N94-21336

GENERATING AND EVALUATING ALTERNATE SCHEDULES

Surender D. Reddy
Computer Sciences Corporation, Beltsville, Maryland

Summary

The demand for and availability of Space Network resources are subject to short-term fluctuations and long-term changes. Generation of acceptable schedules under changing demand and resource availability will require the use of different scheduling policies. This paper identifies several such scheduling policies. It defines metrics for evaluating schedules using the criteria directly related to these scheduling policies. Then it applies the metrics to compare several schedules generated for a scenario representative of 1998 SN demand and resources. Finally, the paper describes a method for using these metrics to evaluate schedules based on multiple criteria.

1. Introduction

The demand for and availability of Space Network (SN) resources are subject to short-term fluctuations and long-term changes. Increased demand during shuttle flights is an example of short-term demand fluctuation. Resource unusability because of repair or maintenance of related ground systems is an example of a short-term fluctuation in resource availability. Growth in number of customers and their demand for SN resources is an example of a long-term demand change.

A scheduling policy that results in acceptable schedules under one demand-resource scenario may lead to unacceptable and inappropriate schedules under a different demand-resource scenario. For example, allocation of resources in strict priority order, without regard to other factors, is likely to result in acceptable schedules when the demand to resource availability ratio is low to moderate. The same policy is likely to result in totally unacceptable schedules for lower priority customers when the demand to resource availability ratio is high. Often, which scheduling policy is appropriate for a given demand-resource scenario is unclear. Furthermore, although a scheduling policy generally implies a unique schedule effectiveness criterion, acceptability of a schedule depends on multiple schedule effectiveness criteria.

Therefore, to develop acceptable operational schedules for the use of SN resources, the Network Control Center must:

- Create alternate schedules using alternate scheduling policies
- Evaluate alternate schedules based on multiple schedule effectiveness criteria

This paper identifies alternate scheduling policies and discusses evaluation of the alternate schedules they produce, based on criteria associated with these policies.

2. Alternate Scheduling Policies

Many parameters are associated with the SN resource scheduling problem. Some more important parameters are

- NASA-assigned customer priorities
- Service duration flexibility in terms of nominal and minimum duration
- SN commitments in terms of specified success rates

Potential interdependencies among these parameters imply many possible mutually exclusive scheduling policies:

- Maximize nominal duration events in priority order
- Maximize near-nominal duration events in priority order
- Maximize minimum duration events in priority order
- Satisfy specified event success rates at near-nominal duration in priority order
- Satisfy specified event success rates at nominal duration in priority order
- Satisfy specified time scheduled success rates at nominal duration in priority order
- Satisfy specified time scheduled success rates at near-nominal duration in priority order
- Balance event success rates at near-nominal durations
- Balance event success rates at nominal durations
- Balance schedule time success rates

A scheduling algorithm used for generating schedules should be consistent with the applicable scheduling policies. In addition, even when scheduling policy specific algorithms are available, multiple schedules based on different scheduling policies may need to be generated because the choice of the appropriate scheduling policy for a given demand-resource scenario is not clear.

When scheduling algorithms which are not explicitly consistent with applicable scheduling policies are used, generation of multiple schedules with different scheduling heuristics, heuristic control parameters and/or algorithms is essential. Resulting multiple schedules must then be compared on the basis of the criteria relevant to the applicable scheduling policies. The following paragraphs describe those criteria, define metrics for comparing schedules based on those criteria, and describe the use of the metrics for comparing schedules based on sets of the criteria.

3. Evaluation of Alternate Schedules

Scheduling success rate (that is, percent or total number of requests scheduled) has been the traditional criterion for schedule effectiveness. Scheduling success rate is an adequate measure of schedule effectiveness when the initial scheduling policies are unimportant or do not affect the resulting schedule, for example, in a demand-resource scenario in which every request is successfully scheduled at the requested duration. However, scheduling success rate alone does not adequately measure schedule effectiveness under demand-resource scenarios that requires rejection or scheduling of many requests at reduced durations.

For example, consider two schedules: Schedule X and Schedule Y. Figure 1 shows plots of the cumulative number of requests scheduled versus priority for the two schedules. Schedule X is biased in favor of higher priority users, whereas Schedule Y is biased in favor of lower priority users. Clearly, Schedule Y should not be selected if the primary scheduling policy is to maximize number of scheduled events in priority order. If the policy is to maximize the scheduled requests, then Y, which has more requests scheduled than X, should be selected.

Which scheduling policy is appropriate for a given demand-resource scenario? Acceptability of a schedule often depends on multiple schedule effectiveness criteria, even though the initial scheduling policy implies a unique schedule effectiveness criterion. Some criteria relate directly to the various initial scheduling policies; the rest relate to other desirable qualities in an effective schedule.

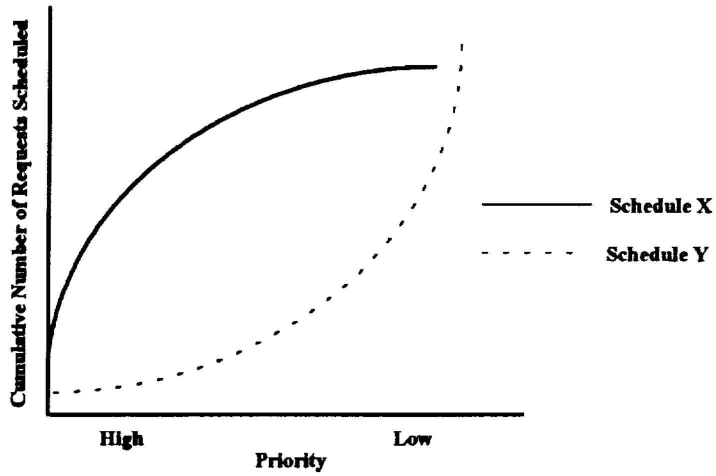

Figure 1. Schedule X Satisfies Priorities Better Than Schedule Y

3.1 Evaluation Criteria

The criteria for schedule evaluation discussed in this paper are categorized as follows:

- Maximizing success rates in priority order
- Satisfying specified success rates in priority order
- Balancing success rates among customers
- Minimizing the impact of undesirable gaps between successive scheduled requests

None of these criteria involves the success of any specified individual requests. Specific requirements for the success of individual requests are assumed to be ensured by means of appropriate algorithm and problem definition, that is, correct definition of requests and enforcement of constraints.

3.1.1 Maximizing Success Rates in Priority Order

Scheduling success may be defined as the percentage of events scheduled or as the percentage of requested time scheduled. Maximizing success rates in priority order means maximizing the success rate of a given customer without considering the success rates of lower priority customers.

These definitions lead to two distinct criteria:

- Maximization of event success rates in priority order
- Maximization of requested time success rates in priority order

For schedule comparisons based on one of these criteria, comparing the success rates in priority order is sufficient. A schedule that has the maximum success rate for the highest priority best satisfies the criterion. However, strict adherence to only one criterion could result in failure to consider schedules that may be nearly as good for the highest priority and superior for lower priorities. Table 1 shows three sample schedules, with the number of events requested and scheduled for each priority.

		Number of Events Scheduled		
Priority	Number of Events Requested	Schedule A	Schedule B	Schedule C
1	100	90	91	80
2	100	80	75	85
3	100	60	65	65
4	100	45	40	40

Table 1. Schedule Statistics by Priority for Three Sample Schedules

Comparing the schedules strictly on the basis of maximization of success rates in priority order results in selecting schedule B to be the best: Schedule B has 91 priority-1 scheduled events, as compared to 90 priority-1 scheduled events for Schedule A. However, such a comparison does not indicate the nearness of Schedules A and B, and it fails to recognize that Schedule A has gained five priority-2 events and lost only one priority-1 event and five priority-3 events. Such gains and losses between adjacent priorities among different schedules make the near equivalence among schedules difficult to recognize from the raw information (as shown in Table 1), especially when the number of priorities is large. Near equivalence of schedules is more easily recognizable from a table of cumulative events scheduled in priority order. Table 2 presents the cumulative events scheduled for the example shown in Table 1.

		Cumulative Number of Events Scheduled		
Priority	Cumulative Number of Events Requested	Schedule A	Schedule B	Schedule C
1	100	90	91	80
2	200	170	166	165
3	300	230	231	230
4	400	275	271	270

Table 2. Cumulative Events Scheduled, by Priority for Sample Schedules A, B, and C

Table 2 shows the following:

- Schedules A and B are nearly the same and Schedule C is substantially worse when the number of events scheduled for priority 1 alone are compared.
- Schedule A is substantially better than Schedules B and C when cumulative number of events scheduled for priorities 1 and 2 are compared.
- All three schedules are nearly the same when cumulative number of events scheduled for priorities 1 through 3 are compared.
- Schedule A is substantially better than Schedules B and C when cumulative number of events scheduled for priorities 1 through 4 are compared.

This analysis suggests that schedule A is either nearly the same as or better than Schedules B and C at every priority level and that Schedule A is better than Schedule B, even though Schedule B best satisfies maximization of event success rates in strict priority order. Figure 2 presents the information from Table 2 in graph form and illustrates the desirability of Schedule A over Schedules B and C.

The preceding analysis led to a definitive conclusion on a simple example with four priorities and three schedules. However, it would be difficult to come to any definitive conclusion using tables like Table 2 and charts like Figure 2 when the number of priorities and the number of schedules are large. The chart would have crisscrossing plots, rendering visual recognition of an overall better schedule a difficult task. Furthermore, this type of analysis based on subjective judgment is not suitable for automated comparison of schedules. Automated comparison of schedules based on maximization of success rates in priority order requires a more comprehensive method involving numerically computable evaluation metrics. The following few paragraphs suggest such metrics.

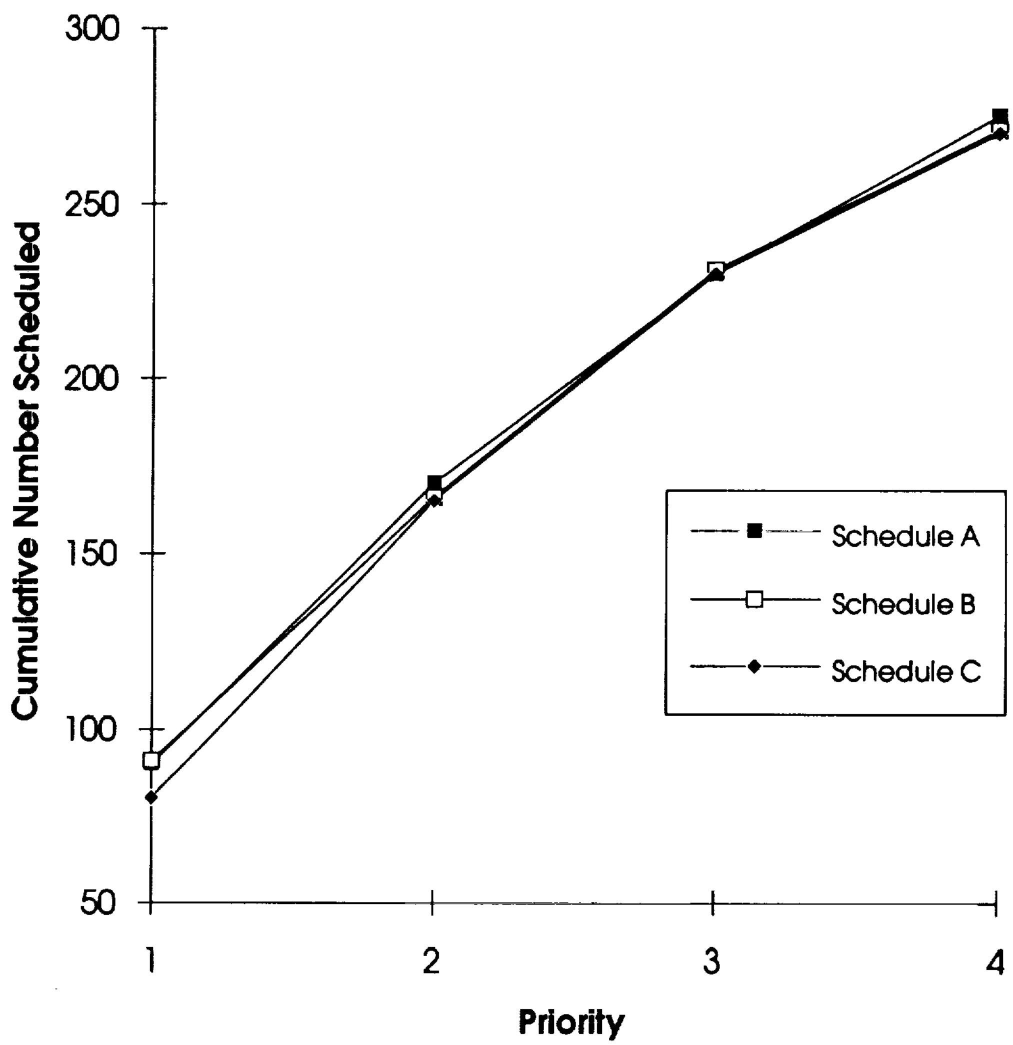

Figure 2. Cumulative Number of Events Scheduled Versus Priority

Maximization of event success rates in strict priority order implies that each event for a given priority is ***more important*** than all lower priority requests combined. In this scenario, request for a given priority has a value that is the sum of the values of all requests for lower priorities. If v_i is the value of an ith priority event, and n_i is the number of requested ith priority events, then

$$v_p = 1$$

$$v_i = \sum_{k=i+1}^{p} v_k . n_k \text{ for } i < p$$

Values computed in this way grossly overstate the differences in values of events of adjacent priorities; they also disallow compensating for any loss of higher priority events with a gain in the number of lower priority events scheduled (as in the example.) In reality, the differences in relative values of events of adjacent priorities are much smaller than those given here.

The true differences can be expected to be consistent with compensation of loss for higher priority events by a larger gain for adjacent lower priority events. A practical comparison of schedules should be based on values of events that more reasonably represent the true differences between adjacent priorities, without understating the differences between widely separated priorities. Setting event values proportional to the total number of requested events for lower priorities is suggested here as a more reasonable representation of the true differences between values of events of different priorities. Mathematically, if N_i is the value of an ith priority event, and n_i is the number of requested ith priority events, then

$$N_p = 1$$

$$N_i = \sum_{k=i+1}^{p} n_k \text{ for } i < p$$

where

p = number of priorities

Therefore, the following metric, based on the weighted sum of scheduled events with weights equal to N_i, is a reasonable suggestion for comparing schedules based on maximization of scheduled event success rates in priority order:

$$\rho = \frac{\sum_{i=1}^{p} s_i N_i}{\sum_{i=1}^{p} n_i N_i}$$

where

ρ = degree of maximization of event success rates in priority order

s_i = number of requests scheduled for ith priority

The denominator in the formula represents the maximum possible value of a schedule and is used to limit the value of the metric ρ between 0 and 1.

Similarly, assuming that the value of a unit of requested time for a given priority is proportional to the combined requested time for all lower priorities, the metric for comparing schedules based on maximization of requested time success rates in priority order can be defined as:

$$\tau = \frac{\sum_{i=1}^{p} z_i T_i}{\sum_{i=1}^{p} t_i T_i}$$

where

τ = degree of maximization of requested time success rates while enforcing priorities

z_i = time scheduled for ith priority

t_i = time requested for ith priority

$T_p = 1$

$T_i = \sum_{k=i+1}^{p} t_k$ for $i < p$

Schedules A, B, and C from Table 1 by applying the previously defined metric ρ for maximization of event success rates in priority order produces the schedule rankings shown in Table 3. The resulting rankings clearly indicate that Schedules A and B are nearly equally preferable; however, the slightly higher metric value of Schedule A appears to have captured the advantage of the additional five priority-2 events over the loss of a single priority-1 event.

	Schedule A	Schedule B	Schedule C
Value of ρ	0.816	0.812	0.791
Rank	1	2	3

Table 3. Metric Values and Rankings of Sample Schedules A, B, and C

3.1.2 Satisfying Specified Success Rates in Priority Order

The metrics for comparing schedules using the criterion of satisfying specified success rates in priority order can be based on a rationale very similar to that used for the metrics for maximization of success rates in priority order.

In terms of satisfying specified event success rates in priority order, scheduling more than the specified event success rate for any priority has no value. That is, any events scheduled over and above the minimum specified success rates have *zero* values. Therefore, the maximum number of events that must be scheduled to meet the criteria for ith priority with a specified success rate of q_i is $n_i q_i$. Assuming that a value W_i of an ith priority event is proportional to the number of events required to satisfy the specified success rates at all the lower priorities,

$$W_p = q_p$$

$$W_i = \sum_{k=i+1}^{p} n_k q_k \text{ for } i < p$$

where

$$0 \leq q_i \leq 1$$

The following metric based, on the weighted sum of the *nonzero*-value scheduled events with weights equal to W_i, is a reasonable suggestion for comparing schedules based on satisfying specified event success rates in priority order:

$$\gamma = \frac{\sum_{i=1}^{p} Min(s_i, n_i q_i)\, W_i}{\sum_{i=1}^{p} n_i q_i W_i}$$

where

γ = degree of satisfying specified event success rates in priority order

Similarly, from the perspective of satisfying specified time scheduled success rates in priority order, scheduling more than the specified requested time success rate for any priority has no value. That is, any time scheduled over and above the minimum specified success rates has *zero* value. Therefore, the maximum time that must be scheduled to meet the criteria for ith priority with a specified success rate of q_i is $t_i q_i$. Assuming that a value U_i of an ith priority event is proportional to the time required to satisfy the specified success rates at all the lower priorities,

$$U_p = q_p$$
$$U_i = \sum_{k=i+1}^{p} t_k q_k \text{ for } i < p$$

The metric, based on the weighted sum of the valuable (*nonzero* value) scheduled events with weights equal to U_i for comparing schedules based on satisfying specified scheduled time success rates in priority order is

$$\varphi = \frac{\sum_{i=1}^{p} Min(z_i, t_i q_i) U_i}{\sum_{i=1}^{p} t_i q_i U_i}$$

where

φ = degree of satisfying specified scheduled time success rates in priority order

3.1.3 Balancing Success Rates Among Customers

The scheduling criterion of balancing success rates among customers implies a goal to generate a schedule with equal success rates for all customers. This criterion, when applied literally, is fully satisfied even by a schedule with zero success rates for all customers. Hence, this criterion should be interpreted to imply balanced success rates among customers with maximum combined success rate of all customers.

The combined event success rate c of a schedule can be calculated as

$$c = \frac{\sum_{i=1}^{p} s_i}{\sum_{i=1}^{p} n_i}$$

The success rates for all customers are balanced when the success rate for each customer equals c. Therefore, scheduling more events for a customer than are necessary to provide a success rate of c has *zero* value from the standpoint of balancing event success rates. Furthermore, scheduling additional events for a customer with a success rate greater than or equal to c could have an adverse impact on customers with success rates lower than c. Therefore, an event scheduled for any customer over and above the required number to provide a success rate of c can be assumed to have *zero* value from the standpoint of balancing event success rates. Hence, the number of *nonzero*-value events scheduled for the ith priority customer can be calculated as $Min\ (s_i, cn_i)$. The total number of *nonzero*-value events scheduled can be calculated as

$$\sum_{i=1}^{p} Min(s_i, cn_i)$$

It naturally follows that the total number of *nonzero*-value events scheduled is a reasonable basis for a metric to compare the schedules based on the criterion of balancing event success rates among customers. Hence, the metric α for comparing schedules based on the criterion of balancing event success rates among customers can be defined as

$$\alpha = \frac{\sum_{i=1}^{p} Min(s_i, cn_i)}{\sum_{i=1}^{p} n_i}$$

From the perspective of balancing time scheduled success rates, scheduling more time for a customer than necessary to provide a success rate d, where

$$d = \frac{\sum_{i=1}^{p} z_i}{\sum_{i=1}^{p} t_i},$$

has a *zero* value from the standpoint of balancing time scheduled success rates. Therefore, the total *nonzero* value time scheduled can be calculated as

$$\sum_{i=1}^{p} Min(z_i, dt_i)$$

As a result, the metric β for comparing schedules based on the criterion of balancing time scheduled success rates among customers can be defined as

$$\beta = \frac{\sum_{i=1}^{p} Min(z_i, dt_i)}{\sum_{i=1}^{p} t_i}$$

3.1.4 Minimizing the Impact of Undesirable Gaps Between Successive Scheduled Requests

A schedule that includes an event for every customer request is obviously the schedule that best satisfies the customer. The time gaps between each consecutive pair of events (i.e., between events 1 and 2, 2 and 3, 3 and 4, and so forth) in such a schedule are within the maximum gap size implicitly intended by the customer, based on the requested start times and tolerances. When some of the customer's requests are not scheduled, these gaps exceed the customer's intended maximum gap size. Figure 3 shows an excessively long gap could mean loss of data for the customer.

A customer is less likely to have adverse impacts when the customer's scheduled events and declined events are more evenly interspersed in time than when otherwise. For example, a customer who has requested six events is more likely to have an adverse impact when the last three requested events are declined than when every other requested event is declined, even though the number of declined events is three in both cases . Generally, most customers can endure one declined event between two scheduled events without a significant loss of data; however, customers who have two or more declined events in succession are likely to experience a significant loss of data. Therefore, the amount of data loss in an excessively long gap can be expected to be proportional to the length of the gap in excess of twice the sum of the duration and the intended gap. Further, the duration of the event may be assumed to represent the amount of data transmitted during the requested event.

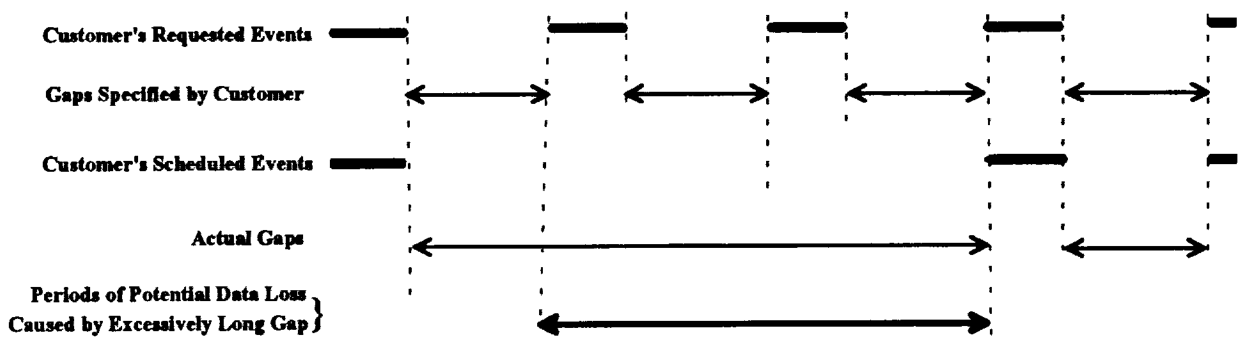

Figure 3. Potential Loss of Data Because of Excessively Long Gaps

Mathematically, if t is the average duration of the requested events for a customer, g the maximum gap intended by the customer, and e the length of the gap between any two successive scheduled events for the customer, the loss of data l because of gap e can be approximated by

$$l(e) = \frac{Max\{0, e-2(g+t)\}t}{(g+t)}$$

This formula can be used to estimate the data loss associated with each gap for each customer. The total data loss associated with a schedule, λ, is estimated by summing the data losses associated with all gaps for every customer. Among all schedules, the schedule having the minimum total data loss is the most desirable.

3.2 Schedule Evaluation Example

Metrics defined in the preceding sections were used to evaluate schedules for a scenario representative of the expected 1998 SN demand, assuming availability of 4 single access antennas. Table 4 summarizes the demand scenario which represents support of one week's demand for 11 customers requesting a total of 1,555 events and 42,084 minutes. All the unmanned flight customers are assumed to have duration flexibility. In addition, some customers are assumed to have start time flexibility. The flexibility parameters used in the scenario are based on discussions with the associated customers and, therefore, are representative of the true flexibility available to these customers.

Customer	Priority	Events Requested	Nominal Dur. (Min)	Minimum Dur. (Min)	Flexibility (%)	Requested Time (Min)	Specified Succ. Rate(%)
STS-121	1	208	50	50	0	9,119	90
SSFreedom	2	202	50	50	0	9,305	60
SIRTF	3	195	22	12	0	4,290	60
AXAF-S	4	194	22	12	0	4,268	60
EOS-AM1	5	218	15	10	67	3,270	60
HST	6	195	18	15	0	3,510	60
GRO	7	52	30	16	40	1,560	60
TOPEX	8	28	18	14	100	504	60
XTE	9	50	25	15	100	1,250	60
Landsat-7	10	109	24	14	0	2,616	60
TRMM	11	104	23	20	87	2,392	60
Total		1,555				42,084	

Table 4. Demand Scenario for Schedule Evaluation Example

Seven different schedules, P through V, were generated for the example using different scheduling strategies. Computer Sciences Corporation's Space Network Scheduling Prototype generated the schedules. Table 5 shows the number of events scheduled for the seven schedules. Table 6 shows the amount of time scheduled for the seven schedules.

Table 7 shows the values of metrics for the seven different schedules for the example. Table 8 shows the ranking of the seven schedules for each of the metrics. As is apparent from Table 8, different schedules are ranked number 1 for different metrics. For example, Schedule U is the best schedule in terms of maximizing events in priority order (ρ), whereas Schedule Q is the best in terms of maximizing scheduled time in priority order (τ). Even though Schedule Q maximizes scheduled time in priority order, it ranks very low in terms of maximizing the number of events scheduled (see Table 8).

Customer	Schedule						
	P	Q	R	S	T	U	V
STS-121	206	207	192	189	143	207	207
SSFreedom	202	202	186	185	140	202	202
SIRTF	109	152	141	148	146	187	164
AXAF-S	106	154	151	150	157	182	166
EOS-AM1	119	140	148	155	175	214	188
HST	66	92	117	104	149	155	133
GRO	7	27	34	28	40	47	35
TOPEX	25	25	25	26	26	28	28
XTE	25	36	42	39	47	47	44
Landsat-7	14	45	58	56	89	89	69
TRMM	15	47	63	59	87	91	69
All	894	1127	1157	1139	1199	1449	1308

Table 5. Number of Events Scheduled for the Example

Customer	Schedule						
	P	Q	R	S	T	U	V
STS-121	9057	9090	8372	8282	5991	7334	8617
SSFreedom	9305	9305	8506	8476	5972	7237	8838
SIRTF	2398	3111	2844	3058	2998	2930	2744
AXAF-S	2332	3119	2993	3081	3200	3104	2835
EOS-AM1	1785	2010	2138	2245	2513	2696	2394
HST	1188	1524	1977	1746	2537	2318	1935
GRO	210	608	817	680	981	932	764
TOPEX	450	446	443	464	460	460	462
XTE	625	815	951	906	1086	922	850
Landsat-7	336	820	1076	1057	1882	1567	1300
TRMM	345	808	1147	1100	1754	1626	1203
All	28031	31656	31264	31095	29347	31126	31942

Table 6. Time Scheduled for the Example

Metric		Schedule						
		P	Q	R	S	T	U	V
Maximize events in priority order	ρ	0.7411	0.8438	0.8124	0.8102	0.7400	0.9645	0.9072
Maximize time in priority order	τ	0.8388	0.8874	0.8327	0.8328	0.6761	0.7677	0.8512
Satisfy specified event success rates in priority order	γ	0.9509	0.9906	0.9999	0.9949	0.8518	1.0000	1.0000
Satisfy specified scheduled time success rates in priority order	ϕ	0.8470	0.8932	0.8372	0.8368	0.6751	0.7697	0.8561
Balance event success rates	α	0.4584	0.6343	0.6871	0.6716	0.7460	0.9018	0.7910
Balance scheduled time success rates	β	0.5186	0.6428	0.6649	0.6620	0.6741	0.7085	0.6746
Loss of data	λ	4210	1119	3693	4327	14775	230	397

Table 7. Evaluation Metrics for Schedule Evaluation Example

Metric		Schedule						
		P	Q	R	S	T	U	V
Maximize events in priority order	ρ	6	3	4	5	7	1	2
Maximize time in priority order	τ	3	1	5	4	7	6	2
Satisfy specified event success rates in priority order	γ	6	5	3	4	7	1	1
Satisfy specified scheduled time success rates in priority order	ϕ	5	4	1	3	7	6	2
Balance event success rates	α	7	6	4	5	3	1	2
Balance scheduled time success rates	β	7	6	4	5	3	1	2
Loss of data	λ	5	4	2	6	7	1	3

Table 8. Schedule Rankings Based on the Evaluation Metrics for the Example

Therefore, selecting the best schedule depends on whether a single criterion or multiple criteria are to be used. Multiple criteria requires a more complex selection procedure. First, weights proportional to the level of importance must be assigned to each relevant criteria. Then, the schedule that minimizes the weighted sum of ranks is the most satisfactory schedule, based on the selected multiple criteria.

For example, if it is equally important to maximize time (τ) and maximize events (ρ) in priority order, then the weights assigned for maximizing events in priority order and for maximizing scheduled time in priority order should each be set at 0.5. Schedules Q and V are the best schedules in this case. Because in this case Schedules Q and V are equivalent from the perspective of maximizing events in priority order and maximizing scheduled time in priority order, the two schedules could be further examined on secondary criteria for example, balanced success rates. In such a case, Schedule V would be superior to Schedule Q.

4. Conclusion

This paper has described a formal mathematical procedure which allows automated comparison of schedules based on scheduling policy specific criteria.

Evaluation of schedules based solely on the total number of scheduled events has been shown to be inadequate to fulfill the current SN priority order scheduling policy. As the demand for SN resources increases, evaluation of schedules based on total number of scheduled events will become even more inadequate and the SN will need to use alternative criteria. This paper identified several alternative scheduling policies and derived schedule evaluation metrics which are directly related to those policies as well as the currently used priority order scheduling policy. It described a method to use these metrics in comparing schedules based on multiple criteria. The method which is suitable for automated schedule comparison involves the following steps:

- Calculation of the metrics for all possible criteria (α, β, γ, etc.)
- Ranking the schedules based on each of these criteria
- Assigning weights to each of these criteria based on relative importance each criterion
- Finding the weighted sum of the criteria specific ranks
- Selection of the schedule which minimizes the weighted sum of ranks.