Detecting Multiple failures in GPS/INS integrated system: A Novel architecture for Integrity Monitoring

Umar Iqbal Bhatti and Washington Y. Ochieng
Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London.

Abstract

GPS is a widely used satellite navigation system. By design, there is no provision for real time integrity information within the Standard Positioning Service (SPS) which is available to the civilian community. However, in safety critical sectors like aviation, stringent integrity performance requirements must be met. This can be achieved using the GIC (GPS Integrity Channel) or RAIM (Receiver Autonomous Integrity Monitoring) or both. RAIM, the most cost effective method relies on data consistency, and therefore requires redundant measurements for its operation. An external aid to provide this redundancy can be in the form of an Inertial Navigation system (INS). This should enable continued performance even during RAIM holes. RAIM algorithms have traditionally been designed for the situation when only one failure occurs at a time. However, due to tighter alert limits and usage of GPS in urban environments there is now a focus on extending the RAIM concept to include multiple failures. Furthermore, in aviation, detection of simultaneous multiple slowly growing errors (SGE) is very challenging particularly in the case of integrated GPS/INS systems. This paper provides a detailed survey of RAIM approaches used to detect multiple failures proposed by navigation and geodesy communities. Furthermore the paper extends a previous algorithm proposed by the authors for the detection of a single SGE to the simultaneous multiple failure case for stand-alone GPS and integrated GPS/INS systems. Simulated and real data results attest to the effectiveness of the approach proposed. The developed algorithm is successful in the detection of multiple failures in GPS as well as in the INS. Furthermore isolation of the faulty sensor (GPS or INS) is possible with the same parallel filter structure. Hence, this approach will enhance the integrity of a GPS/INS integrated system installed on an aircraft.

Keywords: GPS, INS, Integrity, Multiple failures

1 Introduction

The Global Positioning System (GPS) today is the only fully operational satellite based navigation system. However, due to the recent shift in focus of worldwide aviation from ground based to space based navigation systems, the safety of use of GPS for such purposes is drawing much attention in current global research. The reason being that after switching off of Selective Availability (SA) in 2000, the capability of the SPS has increased dramatically. In order to use GPS for aviation, stringent standards, established by the International Civil Aviation Organization (ICAO), have to be met (ICAO SARPS, 2004). One of the requirements is integrity, a measure of the degree of trust that can be placed in the correctness of navigation information. However, the GPS Standard Positioning Service (SPS) performance standard (US DoD, 2001) does not have a requirement for the provision of real time integrity information. It must be noted here that PPS (Precise Positioning Service) is not available for civilian use hence this paper consider only the SPS. Hence, for safety critical applications like aviation, GPS signals must be monitored. The vulnerability of GPS signals has been investigated for example by Ochieng et al. (2003) and the Volpe Report (DoT, 2001). Furthermore, other recent research activities are focussed on the quantification of the failure modes of GPS (Bhatti and Ochieng, 2007a; Ochieng et al., 2003; Van Dyke et al., 2003, 2004; Walsh et al., 2004). These approaches are based on the exhaustive search for potential failure modes that can affect GPS navigation performance. In this regard research on Failure mode and Effect Analysis (FMEA) for the complex and multi-segmented GPS is still ongoing.

GPS augmentations like GBAS (Ground Based Augmentation Systems) and SBAS (Satellite Based Augmentation Systems) monitor GPS signals in real time. They relay integrity information by signals which are vulnerable to jamming and interference, a principal failure mode of GPS. A potentially effective method to the
exposure to such risks is to integrate GPS with other navigation systems such an Inertial Navigation System (INS). The INS is a self contained system with high short term stability, immune to jamming as well as interference. However, high grade systems are very expensive. The emergence of INS sensors exploiting MEMS (Micro-Electromechanical Systems) technology is creating the potential for affordable integrated GPS/INS architectures if the problems associated with performance could be overcome. This has the potential to offer a cost effective alternative to other forms of augmentations depending on the user (operational) requirements (Other augmentations include Wide Area Augmentation System (WAAS), Local Area Augmentation System (LAAS) or GRAS (Ground Based Regional Augmentation System). Long term use of the integrated system is possible if the INS can be calibrated periodically (Lee & O’Laughlin, 1999).

INS can be integrated synergistically with GPS so that short term and long term stabilities of INS and GPS respectively, can be exploited. The traditional integration method is the usage of a Kalman filter. However, in order to realise an optimal integrated system, a number of issues need to be considered. These include the type of INS and the integration architecture. Various types of integration methods are available, broadly classified as loosely coupled, tightly coupled and ultra-tight/deep (Gautier, 2003). Loosely coupled systems combine processed measurements of the two systems while tightly coupled systems generally carry out the integration at the raw measurement level. Ultra-tight systems generally have feedback loops between the INS and GPS.

In most of integrity algorithms for GPS as well as those for the integrated system, single satellite failure is assumed in general by the navigation community (Brown et al., 1992). The algorithms are hence designed to cope with one satellite failure at a time. However, more stringent alert limits such as for aircraft precision approaches and the use of GPS in urban environments calls for autonomous integrity monitoring techniques such as RAIM to have the capability to detect multiple failures. It is noteworthy that in addition to those associated with GNSS such as GPS, consideration of multiple failures is important because their probability of occurrence becomes much greater when two systems (in this case GPS and INS) are integrated. It can be seen from the failure mode analysis in Bhatti and Ochieng (2007a) that in addition to the failure modes of the individual systems, a number of failure modes also arise due to the coupling mechanism of the two systems. An early attempt towards solving the problem of multiple failure detection with respect to satellite based navigation was by Brown (1997). In this approach the traditional parity space method and slope-max concept is extended to include two failures at a time. This algorithm introduced the slope max-max concept defined as the maximum slope for a pair of failed satellites. Subsequent work by Lee (2004) introduced the concept of extended RAIM, which is based on the maximum of the protection levels determined for the single and the multiple failure cases. It is noteworthy, however, that multiple failure detection has long been recognised in the geodesy community with the main technique being the so-called data snooping procedure (Baarda, 1968) and is largely the basis for subsequent methods applied in navigation.

This method is developed further and formalized in the form of Detection, Identification and Adaptation (DIA) procedure (Teunessin and Kleusberg, 2005; Hewitson and Wang, 2005).

In this paper the approaches by the two communities are compared. Furthermore, previously introduced rate detector algorithm (Bhatti et al., 2007c) is extended to address multiple slowly growing errors. A new architecture called the piggy back architecture is introduced that has the capability of detecting and isolating a fault in the GPS as well as INS.

This paper is arranged in seven sections. Section 1 introduces the paper. Section 2 presents a review of RAIM for multiple failures. Section 3 describes the efforts of navigation and geodesy community for the development of RAIM methods. Section 4 is devoted to the description of the proposed piggy back architecture. Section 5 provides simulation and real data results for the proposed architecture. Section 6 concludes the paper.

2 Extension of RAIM to multiple failures

Typically RAIM formulae designed in the eighties based on the assumption that there can be only one faulty satellite at a time. This assumption needs to be relaxed in the design of future algorithms for the following reasons:

In the future, following the modernization phase of GPS and the launch of GALILEO, the accuracy of the position solution of GNSS will improve by an order of magnitude. This will be due to the use of multiple frequencies and the availability of a greater number of satellites at a given location. This will not only improve RAIM performance but will also enable the users to have tighter protection limits. However, the probability of errors previously assumed to be insignificant can now be considered significant for lower alert limits. Hence, the probability of multiple faults will increase, and this impact has yet to be quantified (Hwang and Brown, 2005a).

The system reliability figure of 3 satellite failures per year is given by the nominal operation of the GPS and only specified for Signal In Space in GPS SPS performance standard (US DoD, 2001). But in actual practice when GPS is used in harsh urban environment, the probability of multiple failures becomes much higher e.g due to strong multipath errors.
In the context of this paper, treatment of multiple failures is important because the probability of multiple failures becomes much greater when two systems are integrated. It can be seen from failure mode analysis in Bhatti and Ochieng (2007a) that in addition to the failure modes of the individual systems, a number of failure modes also arise only due to the coupling mechanism of the two systems.

In the existing literature, there are two streams of RAIM methods to address multiple failures; developed by the navigation community and that by the surveying/geodesy community. Each of these two approaches will be described next.

**Multiple Failure Detection By Geodesy Community**

The efforts of the geodetic community are concentrated on finding the performance capabilities of space based navigation systems and multiple failure detection algorithms as far as RAIM is concerned. In this analysis, the satellite measurements are treated as measurements from a network. This is because in geodesy, measurements from a surveying network are treated generally. In this respect two recent examples are by Ochieng et al. (2002) and Hewitson et al. (2006). These are in fact offline simulations that are effective for prediction of RAIM availability all over the globe. Hence, these provide performance capability analysis of the space based navigation systems beforehand.

The detection of multiple failures as proposed in the geodesy literature known as the Detection, Identification and Adaptation process, is reviewed in the next section. This is followed by a review of multiple failure detection methods presented by the navigation community.

**Detection, Identification and Adaptation (DIA)**

The problem of multiple outlier detection has been well known in the surveying community for the last couple of decades. This is also known as data snooping method (Baarda, 1968). The need for outlier detection arose in the surveying of large networks. During the post processing phase, it is necessary to identify the outliers to avoid surveying errors. Significant contributions were then made by Pope (1975), Forstener (1983), Cross and Nicolai (1994) and Teunissen and Kleusberg (1998). In Teunissen and Kleusberg (1998), various methods are presented for quality control in GPS surveying. A review of RAIM algorithms in this context is provided by Hewitson and Wang (2006). In this method failures are detected by removing one measurement at a time by comparing the test statistic with a threshold determined by chi-squared statistics. Subsequently, failed measurement is identified in the identification phase. In the adaptation phase, either the measurement is discarded or threshold is elevated to accommodate the failed measurement.

In the context of this paper, it should be noted that in contrast to the traditional RAIM algorithms proposed by the navigation community, quality control in GPS surveying has included the consideration of multiple failures for a long time. These methods are structured under a general procedure known as Detection, Identification and Adaptation (DIA) (Teunissen and Kleusberg, 1998, Hewitson and Wang, 2006).

**Multiple Failure Detection as proposed by the Navigation Community**

In the failure detection methods pursued by the navigation community it is assumed that only a single measurement is faulty. When a typical RAIM method designed for a single failure assumption (e.g Brown, 1992) is used for multiple failure detection the following problems occur.

If the test statistic is greater than the threshold it is not possible to judge whether this is due to a single satellite failure or multiple satellite failures

It is also possible that due to multiple failures the test statistic does not cross the threshold even if the presence of only one of the bias failures would cause the test statistic to exceed the threshold. This can be referred to as masking effect.

In view of the above, in order to deal with multiple failures, a multiple solution separation approach may be utilized (Escher et al., 2002). In this configuration a full solution is formed alongside sub-solutions that are obtained by removing one measurement at a time. To modify this procedure in order to cater for multiple failures a further level of sub-solutions is to be formed for dual satellite failures. This concept of forming sub-solutions can be extended in a similar way to address multiple failures. The test statistic in this case is formed from the difference between the full position solution and the sub-solution. Similarly, another test statistic is formed from the difference between the first level sub-solutions and the second level sub-solutions. The test statistics are compared with the threshold. A dual failure situation may be identified if all of the test statistics (between full set
and first level sub-filter) are below the threshold except the one which excludes two faulty measurements.

In Wang (2005), multiple fault situation is considered for combined GALILEO/GPS case. Variation of availability of position solution with the number of faults is studied. The issue of effect of multiple failures on the test statistic is also discussed in Macabiau et al. (2005). It is shown analytically that the linear subspace of those errors that do not affect the test criterion has dimensions 4-(N-p) where N is the number of satellites and p is the number of faulty pseudorange measurements. If this value is negative, then a fault can be detected.

In Feng and Ochieng (2006), another method is presented to handle multiple failures. It is called the group separation (GS) method (it differs from the GS method proposed by Lee et al. (2005). This method is based on classifying the measurements on the basis of common mode faults. For example, measurements of one constellation are treated separately from that of another (e.g measurements that belong to Rubidium based satellites clocks belong to a different group from those derived from Cesium clocks based satellites). This method is especially beneficial to multiple failure detection as it substantially decreases the number of subfilters that are needed to be checked for failures.

Martini et al. (2006) proposed an error reconstruction strategy from the test statistic to detect particular multiple failure situation. But it is limited to special cases of high magnitude failures (e.g 5 km). It can be concluded that the method of multiple failures detection by the navigation community (Brenner, 1995; Escher et al., 2002) is similar to the DIA procedure presented by the geodesy community. Hence, in fact the essence of the method is to form multiple subsets (by exclusion of one or more measurements) and then examine various test statistics which can be compared with a pre-determined threshold (vertical protection limit is also calculated on the similar lines). In contrast to multiple failure handling by the geodesy community, protection limit calculations are also needed by the users in Aviation. A review in this respect is presented below.

**Calculation of HPL as proposed by the Navigation community**

In the context of multiple failures, there are quite a few approaches presented by navigation community recently. These are discussed below

**Slope Max-Max concept**

The issue of multiple failures for GPS was treated in by Brown in 1997. The slope-max concept (discussed in Brown, 1995) is extended to give the slope-max-max concept. Firstly, it is assumed that two satellites are faulty simultaneously. The condition that two multiple failures can be detected requires that sufficient redundancy is present in the available geometry. Hence the solution of the slope calculated is only possible when the number of assumed failures is (n-4) or less (n is the number of available measurements). In other cases, when number of failures is higher than this value, the test statistic becomes zero, hence making slope infinite.

From the available satellites, a pair is selected. The errors due to all the other satellites are assumed to be negligible. Then slope-max is calculated for this pair. In the case of the single satellite failure assumption, the slope is simply calculated by the usage of geometry matrix components and parity vector components relevant to the satellite as described by Brown (1995). The maximum value of slope among the available satellites is termed as slope-max.

The situation becomes complicated in the case of dual faults. This problem is posed as a maximization problem to derive the maximum slope. As the slope is defined by horizontal position error divided by modulus of the parity vector, an assumption is taken to simplify the analysis. It is assumed that modulus of p is unity and horizontal position error is to be maximized. The unity constrained modulus of the parity vector is multiplied by a Langrangian vector and its dot product with the horizontal position error is chosen as the objective function. When the derivative of this objective function is calculated and equated to zero it is found that this is a generalized eigenvalue problem. The calculation of the square root of the maximum eigenvalue gives the maximum slope for this pair. To calculate slope-max-max, this procedure is repeated for other pairs of satellites systematically choosing two at a time. The intuition that slope-max-max for a particular geometry is greater than slope-max is confirmed by a numerical example given in Brown (1997). The key assumption that the satellite with the maximum slope is the most difficult to detect in basic RAIM algorithm is also followed in this approach. This concept will be used further in the NIORAIM approach discussed later in this section.

**The HMAX concept**

The traditional slope max approach was analysed for the presence of multiple faults in the future GNSS scenario (when GALILEO also becomes available along with the modernized GPS) by Lee (2004). It is conjectured that due to the future availability of multiple frequencies and almost double the number of satellites, the navigation capability of GNSS as a whole will improve. This improvement will result in tighter alert limits as civilian users will want RAIM to be available for applications that demand better accuracy. In this case, the use of RAIM algorithms that are based on a single failure assumption is insufficient. This is especially true in the case when there are multiple satellite faults where none of the fault is
above the threshold to classify it as a Hazardously Misleading Information (HMI). When there are two faults present, neither of which is greater than the threshold, the ratio of these two faults is important for their detection by a traditional algorithm. Simulated plots for different ratios of the dual faults with their effect on the position error are presented by Lee (2004).

Although Lee (2004) only considered the vertical position error, horizontal position error is similar in nature and these techniques are translatable. From the simulation, it is found that dual-fault maximum slope is always larger than the single fault maximum slope. While this result is intuitively clear it is important to determine the magnitude of the difference in the two types of slope for a typical configuration. In a GPS constellation with 24 satellites there can be a ratio of dual over single maximum slope as large as 70. Hence the use of an algorithm based on the single fault assumption is inappropriate if the probability of multiple faults is not negligible. However, no such probability is available as part of the GPS standard (US DoD, 2001). For the case of a randomly selected numerical example, this ratio is around 7 (Brown, 1997).

A method called the HMAX method presented by Lee (1988) is later extended to cater for dual failures. The modification proposed is the inclusion of two terms in HPL formula a) bound of the position error due to a multiple satellite fault and b) 99.99% bound in a fault free case.

However, as given in Lee et al. (2005), this method which is later termed as group separation (GS) is not very efficient in terms of availability. Another method that specially addresses availability is the NIORAIM method.

Novel Integrity-Optimized RAIM (NIORAIM)

A new RAIM procedure known as Novel Integrity-Optimized RAIM (NIORAIM) was present by Hwang (2005). The motivation behind this effort is to increase the availability of the conventional RAIM algorithm. However, this method does not contribute towards detection of multiple failures and is limited to the calculation of protection limits.

As conventional RAIM is based on the slope-max concept which suffers a limitation that even if the most difficult to detect satellite is not the faulty one, its slope (which is the maximum slope) is used to calculate the protection levels. The new approach in NIORAIM proposes the use of a non-uniform weighted least square algorithm in place of the uniformly weighted least squares position solution. Hence, the impact of each satellite measurement on the position solution is different based on its weighting. These weights are initialized as uniform and then adjusted using an optimization algorithm. The criterion for the change of weight is inversely proportional to the integrity limit provided by each satellite. In this way after certain number of iterations, the weights are adjusted such that the integrity limits (or slopes) of the satellites become nearly equal. Although this method results in lowering the protection limit, it suffers from the fact that the position accuracy decreases. This is due to the use of non-uniform weights. As the position accuracy of GPS is improving as modernization takes place, this constraint is becoming less of a concern. A complication that arises in this case is that the covariance of position error and parity vector no longer remains zero with the use of non-uniform weights. The mutual correlations between the position error and the parity vector are zero in the case of uniform weights. However, in general, in the case of non-uniform weighting these become non-zero. A flowchart of this method is given in Figure 13.

Recently, a theoretical analysis of RAIM in the presence of multiple failures is presented by Liu et. al. (2007). Elegant and simple to calculate formulae are derived therein. This approach provides value of maximum slope in the presence of two satellite failures. In terms of integrity monitoring, however, use of maximum slope may be sub-optimal as suggested by the NIORAIM method.

Summary

The conclusions drawn from the literature regarding multiple failure RAIM algorithms are as below

RAIM needs to be extended to incorporate multiple faults.

The concept of HMAX might provide the solution for multiple failure detection but the availability of the method is limited. This essentially means horizontal protection limit for the algorithm has typically high value compared to a typical horizontal alarm limit.

NIORAIM is an effective method for increasing the availability as compared to conventional RAIM and has also been extended to include the case of multiple failures (Hwang and Brown, 2005b). This method is effective for protection limit calculations but does not address the problem of multiple failures detection.

The lookup table approach for NIORAIM is approximate and some of the availability limits can be in error as compared to the exhaustive Monte Carlo approach.

The weight computation method for the NIORAIM method although ad-hoc has been shown to have good performance.

The assumption that the faulty satellite is the one with the maximum slope is not always valid and can result in reduced availability (Walsh et al., 2005).

Detection of multiple failures has not been attempted except in Escher et al. (2002) and Feng and Ochieng (2006), as much of the effort is primarily directed towards the calculation of protection limits for the algorithms.
Significant research effort has been dedicated to the protection limits offered by the integrity algorithms in the presence of multiple failures (Lee, 2004, Hwang and Brown, 2005a). The detection of multiple failures is presented for RAIM availability analysis (for example Hewitson et al., 2006) by geodesy community. These methods are also applicable to online RAIM algorithms required for aviation and are similar in concept to the methods presented by the navigation community. However, the terminology is different.

After the discussion on GPS integrity monitoring, integrity of the integrated system is discussed. RAIM methods for tightly coupled and deeply integrated system are discussed next (see also Bhatti, 2007). These methods are not applicable to loosely coupled systems because access to the measurements in such systems is not provided, in general.

3 Treatment of Multiple Failures for tightly coupled GPS/INS integrated system

There are two approaches in the literature, one for failure detection and the other for the calculation of HPL in the case of multiple failures. These are described below.

Detection

In the context of GPS/INS integrated systems, the detection of multiple failure was addressed by Escher et al. (2002). The theoretical approach used is that of Brenner (1995). However, this approach is not based on simultaneous fault assumption. It is assumed by Escher et al. (2002) that multiple faults only occur at different epochs. Hence, the second fault can be dealt with after the exclusion of the first fault if method by Escher et al. (2002) is used. However, in order to address simultaneous fault detection, two or more simultaneous faults should be detected. In essence, this is a single fault detection algorithm. As discussed earlier, NIORAIM method is also used for GPS/INS integrated system.

Calculation of HPL

The NIORAIM method is applied to a GPS/INS integrated system in Hwang (2005). Weights calculated using the weight search techniques are applied to the matrices of the Kalman filter by using the following formulae

\[ R_k = wR_k (w^{-1})^T \]  

where \( R \) is the measurement matrix of the Kalman filter

\( k \) is current epoch, \( w \) is the new matrix of weights to be applied to the satellite measurements.

The results presented in Hwang (2005) show that the introduction of the non-uniform weights change the behaviour of the system to yield lower protection levels. However, when there is a constellation change, sudden weight changes can induce a large transient. This problem is solved by the choice of scaling factors for the weights. The results for the application of NIORAIM algorithm will be discussed later.

In the integrity literature, a method for integrity monitoring of deeply integrated systems (or ultra tightly coupled systems) is also discussed. This is presented in the next section.

The GI-RAIM deep integration Integrity monitoring algorithm

A RAIM method suggested for ultra-tightly coupled systems is the GI-RAIM (GPS Inertial RAIM) method (Gold and Brown, 2004). It is based on the BOPD (BOunded Probability of missed Detection) concept. Based on a pre-filter, it is anticipated that a certain satellite is faulty. The pre-filter is an algorithm that implements reasonableness checks to detect blunders in the data. By excluding this satellite, a position solution is computed. From the comparison of this solution with the full solution, the contribution of the faulty satellite to the radial position error is estimated with a high probability.

The treatment is presented for a single failure case and multiple failures situation is not discussed, hence this method is not discussed further.

In the case of integrated GPS/INS systems, extension of RAIM has been considered. However, in this approach there is little emphasis on the problem of detecting a fault in the INS should one occur. A very recent example of this is presented by Curt et al. (2006) where failure in INS is not considered alongside GPS integrity monitoring in an integrated GPS/INS system. It is argued by Lee and O’Laughlin (1999) that due to the very nature of the INS based Kalman filter, it is a great challenge to design an algorithm to cater for the errors in the INS. This is because the Kalman filter adapts itself to the slowly growing nature of the nominal errors in the INS. Hence slowly growing errors in a Kalman filter become very difficult to detect. However, this is important because most of the errors in the INS grow slowly over time (see Bhatti, 2007a).

This paper proposes a new architecture that has the capability to detect GPS and INS faults and can isolate the faulty measurement.
4 A new architecture for multiple failure detection for GPS/INS integrated system

In the traditional form of the tightly coupled architecture, the differences of the available satellite measurements from their predicted counterparts are formed. This prediction is obtained by the use of the receiver position estimated from INS measurements (see Brenner, 1995). However, this method suffers from the fact that an error in the INS affects all the components of the measurement vector. Hence, INS error cannot be excluded in any of the sub-filters. This can be accepted when using a very good quality INS with a large Mean Time Between Failure (MTBF) value. In effect, the INS literally acts as a reference to the Kalman filter in such configurations. However, this method is not suitable for the situations when a low cost INS with a lower MTBF value is utilized (for example a typical MEMS based INS).

This situation is discussed in Greer et. al. (2006). However, it must be mentioned here that the case when a fault occurs in the INS, no algorithm is shown which is able to exclude INS.

Furthermore, in Bruggeman et. al. (2007) it is shown that integrity monitoring of an integrated system may be improved by use of a dynamic model of the aircraft. However it must be mentioned that this (inclusion of aircraft dynamic model) essentially improves the coasting performance of the INS when there is an outage of the satellite navigation system and no improvement is integrity monitoring may be achieved. The integrity monitoring methodology is that of essentially of solution separation method (Young and McGraw, 2003).

Hence, a method is needed which exploits the advantages of INS but also allows it to be excluded for the purpose of fault isolation. Due to the manifold increase in the number of commercial aircraft and severe competition between airlines, there is a drive to keep total service provision costs down. In terms of aircraft navigation systems, therefore, significant research and development activities are aimed at the use of low quality Micro-electromechanical Systems (MEMS) based INS (Strachan, 2000). In this case, the INS cannot be used as a reference due to performance limitations. For this purpose, a scheme is presented in this paper, in which it is possible to isolate an INS if it fails. The basic configuration for the proposed architecture which is referred to in this paper as the piggy back architecture is that of the GPS receiver Kalman filter. In a GPS receiver, pseudoranges are fed to a Kalman filter which incorporates a dynamic model of the receiver, and a new position calculated every epoch (Parkinson and Spilker, 1996). The piggyback architecture extends the power of solution separation method of detection as well as exclusion to GPS/INS integrated system which was previously only used for GPS. In this architecture INS solution can also be separated from the final solution in the case of fault in the INS.

4.1 Configuration of the “Piggy Back” Tightly coupled architecture

For the case of integrity monitoring of an integrated GPS/INS system with a low cost INS, a configuration similar to a GPS receiver Kalman filter can be utilized. In a typical receiver, a Kalman filter estimates the position of the aircraft (or host vehicle) on the basis of satellite measurements. The receiver accepts the satellite measurements and updates the position solution at each epoch. A new approach is presented here in which a position derived from INS measurements can be used in the GPS Kalman filter configuration by treating it as a fictitious satellite measurement. This idea is similar to the Non Line of Sight (NLOS) concept used in Wireless Broadband Communications (WBC) (Correia and Prasad, 1997). In this case, the INS derived position is used to predict an extra pseudorange measurement for a satellite for which orbital information is available in the GPS broadcast message. It should be noted here that INS measurement is not undervalued in any sense but it is used in a way that it represents a fictitious satellite.

This new configuration is referred to as a piggy back tightly coupled architecture because it is based on the idea that the INS measurements piggy back on the GPS range measurement. A high level schematic of this architecture is shown in Figure 1. As shown, the tightly coupled Kalman filter accepts pseudorange measurements from the GPS. The INS position is converted to an additional pseudorange measurement by the use of broadcast ephemeris data. Then the Autonomous Integrity Monitoring by Extrapolation (AIME) method proposed by Diesel and Dunn (1996) is applied to the output of the Kalman filter. The AIME test statistic can then be used to monitor for faults in the measurements (whether from GPS or INS). AIME is a sequential algorithm in which the measurements used are not limited to a single epoch. The test statistic is a weighted average of the Kalman filter innovation over the past measurements. The weight matrix used in the test statistic is the inverse of the innovation covariance matrix of the Kalman filter. The tests statistic exhibit central and non-central chi-square distributions for the no-fault and fault cases respectively. Three test statistics are formed \( s_1, s_2 \) and \( s_3 \); averaged over 150 sec, 10 min and 30 min respectively. The decision threshold is also based on chi-square distribution. This is selected on the basis of a false alert rate of \( 10^{-5} \) per hour in a fault free environment. In practice, the rate detector algorithm can be implemented alongside the AIME algorithm to detect the slowly growing errors early (Bhatti et al., 2007b,c). The steps of the algorithm are as shown in Figure 2. After the initialization of the Kalman filter according to the host vehicle configuration, INS
measurement is converted to a GPS satellite measurement. This extended pseudorange set is then used in the Kalman filter. Then the test statistic is formed using Kalman filter innovation.

\[ T_k V_k = H_k \tilde{P}_k H_k^T + R_k \]  

where \( H_k \) is the measurement matrix, \( \tilde{P}_k \) is the a priori covariance matrix and \( R_k \) is the measurement covariance matrix.

Therefore, the test statistic is given by

\[ TS_k = T_k V_k^{1/2} \]  

where \( r_k \) is the innovation vector

\[ \begin{bmatrix} p \\ v \\ a \\ b \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\alpha & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p \\ v \\ a \\ b \end{bmatrix} \]  

where \( p \) is the position state, \( v \) is the velocity state for the test statistic, \( a \) is the acceleration state, \( \alpha \) is the correlation coefficient and \( b \) is the bias state (this is new state added as compared to the dynamic model shown in Bhatti et al. (2007c)).

The new measurement matrix is given by

\[ H = \begin{bmatrix} 1 & 0 & 0 & -1 \end{bmatrix} \]  

The new output matrix is given by

\[ C = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \]  

In this way, the dynamic model for the test statistics takes care of the residuals by modelling the bias. Hence, this is the final dynamic model proposed for the rate detector algorithm. If this algorithm (with the modification suggested above) is applied to the simulation model in Bhatti et al. (2007c), the estimated bias state would have been zero and the results would remain the same. This is because the test statistic (in the simulations in Bhatti et al., 2007b) exhibited a mean near to zero.
Kalman Filter Operation
1. Propagate state variables through time
2. Propagate state covariance through time
3. Calculate Kalman Gain
4. Perform update step
5. Calculate Innovation and its covariance
6. Calculate velocity of the test statistic

Initialize Rate Detector Kalman Filter
1. Initialize State Variables
2. Initialize State Estimate Covariance Values
3. Define Measurement Noise Matrix
4. Define Dynamic Matrix
5. Define Sample Time

Kalman Filter Operation
1. Propagate state variables
2. Perform Time update
3. Accept measurements from GPS and INS
4. Perform measurement update

Test Statistics Calculation Using AIME method

Measurement Processing
1. Accept Pseudorange Measurements from GPS receiver
2. Convert INS position to predicted pseudorange by using lever arm correction
3. Form measurement of the Kalman filter

Initialize Main Navigation Kalman filter
1. State Variables
2. State Estimate Covariance matrix
3. Measurement Noise Matrix
4. Dynamics Matrix
5. Define sample time

Velocity of the test statistic > velocity threshold
Yes
Integrity Flag

Fig 3: Flowchart for implementation of the Rate Detector Algorithm with the piggy back architecture
A single SGE can be detected using the rate detector algorithm with the piggy back architecture. In this case, the measurements are fed to the piggyback architecture, and the rate detector algorithm applied to the test statistic formed from its output to generate an integrity flag. The flowchart for this implementation is presented in Figure 3. There are two types of Kalman filters used in the configuration. The first type implements the integration algorithm for INS and GPS. The second Kalman filter is used to estimate the rate of change of the test statistic. By estimating the rate or velocity it is possible to detect failures early.

To detect multiple SGEs, multiple configurations of this type are required. The detection of a satellite fault (or an INS fault) is achieved by the use of sub-filter hierarchical type are required. The detection of a satellite fault (or an INS fault) is achieved by the use of sub-filter hierarchical type are required. The first type implements the integration algorithm for INS and GPS. The second Kalman filter is used to estimate the rate of change of the test statistic. By estimating the rate or velocity it is possible to detect failures early.

Fig 4: The hierarchy of sub-filters for detecting multiple failures

The practical configuration required for multiple failure detection is in the form of parallel filters. A high level block diagram for the practical implementation of parallel filters is shown in Figure 4. As shown, the full set solution consists of a Kalman filter that is formed by all the available measurements (five measurements are assumed to be available) and a measurement that is predicted by the use of the INS (represented by INS) are shown. There is a rate detector filter at the output of each of the filter or sub-filter. Further levels of sub-filters are formed using subset of full set of measurements. In order to detect additional failures (more than two), further levels of sub-filters are required.

4.2 Benefits of Piggy Back Architecture

There are many potential benefits of the use of the piggy back architecture, some of these are given below:

Existing GPS positioning software can be utilized (which are based on a Kalman filter configuration) with minor modifications. In the filter, only one addition is required. This is the computation of a fictitious satellite range using the INS derived antenna position and lever arm correction.

The various existing RAIM methods for GPS can be directly used for the purpose of monitoring the integrity of the integrated GPS/INS systems.

Slowly growing errors in the INS are treated in the same way as errors in GPS satellite measurements. This also answers an issue raised in the RTCA MOPS (RTCA, 2001) that there are no guidelines for single string (no redundancy) detection of failure/s in INS. “Single string” essentially means the standalone operation of an INS. Hence in this architecture, GPS integrity guidelines are applied for INS integrity.

The fictitious satellite measurement (derived from the INS position) can be chosen such that it makes the geometry of the satellites taking part in the position solution stronger. If measurements from more than one INS are available, this method is readily extendible. Another fictitious satellite measurement can be added to the position solution.

This method essentially treats INS not as a continuous system but as a system like GPS in which at each epoch, a new measurement is obtained independent of the previous measurements. Although this method is applicable to all classes of INS, it is more suited to a low quality INS.

DOP Improvement by the use of Piggy Back Tightly Coupled Architecture

In the case of the piggy back architecture, the INS is used to predict the position of a satellite for which data are available in the broadcast ephemeris irrespective of the existence of a line of sight between the GPS antenna and the satellite.

The immediate effect of this is the possibility of using those satellites for prediction that enhance the dilution of precision offered by the current geometry. For example, if a satellite in view is directly overhead, a satellite from the broadcast ephemeris (using the values of elevation angles from the aircraft to the relevant satellite) directly below the aircraft on the other side of the Earth can be chosen for maximum benefit. This will minimise the vertical
errors and enhance the value of DOP obtained by the GPS measurements alone

This effect is illustrated in Figure 6. It can be seen that there is an improvement of around 50% with this algorithm in the horizontal dilution of precision. The DOP value is calculated from the trace of the geometry matrix (Kaplan, 2005).

In this way, a potential benefit in accuracy is envisaged. This is because the accuracy of a ranging system depends on the accuracy of the ranges used and their geometry (Ochieng, 2006). Hence, improving the DOP value has the potential to increase the accuracy if the precision of the INS derived range is close to or better than the GPS pseudorange precision.

In this case, the satellite whose measurement is predicted using position derived from the INS is on the other side of the earth and hence the benefit is evident.

The error in the fictitious range measurement derived from the INS position depends on a number of factors including the contribution of the orbital errors. Clearly, there is the possibility to use more precise and accurate sources of orbital information e.g EGNOS (European Geostationary Navigation Overlay System), WAAS (Wide Area Augmentation System) and indeed other celestial bodies.

5 Simulation Analysis of the proposed Piggy Back Architecture

Simulations for the case of a single failure, multiple GPS satellite failure and the case when a single failure occurs in GPS as well as INS are discussed below.

5.1 Single Failure in a GPS measurement

The test statistics (TS) for each filter including the main filter, sub-filter and subsequent level sub-filters are formed as shown in Figure 5. Each of these test statistics is checked to determine whether the TS value rises above the threshold or not (i.e. presence or absence of failure). As given in Bhatti (2007), test static is directly formed as a function of horizontal position error. The issue that whether it faithfully shows the error is discussed therein and in Bhatti (2007c). It is shown by the help of simulation that use of such test statistic for detection of alert limit is appropriate.

For detecting slowly growing errors, each filter/sub-filter is followed by a rate detector Kalman filter (Bhatti et al., 2007c). The arrangement works as follows. A single failure is injected to one of the satellite measurements at 600 sec, for example, the measurement from SV4 (as per the arrangement shown in Fig 5).

At the sub-filter level, the sub-filter without the failed satellite is the only one with the TS below the threshold. This scenario is shown in Figure 7. The dotted line shows the TS of sub-filter e and continuous lines are for the other sub-filters. Hence, in this scenario, sub-filter e is the one with the TS below the threshold. Therefore, all the other sub-filters are not used further, as these contain the measurement from SV4. However, in
the situation where SV4 becomes unavailable and is replaced by any other satellite measurement, other sub-filters may be used subsequently. For the remaining flight time, sub-filter \( e \) becomes the primary filter and its lower levels of sub-filters assume the role of new sub-filters. A dual fault situation in GPS is considered below.

5.2 Dual Failures in GPS Measurements

A failure of magnitude 2 \( \text{m/s} \) is introduced in SV3 at 300 seconds. For clarity, all the test statistics are not shown in Figure 8. However, the conventional measurement domain test statistic (Equation 3) is shown. It can be seen that one of the sub-filters that contains the measurement from SV3 crosses the threshold. The sub-filter with SV3 excluded now becomes the primary filter. Another failure of 3 \( \text{m/s} \) is injected into SV2 at 15 minutes and this is detected at 21 minutes at a second detection point. The second detection is achieved by the TS of the sub-filter level that contains the measurement from SV2. It should be noted also from Figure 8 that detection point 1 was achieved earlier than the injection of the second error. This is because the sub-filter that is without the measurement from SV3 contains the measurement from SV2 (in which an error is injected). For such a scenario, level 2 filter is used (which excludes two measurements). The dotted line shows the TS of the level of the sub-filter that does not contain the measurements either from SV2 or SV3 and hence is below the threshold for the entire period. Note the transition of threshold due to the decrease in the number of available satellites (see Equation 4). Hence, detection of multiple failures in GPS is possible by using the sub-filter architecture. The case of a failure occurring in GPS as well as the INS is discussed below.

5.3 Case with a failure in INS as well as GPS

In the proposed piggy back tightly coupled architecture, the INS position is represented by a fictitious satellite range. This is used in a manner similar to the other satellite measurements. Hence, if there is a failure in the INS, sub-filter \( a \) is chosen as the primary filter for the rest of the flight. When there is an error in the INS, the test statistic grows due to this error and goes beyond the threshold. The INS fault may be identified because only the test statistic from the sub-filters that have excluded the INS will be below the threshold. In Figure 9, the INS is declared faulty after around 8 minutes due to the errors that are representative of a typical automotive grade INS. It should be noted that green colour is used for the test statistic of the filter that exhibit the first failure and blue for that test statistic that detects the subsequent failure. For a typical automotive grade INS, gyroscope biases are chosen as \( 1 \text{ deg/hr} \), \( 2 \text{ deg/hr} \) and \( 1 \text{ deg/hr} \) respectively for \( x \), \( y \) and \( z \) axes. The biases for \( x \), \( y \) and \( z \) accelerometers used are \( 1 \text{ milli-g} \), \( 2 \text{ milli-g} \) and \( 3 \text{ milli-g} \). An error of magnitude 3 \( \text{m/s} \) is injected in the GPS satellite measurement from SV3 at 25 minutes and is detected shortly afterwards at detection point 2.

Hence, the piggy back architecture is successful in the detection of multiple failures even if the fault is present in an INS. The GPS/INS algorithms in current literature can detect whether the integrated system is faulty or not (eg. Brenner, 1995). But only the piggy back architecture can detect that the fault is in the INS and it can be isolated and the algorithm continues. In the simulations presented, multiple failures are shown occurring one after the other and are subsequently detected. However, the piggy back architecture has the capability to detect multiple errors occurring simultaneously by monitoring the test statistics of all the filters/sub-filters. The case of simultaneous
onset of failures is shown in Figure 10. The onset time is the same for both the GPS and INS failures (10 minutes). However, since the INS error growth is less than the corresponding growth of the GPS failure (3 m/s), the INS failure is detected after the GPS failure at detection point 1 and 2 respectively.

5.4 Handling of multiple SGEs in real Data

The performance of the proposed piggy back architecture proposed was demonstrated by simulation developed in Bhatti et al., 2007c (see also Bhatti, 2007). In this section, the algorithm is subjected to real data. The error in the INS predicted pseudorange grows due to errors in the INS. The ‘failure’ of INS derived pseudorange measurement occurs after 20 minutes and is detected when the test statistic of one of the sub-filters containing that range exceed the threshold. This is because INS is not being calibrated. This effect is shown around 20 minutes in Figure 11. A test statistic of another filter is below the threshold till 22 minutes. It is the sub-filter that does not contain the INS predicted measurement. At 22 minutes, an error of 3 m/s is injected in a GPS satellite measurement leading to detection in 25 seconds (Detection point 2). However, the dotted line (near the X-axis) representing the test statistic for a fault free sub-filter is always below the threshold.

Figure 12 shows the case when failures are introduced in INS and GPS at the same time. A gyro fault in the azimuth gyro of 1 deg/hr is introduced in the INS and a range error of 3 m/s is injected in a satellite measurement at 5 minutes. The INS is declared faulty in 16 seconds while GPS measurement (blue line) is declared faulty in 30 seconds at detection point 1 and detection point 2 respectively. Although the fault is introduced at the same time, the detection time is different because of the different growth rates.

Since an integrity algorithm is not only characterized by its detection performance but also by its protection limit. The latter is considered below. Note that the Vertical Protection Level (VPL) is not considered because the INS is not stable in the vertical domain (an effect well known in the navigation community). The HPL is compared to the Horizontal Alert Limit (HAL) to make operational navigational support. The NIORAIM method (Hwang, 2005a) has been adopted in this paper to compute HPL. This method has the ability to reduce HPL (by optimization) with consideration of multiple failures. However, it should be noted that the NIORAIM method is limited to protection level calculations and does not provide solution to detection of failures. The HPL performance is described in the next section.
5.5 NIORAIM-based Horizontal Protection Level

The HPL is used to determine whether the position solution can be used for the particular phase of flight or not (whether HPL < HAL). In Bhatti et al. (2007c), respective values of HPL are plotted for the two existing integrity algorithms, the MSS (Multiple Solution Separation) and AIME (see Brenner, 1995 and Diesel and Dunn, 1996). The former is based on forming the solution using different sub-filters by removing one measurement at a time and comparing it with the full solution (Brenner, 1995). In effect it is a snapshot method which uses only the measurements at the current time. The test statistics are formed using the horizontal measurements of the full and the sub-solution. This is assumed to follow zero mean Gaussian distribution in the no fault case and non-zero mean Gaussian distribution in the faulty case. The decision threshold (to compare the test statistic with) is chosen based upon the maximum probability of false alert. Another threshold is computed for which it is assumed that if there occurs a fault it should not cross the threshold except with the specified missed detection probability. In contrast to the AIME method, the MSS is a position domain method. These methods assume the occurrence of a single failure at a time. Recently, Hwang and Brown (2005b) proposed the NIORAIM method to determine HPL in the presence of a dual fault scenario. The flowchart for calculation of HPL using the NIORAIM algorithm is presented in Figure 13. In the case of the piggy back architecture, the dual fault scenario covers both the classes of faults whether these are in two GPS measurements or in one GPS measurement and in the INS measurement. This is because the INS position is transformed to a GPS range measurement and is treated as such for the fault scenario. The HPL as calculated by the NIORAIM method (Hwang and Brown, 2005b) is shown in Figure 14. NIORAIM algorithm for typical GPS solution is used for piggyback architecture because of similarity of these two methods.

In this method, weights are applied to the satellite measurements and then optimized to reduce the HPL. The ‘start value’ is the value of the HPL arrived at by the use of initial weights while the ‘trained value’ is determined after the training of the weights. In this method, a dual fault situation was considered. It can be seen (Figure 14) that the value of HPL is very high and is in kilometres. The HPL values are quite high when compared to those shown in simulation results calculated by the use of existing GPS/INS integrity algorithms (Bhatti et al., 2007c). This is because those algorithms assume the case of a single failure. There can be a difference of an order of magnitude when a dual fault scenario is considered as compared to the single failure case (Brown, 1997).

The point to be noted from Figure 14 is that the NIORAIM training method is successful in decreasing the HPL substantially (up to 50%) for a part of the flight. However, this is not true in general as it is always not possible to find the weights that can optimize the HPL value.

This is because of the limitation of numerical optimization methods. The weights that are trained during the iteration process are shown in Figure 15.

The initial value for all of the weights used is unity. After training, these are reduced to significantly smaller values. In Figure 15, only the final values of the weights at each time epoch are shown. However, the important aspect of these weights is their relative value with respect to each other and not their absolute values (Hwang and Brown, 2005b). The change in weights is generally associated with a decrease in position accuracy. However, this accuracy is
degraded at the expense of improved ‘availability’ of RAIM (getting a reduced HPL). Hence, in the case of successful training of the weights, NIORAIM is a very effective method in increasing the availability in a given geometry. However, there are associated problems of non-divergence of weights or excessive time to find the minimum HPL which are common for training methods. In such a case, the HPL is calculated by initial weights (unity). Should be noted that Figure 15 contains the final weights for the simulated satellites while Figure 17 shows weights for satellite measurements in the real data.

The NIORAIM method is also applied to real data used in section 5.4. It can be seen from Figure 16 that a reduction of around 40 metres is achieved in the HPL by using the training algorithm. However, it can also be seen from the right hand side of the plot that training does not always converge and in some cases the initial values of HPL (with unity weights) is used. The respective trained weights arrived at by the training algorithm are shown in Figure 17 for the six available satellite measurements. It

6 Conclusion and Future work

This paper has reviewed methods presented for detection of multiple failures for integrated GPS/INS system. Furthermore the paper has introduced a new architecture for failure detection in integrated GPS/INS systems. It has the ability to detect multiple failures, whether these are in GPS, INS or both. Furthermore, it possesses a novel ability to show whether the fault is in the INS or GPS or both. In this way the faulty sensor can be isolated subsequently. It has been shown by simulated and real data that multiple failures in a GPS/INS integrated system can be detected especially for the case of failures that grow slowly over time. Furthermore, a new architecture in the form of parallel filters is introduced that has the ability to isolate failures in GPS/INS integrated system even for the case when INS fails or its output crosses the system error budget requirements. Future work to consolidate the
results is needed in the form of large number of Monte Carlo simulations.

References

Baarda (1968), A testing procedure for use in geodetic networks, Netherland Geodetic Comission, New Series, No. 4


Department of Transport (2001), Vulnerability Assessment of the Transportation Infrastructure relying on the Global Positioning System, USA.


Institute of Navigation, Kansas City, Missouri, USA, pp 1599-1606.

Gautier, J. (2003), GPS/INS generalized evaluation tool for the design and testing of integrated navigation systems, PhD Thesis, Stanford University, USA.


Parkinson, B. W. & Spilker, J. J. Jr. (1996), Understanding GPS: Principles and Applications, Volume 1, American Institute of Aeronautics and Astronautics, Virginia, USA


Wang, Yongchao (2005), *FDE and Positioning Performance in a Multi-Fault Scenario Using GPS and GALILEO*, ION GNSS 18th International Technical Meeting of the Satellite Division, 13-16 September 2005, Long Beach, CA, USA.


The corresponding author

**Umar Iqbal Bhatti** ([uiqbal3@hotmail.com](mailto:uiqbal3@hotmail.com)) did his BSc from University of Engineering and Technology, Lahore, Pakistan in 1994 in Electrical Engineering. He completed his MSc from King Fahad University of Petroleum and Minerals, Dhahran, Saudi Arabia in 1997 in Systems Engineering. He obtained his PhD in Navigation from Imperial College London in 2007. His PhD thesis is titled 'Integrity Algorithms for GPS/INS integrated systems in the presence of slowly growing errors'. He is currently working at Design Engineering Centre in Islamabad, Pakistan. Furthermore, he has taught courses in Inertial Navigation as a visiting faculty member. His research interests are GPS, INS, Integrity and Initial Alignment.