) this visibility time-window.

The transmission may happen in different ways and through two different channels, indicated as channel 1 and channel 2. Both of them are capable of transmission at a given bit-rate and the two channels can be used independently.

Each station may receive information on one or two channels simultaneously; this is a given characteristic depending on the station.

The DLOs of different stations for a same satellite may overlap, in which case the satellite can communicate with both stations simultaneously.

Transmission Modes. Each image is associated to a set of ground stations and to a transmission mode. The transmission mode can be either "AND" or "OR". If the transmission mode is "AND", the image must be transmitted to every ground station associated with it; if the transmission mode is "OR", the image must be transmitted to one associated ground station.

If an image is segmented in more than one file, the image segment files can be transmitted separately and independently and, if the transmission mode is "OR", they can be transmitted to different ground stations.

Each image segment file is indivisible: it must be transmitted completely and without interruption on the same channel to the same station.

The Transmission Device. The transmission device on board of each satellite is responsible for storing files in memory and for transmitting them to the ground stations.

Storage and transmission can be done simultaneously but they share a common resource, that is a bus.

Each image has a given acquisition bit-rate associated with it: it is equal to the ratio between the size of the image and the duration of its acquisition.

The maximum bit-rate available to store information, B_{mem}, is given; the maximum bit-rate available on each of the two transmission channels, B_{transm}, is also given; the common resource amounts to the capacity of channel 2.

Channel 1 is always available for transmission. Either channel 2 is used for transmission and the bit-rate to store information is limited to $B_{mem} - B_{transm}$ or only channel 1 is used for transmission and the bit-rate to store information may be up to B_{mem}.

As a particular case, the file transmitted on channel 1 may be the same currently acquired by the SAR ("pass-through" mode).

There is no need for synchronization between acquisition and transmission, not even in pass-through mode.

Unavailability Periods. Both satellites and ground stations may be unavailable for some given time periods. During a period in which a station is unavailable no transmission can be scheduled to that station. During a period in which a satellite is unavailable no acquisition, transmission or set-up change can be planned for that satellite.

Some unavailability periods for the satellites are artificially imposed because such periods of inactivity are used to replace an old plan by a new one. This replacement operation can be done once or twice a day. In this case the unavailability periods are the same for all satellites.

2.4 Objective Function

Two different objective functions have been taken into account. The standard objective is to maximize the total value of the satisfied requests. A request is satisfied when all its images have been taken (before the deadline) and transmitted. On demand the objective is switched to a second function and consists of taking and transmitting as soon as possible all requests marked as "very urgent". This second objective is taken into account by simply setting the priority of all very urgent requests to "HIGH" in order to ensure they have to be taken. Since our algorithms build the plan by making choices in chronological order, the very urgent requests are inserted into the plan of the constellation at their earliest occurrence.

3 Algorithms

We addressed the planning and scheduling problem defined above by designing greedy constructive algorithms, enriched with look-ahead and backtracking capabilities.

Greedy constructive (randomized) algorithms have the following advantages. First, their computational complexity is linear in the number of DTOs and DLOs and this allows to deal with very large problem instances. Second, they are not a single algorithm but rather a set of algorithms, one for each policy that can be inserted in the decision routines; this allows trade-off analysis on different scenari, which is useful to put in evidence bottleneck activities and slacks in resources.

Greedy constructive (randomized) algorithms are as follows: a DTO list and a DLO list for each satellite are sorted chronologically, according to the starting time of each DTO and DLO.

The state of each satellite is initialized with its position and set-up and with the files currently in its memory.

An acquisition time and two transmission times are defined for each satellite. The acquisition time t_i^0 for satellite i is the time instant in which satellite i can start the acquisition of the next DTO; the transmission time t_i^τ for satellite i and channel $\tau = 1,2$ is the time instant in which satellite i can start the transmission on channel τ. A decision time t_i is defined for each satellite i as $t_i = \min\{t_i^0, t_i^1, t_i^2\}$. The satellite with the minimum decision time is iteratively selected as the *active* satellite.

Each time the decision time corresponds to the time instant in which a new DTO can be taken, a check for feasibility is made. The following conditions are checked:

(a) the constraints on the operational profiles must be satisfied if the image is taken;

(b) if the image results from target splitting and it is not the first acquisition related to that target, its look-angle must be compatible with those of the already acquired images related to the same target;

(c) there must be sufficient memory space in one of the two memory devices of the satellite, to store the image;

(d) the number of stored files must not exceed the limit, if the image is taken;

(e) the algorithm also looks ahead to the next DTOs to check whether there are DTOs corresponding to higher priority requests incompatible with the DTO to be taken;

(f) if the acquisition requires bus capacity greater than $B_{mem} - B_{transm}$, channel 2 must not be in use.

Low priority DTOs must pass tests (a), (c), (d), (e) and (f) to be declared feasible. Medium priority DTOs must pass tests (b) and (e) and they must pass tests (c) and (d) through backtracking, to be declared feasible. High priority DTOs must pass test (b) and they also must pass tests (c) and (d) through backtracking, to be declared feasible.

In case tests (c) or (d) fail, that is the amount of available memory or the number of storable files is not sufficient, and the DTO is not a low priority one, the algorithm *backtracks*: backtracking involves two phases, named *Delete* and *Restore*.

In the *Delete* phase the most recently taken images are scanned in reverse chronological order and they are tentatively eliminated from the plan, until enough memory resources become available to store the DTO to be taken. During this phase an acquisition cannot be eliminated if it has already been completely transmitted. Moreover images can be deleted only if they are related to requests whose priority level is strictly lower than that of the DTO to be taken. The *Delete* phase stops going back in time when the starting time of the plan is reached. If this stop criterion prevents the backtracking routine from making enough memory resources available, the *Delete* phase fails, the DTO to be taken is considered infeasible and no modification is made to the plan. On the contrary, if the *Delete* phase succeeds, the DTO is considered as feasible and the *Restore* phase starts.

In the *Restore* phase the list of temporarily deleted acquisitions is scanned again in chronological order and all the acquisitions which do not make the DTO infeasible are restored.

After backtracking, if the DTO has been made feasible, it is taken; otherwise it is skipped.

If a low priority DTO is feasible, a *decision policy* is applied, to decide whether to take it or to skip it. The decision policy may take into account different deterministic or probabilistic criteria, depending on the value of the image, on its urgency, on the station to which it must be transmitted, etc. If a medium or high priority DTO is declared feasible, it is always taken.

The acquisition time or transmission time is updated accordingly. The decision time of the active satellite is also updated.

The routine which executes a downlink operation is executed whenever the decision time of the active satellite corresponds with a transmission time on one of the two channels. This means that the decision time is certainly inside one or more DLOs for its satellite. Since the DLOs may overlap, it is possible that more than one station is available for transmission when the routine is executed.

The transmission routine searches for a stored image segment file that can be transmitted to one of the ground stations available at that moment. The file must have a size such that its transmission ends before the DLO of the corresponding station. If the transmission channel is channel 2 a further condition must be satisfied: the algorithm looks-ahead and bounds the available transmission time if a DTO is encountered whose priority is high and such that its acquisition requires more than $B_{mem} - B_{transm}$ units of capacity on the bus of the transmission device.

Files are searched according to an order, depending on a *sorting policy* that can take into account different deterministic or probabilistic criteria, such as file size, image value, aging on board, urgency with respect to the deadline, acquisition time, etc. While scanning the ordered list, the first file which is found to satisfy the above constraint is selected for transmission.

When a decision is taken to transmit a file on channel 2, every low priority DTO overlapping with the transmission period and whose acquisition bit-rate is greater than $B_{mem} - B_{transm}$ is deleted from the DTO list. This automatically satisfies condition (f) above.

4 Results

Our computational tests were performed on simulated scenari with up to 3000 requests per day on a planning horizon of sixteen days, corresponding to more than one million DTOs.

Changing the decision policy we could observe that the algorithm can be guided to choose more images with lower value or fewer images with higher value.

Our simulations allowed us to recognize which among the many constraints of the problem was the real bottleneck, and to compare different possible ground stations scenari. For each of them we could measure the percentage of use of each ground station as well as of all system resources, such as the memory on board of each satellite. We could also have a quantitative estimate of the access time and the transmission time of the taken images and the percentage amount of time spent on acquisition and set-up operations for each orbit and each satellite.

Owing to the features of the greedy algorithms, the observed computational times were largely inferior to the imposed limits, since they never exceeded a few minutes on the largest problem instances.

5 Future Developments

The quality of the solutions computed by the described algorithms can be evaluated by solving a suitable combinatorial or linear relaxation of the problem. Our preliminary results suggest that the former approach is efficient but it gives very loose bounds; the latter is more precise but it is impractical when the instance size is large. Therefore a significant term of comparison is still needed to assess the effectiveness of the optimization approaches developed so far. To obtain a significant upper bound a suitable *relaxation* must be identified such that it can be solved in some efficient way, and it provides a bound as tight as possible to the optimal value.

Mathematical programming offers a very wide arsenal of algorithmic techniques to cope with difficult combinatorial optimization problems. Many of these techniques may be successfully applicable to the COSMO-SkyMed or similar optimization problems. In particular we are currently developing *column generation* algorithms.

Apart from the main objective function to be optimized, that is the overall value of the satisfied requests, it may be convenient to take into account secondary objective functions such as response time, balance in the use of the ground stations, and others. In such case, an optimization algorithm in not enough; rather a *decision support system* is recommendable, allowing for a trade-off analysis between different non-dominated (Pareto-optimal) solutions, according to dynamically adjustable criteria that can depend on specific temporary needs or operating conditions.

Besides optimizing the operations of a system, in the respect of technological constraints, Operations Research can also be used to optimize the management of a system. *Revenue management* is concerned with the maximization of the revenue that can be obtained when selling a product or a service whose availability has already been established, so that the costs are more or less fixed. In the case of the COSMO-SkyMed project the negotiation with the customers, the definition of a price for each image, the possible definition and management of users' quotae are very important aspects that need to be empowered with revenue management tools. This is especially important in consideration of the very high fixed cost necessary to build and operate the constellation.

References

E. Bensana, M. Lemaitre, G. Verfaillie, *Earth observation satellite management*, Constraints 4 (1999) 293-299

J.F. Cordeau, G. Laporte, *Maximizing the value of an Earth observation satellite orbit*, Technical report, University of Montreal, 2003.

J. Frank, A. Jonsson, R. Morris, D.E. Smith, *Planning and scheduling for fleets of Earth observing satellites*, NASA Ames Research Center, 2001

A. Globus, J. Crawford, J. Lohn, A. Pryor, *Scheduling Earth observing satellites with evolutionary algorithms*, International Conference on Space Mission Challenges for Information Technology, Pasadena, California, 2003

S.A. Harrison, M.E. Price, M.S. Philpott, *Task scheduling for satellite based imagery*, DERA - Defense Evaluation and Research Agency, United Kingdom, 1999

M. Lemaitre, G. Verfaillie, F. Jouhaud, J.M. Lachiver, N. Bataille, *Selecting and scheduling observations of agile satellites*, Aerospace Science and Technology 6 (2002) 367-381

M. Vasquez, J.K. Hao, *A "logic-constrained" knapsack formulation and a tabu algorithm for the daily photograph scheduling of an Earth observation satellite*, Journal of Computational Optimization and Applications 20 (2001) 137-157

M. Vasquez, J.K. Hao, *Upper bounds for the SPOT5 daily photograph scheduling problem*, Journal of Combinatorial Optimization 7 (2003) 87-103

G. Verfaillie, G. Lemaitre, N. Bataille, J.M. Lachiver, *Management of the mission of Earth observation satellites - Challenge description*, www.roadef.org, 2002

J.W. Wolfe, S.E. Sorensen, *Three scheduling algorithms applied to the Earth observing systems domain*, Management Science 46 (2000) 148-168