Chapter 12: Advanced RAIM

Introduction

GNSS is currently undergoing major upgrades: new constellations are being launched (Galileo and Beidou) and expected to be fully operational by 2025. GPS and the new constellations will have signals in both L1 and L5, allowing users to remove the ionospheric delay affecting the pseudorange errors. Finally, improvements in both the on-board clocks and the ground segments are reducing the errors due to clock and ephemeris to standard deviations below a meter (at least for GPS). As a consequence, users will have more accurate pseudoranges, stronger geometries, and much more redundancy. RAIM, as described in Chapter XX, performance is a direct function of the satellite geometry and the expected accuracy of the pseudorange measurements. This has naturally led to consider the use of new signals and new constellations in RAIM, with the hope that it could provide enhanced horizontal guidance, and even vertical guidance.

RAIM algorithms used in aviation today were designed: first, for GPS L1 C/A only, second, for horizontal guidance only. On the first point, these algorithms make assumptions on the signals that may not be adequate with for new constellations. For example, RAIM algorithms protect users against single faults only. With up to thirty satellites in view, the single fault assumption may not hold anymore, especially if new constellations turn out to have large probabilities of fault. On the second point, assumptions that may have been good for horizontal guidance will not necessarily be sufficient for vertical guidance, which has a higher criticality (see Appendix).

Because of these necessary updates, this envisioned extension of RAIM is known as Advanced RAIM. Note that there is some ambiguity in the literature in the use of the term ARAIM: it can either designate the system architecture enabling the use of new signals for aviation where the user exploits redundancy, or a type of user algorithms that exploits redundancy in a way that is more general than currently used RAIM algorithms (and can be used in many more settings). In this chapter we discuss both meanings of Advanced RAIM.

We start by asking ourselves what we need from the system, so in the first section we will describe navigation requirements for vertical guidance, in particular the integrity requirements. Once we know what we want, we need to get acquainted with out we fear, that is, what is that could make the navigation larger than we want it. This, of course, is the list of error sources, both nominal and faulted. The result of this section will be a parametrization of the threat model that can be used by the user receiver to assess whether the requirements are met. In the third section we will turn our attention on the user algorithm. The goal here will be twofold: first, to describe the principle of RAIM as simply as possible, and then to provide readers with a baseline ARAIM user algorithm that offers very good performance while remaining practical. In section 4, we will describe the how to apply this generic algorithm in the ARAIM concept. Finally, the last section will deal with the most critical part of ARAIM, which is the determination of the threat model parameters.

This chapter is based on the references [].

1. Requirements

The first question we need to ask ourselves is what we want. One of the most critical steps in the design of ARAIM is to have a clear and practical formulation of the navigation requirements. Here we will focus on the main goal of ARAIM for aviation, which is to provide worldwide vertical guidance. The ICAO International Civil Aviation Organization (ICAO) GNSS standards and recommended practices (SARPs) contained in [3], which describes the navigation requirements for radio navigation aids, constitute the basis for these requirements. These requirements were specifically derived for SBAS, and based on SBAS performance, so there is some uncertainty on how to apply them to a new system like ARAIM. However, this uncertainty will most likely only affect the particular numerical values, not the formulation of the requirements.

1.1 Integrity, Alert Limits, and Protection Levels

In order to formulate the integrity requirement, we must first define the Probability of Hazardously Misleading Information (PHMI) (also called integrity risk (IR)). The PHMI is the probability that the position error exceeds the Alert Limit (AL) and there is no timely alert, that is, within the Time-to-Alert (TTA). The integrity requirement states that the PHMI should be below a predefined value. For example, for LPV-200 operations, the Vertical Alert Limit is 35 m, The Horizontal Alert Limit is 40 m, the TTA is 6 s, and the PHMI should be below 2 x 10^{-7} per approach [8]. If we ignore the TTA, the integrity requirement can be written as follows:

$$\text{Prob}\left(\left(VPE > VAL \text{ or } HPE > HAL\right) \text{ and } \left(\text{No Alert}\right)\right) \leq 2*10^{-7} / \text{approach} \qquad (1)$$

where VPE is the Vertical Position Error and HPE is the Horizontal Position Error.

It is also possible, and practical, to formulate the integrity requirement using the Protection Levels (PLs). The Vertical Protection Level (VPL) and the Horizontal Protection Level (HPL) are defined such that we have:

$$\text{Prob}\left(\left(VPE > VPL \text{ or } HPE > HPL\right) \text{ and } \left(\text{No Alert}\right)\right) \leq 2*10^{-7} \text{ / approach} \quad (2)$$

As a consequence, if we have VPL<VAL, and HPL<HAL, the requirement formulated in (1) will be met. One may wonder why we would ever use PLs instead the PHMI. One of the advantages is that receivers that output PLs do not need to be aware of the Alert Limit, which may vary depending on the operation. Another, more subtle reason, is due to the differences in treatment of continuity compared to integrity.

1.2 Continuity

We start by citing the definition of continuity given in [MOPS229]: "The continuity of a system is the ability of the total system (comprising all elements necessary to maintain aircraft position within the defined airspace) to perform its function without interruption during the intended operation. More specifically, continuity is the probability that the specified system performance will be maintained for the duration of a phase of operation, presuming that the system was available at the beginning of that phase of operation, and predicted to exist throughout the operation." The continuity requirement varies depending on the operation.

As an important example, the ICAO continuity requirement for CAT I (and LPV-200) is 8×10^{-6} per 15-second interval, which applies to losses of service from all causes, including those external to the receiver processing [3]. For ARAIM it means that the thresholds used to determine availability should not trip at a higher rate than a fraction of this requirement. In [6], it was assumed that the required probability for the airborne algorithm was half of that of the ICAO requirement, that is, 4×10^{-6} per 15-sec interval.

1.3 Accuracy

In addition to the integrity and continuity requirement, we also want the position error to be mostly close to zero. Since this is not expressed explicitly in the integrity requirement, we need an additional requirement constraint in the distribution of position errors. For Cat. I and LPV-200, the accuracy requirement is that the 95% bound on the vertical position error must be below 4 m for any given geometry. Most positioning algorithms are locally linear so the formula for the standard deviation of the error for a given geometry is straightforward, since it uses the nominal distribution of errors, which is supposed be known (more on this later).

2. Threat model

In this section we want to answer the question: what do we fear? This is the set of events that could affect our assessment of the requirements formulated in the first section. This set is called the *threat model* or *list of feared events*, and has been presented in Chapter YY(SBAS). What is important for ARAIM (at least for aviation users) is that the threat model can be partitioned in three categories: nominal errors, narrow faults, and wide faults.

2.1 Nominal errors

The nominal errors are such that they can be characterized by known probability distributions. They are not necessarily present, but they are assumed to be always present. These include both errors due to the signal in space (clock and ephemeris, nominal signal deformation, etc), propagation errors (residual tropospheric delay, residual ionospheric delay), and local effects (multipath and receiver noise). They are almost always characterized by a gaussian distribution with an unknown, but bounded bias.

2.2 Narrow faults

For the other two categories, it is not practical to characterize them by a known distribution, as it would result in extremely wide distributions that lead to useless position error bounds. Of these two categories, narrow faults are the faults that affect satellites independently. The best example of a narrow fault is a clock run-off, because it is assumed that the occurrence of a clock run off will not impact the probability of clock run off in another satellite. In general, faults due to satellite malfunctions are considered to be narrow faults. For clock and ephemeris faults, it is necessary to determine the root cause of the fault before classifying it as a narrow fault.

2.3 Wide faults

These are the faults that can affect more than two satellites simultaneously in one constellation (where we defined the constellation is a system of satellites run by the same ground segment, like GPS, Galileo, GLONASS or BeiDou). Wide faults are the category of faults that is the most problematic in RAIM. This is because in the worst case they can produce errors (within one constellation) that are compatible with an arbitrary shift in user position. These errors produce measurements that are faulty, but self-consistent, which makes them undetectable to a redundancy check. Luckily, wide faults across constellations have been deemed sufficiently unlikely to be neglected in aviation applications.

Wide faults have been observed in two instances in GLONASS. In one case, the fault was due to the successive adjustment of the clock reference, which caused the constellation to have some satellites on one reference and the remaining ones on another one [reference]. Another likely mechanism for these faults, is the corruption of the Earth Orientation Parameters, which are common to all satellites [reference].

2.4 Mapping the threat model on the Integrity Support Message

In the ARAIM concept, the threat model information is encapsulated in the ISM. For this purpose, these categories are concisely characterized as illustrated in Table 2.1. Narrow faults are characterized by $P_{sat,i}$, the probability that satellite i has a fault during the operation. Wide faults are characterized by $P_{const,j}$, the probability that constellation j has a fault during the operation. The errors due to the nominal signal in space are characterized by the standard deviation $\sigma_{URA,i}$ and the nominal bias $b_{nom,i}$. These parameters constitute the Integrity Support Message (ISM) (or at least can be derived from the information contained in the ISM). In the current concept [ref], $\sigma_{URA,i}$ will be a multiple of the broadcast URA, that is, the ISM will include a factor α such that:

$$\sigma_{URA,i} = \alpha \sigma_{Broadcast\_URA,i} \tag{3}$$

Table 2.1 also shows the error models characterizing the nominal errors due to the residual tropospheric delay ($\sigma_{tropo,i}$) and the receiver noise and multipath ($\sigma_{airborne,i}$). For single frequency users, we would add the residual ionospheric delay error model ($\sigma_{iono,i}$).

	Nominal	Narrow fault	Wide fault
Clock and Ephemeris	Orbit/clock estimation and prediction and broadcast limits	Includes clock runoffs, bad ephemeris, unflagged manoeuvres	Erroneous EOPP, inadequate manned ops, ground-inherent failures
Signal Deformation	Nominal differences in signals due to RF components, filters, and antennas waveform distortion $\sigma_{URA,i}$	Failures in satellite payload signal generation components. Faulted signal model as described in ICAO	N/A
Code-Carrier Incoherence	e.g. incoherence observed in IIF L5 signal or GEO L1 signals	e.g. incoherence observed in IIF L5 signal or GEO L1 signals	N/A
Interfrequency Biases	Delay differences in satellite payload signal paths $b_{nom,i}$	Delay differences in satellite payload signal paths TBC $P_{sat,i}$	N/A $P_{const,j}$
Satellite Antenna Bias	Look-angle dependent biases caused at satellite antennas	Look-angle dependent biases caused at satellite antennas	N/A
Ionosphere	N/A	Scintillation	Multiple scintillations at solar storms
Troposphere	Nominal troposphere error (after applying SBAS MOPS model for tropo correction) $\sigma_{tropo,i}$	N/A	N/A
Receiver Noise and Multipath	Nominal noise and multipath terms in airborne model (TBC Gailleo BOC(1,1) and [illegible]) $\sigma_{airborne,i}$	e.g.: receiver tracking failure or multipath from onboard reflector. TBC	e.g.: receiver tracking multiple failure or multipath from onboard reflector. TBC

Table 2.1. Mapping of the threat model onto the ISM

3. A generic ARAIM User receiver algorithm

Our goal in this section is to provide a baseline algorithm for Advanced RAIM that is both easy to implement and for which there is a straightforward integrity analysis. In order to focus on the most important elements, we have simplified several assumptions. In particular, we will only deal with one coordinate, and we will do away with the nominal biases. In the last section, we will indicate how to account for these assumptions. We start this section with a simple example that illustrates the main ideas of RAIM algorithms.

3.1 Main idea in RAIM algorithms

The core idea in autonomous integrity monitoring is to make sure that the measurement residuals (the difference between the measurements and the pseudoranges that would be expected at the position solution) are consistent. To illustrate this idea, let us consider a simplified set of measurements:

$$y_i = x + n_i \tag{4}$$

where:

y_i are scalar measurements

x is the unknown position (one scalar) which we wish to determine

n_i is the noise affecting each of the measurements i

Let us assume that we have two such measurements (i=1,2), and that we know that when the errors are nominal (that is, when there is no fault), they follow a unit normal distribution:

$$n_i \sim N(0,1) \tag{5}$$

If the errors are nominal, the best estimate of the position is given by the least squares solution, which in this case is simply the average of the two measurements:

$$x = \frac{1}{2}(y_1 + y_2) \tag{6}$$

Let us now assume that we have the following measurements:

$$y_1 = 0,\ y_2 = 10 \tag{7}$$

The least squares estimate would be:

$$x = 5 \tag{8}$$

However, the error in this estimate probably exceeds what would be expected if the errors were nominal, because the measurements are not consistent. It is likely that there is a fault in at least one of the measurements, as we would not expect the difference between these two measurements to be so large when the errors are nominal. More precisely, we can form the statistic:

$$t = \frac{1}{\sqrt{2}}\left(y_2 - y_1\right) = \frac{1}{\sqrt{2}}\left(n_2 - n_1\right) \tag{9}$$

This statistic is independent of the true position *x*, and if the errors were nominal, we would expect it to follow a unit normal distribution. Under nominal conditions, a reading of:

$$t = \frac{1}{\sqrt{2}}\left(y_2 - y_1\right) \sim 7 \tag{10}$$

would therefore be very unlikely. Intuitively, it would make sense to place a threshold on this statistic and declare that there is a fault if the statistic exceeds the threshold. RAIM is based on this idea.

This simple example also illustrates some of the most important features of RAIM (and of any detection algorithm). The first one is the tradeoff between the loss of integrity risk and the false alert rate. If we make the threshold small, then we will be able to detect smaller faults. However, we will also increase the chance of exceeding the threshold under nominal conditions. The second one is the fact that any redundancy check relies on prior assumptions on the fault probabilities. For example, if both measurements were faulted simultaneously and with the same bias in each measurement, the test statistic distribution would be the same as in the nominal case, and therefore the fault would be undetectable.

We now generalize these ideas.

3.2 Measurement model

We will use the linear model for position estimation as described in [EngeMisra]. It is given by the linear relationship:

$$y = Gx + n \tag{11}$$

where

y is the vector of linearized GNSS measurements. In most cases, it will refer to the carrier-smoothed code [reference] using the ionospheric free combination.

G is the geometry matrix [reference]

x is the vector representing the position and the clock states of a user. In most cases, it will refer to the position and clock states at the time of the measurement.

n is the vector representing the noise affecting the measurements

This model is very general, since the vector of measurements y can include measurements made at different time steps.

3.3 Pseudorange Error Model

Nominal error model

It is practical to characterize the nominal pseudorange error by a zero mean Gaussian distribution. Although the actual distributions are not necessarily Gaussian distributions, we assume that one can always find a Gaussian distribution that is an upper bound (in a certain sense) of the actual distribution [reference overbounding]. For the nominal case, we then have:

$$n \sim N(0, C) \tag{12}$$

where C is the covariance of n. In aviation, this covariance is usually diagonal, but the algorithms described below do not require it.

Fault error model

The faults can be modeled by the addition of a new state to the nominal model. Each fault mode (or *hypothesis*) H_k is characterized by a known matrix F_k and an unknown state $x_{fault,k}$. Under hypothesis H_k, the measurement model is given by:

$$y = Gx + F_k x_{fault,k} + n \tag{13}$$

For example, for a single satellite fault, F_k is a column vector with a one in the index corresponding to the satellite assumed to be affected and zeros elsewhere, and $x_{fault,k}$ is the magnitude of the fault.

It is essential that the set of hypothesis H_k (plus the nominal mode H_0) form a partition, that is, the error model follows one of the states in Equation (13) and only one. This is not a limitation, it only means that if faults occur simultaneously, a new hypothesis needs to be created and accounted for separately. Each fault mode k has a known prior probability p_k. If we note p_0 the nominal error model, we have:

$$\sum_{k=0}^{K} p_k = 1 \tag{14}$$

In the paragraph will outline methodologies to determine the prior probabilities p_k based on the properties of the constellations in the ARAIM architecture.

3.4 Integrity requirement for fault detection

The main objective of the integrity monitor is to make sure that the probability that the position error exceeds a certain limit (the Vertical Alert Limit (VAL) for the vertical coordinate and the Horizontal Alert Limit (HAL) for the horizontal coordinates) and there is no alert is below the allowed Probability of Hazardously Misleading Information (P_{HMI}). The design of the RAIM algorithm consists in deciding two elements: which position fix $\hat{x}$ to choose, and when to declare an alert as a function of the measurements. Deciding when to declare an alert is equivalent to the determination of a region R (in the multidimensional vector space of measurements) where the measurements are deemed to be consistent. We can then formulate the integrity requirement as follows:

$$P\left(\left|x_v - \hat{x}_v\right| \geq VAL, \left|x_h - \hat{x}_h\right| \geq HAL, y \in R\right) \leq P_{HMI} \tag{15}$$

where:

$\hat{x}$ is the position and clock solution

$|\cdot|$ designates the L_2 norm (absolute value in the case of a scalar)

$\cdot_h$ designates the two components of the horizontal position

$\cdot_v$ designates the vertical component:

$$x_v = e_v^T x \tag{16}$$

Where the vector e_v extracts the vertical component. For example if the vertical component is on the third component and x has four components:

$$e_v = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^T \tag{17}$$

For the description of the algorithm, we will limit the integrity requirement to the one component of the error:

$$P\left(\left|x_v - \hat{x}_v\right| \geq VAL, y \in R\right) \leq P_{HMI} \tag{18}$$

A straightforward generalization to the horizontal error can be found here [ARAIM report reference]. It is important to point out that Equation (15) expresses an instantaneous requirement (as opposed to the requirement as described in section 1.1). For many applications, it is adequate to use the value per operation for the instantaneous requirement (like in vertical guidance [reference]).

3.5 Continuity requirement

One of the causes for loss of continuity is the declaration of an alert under nominal conditions. This can be expressed by the following equation:

$$P\left(y \notin R \mid H_0\right) \leq P_{FA} \tag{19}$$

As for the integrity, this is now an instantaneous requirement.

3.6 Position solution

The position solution as described in [EngeMisra] is, locally, a linear estimate of the measurements. That is, we can write:

$$\hat{x} = Sy \tag{20}$$

For the least squares estimate, S is given by:

$$S = \left(G^T W G\right)^{-1} G^T W \tag{21}$$

We will seen later that the least squares estimate is not the only possible position solution choice, and that for ARAIM it does not necessarily provide the best performance.

3.7 Determination of the detection region *R*

Ideally, we would like to find an optimal detection region in the sense that it provides the smallest integrity risk for a fixed false alert. That is, we would like to have a region R that solves the minimization problem:

$$\begin{array}{l} \text{minimize } P\left(\left|x_v - \hat{x}_v\right| \geq VAL, y \in R\right) \\ \text{such that } P\left(y \notin R \mid H_0\right) = P_{FA} \end{array} \tag{22}$$

This is in general a very difficult problem to solve, and this difficulty accounts partly for the variety of RAIM approaches. To obtain a practical algorithm, the detection region must be such that it can be easily described (so that the test statistic is easy to compute), and that leads to a simple expression of the integrity risk (or an upper bound on the integrity risk). There are at least two type of test statistics (or equivalently, detection regions) that have been extensively used in RAIM: the weighted sum of squared residuals and the solution separation statistics. The solution separation statistic was chosen to form the basis of the reference algorithm for ARAIM to evaluate the concepts developed in [ARAIM reports]. This approach was chosen because it is relatively simple, it is robust (to the error models), provides good performance, and it has optimality properties. For this reason, this is the one that we will describe.

Contribution of each fault mode to the integrity risk

By applying the chain rule of probability to the integrity requirement, we get:

$$\begin{aligned} & P\left(\left|x_v - \hat{x}_v\right| \geq VAL, y \in R\right) \\ & = \sum_{k=0}^{K} P\left(\left|x_v - \hat{x}_v\right| \geq VAL, y \in R \mid H_k\right) p_k \end{aligned} \tag{23}$$

Each one of these terms depends on a state $x_{fault,k}$, which according to our model, can take any value. For this reason, to compute the worst case integrity risk, we write:

$$P\left(\left|x_v - \hat{x}_v\right| \geq VAL, y \in R \mid H_k\right) = \max_{x_{fault,k}} P\left(\left|x_v - \hat{x}_v\right| \geq VAL, y \in R \mid H_k, x_{fault,k}\right) \tag{24}$$

Subset solutions

Let us now consider $\hat{x}_v^{(k)}$, the least squares position solution that is not affected by fault H_k. To do this, we solve the least squares problem defined by Equation (13), which estimates the fault parameter $x_{fault,k}$ in addition to the position and clock.

In many practical cases, the matrix F_k defines a fault in a subset of satellites where each new state is an added bias in one satellite. In this case, it can be shown that $\hat{x}^{(k)}$ is the least squares solution corresponding to the subset of measurements that is not affected by the fault:

$$\hat{x}^{(k)} = \left(G^T W^{(k)} G\right)^{-1} G^T W^{(k)} y \tag{25}$$

where

$$W_{ii}^{(k)} = 0 \text{ if measurement } i \text{ is affected by the fault}$$

$$W_{ii}^{(k)} = W_{ii} \text{ otherwise} \tag{26}$$

Solution separation statistics

Since $\hat{x}_v^{(k)}$ is not affected by fault *k*, a fault will cause our all-in-view position $\hat{x}_v$ to diverge from our fault tolerant position $\hat{x}_v^{(k)}$. To monitor the presence of the fault *k*, it is therefore reasonable to make sure that our all-in-view solution $\hat{x}_v$ is close to $\hat{x}_v^{(k)}$, the position solution that is not affected by the fault. This is exactly what we did in our simple example in paragraph 3.1. Now, if we consider all the fault modes, the detection region is then defined by:

$$R = \left\{ y \mid \left| \hat{x}_v - \hat{x}_v^{(k)} \right| \le T_k \text{ for all } k \right\} \tag{27}$$

where T_k are the detection thresholds (and we will be defined later).

Not only is it reasonable, but it turns out that in the case of one fault and when the all-in-view solution is given by (21), this is the best possible test, in the sense that it minimizes integrity for a given false alert requirement, that is, it is the exact solution of the problem (22) [referenceOptimaldetection]. Also, in all cases, the best estimate of the position error under the assumption of H_k is the difference between $\hat{x}$ and $\hat{x}^{(k)}$[reference ARAIM algorithm].

To make the dependency on the measurements y explicity, we can write:

$$R = \left\{ y \mid \left| \left(s_v^T - s_v^{T(k)} \right) y \right| \leq T_k \text{ for all } k \right\} \tag{28}$$

Where:

$$s_v^{T(k)} = e_v^T S^{(k)} \tag{29}$$

At this point, Equation(28) only defines a class of regions R, as we have not mentioned yet how to choose the thresholds T_k.

3.8 Determination of the solution separation thresholds T_k

The thresholds T_k must be set to meet the false alert requirement (Equation (19)). To do this, we exploit the fact that, under nominal conditions, the distribution of each of the solution separation statistics is known, because we have:

$$\left(s_v^T - s_v^{T(k)} \right) y = \left(s_v^T - s_v^{T(k)} \right) n \tag{30}$$

Therefore:

$$\left(s_v^T - s_v^{T(k)} \right) y \sim N\left(0, \sigma_{v,ss}^{(k)} \right) \tag{31}$$

where:

$$\sigma_{v,ss}^{(k)} = \sqrt{\left(s_v^T - s_v^{T(k)}\right) C \left(s_v - s_v^{(k)}\right)} \tag{32}$$

As a consequence, the probability that one of the solution statistics exceeds the threshold under nominal conditions is given by:

$$P\left(\left|\left(s_v^T - s_v^{T(k)}\right) y\right| > T_k\right) = 2Q\left(\frac{T_k}{\sigma_{ss,k}^{(k)}}\right) \tag{33}$$

Now, using the property that the probability of at least one event out of a set of events is bounded by the sum of the probabilities of each of the events, the false alert probability is bounded as follows:

$$\begin{aligned} P\left(y \notin R \mid H_0\right) &= P\left(\bigcup_k \left(\left|\left(s_v^T - s_v^{T(k)}\right) y\right| > T_k\right)\right) \leq \sum_{k=1}^{K} P\left(\left|\left(s_v^T - s_v^{T(k)}\right) y\right| > T_k\right) \\ &= 2\sum_{k=1}^{K} Q\left(\frac{T_k}{\sigma_{ss,k}^{(k)}}\right) \end{aligned} \tag{34}$$

Where:

Q is the right hand side cdf of a normal standard distribution.

Equation (34) imposes a constraint on the choice of the thresholds T_k. A simple choice (and often sufficient [reference RAIM int and cont]) is to distribute the false alert evenly across the K hypotheses:

$$T_k = Q^{-1}\left(\frac{P_{fa}}{2K}\right) \tag{35}$$

3.9 Integrity Equation

At this point, we have defined the detection region, but we have not shown yet how to evaluate the integrity risk. The integrity equation will be derived directly from the integrity requirement expressed in Equation (23). Given the definition of the region *R*, the following inclusion is always true for any *k*:

$$R \subset \left\{ y \mid \left| \hat{x}_v - \hat{x}_v^{(k)} \right| \le T_k \right\} \tag{36}$$

As a consequence, the contribution to the integrity risk of each hypothesis H_k (Equation (24)) can be bounded as follows:

$$\max_{x_{fault,k}} P\left(\left| x_v - \hat{x}_v \right| \ge VAL, y \in R \mid H_k, x_{fault,k} \right) \le \max_{x_{fault,k}} P\left(\left| x_v - \hat{x}_v \right| \ge VAL, \left| \hat{x}_v - \hat{x}_v^{(k)} \right| \le T_k \mid H_k, x_{fault,k} \right) \tag{37}$$

In the case where the position solution is computed using the least squares solution (Equation (21)), it is possible, and even practical, to compute numerically the left side of the above inequality [refARAIM2012]. However, this formula relies on the independence of the position estimation error and the test statistic, which is difficult to guarantee when the error distributions do not strictly follow the Gaussian assumptions.

Simple upper bound of integrity risk

A simpler and more robust approach (in the sense that it is more robust to the error model) can be obtained by removing the dependency on the fault parameter $x_{fault,k}$ [refARAIM2012]. Specifically, we have:

$$\begin{aligned} &\max_{x_{fault,k}} P\left(\left| x_v - \hat{x}_v \right| \ge VAL, \left| \hat{x}_v - \hat{x}_v^{(k)} \right| \le T_k \mid H_k, x_{fault,k} \right) \le \\ &P\left(x_v - \hat{x}_v^{(k)} \ge VAL - T_k \mid H_k \right) + P\left(x_v - \hat{x}_v \ge VAL \mid H_0 \right) \end{aligned} \tag{38}$$

This inequality (shown [ARAIM algorithm reference]) simplifies the integrity calculation, because the distribution of the random variable $x_v - \hat{x}_v^{(k)}$ is known (it is independent of the fault state), more precisely:

$$x_v - \hat{x}_v^{(k)} \sim N\left(0, \sigma_v^{(k)}\right) \tag{39}$$

Where:

$$\sigma_v^{(k)} == \sqrt{s_v^{(k)T} C s_v^{(k)}} \tag{40}$$

Again, the key property here is that this distribution holds even in the case of fault hypothesis H_k, As a consequence, we have:

$$P\left(x_v - \hat{x}_v^{(k)} \geq VAL - T_k \mid H_k\right) = Q\left(\frac{VAL - T_k}{\sigma_v^{(k)}}\right) \tag{41}$$

Finally, a practical upper bound of the loss of integrity risk is given by:

$$\begin{aligned} P\left(\left|x_v - \hat{x}_v\right| \geq VAL, y \in R\right) &\leq 2 p_0 Q\left(\frac{VAL}{\sigma_v^{(0)}}\right) + \sum_{k=1}^{K} \left(Q\left(\frac{VAL - T_k}{\sigma_v^{(k)}}\right) + Q\left(\frac{VAL}{\sigma_v^{(0)}}\right)\right) p_k \\ &\leq 2Q\left(\frac{VAL}{\sigma_v^{(0)}}\right) + \sum_{k=1}^{K} Q\left(\frac{VAL - T_k}{\sigma_v^{(k)}}\right) p_k \end{aligned} \tag{42}$$

where Q is the right hand side cdf of a unit Gaussian:

$$Q(t) = \int_{u=t}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \tag{43}$$

3.10 Protection level equation

From the integrity equation, we can deduce a formula for a valid Protection Level. If we go back to section 1.1, the Protection Level is defined as an Alert Limit that exactly meets the integrity requirement. That is:

$$2Q\left(\frac{VPL}{\sigma_v^{(0)}}\right)+\sum_{k=1}^{K} Q\left(\frac{VPL-T_k}{\sigma_v^{(k)}}\right) p_k = P_{HMI} \tag{44}$$

This formula is the perhaps the most important in this chapter, because it concisely integrates all the elements of ARAIM: the threat model (through the probabilities p_k), the nominal error model (through the sigmas), the false alert (through the thresholds), and the integrity requirement (through the P_{HMI}).

Another interpretation of the PL

We have defined the *PL* as the AL for which the integrity is exactly met. After extending its definition, we provide another interpretation of the *PL*, which turns out to be convenient. If the test statistics are below the thresholds T_k, then the *PL* is defined by Equation (44), if not, then the *PL* is set to $+\infty$. With this definition, the *PL* is dependent on the actual measurements, and it has the following property:

$$P\left(\left|x_v-\hat{x}_v\right| \geq PL\left(y\right)\right) \leq P_{HMI} \tag{45}$$

In this form, the test results are integrated in the *PL* definition. This interpretation of the *PL* turns out to be useful when defining exclusion algorithms [reference ARAIM ADD online].

3.11 Exclusion

in an ARAIM algorithm consists in finding a set of measurements for which the test statistics pass [references]. The receiver can then output the Protection Level corresponding to this set. It is however important to keep in mind that the integrity risk available must be shared across all the possible exclusion options [references]. The allocation of this integrity risk is very dependent on the assumptions on the causes of loss of continuity.

3.12 Interpretation of the PL equation

In addition to being practical, the PL equation (44) has a simple interpretation. Let us consider one term in the summation:

$$Q\left(\frac{VPL - T_k}{\sigma_v^{(k)}}\right) \tag{46}$$

This is the probability that the position error of a subset position solution located T_k from our all-in-view position will exceed the *VPL*. In other words, the *PL* tracks the subset solutions and adds an integrity margin that is dependent on the prior probability p_k.

In fact, it is possible to formulate a measurement dependent PL (in the sense described in 3.11) that does not use thresholds [references].

3.13 Error model refinements and resulting modified integrity equation

Nominal biases

To account for unknown, but bounded, biases, it is possible to add a bias to this distribution and assert that the norm of the bias is bounded by a known maximum value. This modifies the PL equation as described in [reference]. This is actually one of the advantages of solution separation over using the sum of squared residuals.

Integrity error models vs accuracy error models

In addition, it is useful to have two error models for each pseudorange: a conservative one and a realistic one. As discussed in [5] the different requirements target different probability levels and different levels of hazard severity. For this reason, there are at two different pseudorange error models. One error model, which will be labeled the integrity nominal error model, is applied for the requirements that are in the Hazardous category. This error model is only used in the PL terms that guarantee the integrity of the error bound.

The other error model, which is labeled the accuracy nominal error model, is applied in the requirements that are in the Major category (which require less scrutiny). This error model is used to compute the terms in the PL that only affect continuity (the false alert rate), the EMT requirement, and the accuracy. Both error models are characterized by a Gaussian overbound with a maximum nominal bias. For a given geometry, the covariance of the measurements will be designated by C_{int} for the integrity error model and by C_{acc} for the accuracy error model. Both covariances are usually diagonal, and the formulas to compute each term is given in [2]: they include the nominal satellite clock and ephemeris error, the tropospheric delay error, and the code noise and multipath error. The maximum biases are designated by b_{int} and b_{acc} respectively.

4. User Algorithm in the ARAIM concept

In section 3 we have presented a generic Advanced RAIM algorithm that can be used for many applications. The description above assumes for example that we know how to compute the probabilities p_k and the covariance C. In this section we bridge this gap by explaining first how to form the nominal covariance and then how to derive the probabilities p_k from the ISM parameters P_{sat} and P_{const}.

4.1 Nominal covariance

The covariance C includes the effect of the nominal errors: clock and ephemeris error, multipath and receiver noise, residual tropospheric delay, (and residual ionospheric delay in the case of single frequency). Since these errors are independent, we can write:

$$C(i,i) = \sigma_{URA,i}^2 + \sigma_{tropo,i}^2 + \sigma_{airborne,i}^2 \tag{47}$$

Where:

$\sigma_{URA,i}^2$ is the standard deviation of the clock and ephemeris error

$\sigma_{tropo,i}^2$ is the standard deviation of the residual tropospheric delay

$\sigma_{user,i}^2$ includes the effect of the multipath and the receiver noise

For single frequency GPS aviation users, these errors are well understood and the corresponding error models have been used for several years. For L5 signals and other constellations, they are still being developed for L5 [ref ARAIM report].

4.2 Fault modes

The ISM provides:

- $P_{sat,i}$: probability of a fault on satellite i
- $P_{const,j}$: probability a fault affecting two or more satellites within a constellation j

In addition we define:

- N_{sat} : number of satellites in view
- N_{const}: number of constellations in view

Throughout this section, “satellites in view” and “all-in-view” will refer to all the satellites that are selected by the receiver given its limitations, the ISM, and any additional constraints (like the approval of GNSS elements by States).

The N_{sat} + N_{const} single fault events characterized by the ISM should be treated as independent events. In particular, they are not exclusive. Therefore, in the integrity risk assessment, the probability of having simultaneous faults must be accounted.

This assessment should be done by considering the set of jointly exhaustive and mutually exclusive fault modes indexed by k, (where k=0 will refer to the fault free mode) formed of all the possible combinations of the events specified in the ISM, of which there are $2^{Nsat+Nconst}$ (the number of subsets in a set of size $N_{sat} + N_{const}$). To lighten the notations, we define:

$$\begin{aligned} P_{event,i} &= P_{sat,i} \\ P_{event,Nsat+j} &= P_{const,j} \\ N_{events} &= N_{sat} + N_{const} \end{aligned} \tag{48}$$

For example, k=1 could refer to a fault in satellite 1, and no fault in the other satellites or constellations. The probability of fault mode 1 would be given by:

$$P_{sat,1}\prod_{i=2}^{N_{sat}}\left(1-P_{sat,\mathrm{i}}\right)\prod_{j=1}^{N_{const}}\left(1-P_{\mathrm{const,j}}\right) \tag{49}$$

More generally, the probability of fault mode k is given by:

$$p_{fault,k} = \prod_{i=1}^{N_{events}} P_{event,\mathrm{i}}^{B_{i,k}}\left(1-P_{event,i}\right)^{1-B_{i,k}} \tag{50}$$

where $B_{i,k}$ is equal to one if event i is in fault mode k and zero otherwise. As will be seen in the reference algorithm in section 4.3, it is not necessary to compute all fault modes.

For the integrity risk computation, it should be assumed that a fault is the addition of an arbitrary bias to the affected satellite, or an arbitrary vector of biases in a group of satellites within a given constellation.

4.3 Subset fault selection

The ISM does not specify explicitly which fault modes need to be monitored or their corresponding prior probabilities. This determination must be made by the receiver based on the contents of the ISM, which specifies the probabilities of events that can be treated as independent.

This paragraph provides a method to establish a list of event combinations (the fault modes) to be monitored. The objective is to make sure that the sum of the probabilities of the modes that are not monitored do not exceed a pre-defined fraction of the total integrity budget (P_{THRES}). The list of fault modes that need to be monitored described here is only sufficient (there could be shorter lists that also meet the integrity requirements). The approach consists on moving fault modes from the list of not-monitored to the monitored list one by one until the remaining modes have a total probability below a pre-defined threshold. We want:

$$\sum_{k \text{ not monitored}} p_{fault,k} \le P_{THRES} \tag{51}$$

This approach is practical because we know that the sum of all the probabilities is one:

$$\sum_{k=0}^{2^{Nevents}} p_{fault,k} = 1 \tag{52}$$

The condition expressed in Equation (51) can therefore be written:

$$\sum_{k \text{ monitored}} p_{fault,k} \geq 1 - P_{THRES} \tag{53}$$

This way, it is only necessary to compute the probabilities (using Equation(49)) of the modes that will be monitored. We then need to decide the order in which the faults are considered.

The order is defined as follows:

- From smallest degree to larger
- Within one degree, from larger to smaller $p_{fault,k}$

where the degree is the number of primary events forming the composite fault mode. If a fault cannot be monitored, it is not included in the list of fault modes and we move to the next one. Each fault mode k is characterized by the set of indices corresponding to the measurements that are not affected by the fault, which will be noted idx_k. The set idx_0 corresponds to the full set of indices.

The integrity risk from the fault modes that are not monitored is bounded by $\bar{P}_{fault,not\ monitored}$, which is defined as:

$$\bar{P}_{fault,not\ monitored} = \sum_{k\ not\ monitored} p_{fault,k} \tag{54}$$

Fault consolidation

After establishing the initial list above, the algorithm consolidates multiple satellite faults from the same constellation with the constellation wide fault. This is done as follows: for each constellation j, we note k_j the fault mode corresponding to the fault of constellation j only, and C_j the set of fault modes that are formed of satellite faults included in constellation j (and included in the list established above). If the following inequality holds:

$$\sum_{k \in C_j} p_{fault,k} \leq F_C p_{fault,k_j} \tag{55}$$

where F_C is a fraction of 1, the fault modes in Cj are removed from the list and the probability of fault mode k_j is updated as follows:

$$p_{fault,k_j}^{(updated)} = p_{fault,k_j} + \sum_{k \in C_j} p_{fault,k} \tag{56}$$

Filtering the subsets

Among the subset faults determined in the previous section, there could be some that cannot be monitored (because the remaining satellites do not allow the receiver to compute a position). In this case, these events must be removed from the list of faults (and their integrity risk subtracted from the available budget). This is true, for example, of all subsets with three satellites or less belonging to one constellation, or four satellites or less belonging to two constellations. We note $P_{unobservable}$ their total probability and therefore an upper bound on their contribution to the integrity risk. An upper bound on the total integrity risk of the modes that are not monitored is given by:

$$P_{fault,not\ monitored} = \bar{P}_{fault,not\ monitored} + P_{unobservable} \quad (57)$$

Once we have determined the list of fault modes to be monitored and their associated probabilities p_k, we can apply the algorithm as described in section 3. However, we must make sure that we adjust the integrity budget, since we have already used part of it to account for the modes that are not monitored.

5. Determination and broadcast of the threat model parameters (ISM)

As we can see in Equation (44), the Protection Levels (or equivalently, the integrity risk within the Alert Limit) will be directly dependent on the threat model parameters (which are encoded in the ISM. Both the integrity and continuity will therefore be only maintained if these parameters describe the properties of the signals accurately (or at least conservatively). The source of the ISM, its latency, and the broadcast method have been under discussion for several years. Initially, the ISM was seen as a way for the ANSP to modify the characterization provided by the CSP [reference critical elements]. In the current projected architecture, at least for horizontal guidance, each CSP (or a closely associated entity) determine the ISM according to the standards set by ICAO. This makes sense because it is extremely difficult for an outside entity to make assumptions on the future performance of a constellation. In this section we examine the different elements that will help determine the content of the ISM: service history, performance commitments, and offline monitoring.

5.1 Service history

For a core constellation to be included in the ARAIM position solution, it would be essential that a good service history has been demonstrated. Any event that could have caused serious integrity risk in an ARAIM user would cause a constellation not to be deemed suitable for ARAIM, and therefore not included in the ISM. Trust in the constellations included in ARAIM would be partly acquired through the analysis of service history. For this reason, service history should be documented to a much more precise and unambiguous extent than today. For example, there are still faults that are not verified, there are gaps in the pseudorange measurements, and signal deformation effects should be quantified further in the standalone case (and for dual frequency users). It is worthwhile mentioning here that one of the most important pieces in the approval of WAAS LPV-200 operations was the analysis of service history over three years.

5.2 Performance commitments

Performance commitments are essential because even if stationarity were assumed in the fault statistics, the service history would not be sufficient to guarantee low bounds on the prior probabilities. A constellation could only be included in the ARAIM solution if performance commitments are published, sufficient, and met. In particular:

- onset probabilities of fault must be bounded,
- under nominal conditions position errors must be bounded by the error bound deriving from the broadcast SISA/URA,
- faults must be removed within a specified time, and
- signal deformation and code carrier coherence bounds should be included.

It is unlikely that a CSP would accept liability if the performance commitments were not met. The ANSP generating the ISM would have to decide whether the performance commitments of a given CSP are trustworthy.

5.3 Offline Monitoring

The offline monitoring would follow a process similar to the offline monitoring done in WAAS [4], which is reported quarterly and can lead to modifications in the constants assumed by the WAAS safety algorithms (these include bounds on the nominal ephemeris error and the probability of satellite fault). The type of analysis that would be necessary would be very similar to the ones performed by the FAA TEC center and reported in the Performance Analysis Reports for both GPS and WAAS [5].

While most of the analysis would rely on automatic tools (just like they rely on automatic tools for the safety analysis of WAAS [6]), there would be room for human intervention to handle exceptions. For example, if a receiver used in the generation of precise ephemeris appeared to be faulty, the list of receivers would be updated to exclude it or replace it. Also, in the case a fault had different effects depending on the type of receiver, it might be necessary to initiate an investigation. As a general rule, human intervention would only be necessary in ambiguous cases.

Reference network

The reference network used to support the ISM would need to be global, as the ISM is meant to be valid anywhere on the globe. Because of the long latency nature of the ISM, there are no stringent requirements on the latency of the measurements, as there are with SBAS. For this reason, a global network based on already existing networks as IGS, NGA or CORS could be appropriate. The data would have to be of good quality –and the stations chosen carefully-, but it would not need to be at the level of the SBAS reference receivers in terms of reliability. Instead, the reliability of the overall set of measurements would be obtained through redundancy, as the cost of additional receivers would low (mostly the cost of downloading the data). Because the measurements might not be under the control of the ANSP, it would be essential to make sure that they come from a variety of independent and trustworthy sources.

APPENDIX

Criticality of operations

The violation of an error bound will have different consequences depending on the operation. These consequences can be categorized following the severity levels listed in Table 1 ([8],[10]). For en-route or non-precision approach, the violation of an error bound would constitute a Major event, whereas it would be Hazardous for precision approach. In practice, this means that the generation of the position error bound must go through much more scrutiny. The new algorithms and assumptions that could provide vertical guidance have been labeled Advanced RAIM (ARAIM) to distinguish them from the current RAIM

Severity	**Description**	**Required DAL**
Catastrophic	Failure may cause a crash. Error or loss of critical function required to safely fly and land aircraft	A
Hazardous	Failure has a large negative impact on safety or performance, or reduces the ability of the crew to operate the aircraft due to physical distress or a higher workload, or causes serious or fatal injuries among the passengers. (Safety-significant)	B

Major	Failure is significant, but has a lesser impact than a Hazardous failure (for example, leads to passenger discomfort rather than injuries) or significantly increases crew workload (safety related)	C
Minor	Failure is noticeable, but has a lesser impact than a Major failure (for example, causing passenger inconvenience or a routine flight plan change)	D
No effect	Failure has no impact on safety, aircraft operation, or crew workload.	E

Table A.1 *Severity levels, description, and corresponding Design Assurance Level (DAL).*

REFERENCES

Global Positioning System (GPS) Standard Positioning Service (SPS) Performance Analysis Report available at http://www.nstb.tc.faa.gov/reports/PAN82_0713.pdf.

Phase II of the GNSS Evolutionary Architecture Study, February 2010 http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/library/documents/media/GEASPhaseII_Final.pdf

WAAS Minimum Operational Performance Specification (MOPS), RTCA document DO-229D.

ICAO, Annex 10, Aeronautical Telecommunications, Volume 1 (Radio Navigation Aids), Amendment 84, published 20 July 2009, effective 19 November 2009, GNSS Standards and Recommended Practices (SARPs) are contained in Section 3.7 and Subsections, Appendix B, and Attachment D.

Working Group C, ARAIM Technical Subgroup, Milestone 3 Report, February 26, 2016. Available at:

http://www.gps.gov/policy/cooperation/europe/2016/working-group-c/

http://ec.europa.eu/growth/tools-databases/newsroom/cf/itemdetail.cfm?item_id=8690

Working Group C, ARAIM Technical Subgroup, Milestone 2 Report, Issue 1.0, February 11, 2015. Available at:

http://www.gps.gov/policy/cooperation/europe/2015/working-group-c/

http://ec.europa.eu/growth/tools-databases/newsroom/cf/itemdetail.cfm?item_id=8191

Working Group C, ARAIM Technical Subgroup, Interim Report, Issue 1.0, December 19, 2012. Available at:

http://ec.europa.eu/enterprise/newsroom/cf/_getdocument.cfm?doc_id=7793

http://www.gps.gov/policy/cooperation/europe/2013/working-group-c/ARAIM-report-1.0.pdf

Walter, Todd, Gunning, Kazuma, Phelts, R. Eric, Blanch, Juan, "Validation of the Unfaulted Error Bounds for ARAIM", *NAVIGATION, Journal of The Institute of Navigation*, Vol. 65, No. 1, Spring 2018, pp. 117-133.

Blanch J., Walter, T., and Enge, P.," Protection Levels after Fault Exclusion for Advanced RAIM," *NAVIGATION, Journal of The Institute of Navigation*, Vol. 64, No. 4, Winter 2017, pp. 505-513.

Blanch, Juan, Walter, Todd, Enge, Per, "Theoretical Results on the Optimal Detection Statistics for Autonomous Integrity Monitoring", *NAVIGATION, Journal of The Institute of Navigation*, Vol. 64, No. 1, Spring 2017, pp. 123-137.

Blanch, J., Walter, T., Enge, P., Kropp, V.,"A Simple Position Estimator that Improves Advanced RAIM Performance," IEEE Transactions on Aerospace and Electronic Systems Vol. 51, No. 3, July 2015.

Blanch, J., Walter, T., Enge, P., Lee, Y., Pervan, B., Rippl, M., Spletter, A., Kropp, V., "Baseline Advanced RAIM User Algorithm and Possible Improvements," *IEEE Transactions on Aerospace and Electronic Systems*, Volume 51, No. 1, January 2015.

Blanch, J., Walter, T., and Enge, P.," Optimal Positioning for Advanced RAIM," *NAVIGATION*, Vol. 60, No. 4, Winter 2013, pp. 279-290.

Blanch, J., Walter, T., Enge, P. "RAIM with Optimal Integrity and Continuity Allocations Under Multiple failures," *IEEE Transactions on Aerospace and Electronic Systems* Vol. 46, No. 3, July 2010.

Phelts, R.E., Blanch, J., Walter, T., Enge, P., "The Effect of Nominal Signal Deformation Biases on ARAIM Users," *Proceedings of the 2014 International Technical Meeting of The Institute of Navigation*, San Diego, California, January 2014, pp. 56-67.

Macabiau, C., Milner, C., Tessier, Q., Mabilleau, M., Vuillaume, J., Suard, N., Rodriguez, C., "Impact of Nominal Bias Bounding Techniques on Final ARAIM User Performance," Proceedings of the 2014 International Technical Meeting of The Institute of Navigation, San Diego, California, January 2014, pp. 68-77.

Blanch, J., Walter, T., Enge, P., "Exclusion for Advanced RAIM: Requirements and a Baseline Algorithm," *Proceedings of the 2014 International Technical Meeting of The Institute of Navigation*, San Diego, California, January 2014, pp. 99-107.

Walter, T., Blanch, J., Enge, P., "Reduced Subset Analysis for Multi-Constellation ARAIM," *Proceedings of the 2014 International Technical Meeting of The Institute of Navigation*, San Diego, California, January 2014, pp. 89-98.

Joerger, M., Pervan, B., "Solution Separation and Chi-Squared ARAIM for Fault Detection and Exclusion," P*roceedings of IEEE/ION PLANS 2014*, Monterey, CA, May 2014, pp. 294-307.

Blanch, J., Walter, T., Enge, P., Wallner, S., Fernandez, F., Dellago, R., Ioannides, R., Pervan, B., Hernandez, I., Belabbas, B., Spletter, A., and Rippl, M., "Critical Elements for Multi-Constellation Advanced RAIM for Vertical Guidance," *NAVIGATION*, Vol. 60, No. 1, Spring 2013, pp. 53-69.

Walter, T., Blanch, J., Choi, M. J., Reid, T., Enge, P., "Incorporating GLONASS into Aviation RAIM Receivers," *Proceedings of the 2013 International Technical Meeting of The Institute of Navigation*, San Diego, California, January 2013, pp. 239-249.

Blanch, J., Walter, T., Enge, P., Pervan, B., Joerger, M., Khanafseh, S., Burns, J., Alexander, K., Boyero, J. P., Lee, Y., Kropp, V., Milner, C., Macabiau, C., Suard, N., Berz, G., Rippl, M., "Architectures for Advanced RAIM: Offline and Online," *Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014)*, Tampa, Florida, September 2014, pp. 787-804.

Lee, Y., "New Advanced ARAIM with Improved Availability for Detecting Constellation-wide Faults, Using two Independent Constellations," *Proceedings of the 2012 International Technical Meeting of The Institute of Navigation*, Newport Beach, CA, January 2012, pp. 1708-.