Opportunity-Adaptive QoS Enhancement in Satellite Constellations: A Case Study

Ann T. Tai   Kam S. Tso
IA Tech, Inc.
10501 Kinnard Avenue
Los Angeles, CA 90024

Leon Alkalai   Savio N. Chau
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

William H. Sanders
ECE Department
University of Illinois
Urbana, IL 61801

Abstract

Systems that are formed by massively distributed mobile resources, such as satellite constellations, often provide mission-critical functions. However, many existing quality-of-service (QoS) management concepts cannot be applied to those systems in a traditional way, due to the continuously changing readiness-to-serve of their mobile resources. In this paper, we describe a case study that investigates a method called "opportunity-adaptive QoS enhancement (OAQ)." Driven by an application-oriented QoS objective, the method focuses on a solution that permits a structurally degraded constellation to deliver signal-position-determination (geolocation) results with the best possible quality. More specifically, the OAQ algorithm enables iterative geolocation accuracy improvement by letting neighboring satellites coordinate, and by progressively expanding the scale of this coordination in the window of a dynamically determined opportunity. For effectiveness demonstration, we define a QoS measure and solve it analytically. The results show that the OAQ approach significantly enhances a constellation's ability to deliver service with the quality at the high end of a QoS spectrum, even in presence of structural degradation.

1 Introduction

As micro-electro-mechanical systems and wireless networking technologies advance, it is anticipated that many critical applications will rely on a class of systems that are composed of massively distributed mobile resources. Examples of such systems include micro-UAV (unmanned aerial vehicle) swarms that perform coordinated actions in hazardous environments for damage control or monitoring, and micro-satellite constellations in which hundreds of nodes coordinate for formation flying, surveillance, and communication. While their quality of service (QoS) is usually mission-critical, they are often vulnerable to failures caused by adverse space-environment conditions, physical or other types of inadvertent faults, and malicious attacks. In addition, due to their mobile nature, the *readiness-to-serve* [1] of individual computing resources in those systems changes dynamically and continuously, making traditional redundancy-based fault tolerance schemes and QoS management concepts difficult to apply.

In spite of their importance, fault tolerance and QoS management for systems built on massively distributed mobile resources have not yet received enough attention. To the best of our knowledge, aside from the efforts concerning reliable inter-satellite and ground-to-satellite communications (see [2, 3], for example), no significant work has been devoted to method development for mitigating the effects on application-oriented QoS of satellite-failure-caused, constellation-structure degradation.

With the above motivation, we carry out a case study to investigate a framework that allows us to exploit mobile resource redundancy to mitigate the effects of a constellation's structural degradation on geolocation accuracy (i.e., the accuracy of locating a signal-emitter on the earth). Accordingly, our QoS objective is to guarantee the timely delivery of geolocation results with the best possible accuracy. As this QoS objective necessitates a cohesive formulation of fault-tolerant satellite constellation operation, our method derivation is based on the integration of concepts and techniques across the areas of satellite constellation and fault-tolerant computing. Specifically, it has been shown in the satellite research literature that sensor measurements accumulated by neighboring satellites that successively fly over a signal emitter can support an iterative weighted least-square algorithm and thereby enable a mechanism called *sequential localization* to reduce errors in signal-position determination [4, 5]. Although the original purpose of this mechanism was to circumvent the difficulties caused by satellite capacity inadequacy (e.g., an insufficient number of onboard sensors) or noisy space environments, the synergy between the theoretical basis of sequential localization and the concepts of data diversity [6] and environment diversity [7] associated with fault-tolerant computing suggests that sequential localization can be judiciously exploited for tolerating the effects of failure-caused loss of satellites on geolocation quality. We thereby develop an algorithm which lets two or more surviving satellites that *con-*

secutively revisit a signal location coordinate for iterative geolocation-accuracy enhancement, in a situation in which satellite failures reduce a constellation's "density" and make it no longer possible to let multiple satellites *simultaneously* "co-visit" the location to ensure result accuracy.

Moreover, the highly dynamic nature of satellite constellations leads us to introduce to the algorithm a concept called *opportunity-adaptive QoS enhancement* (OAQ). Accordingly, the algorithm permits the coordinated, iterative geolocation-accuracy enhancement to be carried out in an aggressive fashion, by continuously expanding the scale of the coordination among peer satellites within a "window of opportunity." From temporal perspective, the window of opportunity is dynamically determined by the alert-message delivery deadline and signal duration. From spatial perspective, the opportunity is characterized by the number of mobile resources that are able to join the coordinated iterative geolocation computation. More specifically, those resources include 1) the satellites that happen to be in the range that allows their footprints[1] to cover the signal location at the initial detection, and 2) those satellites whose routine traveling patterns bring their footprints to the target location subsequent to the initial detection and within the window of opportunity.

The central purpose of this paper is to demonstrate the effectiveness of the OAQ framework. Hence, in addition to describing the algorithm, we conduct a model-based quantitative evaluation to analyze the QoS gain from the use of the OAQ algorithm. The model is constructed based on a reference satellite constellation that is designed for detection and position localization of radio-frequency (signal) emitters [8]. Through analyzing the evaluation results, we show that the OAQ framework significantly enhances the system's ability to deliver service with the quality at the high end of an application-oriented QoS spectrum, even after a significant number of satellites are lost due to faults.

The remainder of the paper is organized as follows. Section 2 provides background information. Section 3 describes the OAQ framework in detail, followed by Section 4 which presents an analytic model and discusses the evaluation results. Concluding remarks are given in Section 5.

2 Degradable QoS in Satellite Constellations

Since our objective is to investigate fault tolerance and QoS issues in the systems that are formed by massively distributed mobile resources, the types of satellite constellations we are concerned with are LEO (low earth orbit) constellations that comprise a large number of small satellites. Moreover, we focus on tactical and strategic applications. Hence, we view accuracy of signal-position determination as a crucial QoS property of a satellite constellation.

For clarity of illustration, we use the constellation shown in Figure 1 as the reference constellation. However, the OAQ framework will also be applicable for other systems of similar types, and is anticipated to be more effective for systems built on very large populations of nodes, such as pico-satellite constellations. As mentioned in Section 1, this reference constellation is designed for geolocation of radio-frequency (RF) emitters for surveillance applications.

The constellation is formed by seven orbital planes. (Informally speaking, an orbital plane is a ring-shaped trajectory along which satellites travel around the globe.) Each of the planes consists of 14 micro-satellites that are intended to be active in service, and two in-orbit spares that can be deployed to replace any failed satellites in the same orbital plane. Therefore, the constellation consists of 98 active satellites and 14 in-orbit spares (for a total of 112 satellites).

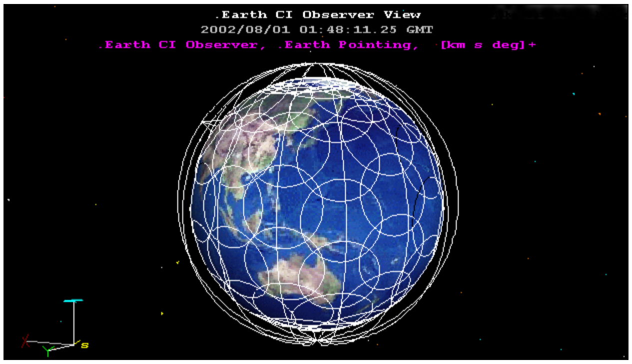
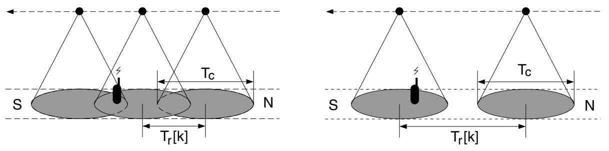


Figure 1: The Reference RF Geolocation Constellation

Figure 1 shows that when the constellation has 98 operational satellites, it offers a full earth coverage. Furthermore, every earth location will be covered by at least one satellite and a large portion of the globe (especially in the areas of high latitude) is covered by overlapped footprints. However, the geometry of the constellation will change if satellites are lost due to physical failures or malicious attacks. Specifically, when an orbital plane loses satellites after exhausting its spares, the surviving satellites will undergo a phasing adjustment so that they can be evenly distributed in the plane again. As a result, the overlapped portion of the footprints of adjacent satellites will shrink, which makes it less likely that a target will be captured simultaneously by multiple satellites. When more satellites fail, the footprints of surviving satellites will eventually become detached (we use the term "underlapping" to refer to this case in the remainder of this paper). Figures 2(a) and 2(b) illustrate the types of geometric orientation a plane may exhibit. In the figures (where we rotate the axis of the earth 90° clockwise), the top dashed line indicates an orbital plane, while the small solid dots represent the satellites traveling in that plane; the shaded ovals are the satellites' footprints and the cellular phones emitting RF signals are the assumed targets.

---

[1] The area on the earth that is covered by a satellite is referred to as the *footprint* of that satellite.



(a) Overlapping ($T_r[k] < T_c$) (b) Underlapping ($T_r[k] \geq T_c$)

Figure 2: Node-Failure-Caused Structural Degradation

As illustrated in Figures 2(a) and 2(b), we define *revisit time*, $T_r[k]$, as the time interval from the instant the center of a satellite's footprint passes a location on the earth to the instant the center of the footprint of the next satellite (in the same plane) passes the same location, given that the plane has k active operational satellites. ($T_r[k]$ can also be viewed as the distance, measured in time units, between the two satellites.) Note that K, the number of operational satellites that are actively in service in an orbital plane, is a random variable, since satellites in the plane fail over time. Further, we use the term "coverage time," denoted by T_c, to refer to the maximum amount of time that a location on the earth can be covered by the footprint of a single satellite. Note that the length of T_c can be "visualized" as the diameter of a footprint, as shown in Figures 2(a) and 2(b). From the definitions of $T_r[k]$ and T_c, it follows that the geometric orientation of the footprint trajectory of an orbital plane can be determined by the relations between $T_r[k]$ and T_c. More precisely, $T_r[k] < T_c$ and $T_r[k] \geq T_c$ imply footprint overlapping and underlapping, respectively.

The geometric orientation changes will affect the QoS of geolocation computation. In particular, when footprints overlap, it is possible that a target will be covered simultaneously by the footprints of adjacent satellites, which we call *simultaneous multiple coverage*. When two or more satellites observe a target at the same time, a measurement collection that is significantly more extensive than that from a single satellite can be obtained. With the added measurements, the ambiguity problem will practically disappear, resulting in a dramatic improvement of positioning accuracy [4]. Nonetheless, even when all satellites in the constellation are functioning, it is still possible that a target is covered by only a single satellite, as the earth is not completely covered by overlapped footprints.

When a constellation successively experiences structural degradation due to loss of satellites, footprints will eventually become underlapping, as shown in Figure 2(b). If that happens, a target will be covered by a single footprint at a time at best, thus preventing geolocation results from having high accuracy. In the worst case, a target could escape from surveillance, if 1) the signal starts when its location is not covered by any footprints, and 2) the signal stops before the nearest footprint moves to that location.

The above discussion implies that a constellation's structural degradation will lead to its QoS degradation. Furthermore, since the readiness-to-serve of each surviving satellite varies over locations and time, and since signal occurrence and duration are unpredictable, the extent to which we can pursue QoS enhancement in a structurally degraded constellation cannot be determined even if the geometric orientation of the constellation is known. In turn, those factors collectively suggest that an effective solution for QoS optimization should be opportunity-adaptive. Accordingly, we develop a framework as described in the next section.

3 OAQ Framework

3.1 Overview

It has been shown in the research literature that information from diverse sources can help resolve ambiguity in signal position determination. Those information includes earlier measurements and previously calculated position coordinates. Further, delayed position determination (termed as sequential localization) may help reduce errors in calculation because another satellite may appear in the range in time to cover the target, and additional measurements can thus be accumulated to support an iterative weighted least-square algorithm [4, 5]. Although the original purpose of sequential localization was to circumvent the difficulties associated with satellites that are not adequately equipped (with respect to quantity and capability of sensors) or to tolerate noisy space environments, the mechanism can be judiciously exploited for mitigating the effects of a constellation's structural degradation on geolocation accuracy. Specifically, we can let two surviving satellites that consecutively revisit the target coordinate for iterative position determination, in the circumstance where satellite failures reduce a constellation's "density" so that footprints become underlapping.

We can take a similar approach to QoS enhancement in the situation where the constellation has enough operational satellites such that an appreciable portion of its earth coverage is made up by overlapped footprints. Specifically, if a signal is initially detected by a single satellite, we can withhold the preliminary result and wait to see whether overlapped footprints will arrive at that location before reaching the deadline for alert-message delivery[2]. If so, simultaneous multiple coverage will ensure a high-accuracy geolocation result which requires no further satellite coordination; otherwise the preliminary result will be enclosed in the alert message and sent to the ground.

While reaching a simultaneous coverage in the overlapping case implies the attainment of a geolocation result with the best quality and thus marks the completion of QoS

---

[2] The deadline means the latest time (measured from the initial detection of a target) by which the alert message must be sent.

optimization, the iterative QoS enhancement based on sequential localization in the underlapping case can be carried out progressively. Informally speaking, as additional information from diverse sources enables further accuracy-improvement iterations, we can continue to exploit the satellites that consecutively revisit the signal location until 1) the estimated error of the geolocation result drops below a threshold, 2) the alert-message delivery deadline becomes too close to allow another iteration, or 3) the signal terminates. Since this mechanism takes advantage of multiple satellites that revisit a signal consecutively, we call it *sequential multiple coverage*.

Our framework is thereby an approach to progressive QoS enhancement via continuously expanding the scale of the coordination among peer satellites, throughout a window of opportunity. While satellite coordination plays an important role in the framework, coordination expansion and termination are enabled by message-passing over crosslinks between neighboring satellites, as described in the next subsection.

3.2 Algorithmic Approach

Figure 3 provides several snapshots of a QoS optimization process, which illustrate how peer satellites coordinate through message-passing at different stages. As shown in Figure 3(a), if S_1, the first satellite that detects the signal, sees further opportunity for QoS enhancement after completing its geolocation computation, it will send a coordination-request message to a peer S_2 that is expected to visit the target next. This message contains the initial measurements and preliminary result. By receiving the message, S_2 will obtain the information it needs for the next iteration of geolocation computation. Consequently, when its footprint moves to the target location, S_2 will be able to generate a result with better accuracy.

The coordination process will continue (see Figure 3(b)) along the chain consisting of satellites that revisit the target one after another[3]. Whereas the coordination will terminate when one of the following conditions becomes true:

TC-1) The estimated error becomes sufficiently small;

TC-2) The elapsed time since the initial detection exceeds a threshold; or

TC-3) The signal stops.

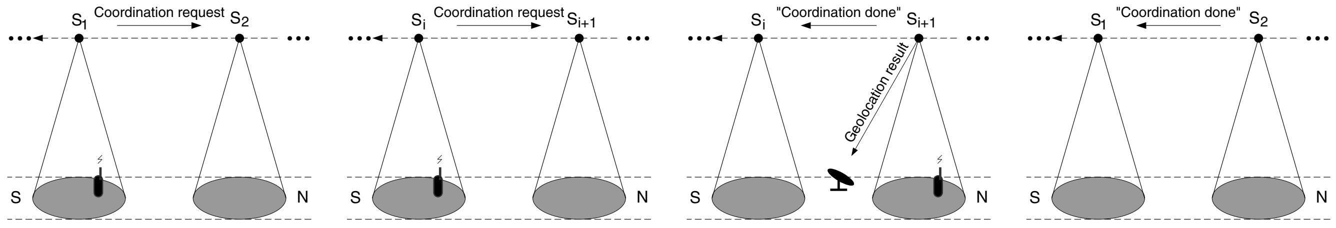
While TC-1 and TC-2 can be routinely checked at the end of each accuracy-improvement iteration so that the satellite that performs the computation can decide whether it should ask another peer to join the coordination, TC-3 can become true after a coordination request is made. Furthermore, the coordinated optimization is highly distributed in nature, meaning that there is no team leader or decision authority. Accordingly, coordination termination is also enabled by message-passing between peer satellites, similar to coordination expansion. More specifically, as shown in Figure 3(c), when a satellite S_{i+1} completes computation and realizes that further coordination for QoS enhancement is impossible or unnecessary because one of the termination conditions holds, the satellite will enclose the final result in an alert message and send it to the ground station. Meanwhile, S_{i+1} will send a "coordination done" message to S_i. The notified S_i will then pass the message to S_{i-1}, and so on. In this manner, S_1, the satellite that performed the initial geolocation, will be notified at the end, as illustrated in Figure 3(d).

Now suppose that S_i does not receive a "coordination done" notification from S_{i+1} when the elapsed time since the initial detection exceeds a threshold which is a function of the alert-message-delivery deadline and S_i's ordinal number i (as described in the next paragraph). Then S_i will assume that S_{i+1} is unable to deliver the alert message because TC-3 becomes true before S_{i+1}'s footprint arrives at that location, as illustrated in Figure 4 (where the shaded cellular phone with no emission represents a terminated signal). Consequently, S_i will treat its result as the final result and send it to the ground. Analogous to the case shown in Figure 3(c), a "coordination done" notification will be sent to S_{i-1} and propagated along the downstream of the chain.

It is important to ensure that all the participating peers, including S_1, receive the "coordination done" notification in time, so that they will not be unnecessarily alarmed. Therefore, the decision (by a satellite that just completes geolocation computation) on whether to request the next arriving peer to join coordination must be made according to whether TC-2 has become true. More specifically, TC-2 is formulated by the expression $getTime() - t_0 > \tau - (n\delta + T_g)$, where t_0 is the point of initial detection, τ is the (system-level) deadline for alert-message delivery, δ is the maximum inter-satellite message-delivery delay, n is the satellite's ordinal number which identifies its position in the coordination chain, and T_g is the maximum amount of time it takes a satellite to perform geolocation computation. Thus the right side of the above inequality can be regarded as the "local threshold" of the elapsed time, according to which S_n will determine whether it should request S_{n+1} to join the coordination. More precisely, if the inequality (i.e., TC-2) holds, then there will be no guarantee that S_{n+1} will be able to complete the next iteration and that all the satellites in the downstream can receive the notification from S_{n+1} in time. In that case, S_n will decide to stop the iterative accuracy improvement and send its geolocation result and "coordination done" message to the ground and S_{n-1}, respectively.

---

[3] For the sake of illustration, we assume here that the target is located near the center line of a plane's footprint trajectory so that the chain of coordinating satellites coincides with a portion of that plane. However, the algorithm itself is general and is not derived based on that assumption.



(a) After initial detection (b) Successive coordination (c) No further opportunity (d) Coordination completion

Figure 3: Coordinated QoS Enhancement

By the same token, if TC-2 does not hold, then S_n will send a coordination request to S_{n+1} and will thereafter wait for the "coordination done" message so long as the condition $getTime() - t_0 < \tau - (n-1)\delta$ holds. If no such message is received from S_{n+1} when time expires, S_n will assume that S_{n+1} is unable to complete computation due to TC-3 or S_{n+1} becomes fail-silent, and thus S_n will send its geolocation result and "coordination done" message to the ground and S_{n-1}, respectively, as shown in Figure 4 and described earlier.

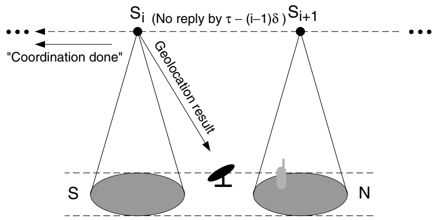


Figure 4: Guaranteed Geolocation Report

Alternatively, we may let S_{n+1}, the satellite that receives a coordination request but is unable to carry out the computation successfully, be responsible for sending the result received from S_n to the ground. This would eliminate the need for the "coordination done" message-passing along the downstream of the chain. However, with the backward-messaging scheme, the delivery of the alert message will be guaranteed even if S_{n+1} becomes fail-silent in the middle of computation.

3.3 Discussion

The opportunity-adaptive nature of our approach thus permits us to strive for the best possible QoS, while guaranteeing that in the worst case, with high probability the preliminary geolocation result will be delivered in a timely fashion. Therefore, the OAQ framework shares a conceptual basis with the imprecise computation scheme [9], which was motivated by the fact that one could often trade off precision for timeliness. Imprecise computation techniques prevent missed deadlines and provide graceful degradation under a transient overload. With the imprecise computation scheme, optional tasks can be scheduled to refine the result and can also be left unfinished at its deadline, if necessary, lessening the quality of the computation. In contrast, the thrust in the OAQ framework is to progressively exploit peer coordination in a highly distributed environment. The OAQ algorithm thus does not rely on a scheduling authority or explicit task decomposition. Instead, the sequence and extent of result refinement depend upon a dynamically determined opportunity.

The difference between our approach and the opportunistic scheduling framework [10] is that the matchmaker in their framework focuses on system throughput, whereas the QoS objective of the OAQ approach is a function of multiple system attributes and our algorithm requires no team leader or decision authority. More importantly, since our approach is intended to exploit mobile resource redundancy, the derivation of our algorithm is driven by the resources' readiness-to-serve, rather than the traditionally defined resource availability.

4 Model-Based Evaluation

4.1 Assumptions

In order to assess the effectiveness of the OAQ framework, we conduct a model-based evaluation. The analytic model is constructed according to the RF constellation described in Section 2. We choose to use this constellation for the quantitative study because 1) the design is conducted in-house at JPL, and an interactive simulation model for visualization and coarse-grained quantitative measures (e.g., coverage time) is available, and 2) while its relatively small size allows a closed-form solution of and efficient evaluation experiments for the QoS measure, the design principle of this constellation is consistent with those for constellations that have massive numbers of nodes. Therefore, evaluation of this system suffices to serve the purpose of effectiveness demonstration.

We assume that the constellation is protected by scheduled and threshold-triggered ground-spare deployment policies. By "scheduled ground-spare deployment policy," we mean a policy in which ground spares will be launched according to a predetermined schedule to restore the constellation to its original capacity (so that it will again be equipped with a total of 112 satellites). In contrast, "threshold-triggered ground-spare deployment policy" refers to a policy in which ground spares will be launched to restore an orbital plane to its original capacity (i.e., 14 active satellites plus 2 in-orbit spares), when the number of operational satellites in the plane drops to a threshold.

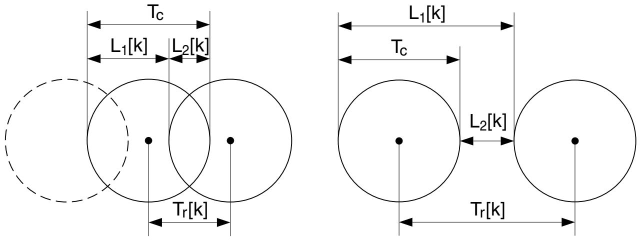
As shown in Figure 1, due to the spherical shape of the earth, the ratio of the total area covered by overlapped footprints to that covered by single footprints changes across different latitudes. In particular, the ratio is the lowest at the equator and the highest at the poles. It follows that in our assumed area of interest, which is around 30° north latitude, the ratio is moderately high. Further, as shown by Figure 1 and the interactive simulation generated by the Satellite Orbit Analysis Program (SOAP), at around 30° north (or south) latitude, a location on or near the center line of a footprint trajectory will be least likely to be covered by overlapped footprints, relative to the locations at the two sides of the trajectory (which are heavily overlapped with the footprint trajectories of the adjacent planes). Hence, the situations in which a signal is located at or near the center line of a footprint trajectory can be regarded as the worst case, given that the emitting source is around 30° latitude. In order to be conservative and keep the complexity of the analytic model manageable, we let the QoS measure be formulated based on this worst case. In addition, we assume that satellite failure will not occur during the interval from the initial signal detection to the completion of the coordinated geolocation computation[4]. Accordingly, we assume the use of the no-backward-messaging option in this evaluation.

4.2 Model

4.2.1 QoS Measure

Since the OAQ framework allows us to strive for the best possible accuracy for position determination with respect to a dynamically determined opportunity, we define a measure that quantifies a system's ability to deliver service in terms of QoS levels. More specifically, if the service delivered by the constellation can be rated by n QoS levels, we can let Y be a random variable that takes its value from the set $\{y \mid y = 1, 2, \ldots, n\}$. We thereby let the QoS measure be the probability that the system will deliver a geolocation result with the quality at level y or above (given that a signal occurs). More succinctly, we choose $P(Y \geq y)$ as the QoS measure. In order to determine a QoS spectrum that enumerates all the QoS levels relevant to the system in question, we analyze the relationships between system behavior and the geometry properties of the constellation as follows.

As described in Section 2, when an orbital plane loses satellites after spare exhaustion, geometric relations between the footprints of adjacent satellites will change; further, the geometric relations between adjacent satellites can be described in terms of $T_r[k]$ and T_c. Since k denotes the number of operational satellites that are actively in service in a plane, we use the term "orbital-plane capacity" to refer to the value of k. Clearly, a decrement of k will result in an increased value of $T_r[k]$. Thus, the initial relation $T_r[k] < T_c$ may change to $T_r[k] \geq T_c$ over time, as shown by Figures 5(a) and 5(b). Since 1) $T_r[k] \approx \theta/k$, where θ is the time required for a satellite to orbit through the plane and equals 90 minutes for the constellation in question, and 2) a satellite's coverage time T_c (as defined in Section 2) in this constellation is 9 minutes, the underlapping scenario will happen when k is dropped to below 11.



(a) $T_r[k] < T_c$ (b) $T_r[k] \geq T_c$

Figure 5: Geometric Relations of Footprints

To facilitate the formulation and solution of the QoS measure, we introduce two auxiliary parameters $L_1[k]$ and $L_2[k]$, as shown in Figures 5(a) and 5(b):

$$\begin{aligned} L_1[k] &= \left(T_r[k] - \frac{T_c}{2}\right) + \frac{T_c}{2} = T_r[k] \\ L_2[k] &= |T_c - L_1[k]| = |T_c - T_r[k]| \end{aligned}$$

In addition, since we wish to distinguish the overlapping case from the underlapping case, we define an indicator variable $I[k]$ as follows:

$$I[k] = \begin{cases} 1 & \text{if } T_r[k] < T_c \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Then if we let $M[k]$ denote the upper bound of the number of satellites that will consecutively capture a signal $\mathcal{S}$, given that the involved plane has k active operational satellites and $I[k] = 0$, $M[k]$ can be expressed as a function of the deadline τ, and auxiliary parameters $L_1[k]$ and $L_2[k]$ (we omit the derivation due to space limitations):

$$M[k] = \begin{cases} 2 + \left\lfloor \frac{\tau - L_2[k]}{L_1[k]} \right\rfloor & \text{if } \tau > L_2[k] \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

---

[4] As explained in Section 4.2.1, the coordination chain in this setting will involve at most two satellites, implying that the likelihood that one or more coordinating satellites will fail during the interval will be negligible.

Since for the constellation in question, the footprints of adjacent satellites in an orbital plane will be underlapping if $k < 11$, Eq. (2) implies that the upper bound for the number of satellites that will consecutively revisit a signal will be 2 (which enables a sequential dual coverage), if deadline τ is less than 9 minutes (which is the case we assume in this evaluation).

Together with the possible scenarios in which a signal may 1) be captured by a simultaneous dual coverage (when footprints overlap), 2) be covered by just a single footprint, or 3) escape from surveillance (when footprints underlap), the above discussion implies that the satellite constellation in question has a 4-level QoS spectrum, as illustrated in Table 1. We have thus completed measure definition and are ready to discuss the method for solution.

Table 1: QoS Levels vs. Geometric Properties

$I[k]$	$Y = 3$ Simultaneous dual	$Y = 2$ Sequential dual	$Y = 1$ Single coverage	$Y = 0$ Missing target
1	√		√	
0		√	√	√

4.2.2 Measure Solution

The relationships between the constellation's structural degradation and QoS levels lead us to choose a decomposition approach for measure solution. Recall that 1) the measure is defined based on the assumption that the signal is located at or near the center line of a footprint trajectory and is far enough from the poles (where the footprint trajectories of adjacent planes heavily overlap), and 2) there are no shared spares between orbital planes. Accordingly, structural variations of neighboring planes will have no effects on the QoS measure and thereby $P(Y \geq y)$ can be evaluated on an individual plane basis. We thus start solution derivation from the following expression:

$$P(Y \geq y) \approx \sum_{Y=y}^{3} \sum_{k=9}^{14} P(Y = y \mid k)\, P(k) \qquad (3)$$

where we neglect the terms concerning the cases in which $k < 9$ because the scheduled and threshold-triggered ground-spare deployment policies make those cases extremely unlikely. Note that Eq. (3) decomposes $P(Y \geq y)$ into two sets of constituent measures:

1) The conditional probability that the system will deliver a geolocation result rated at QoS level y, given that the involved plane has k active operational satellites; i.e., $P(Y = y \mid k)$.
2) The probability that the involved plane has k active operational satellites; i.e., $P(k)$.

Because of the scheduled ground-spare deployment policy, we use *UltraSAN* [11], which supports deterministic activity times, to compute the steady-state probability $P(k)$. Steady-state solutions of $P(k)$ are feasible because the occurrence of a signal is assumed to be a Poisson process, and the probability structure of what a Poisson arrival observes is identical to the steady-state probability structure of the system [12]. Our SAN (stochastic activity network) model represents a plane's behavior concerning its structural degradation and scheduled/threshold-triggered spare deployments, and is built based on the assumption that satellite failures are statistically independent within an orbital plane. The independent failure assumption is reasonable for our application, because within a plane the separation between neighboring satellites is a constant for a particular configuration (i.e., $T_r[k]$) and is adequate to ensure environment diversity.

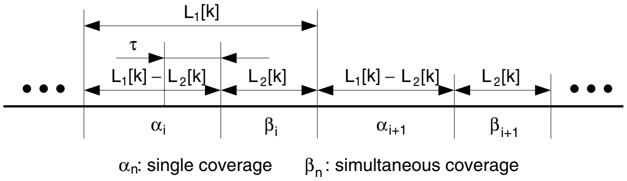
To solve for $P(Y = y \mid k)$, we begin with analyzing the problem based on the two timing diagrams shown in Figure 6. In Figure 6(a), which is intended to support the analysis for the overlapping case, we break the time horizon into intervals α_n and β_n. Relating this timing diagram to Figure 5(a), interval α_n corresponds to the duration (with a length $L_1[k] - L_2[k]$) through which an earth location (at the center line of a footprint trajectory) can be covered by a single footprint, while interval β_n corresponds to the duration (with a length $L_2[k]$) through which an earth location can be covered by the overlapped footprints. The timing diagram in 6(b), which is intended to support the analysis for the underlapping case, is depicted in an analogous way, but interval γ_n corresponds to the duration (with a length $L_2[k]$) through which an earth location will not be covered by any footprints (see Figure 5(b)).

Based on the timing diagrams, we derive the following theorems (but we omit the proofs due to space limitations):

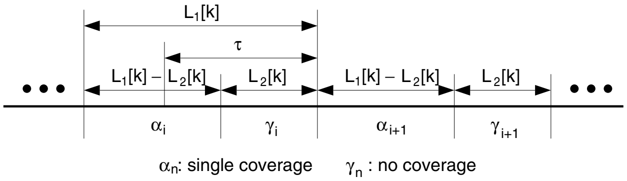
Theorem 1 *When $T_r[k] < T_c$, position determination of a signal $\mathcal{S}$ can be accomplished by a simultaneous multiple coverage only if $\mathcal{S}$ occurs in 1) interval β_i, or 2) interval α_i, with at most τ or $L_1[k] - L_2[k]$ time units, whichever is smaller, away from interval β_i.*

Theorem 2 *When $T_r[k] \geq T_c$, position determination of a signal $\mathcal{S}$ can be accomplished by a sequential multiple coverage only if 1) $\tau > L_2[k]$ and $\mathcal{S}$ occurs in interval α_i with at most $L_1[k]$ or τ time units, whichever is smaller, away from α_{i+1}, or 2) $\tau > L_1[k]$ and $\mathcal{S}$ occurs in interval γ_i with at most $L_1[k] + L_2[k]$ or τ time units, whichever is smaller, away from α_{i+2}.*

Note that the second (alternative) necessary condition in Theorem 2 will never hold for this evaluation, because we



(a) $T_r[k] < T_c$



(b) $T_r[k] \geq T_c$

Figure 6: Timing Diagrams

assume $\tau < 9$ and thus $\tau \leq L_1[k]$ is true for all the underlapping cases (in which $k < 11$). Then, Theorems 1 and 2 lead us to define two more auxiliary parameters to facilitate the solution derivation for $P(Y = y \mid k)$:

$$\hat{L}[k] = \min\{L_1[k] - L_2[k], \tau\}, \quad \tilde{L}[k] = \min\{L_1[k], \tau\}$$

Since the signal occurrence is assumed to be a Poisson process, the distribution of the instants when the signals occur is the same as the uniform distribution of the event over the same interval [13]. Consequently, if we let $G_3[k]$ denote the probability that the system will deliver a geolocation result rated at QoS level 3 (i.e., the position of a signal is determined by a simultaneous dual coverage) given that $I[k] = 1$ holds for the involved plane, we have

$$\begin{aligned} G_3[k] \quad = \quad & \int_0^{\hat{L}[k]} \frac{1}{L_1[k]} W_x[k] \int_0^{\tau - (\hat{L}[k] - x)} h(z) dz\, dx \\ & + \int_0^{L_2[k]} \frac{1}{L_1[k]} \int_0^{\tau} h(z) dz\, dx \end{aligned} \tag{4}$$

where $W_x[k] = 1 - \int_0^{\hat{L}[k]-x} f(y) dy$, which computes the probability that the signal does not terminate before the arrival of the overlapped footprints, while the integrals over h evaluate the probability that the iterative computation completes before the deadline is reached. Note that f and h are the probability density functions of signal duration and iterative geolocation computation time, respectively. We assume that signal duration is exponentially distributed, which is a fairly typical assumption used in performance modeling for telecommunication systems; we also assume that iterative geolocation computation time is exponentially distributed in order to allow the amount of time required for result convergence to be nondeterministic. Note also that the limits of the integrals are defined based on Theorem 1.

Since QoS level 3 can be achieved only if footprints overlap, $P(Y = 3 \mid k)$ has the following expression:

$$P(Y = 3 \mid k) = \begin{cases} G_3[k] & \text{if } I[k] = 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Based on Theorems 1 and 2, we solve for $P(Y = 2 \mid k)$ and $P(Y = 1 \mid k)$ in a similar fashion. Due to space limitations, we omit further discussion of the solution method and proceed to describe the evaluation results in the next subsection.

4.3 Evaluation Results

In order to analyze the QoS gain from the use of the OAQ algorithm, we also compute the measure for the basic fault-adaptive QoS enhancement (BAQ) scheme for comparison. The BAQ scheme refers to the case in which the constellation is equipped with in-orbit spares and protected by the scheduled and threshold-triggered ground-spare deployment policies, but does not apply the opportunity-adaptive algorithm. Thus, when a signal is detected in a constellation under the BAQ scheme, a geolocation result will be delivered after the initial computation based on either a single coverage or simultaneous coverage, implying QoS level 2 is not applicable.

We first evaluate the constituent measure $P(Y = y \mid k)$ using *Mathematica*™, based on the parameter values $\tau = 5$, $\mu = 0.5$, $\nu = 30$ (μ and ν are the signal termination rate and iterative computation completion rate, respectively; time is quantified in minutes by default). As described in Section 2, the values of θ and T_c are 90 and 9, respectively, for the constellation in question.

The results shows that the OAQ scheme is able to push a system's QoS toward the high end, guaranteeing a result with the best possible quality even in presence of severe structural degradation. In particular, even when $k = 12$ (which implies two more satellite failures after spare exhaustion and a total loss of 25% of nodes in the plane), with probability 0.44 the constellation will still be able to deliver a geolocation result rated at QoS level 3. On the other hand, the evaluation reveals that the value of $P(Y = 3 \mid 12)$ is only 0.20 with the BAQ scheme.

Next we evaluate the other constituent measure $P(k)$ using *UltraSAN*. The results are shown in Figure 7, where time is quantified in hours for λ (node failure rate) and ϕ (time to scheduled ground-spare deployment), and η is the threshold for the number of operational satellites in a plane (meaning that ground-spare deployment will be triggered when $k = \eta$). From the curves, we observe that when protected by the scheduled and threshold-triggered ground-spare deployment policies, the full orbital-plane capacity (i.e., $k = 14$) will dominate when node-failure λ rate is low. On the other hand, the threshold capacity (i.e., $k = \eta$) tends to become dominant as failure rate increases. Specifically, the value of $P(10)$ in Figure 7 is very small when $\lambda = 10^{-5}$, but it rapidly increases and becomes dominant as λ increases. The reason is that when satellites become more vulnerable to failure, the capacity of an orbital plane is likely to drop toward the threshold sooner. However, as the threshold-triggered ground-spare deployment policy prevents the scenario in which the plane's capacity drops below the threshold from happening, the likelihood that the system is operating at its threshold capacity becomes dominant when λ is high.

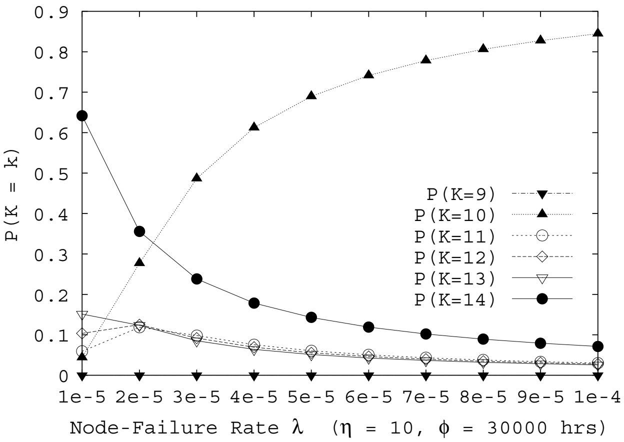


Figure 7: Probability of Orbital-Plane Capacity

Using the results of the constituent measures and Eq. (3), we evaluate $P(Y = y)$ and $P(Y \geq y)$. Figure 8 compares the probabilities that the OAQ and BAQ schemes will deliver a result rated at QoS level 3. For this evaluation experiment, we set η to 12 and let ϕ remain 30000 hours (over 3 years). The curves show that under the OAQ scheme, the system will achieve level-3 QoS with a greater probability as signal completion rate μ decreases (i.e., mean signal duration increases). More specifically, when μ decreases from 0.5 to 0.2, $P(Y = 3)$ increases up to 38% over the domain of λ considered. On the other hand, the same variation does not yield any differences in the behavior of the BAQ scheme. This exemplifies that the QoS gain from the use of the OAQ scheme is due to its awareness and exploitation of conditions that arise in the operational environment, while the BAQ scheme ignores potential opportunities.

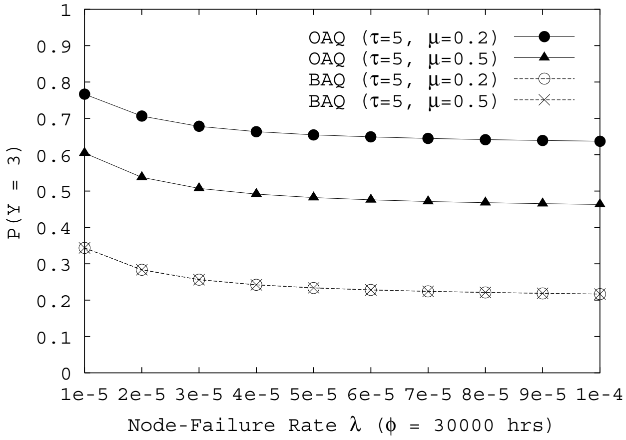


Figure 8: $P(Y = 3)$ as a Function of λ

Figure 9 displays the results of the QoS measure $P(Y \geq y)$. The curves reveal that OAQ always makes the system significantly more likely achieve higher levels of QoS than BAQ does. In particular, when λ equals 10^{-5}, the OAQ scheme enables the system to achieve QoS level 2 or above with a probability of 0.75, while with the BAQ scheme the probability is only 0.33. When λ increases to 10^{-4}, the system will achieve a QoS level of 2 or above with a probability of 0.41 under the OAQ scheme, whereas the probability is decreased to 0.04 under the BAQ scheme. Nonetheless, we can observe that the values of $P(Y \geq 1)$ are always equal for the two schemes (both are equal to 1 over the domain of λ we consider), meaning that OAQ and BAQ perform equally well with respect to guaranteeing the delivery of a result rated at QoS level 1 or above. The results indeed confirm that the major advantage of the OAQ scheme is its ability to push a system's performance to the high end of the QoS spectrum while guaranteeing the timely delivery of results with minimally acceptable quality.

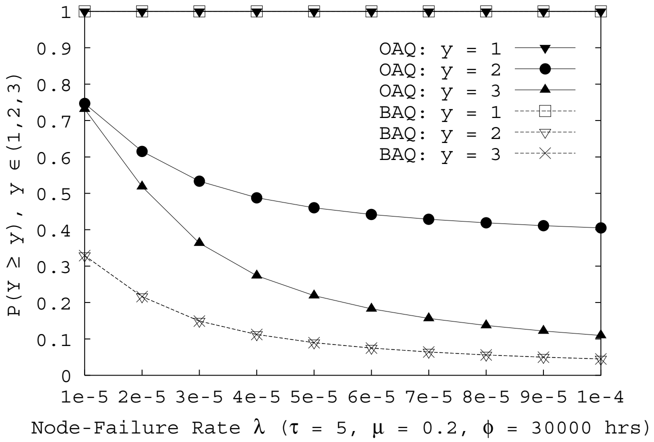


Figure 9: $P(Y \geq y)$ as a Function of λ

We also evaluate the QoS measure as a function of τ. The results illustrate how the OAQ scheme achieves better QoS by taking full advantage of the "time allowance." Finally, when the QoS measure is evaluated as a function of the mean signal duration, we observe that the OAQ scheme is able to responsively treat a longer signal duration as the extended opportunity to achieve better geolocation quality.

5 Concluding Remarks

We have developed an approach to coordinated, progressive QoS optimization in satellite constellations. By letting peer satellites successively join the coordinated geolocation computation as they become ready to serve within the window of a dynamically determined opportunity, the OAQ algorithm is able to guarantee the timely delivery of geolocation results with the best possible quality.

The results of this effort are meaningful. First, the OAQ framework advocates a marriage between satellite constellation and fault tolerance technologies. In particular, although sequential localization was studied and proved to be sound by the satellite research community, the techniques have not been considered as a solution for tolerating the loss of satellites in a constellation. The OAQ framework demonstrates a novel, yet practical application of this satellite technology for fault tolerance in constellations that are vulnerable to structural degradation. Further, we exploit peer-to-peer message passing that is often used for distributed fault tolerance to enable dynamic, progressive peer-satellite coordination, eliminating the need for ground intervention.

Second, this effort shows that while the continuously changing readiness-to-serve of satellites creates many difficulties for fault tolerance, their mobile nature can indeed be exploited to enable novel utility of resource redundancy. More generally, from the perspective of fault tolerance in systems that comprise large populations of mobile resources, the results of our investigation demonstrate the feasibility of adaptation, extension, and generalization of various existing fault tolerance concepts, such as analytic redundancy, data diversity, environment diversity, imprecise computation, and active/passive replication.

It is worth noting that the OAQ framework can be extended and generalized for many other applications. In particular, as researchers in the wireless-networking area have been investigating advanced applications of TCP/IP and multicasting in satellite constellations, it will be feasible to extend the OAQ concept to adapt those well-known fault tolerance schemes, including group membership protocols, for use in micro- and pico-satellite constellations. Accordingly, our current work is directed toward adapting group membership management techniques to the applications in the environments of distributed autonomous mobile computing.

Acknowledgments

This work was performed in part by NASA's Jet Propulsion Laboratory, California Institute of Technology and was supported by DARPA/IXO. The authors are thankful to Vijay Raghavan of DARPA/IXO for his support and encouragement, to James Chase, David Collins and Robert Shotwell of JPL for their untiring help and stimulating discussions, to Jenny Applequist of UIUC for her meticulous proofreading, and to anonymous reviewers for their valuable comments and suggestions.

References

[1] J.-C. Laprie, editor, *Dependability: Basic Concepts and Terminology*, vol. 5 of *Dependable Computing and Fault-Tolerant Systems*. Vienna, Austria: Springer-Verlag, 1992.

[2] M. Gerla *et al.*, "TCP/IP via satellite constellations," in *Proceedings of the 4th European Workshop on Mobile/Personal Satcoms*, (London, UK), pp. 82–91, Sept. 2000.

[3] L. Wood, G. Pavlou, and B. Evans, "Managing diversity with handover to provide classes of service in satellite constellation networks," in *Proceedings of the AIAA International Communication Satellite Systems Conference*, (Toulouse, France), pp. 17–20, Apr. 2001.

[4] N. Levanon, "Quick position determination using 1 or 2 LEO satellites," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, pp. 736–754, July 1998.

[5] Y. T. Chan and J. J. Towers, "Sequential localization of a radiating source by Doppler-shifted frequency measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, pp. 1084–1090, Oct. 1992.

[6] P. E. Ammann and J. C. Knight, "Data diversity: An approach to software fault tolerance," *IEEE Trans. Computers*, pp. 418–425, Apr. 1988.

[7] Y. Huang and C. Kintala, "Software implemented fault tolerance: Technologies and experience," in *Digest of the 23rd Annual International Symposium on Fault-Tolerant Computing*, (Toulouse, France), pp. 2–9, June 1993.

[8] D. Collins, R. Shotwell, L. Alkalai, J. Chase, and T. George, "Conceptual design and analysis of two satellite constellations for the application of a FAST architecture," D-25994, Jet Propulsion Laboratory, Pasadena, CA, Mar. 2003.

[9] J. W. S. Liu, K.-J. Lin, W.-K. Shih, A. C. Yu, J.-Y. Chung, and W. Zhao, "Algorithms for scheduling imprecise computation," *IEEE Computer*, vol. 24, pp. 58–68, May 1991.

[10] R. Raman, M. Livny, and M. Solomon, "Matchmaking: Distributed resource management for high throughput computing," in *Proceedings of the Seventh IEEE Symposium on High Performance Distributed Computing*, (Chicago, IL), pp. 140–146, July 1998.

[11] W. H. Sanders, W. D. Obal II, M. A. Qureshi, and F. K. Widjanarko, "The *UltraSAN* modeling environment," *Performance Evaluation*, vol. 24, no. 1, pp. 89–115, 1995.

[12] S. A. Nozaki and S. M. Ross, "Approximations in finite-capacity multi-server queues with Poisson arrivals," *Journal of Applied Probability*, vol. 15, pp. 826–834, 1978.

[13] L. Kleinrock, *Queueing Systems. Volume I: Theory*. New York: John Wiley & Sons, 1975.