Schedule metrics

Brief preface

We are able to calculate a score for the schedule using any metric that has been implemented (so far, just **Throughput**, **Resource utilization**, and **Workload distribution**).

We have access to the state of the satellite across time. We can calculate a score for the schedule across any arbitrary time range (this is feasible as we store, not the actual state of the satellite, but the change in state - the delta. This also allows for a distributed scheduler system, scheduling concurrently across different periods of time. I digress)

In what follows, 'asset' refers to either a ground station or a satellite. It is a general term that encompasses them both.

Schedule metric calculation overview

For each metric, we calculate the metric values for each individual schedule, and then perform a min-max normalization across all schedules in our population (could be just all schedules we want to compare) to create a "grade" value that is between 0 and 1, indicating how good the schedule is compared to the rest of the population.

Say $M(schedule)$ is a metric that produces a measure for a given schedule of how well it does on that metric. The $grade$ of the schedule $G(schedule)$ is measured simply using min/max normalization over the schedule population of interest S.

$$metric\_measures = \{M(s) | s \in S\}$$

$$G(schedule) = \frac{M(schedule) - min(metric\_measures)}{max(metric\_measures) - min(metric\_measures)}$$

Below are descriptions of how exactly we calculate each of the three metrics - Throughput, Resource utilization, and Workload distribution.

Throughput calculation

Currently, throughput is simply calculated as the sum of the priority of all activities scheduled in that schedule. So if we want the throughput of a schedule, we go through all assets, and for all scheduled activities scheduled to take place on that asset, within our time period of interest, we sum up the priorities of all those scheduled events. So this allows us to describe throughput as the weighted count of the number of scheduled activities, where the weight is the priority of the scheduled activities under consideration.

More formally, where A is the set of all assets in our schedule, and E_{asset} is the set of all events scheduled to take place in on that asset within our time period of interest, and $P(event)$ denotes the priority of an $event$, throughput is defined as follows:

$$T(schedule) = \sum_{a \in A} \sum_{event \in E_a} P(event)$$

Resource utilization calculation

Currently, we only consider storage utilization in calculating resource utilization. However, this will, in the future, be adjusted to incorporate other things such as how much a ground station is being used. This will be measured by how much time in total it spends uplinking/downlinking images - essentially it's busy time. (this should be fairly easy to implement, but is a low priority task at the moment)

Asset resource utilization

We define $R_{asset}(t)$ as the resource utilization of a single asset '$asset$' within a schedule at the specified time t. We also define $R(asset)$ as signifying the resource utilization of an asset, and $R(schedule)$ as signifying the resource utilization of a schedule (which comprises of multiple assets)

Where $\Delta Storage(asset, t)$ is the change in the asset's storage at a given time and $SCAP$ is the storage capacity of the asset, Resource utilization of an asset **as a given time** is defined as follows (we only consider satellites for now):

$$R_{asset}(t) = \begin{cases} type(asset) = sat & \frac{\Delta Storage(asset,t)}{SCAP(asset)} \\ type(asset) = gs & 0 \end{cases}$$

Resource utilization for an asset, $R(asset)$ is defined simply as the average resource utilization for that asset across time:

$$R(asset) = avg(\{R_{asset}(t) | t \in desired\_time\_range\})$$

Schedule resource utilization

The resource utilization of a schedule is simply just the average resource utilization, where the average is taken individually across the different asset classes (average for .

Where SAT is the set of all satellites in the schedule and GS is the set of all ground stations in the schedule, we define resource utilization of a schedule, $R(schedule)$, as such:

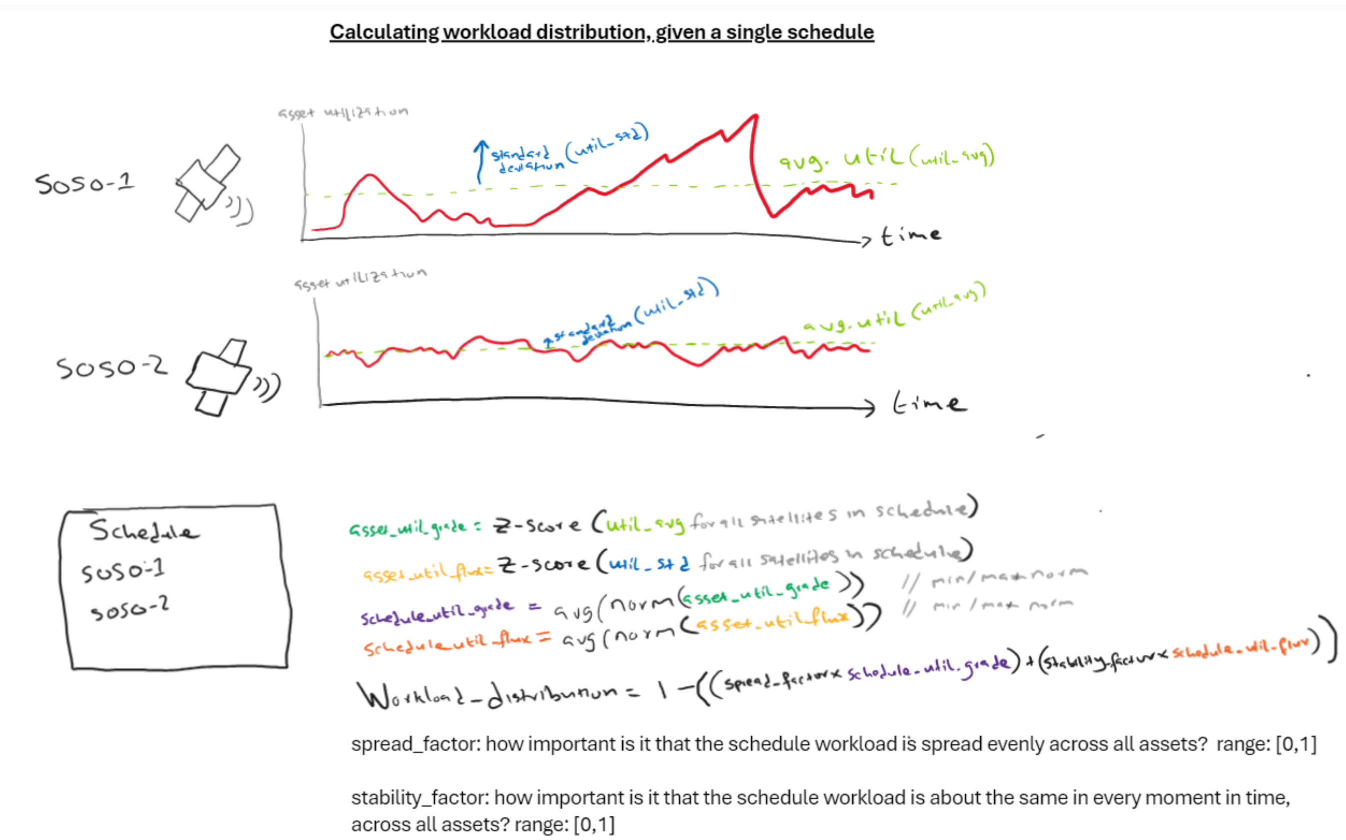
$$R(schedule) = avg(\{R(satellite) | satellite \in SAT\}) + avg(\{R(gs) | gs \in GS\})$$

We separate the averaging of resource utilization across the different assets to account for the possible distributional shift in the values of resource utilization across the different asset classes.

I am not quite satisfied with this formula for schedule resource utilization though. It may need some work in the future, to be honest to refine how it is calculated (maybe weighting by asset type, or including other factors such as power usage). (side note to self: Since this is meant to be a measure, and not a grade, since we only have a population of one "schedule" under consideration, and it is possible to have a population of only one asset of a given type, z_score should be used cautiously, or else the value you get back from the measure $R(schedule)$ may be useless. We still need a way to better account for the possible distributional shift across the value of resource utilization across different asset classes tho, but z_score might not be the way because of this problem.)
This formula works alright for now tho, as we are only considering satellites for the time being.

Workload Distribution calculation?

The diagram below **attempts** to clarify how the workload distribution is calculated. You could perhaps use it as a reference for understanding the explanation that follows.



Workload distribution is calculated based off of the resource utilization of the different assets in the schedule, so we start by remembering that the resource utilization of an asset is represented by the function $R(asset)$.

The resource utilization of an asset **R(asset)** is simply the asset's average utilization. This fails to capture information about outliers - situations where a lot of tasks are assigned to an asset sometimes, but very few at other times. Instead of having it do a lot, we want to either spread out that work across the same satellite, or assign some of its tasks to other satellites (barring extreme situations where it is just necessary that there are those outlier peaks, like if it is only possible to take some image with some specific satellite at some specific time, so inevitably the load is high at that time)

To capture this information, we have another measure - simply the standard deviation
Let us then also define $R_{std}(asset)$ as the standard deviation of the asset's resource usage across time:

$$R_{std}(asset) = \sigma(\{R_{asset}(t) | t \in desired\_time\_range\})$$

We thus have to consider the raw resource utilization $R(asset)$, and also the flux in resource utilization $R_{std}(asset)$, to accurately understand how the workload is distributed across the different assets.

Because assets of different classes have different distributions of resource utilization and utilization flux, we standardize these values by using the z-score function. Hence we have:

$$UtilizationGrade(asset) = \begin{cases} type(asset) = sat & \frac{R(asset) - mean_{sat \in SAT}(R(sat))}{\sigma_{sat \in SAT}(R(sat))} \\ \\ type(asset) = gs & \frac{R(asset) - mean_{gs \in GS}(R(gs))}{\sigma_{gs \in GS}(R(gs))} \end{cases}$$

where SAT is the set of all satellites in the schedule, and GS is the set of all ground stations in the schedule. The use of z-score also, more importantly, helps us better reflect the fact that we are trying to measure the distribution of the resources, so $UtilizationGrade(asset)$ is actually measuring "how much more resources than the average amount of resources is this asset using?"

We also calculate the standardized resource utilization flux of an asset in the same manner, using the z-score function. This is for the same purpose as stated above - to standardize across different asset types, and more importantly, to measure how much more this asset's resource utilization fluxes than the average amount of flux:

$$UtilizationFlux(asset) = \begin{cases} type(asset) = sat & \frac{R_{std}(asset) - mean_{sat \in SAT}(R_{std}(sat))}{\sigma_{sat \in SAT}(R_{std}(sat))} \\ \\ type(asset) = gs & \frac{R(asset) - mean_{gs \in GS}(R(gs))}{\sigma_{gs \in GS}(R(gs))} \end{cases}$$

We now want to measure a single utilization grade and utilization flux for the whole schedule.
To measure a single $UtilizationGrade(schedule)$ value for a schedule, we basically average the

utilization across all assets (we can safely do this, all the values have all been standardized). For reasons we will see later though, we want this average to be guaranteed to be between 0 and 1, so before we take the average, we normalize the values using min/max normalization.

In essence, what we are trying to do is this:

$$UtilizationGrade(schedule) = avg(norm(\{UtilizationGrade(asset) \mid for\ all\ assets\}))$$

But providing the full detail, $UtilizationGrade(schedule)$ is calculated as follows:

$$min\_util\_grade = min_{asset \in A}(UtilizationGrade(asset))$$

$$max\_util\_grade = max_{asset \in A}(UtilizationGrade(asset))$$

$$UtilizationGrade(schedule) = \frac{1}{|A|} \sum_{asset \in A} \frac{UtilizationGrade(asset) - min\_util\_grade}{max\_util\_grade - min\_util\_grade}$$

where A is the set of all assets in the schedule.

We perform the same operation of normalization and averaging to get a final value for $UtilizationFlux(schedule)$:

$$min\_util\_flux = min_{asset \in A}(UtilizationFlux(asset))$$

$$max\_util\_flux = max_{asset \in A}(UtilizationFlux(asset))$$

$$UtilizationFlux(schedule) = \frac{1}{|A|} \sum_{asset \in A} \frac{UtilizationFlux(asset) - min\_util\_flux}{max\_util\_flux - min\_util\_flux}$$

Because both $UtilizationGrade(schedule)$ and $UtilizationFlux(schedule)$ are normalized, their values are always between 0 and 1. This is useful when weighting as we will soon see.

The final equation for workload distribution is as follows:

$$WD(schedule) = 1 - (spread\_factor \times UtilizationGrade(schedule)) + (stability\_factor \times Utilization$$

where we have the constraint that $spread\_factor \in [0, 1]$, $stability\_factor \in [0, 1]$, and

$$spread\_factor + stability\_factor = 1$$

This gives us a 0-1 value showing how well resources are distributed across the assets in the schedule.

The $spread\_factor$ is a weight that signifies how important it is that the average resource utilization is about the same across all assets
The $stability\_factor$ is a weight that signifies how important it is that the variation in resource utilization across time is about the same across all assets

If utilization grade for the schedule is high, that means that on average, the schedule's resource utilization is further from the average resource utilization - we want resource utilization evenly spread out, so we want a low value for utilization grade. It is a similar story for utilization flux. If utilization flux is high, that means that on average, the schedule's resource utilization fluctuates more than average. we want it to be spread out in cases where possible, so we want a low value for utilization flux as well. That is why we have the $1-\ldots$ at the start, to reverse the directionality of the score.