JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 24, No. 3, May–June 2001

Methodology for the Performance Evaluation of a Conflict Probe

Karl D. Bilimoria*
NASA Ames Research Center, Moffett Field, California 94035

A conflict probe is an air traffic controller decision support tool that can predict conflicts well in advance by the use of information on aircraft position, speed, and flight plans, along with forecasts of wind and temperature profiles. A comprehensive methodology is presented to evaluate quantitatively the performance of a conflict probe by the use of real traffic data and expanded separation criteria. The methodology is universal in nature and can be applied to any conflict probe. Several metrics of conflict probe performance have been developed and evaluated. The missed alert rate and false alert rate are primary metrics that quantify the reliability of a conflict probe; these metrics are partitioned by quality of flight intent information. The mean conflict warning time and errors in key conflict prediction parameters, for example, conflict start time, minimum horizontal and vertical separations, are important secondary metrics that quantify the accuracy of a conflict probe. The evaluation methodology developed was exercised by applying it to the Center-TRACON Automation System's Conflict Probe Tool, using real traffic data from the Denver Air Route Traffic Control Center; some preliminary results are presented as an example.

Introduction

As demands on the National Airspace System continue to increase, there is a need for decision support tools (DSTs) that can assist air traffic controllers in meeting the needs of airspace users, for example, increased capacity and efficiency, while maintaining the highest levels of safety.[1] In the context of this paper, a conflict probe is an air traffic controller DST that can predict conflicts well in advance by the use of information on aircraft position, speed, and flight plans, along with forecasts of wind and temperature profiles. Note that a conflict probe used as an air traffic controller DST merely provides advisories to a human controller who retains full authority and responsibility for safe separation of air traffic. Strategic (early) resolution of conflicts has the potential to reduce the cost associated with conflict resolution maneuvers and to enable other traffic management benefits. Such a tool would be especially useful in a free flight environment,[1] which is expected to have a less structured traffic flow compared to the current operating environment.

A complete evaluation of a conflict probe has two aspects: functionality and performance. A functionality evaluation is qualitative in nature and involves testing of conflict probe features and user interface through controller-in-the-loop simulations and field tests. Reference 2 describes a field test of conflict probing and trial planning capabilities. A performance evaluation is quantitative in nature and is directed at the conflict prediction engine that underlies the features and user interface of a conflict probe.

The objective of this work is to develop a comprehensive methodology to evaluate quantitatively the performance of a conflict probe in a manner that preserves all real-world effects such as flight intent errors, wind model errors, aircraft dynamics modeling errors, aeropropulsive modeling errors, navigation errors, and velocity errors due to radar tracker noise. One possible approach would be to use artificial traffic data in a simulation environment that attempts to model aircraft trajectories in the presence of various error sources. However, conflict probe performance degradation is primarily a manifestation of real-world errors, and they are difficult to model accurately. Other approaches to conflict probe performance evaluation may be found in Refs. 3–6. For example, Ref. 4 develops performance metrics using a hybrid approach involving data collection and transformation models applied to a recorded air traffic scenario.

The approach taken in this work is to evaluate the conflict probe using real traffic data to preserve all real-world effects. This is a challenging task because real traffic data includes the effects of controller actions to separate traffic; hence, true conflicts are generally not present in such data. Therefore, the parameters that define a conflict are expanded for the purposes of the evaluation, and the conflict probe is evaluated on its ability to detect these pseudoconflicts.

Conflict probe performance measures are described in the next section. A methodology for conflict probe performance evaluation is then presented. A later section describes a demonstration of this methodology by applying it to the Center-TRACON Automation System (CTAS)[7,8] Conflict Probe Tool,[9,10] using real traffic data from the Denver Air Route Traffic Control Center. Some preliminary results are presented as an example.

Conflict Probe Performance Measures

The performance of a conflict probe can be characterized by its reliability and accuracy (Fig. 1). Reliability refers to the capability of the probe to alert correctly the air traffic controller to impending conflicts. Accuracy refers to the timeliness and quality of conflict information provided by the probe to the air traffic controller.

Reliability

The reliability of a conflict probe is measured by the rate of missed alerts and false alerts. A conceptual definition of missed, correct, and false alerts is presented in Fig. 2. The observed conflicts set corresponds to all conflicts that were actually observed to occur; it is the truth set for reliability analysis. The predicted conflicts set corresponds to all conflicts predicted by the probe.

Correct alerts are predicted conflicts that were actually observed. Missed alerts are observed conflicts that were not predicted by the probe. False alerts (or nuisance alerts) are predicted conflicts that were not observed. Perfect reliability would correspond to a zero rate of false alerts and missed alerts. It is of interest to determine missed and false alert rates as functions of time to conflict.

Accuracy

The accuracy of correct alerts (Fig. 1) is measured by the value of the conflict warning time and the errors in key conflict prediction parameters. Conflict warning time is defined as the time interval

Presented as Paper 98-4238 at the AIAA Guidance, Navigation, and Control Conference, Boston, MA, 10–12 August 1998; received 21 January 2000; revision received 29 December 2000; accepted for publication 30 December 2000.

*Research Scientist, Automation Concepts Research Branch, Mail Stop 210-10; kbilimoria@mail.arc.nasa.gov. Associate Fellow AIAA.

between initial notification of the conflict and observed conflict start.

Examples of conflict prediction parameters appropriate for error analysis include minimum horizontal separation, minimum vertical separation, conflict start time, and aircraft positions at conflict start. A conflict probe typically provides numeric and/or graphic information on these conflict parameters to the air traffic controller. It is of interest to determine conflict parameter errors (for correct alerts) as functions of time to conflict.

Development of Conflict Probe Evaluation Methodology

The methodology developed in this work is universal in nature and can be applied to any conflict probe. It can, therefore, provide a framework for a comparative study of conflict probes. To evaluate the performance of a conflict probe, it should be exercised in an appropriate environment. The approach used here is to exercise the conflict probe using real traffic data to preserve all real-world effects that degrade the performance of a conflict probe. This requires the use of expanded separation criteria.

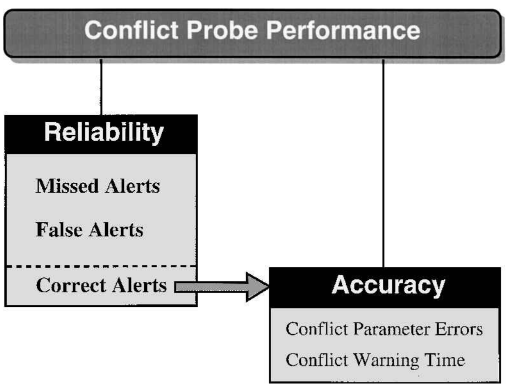


Fig. 1 Measures of conflict probe performance.

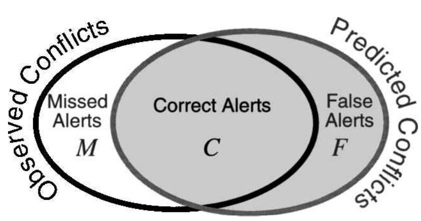


Fig. 2 Conflicts and alerts.

Expanded Separation Criteria

The operational separation criteria for en route flight correspond to a horizontal separation standard of 5 n mile and a vertical separation standard of 2000 ft (1000 ft if either aircraft is below flight level (FL) 290, i.e., 29,000 ft); they are represented by the operational conflict window shown in Fig. 3. An operational conflict occurs when these separation criteria are violated, that is, when two aircraft are horizontally separated by less than 5 n mile and vertically separated by less than 2000 ft (1000 ft if either aircraft is below FL 290). Because real traffic data includes the effects of air traffic controller actions, operational conflicts do not generally exist in such a data set. Therefore, a conflict probe cannot be evaluated with real traffic data while using an operational conflict window.

To evaluate a conflict probe with real traffic data, the concept of a pseudoconflict is introduced. A pseudoconflict occurs when the separation between two aircraft violates a specified set of expanded separation criteria that exceed the operational separation criteria. There are many possible ways to specify expanded separation criteria; some of them are represented by the pseudoconflict windows shown in Fig. 3.

Specifically, Fig. 3 presents three examples of pseudoconflict windows: horizontally expanded, vertically expanded, and vertically offset. It can be seen, for example, that if two aircraft at the same altitude have a horizontal separation of, for example, 8 n mile, they are within the horizontally expanded pseudoconflict window and could be considered as being in conflict for the purpose of conflict probe performance evaluation, even though they are not operationally in conflict. Similarly, if two aircraft (both flying above FL 290) are vertically separated by 4000 ft and have a horizontal separation of, for example, 3 n mile, they are within the vertically expanded as well as the vertically offset pseudoconflict windows and could be considered as being in conflict for the purpose of conflict probe performance evaluation, even though they are not operationally in conflict.

Operational conflicts are generally not observed in real traffic data. Therefore, to evaluate a conflict probe using real traffic data, it is necessary to generate an observable pseudoconflicts set whose characteristic parameters closely match those of the unobservable operational conflicts set. It is clear that many choices can be made for pseudoconflict windows, which result in many possible pseudoconflicts sets. The most appropriate pseudoconflict window for conflict probe evaluation was determined as will be described next.

Selection of Evaluation Window

The performance of a conflict probe is influenced by the characteristic parameters of the conflict itself. For example, a head-on conflict involving cruising aircraft is generally easier to predict compared to a conflict involving climbing/descending aircraft with a small

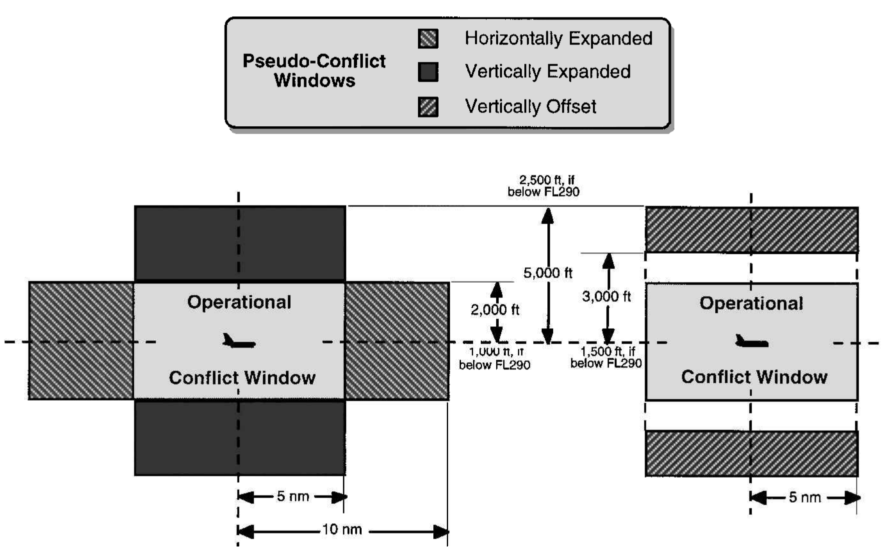


Fig. 3 Conflict windows.

encounter angle. Although there are many parameters that fully characterize a conflict, three key parameters were determined from experience: 1) encounter angle, 2) altitude–rate combination, that is, both aircraft level, one nonlevel, both nonlevel, and 3) minimum horizontal separation.

Different pseudoconflict windows will result in different sets of pseudoconflicts, characterized by different distributions of conflict parameters. Hence, the same conflict probe evaluated with different pseudoconflict windows may yield different performance results for each window. It is, therefore, important to identify a pseudoconflict window for which the conflict probe will yield results similar to those of an operational conflict window.

The following approach was used to determine the most appropriate pseudoconflict window. First, an operational conflicts set and various candidate pseudoconflicts sets (corresponding to various pseudoconflict windows) were generated using a common air traffic database. This database contained 3 h of real traffic data (corresponding to over 1000 aircraft) from the Denver Air Route Traffic Control Center. The operational conflicts set was determined from simulated aircraft tracks constructed with actual flight plans, by the use of actual aircraft birth points as initial conditions. Real track data were used to determine the pseudoconflicts sets. Note that none of these data were generated by a conflict probe.

Next, the characteristic parameters of the candidate pseudoconflict sets were determined and compared against those of the operational conflicts set. Figure 4 shows the results of this comparison for horizontally expanded, vertically expanded, and vertically offset pseudoconflict windows (other windows, such as a horizontally plus vertically expanded window, were also evaluated but are not shown in Fig. 4). The results of this comparative study indicate that the characteristic properties of the vertically offset pseudoconflict set generally match those of the operational conflicts set. The vertically offset pseudoconflict window was, therefore, selected to evaluate the performance of a conflict probe using real traffic data.

In the interest of brevity, only the term conflict will be used from this point forward, except when a distinction needs to be made between an operational conflict and a pseudoconflict.

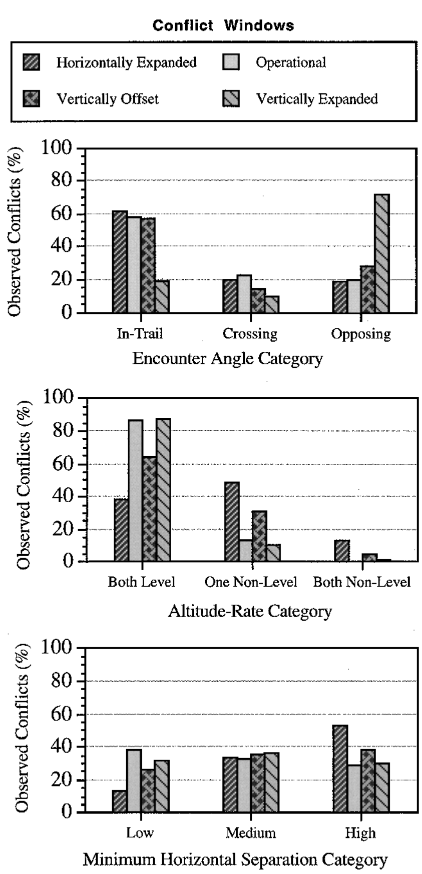


Fig. 4 Evaluation of conflict windows.

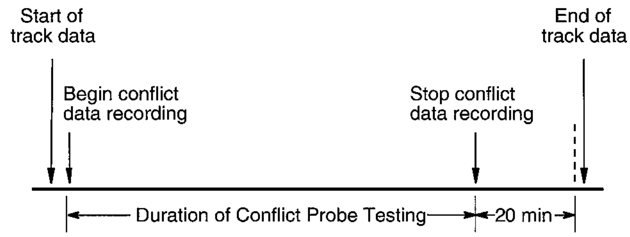


Fig. 5 Data collection procedures.

Data Collection

A conflict probe needs radar track, flight plan, and weather data for conflict prediction. Track data correspond to positions (x and y coordinates), processed from ground-based surveillance radar returns, and pressure altitudes h, measured by the aircraft's air data system and transmitted via mode C transponders. Flight plans contain information about the aircraft's route, assigned altitude, and other flight parameters. Track and flight plan data are available from the host computer system (commonly called the Host) of the air route traffic control center (commonly called the Center). The track data are updated approximately every 12 s; the flight plan data are activated some time before aircraft enter the Center's airspace and are updated when a flight plan amendment is entered into the Host.

Weather data consist of temperature and wind velocity forecasts for a specific time at various isobaric levels (pressure altitudes) over a defined horizontal grid. These data can be obtained over the Internet from the National Meteorological Center (NMC) in the form of a rapid update cycle (RUC) model generated using the Meso-scale Analysis and Prediction System.[11] The RUC model periodically generates files for wind and temperature forecasts.

A consistent set of track, flight plan, and weather data is input to the conflict probe with the appropriate conflict window settings to generate the corresponding conflict prediction data files. The track and flight plan data sets are loaded into the conflict probe, along with the corresponding weather data files. After an appropriate data initialization period has elapsed, the conflict prediction data recording is turned on. If the conflict probe is set for a typical 20-min look-ahead time horizon, then at least 20 min of track data are needed beyond the duration of conflict probe testing to evaluate the conflicts predicted over the last 20 min (Fig. 5); an appropriate data recording tolerance interval is added to accommodate errors in starting and stopping times for the conflict prediction data recording. The weather data files are updated at the appropriate times while the conflict probe is running.

Predicted Conflicts

A conflict prediction data file corresponding to the specified separation criteria and the input set of track, flight plan, and weather data is obtained from the conflict probe (some postprocessing may be necessary, depending on the range of user-selectable conflict window parameters available in the conflict probe). Typically, the output file contains a list identifying all conflicts predicted by the probe, along with the predicted values of various conflict parameters. The predicted conflicts set $\{PC\}$ is generated from this file.

Consider a false/correct alert evaluation for a time-to-conflict value of n minutes. A predicted conflict is admissible for this evaluation only if the conflict was ever predicted to occur with a time to conflict of n minutes ($\pm$some seconds). The $\{PC\}$ is used to determine the appropriate subset $\{PC_n\}$ for a time-to-conflict value of n minutes.

Observed Conflicts

The observed conflicts set (or truth set) is determined by analyzing the track histories of all possible aircraft pair combinations $[n\,(n-1)/2]$ in the track data set. Note that for a given aircraft pair the two track histories do not generally start and end at the same time.

Track histories for each aircraft pair are first checked to determine if a common time interval exists. If there is no common time interval, then the two aircraft were not within the test airspace, for example, a Center at the same time and, therefore, could not be in conflict in the probed airspace.

In actual air traffic control operations, there is an altitude tolerance of ±200 ft for cruise altitudes, for example, a cruising aircraft with an assigned altitude of 35,000 ft is allowed to operate at any altitude from 34,800 to 35,200 ft. For the purposes of conflict determination, an adjusted altitude is defined to reflect this tolerance. Hence, for a cruising aircraft, if the magnitude of the difference between actual (reported) altitude and assigned (current flight plan) altitude is less than or equal to 200 ft, the adjusted altitude is defined equal to the assigned altitude; otherwise, it is defined equal to the actual altitude.

Let Δs and Δh represent the horizontal and vertical (in terms of adjusted altitude) separations between two aircraft. The vertically offset pseudoconflict window dimensions are represented by S_{conf} in the horizontal dimension and $H_{\text{conf}}^{\min}$ and $H_{\text{conf}}^{\max}$ in the vertical dimension (see Fig. 3 for an example). A point-by-point check of the corresponding conflict criterion, $[\Delta s < S_{\text{conf}}$ AND $H_{\text{conf}}^{\min} < \Delta h < H_{\text{conf}}^{\max}]$, is conducted over the common time interval using a small time step ($\sim$1 s). Conflicting aircraft pairs are determined in this manner.

It is appropriate to exclude certain conflicts observed in the track data set. One must first exclude conflicts that no conflict probe could ever record, even in theory; these are conflicts that ended before the conflict probe data recording began, conflicts that began more than 20 min after the conflict probe data recording ended (Fig. 5), and conflicts involving at least one aircraft that entered the test airspace (either by taking off within the test airspace or by flying in from adjoining airspace) after the conflict probe data recording ended. Conflicts in local regions of airspace that are not being probed are excluded. Conflicts of extremely short duration (less than 1 s) are also excluded. The set of conflicts that remain after excluding these categories is the observed conflicts set {OC}. It contains pseudoconflicts (for the vertically offset pseudoconflict window) that were observed in the track data recording and could have been detected by the conflict probe, although not necessarily 20 min in advance.

Consider a missed/correct alert evaluation for a time-to-conflict value of n minutes. An observed conflict with a given time at first loss of separation (TFLS) is admissible for this evaluation only if the conflict probe was in operation and both aircraft were in the test airspace (track data available) n minutes before TFLS, that is, at time (TFLS $- n$). Furthermore, if a conflict probe is typically set to probe above some specified altitude $h_{\min}$ ($\sim$18,000 ft), then both aircraft involved in the conflict must be above altitude $h_{\min}$ at time (TFLS $- n$) for that conflict to be admissible. The {OC} is used to determine the appropriate subset $\{\text{OC}_n\}$ for a time-to-conflict value of n minutes.

Classification of Missed/Correct/False Alerts

A comparison of the corresponding observed and predicted conflicts sets (represented conceptually in Fig. 2) yields the corresponding missed, correct, and false alerts sets. It is of interest to classify these alerts as functions of time to conflict. Figure 6 shows the classification procedure for time-to-conflict values of 5, 10, 15, and 20 min. The origin of the time-to-conflict line is the conflict start time. The sequence of dots represents conflict prediction data points over time; in general, the absence of data points corresponds to missed alerts and the unwarranted presence of data points corresponds to false alerts.

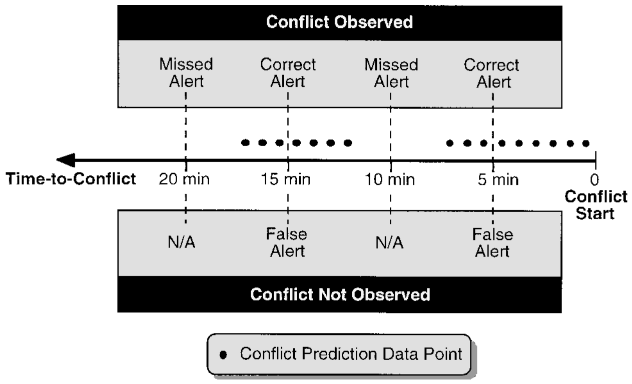


Fig. 6 Classification of alerts.

Consider a conflict that was predicted by the probe over the time intervals shown in Fig. 6 and was actually observed (see upper box in Fig. 6) some time later. The conflict was being predicted at 5 (and 15) min before the observed conflict start time; hence, it is classified as a correct alert for a look-ahead time of 5 (and 15) min. The conflict was not predicted at 10 (and 20) min before the observed conflict start time; hence, it is classified as a missed alert for a look-ahead time of 10 (and 20) min.

Now consider a conflict that was predicted by the probe over the time intervals shown in Fig. 6, but was never observed (see lower box in Fig. 6). The conflict was being predicted with a time to conflict of 5 (and 15) min; hence, it is classified as a false alert for a look-ahead time of 5 (and 15) min.

Flight Intent Information

The term flight intent refers to future changes in the velocity vector. Good knowledge of flight intent yields good trajectory prediction, which generally results in good conflict prediction. Imperfect knowledge of flight intent will generally have an adverse effect on conflict probe performance.

Data typically available for conflict probe performance analysis include position reports (from radar returns), altitude reports (from altitude transponders), and flight plans (from the host computer system). Based on these data, it is possible to determine the quality of flight intent information in the lateral and vertical dimensions. However, it is not possible to determine the quality of flight intent information in the longitudinal (speed) dimension because airspeed data are not typically available. Therefore, the term flight intent information used in this work refers only to the lateral and vertical components.

Good flight intent information is considered to be available for an aircraft if it is following its flight plan over the time interval of interest. For the purposes of this work, an aircraft is not following its flight plan if it is outside certain bounds relative to the nominal lateral or vertical components of that flight plan. The lateral flight plan component is outside bounds if the average (over all position reports within the time interval of interest) lateral deviation from the planned lateral route exceeds 4 n mile (the standard lateral tolerance for jet routes). The vertical flight plan component is outside bounds if the following condition holds for more than 10% of the altitude reports over the time interval of interest: the difference between the reported altitude and assigned altitude exceeds 200 ft (the standard vertical tolerance for assigned altitude), unless the aircraft is below its assigned altitude and climbing, or is above its assigned altitude and descending. An aircraft is considered to be climbing if its altitude rate is greater than 300 ft/min and descending if its altitude rate is less than −300 ft/min.

Nominal flight intent information is available to a conflict probe through the current flight plan obtained from the Center's host computer system. Changes in flight intent information are known to the relevant pilots and air traffic controllers and in most cases are known to the conflict probe through a filed flight plan amendment obtained from the Host. However, changes in flight intent information are not explicitly known to the conflict probe in some cases where flight plan amendments are not entered into the Host (e.g., direct routings or parallel offset routings that begin and end within the same sector); in such cases the aircraft is not following its current flight plan (as found in the Host).

Therefore, a distinction is made between a known change in flight intent after which the aircraft follows an amended flight plan (available to the conflict probe as a flight plan amendment) and an unknown change in flight intent after which the aircraft deviates from its current flight plan. A known flight intent change is a significant and well-defined event, with an associated time that is clearly identifiable from a flight plan amendment entered into the Host. On the other hand, an unknown flight intent change may be a fuzzy event, the definition of which depends on the nature of the tolerances specified for the current flight plan.

A conflict probe uses the current flight plan as the basis for flight intent information. If an aircraft's actual route differs from its flight

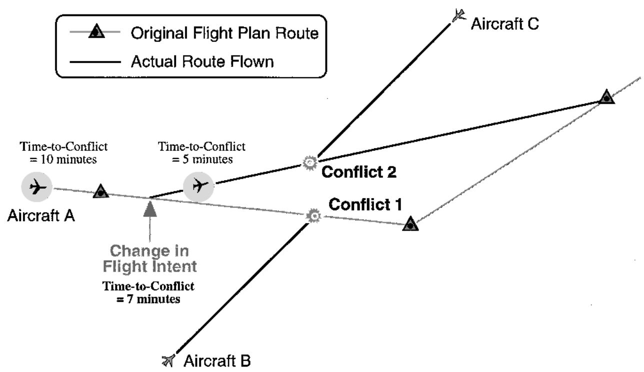


Fig. 7 Flight intent information.

plan route, as shown in Fig. 7, then the classification of missed/false alerts may be affected. Although the following discussion refers to some specific features/numbers shown in Fig. 7, its applicability is general in nature. Figure 7 shows a highlighted aircraft (aircraft A) that experiences a change in flight intent. Aircraft A is involved in conflicts with aircraft B (conflict 1) and aircraft C (conflict 2); for the sake of convenience, it is assumed that both conflicts 1 and 2 have the same conflict start time. The conflicts shown may be pseudoconflicts or operational conflicts and depend on the situation.

Consider conflict 1, predicted by the probe with a time-to-conflict value of 10 min. There is a change in flight intent 3 min later, and aircraft A deviates from its original flight plan route. The probe could not possibly have anticipated this change in flight intent. When the track data are examined in postprocessing, conflict 1 will not be observed. It is possible that the prediction of conflict 1 was a false alert. However, it is also possible that conflict 1 was in fact an operational conflict correctly predicted by the probe and that the controller vectored aircraft A in response to this potential conflict. Yet another possibility is that the controller vectored aircraft A for reasons totally unrelated to the flight path of aircraft B, for example, to give aircraft A a direct routing that cuts a corner off its flight plan route. Although all three scenarios are possible, there is no definitive evidence of what would have happened had there not been a change in flight intent for aircraft A.

After aircraft A changes course, it encounters aircraft C in a pseudoconflict, which is observed in the track data. Clearly, conflict 2 could not have been predicted by the probe before the change in flight intent of aircraft A, for example, for a time-to-conflict value of 10 min, because it is virtually impossible for any conflict probe to anticipate a future change in flight intent. It is certainly possible for the probe to predict conflict 2 after the change in flight intent occurred, for example, for a time-to-conflict value of 5 min, if the change was explicitly known to the probe as a flight plan amendment. Even if the change in flight intent is not explicitly known to the probe, it may be possible for the probe to detect conflict 2 in some cases; a conflict probe typically uses some deviation criteria to determine that an unknown change in flight intent has occurred and then uses some heuristics that attempt to compensate for this change in flight intent.

Changes in flight intent (both known and unknown to a conflict probe) are real-world features routinely encountered in current air traffic operations. On the other hand, one can envision a time in the future when all changes in flight intent are entered into the Host as soon as they happen, or in some cases even before they are scheduled to happen. In an attempt to assess conflict probe performance at both extremes of this spectrum of flight intent knowledge, missed and false alerts are partitioned into two categories: overall and filtered.

Overall and Filtered Categories

The overall category attempts to capture all valid missed/false alerts, which reflects the level of flight intent knowledge available in current air traffic operations. Hence, only missed/false alerts associated with an explicitly known flight intent change, that is, a flight plan amendment sent to the Host, are excluded from the overall category, for those look-ahead time values corresponding to a time before that flight intent change. When this rule is applied to the example of Fig. 7, if conflict 1 was predicted with a time to conflict of 10 min, it would not be counted as a false alert for a look-ahead time of 10 min if a flight plan amendment had been entered into the Host at any time within the preceding 10 min; otherwise, it would be counted. Similarly, even if conflict 2 was not predicted 10 min in advance, it would not be counted as a missed alert for a look-ahead time of 10 min if a flight plan amendment had been entered into the host at any time within the preceding 10 min; otherwise, it would be counted.

The filtered category is a subset of the overall category. It attempts to capture only those missed/false alerts associated with good flight intent information, which reflects a higher level of flight intent knowledge that may be available in future air traffic operations. An overall missed/false alert is excluded from the filtered category if one of the aircraft involved in the conflict had bad flight intent information over the time interval of interest. In the example of Fig. 7, the false and missed alerts described in the preceding paragraph would not be counted if aircraft A was not following its flight plan (as defined in the preceding subsection) during the preceding 10 min; otherwise, they would be counted.

Note that even the filtered category will include missed/false alerts with bad flight intent information in the longitudinal (speed) dimension because it is not possible to identify them using the data typically available for conflict probe performance evaluation. Flight intent information on airspeed is especially important for predicting the trajectories of climbing/descending aircraft.

Definition of Alert Rates

Consider a reliability analysis for a time-to-conflict value of n minutes. Overall and filtered missed and false alerts are determined as functions of time to conflict, following the procedures described earlier. Let M, F, and C denote the number of overall missed alerts, false alerts, and correct alerts, respectively, as shown in Fig. 2.

Figure 8 shows overall missed/false alerts partitioned into their corresponding subsets, based on the quality of flight intent information. The overall missed alert rate R_{MA} is defined as the ratio of the number of overall missed alerts to the total number of admissible observed conflicts. The overall false alert rate R_{FA} is defined as the ratio of the number of false alerts to the total number of admissible predicted conflicts:

$$R_{\mathrm{MA}} = \frac{M}{C + M} = \frac{(m_{\mathrm{GFI}} + m_{\mathrm{BFI}})}{C + (m_{\mathrm{GFI}} + m_{\mathrm{BFI}})} \tag{1}$$

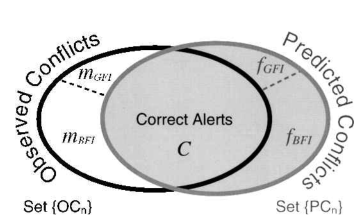
$$R_{\mathrm{FA}} = \frac{F}{C + F} = \frac{(f_{\mathrm{GFI}} + f_{\mathrm{BFI}})}{C + (f_{\mathrm{GFI}} + f_{\mathrm{BFI}})} \tag{2}$$

In Eqs. (1) and (2), m_{GFI} and f_{GFI} are the number of missed alerts and false alerts, respectively, with good flight intent information; similarly, m_{BFI} and f_{BFI} are the number of missed alerts and false alerts, respectively, with bad flight intent information. When it is noted that the filtered category is associated with good flight intent information, expressions for the filtered missed alert rate $\tilde{R}_{\mathrm{MA}}$ and the filtered false alert rate $\tilde{R}_{\mathrm{FA}}$ are obtained by setting $m_{\mathrm{BFI}} = f_{\mathrm{BFI}} = 0$ in Eqs. (1) and (2), which results in the following expressions:

$$\tilde{R}_{\mathrm{MA}} = m_{\mathrm{GFI}}/(C + m_{\mathrm{GFI}}) \tag{3}$$

$$\tilde{R}_{\mathrm{FA}} = f_{\mathrm{GFI}}/(C + f_{\mathrm{GFI}}) \tag{4}$$

Fig. 8 Partitioning of missed and false alerts.



When Eqs. (1–4) are used, overall and filtered missed and false alert rates can be computed at various values of time to conflict, for example, 5, 10, 15, and 20 min.

Conflict Parameter Errors

Examples of conflict parameters include conflict start time, minimum horizontal separation, minimum vertical separation, and position at first loss of separation.

Conflict parameter errors are computed for all correct alerts, that is, predicted conflicts that were actually observed. The errors in conflict parameters are the differences between predicted and observed values; they are computed at various times, for example, 5, 10, 15, and 20 min before the time when first loss of separation was observed.

Predicted values of the conflict parameters are available from the conflict prediction data; these values may vary with time, corresponding to conflict prediction updates. Actual (truth) values of the conflict parameters are computed from point-by-point examinations of the track (position and altitude) data.

The TFLS is the time when the specified horizontal and vertical separation criteria are first violated; the error is the difference between the predicted and actual values. When it is noted that the predicted value of TFLS varies over time, the error ΔTFLS, where $n = 5$, 10, 15, and 20 min, is given by

$$\Delta \mathrm{TFLS} = \{[\mathrm{TFLS}_{\mathrm{pred}}(t = \mathrm{TFLS}_{\mathrm{actual}} - n)] - \mathrm{TFLS}_{\mathrm{actual}}\} \qquad (5)$$

The minimum horizontal separation (MHS) is defined as the lowest value of the horizontal separation over the entire duration of the conflict. The errors are the differences between the predicted and actual values. When it is noted that the predicted value of MHS varies over time, the error ΔMHS, where $n = 5$, 10, 15, and 20 min, is given by

$$\Delta \mathrm{MHS} = \{[\mathrm{MHS}_{\mathrm{pred}}(t = \mathrm{TFLS}_{\mathrm{actual}} - n)] - \mathrm{MHS}_{\mathrm{actual}}\} \qquad (6)$$

Errors in other conflict parameters such as minimum vertical separation and position at first loss of separation can be computed in a similar fashion. A statistical analysis (mean, standard deviation, etc.) of the error data can then be conducted.

Demonstration of Methodology

The methodology presented is demonstrated by applying it to a performance evaluation of the CTAS conflict probe tool, using real traffic data. Algorithms were developed to implement the evaluation methodology presented and software code tailored to the CTAS conflict probe tool was then written based on these algorithms.

The preliminary results presented later in this section illustrate the type of performance data that can be generated by an implementation of this methodology, but they do not represent the actual performance of the CTAS conflict probe tool under operational conditions. To determine the actual performance, additional parameters associated with actual operation of the probe must be incorporated into the evaluation. These parameters include the effects of controller-selected separation buffers, temporary altitude restrictions, conflict display stability filters, and conflict display color coding based on conflict probability. Whereas the inclusion of these factors contributes to the complexity of the evaluation process, this is necessary to gain a complete and accurate picture of conflict probe performance. When these factors are fully incorporated into the evaluation, the actual performance as measured by missed and false alert rates may differ substantially from the examples presented later in this section.

Data Collection

Five sets of track and flight plan data from the Denver Center were used for this numerical study. These data were recorded on five different days, spanning the local time period from 0830 to 1830. The combined recordings contain approximately 15 h of data, representing nearly 5000 aircraft. All relevant 3-h weather forecast data files were obtained from the NMC.

Conflict prediction data files were obtained by running the CTAS conflict probe tool (version 5.3.0 released in May 1998). Because conflict probes are not designed for operation with vertically offset separation criteria, conflict prediction data files were first recorded with vertically expanded separation criteria corresponding to a vertical distance of 5000 ft (2500 ft if either aircraft is below FL 290) and a horizontal distance of 5 n mile. This data set was reduced in postprocessing to create conflict prediction data files for vertically offset separation criteria corresponding to a vertical distance from 3000 to 5000 ft (1500 to 2500 ft if either aircraft is below FL 290). A minimum conflict search altitude $h_{\min}$ of 18,000 ft was used.

Conflict prediction data recording in CTAS was upstream of a conflict display stability filter; therefore, the output files were postprocessed to simulate partially the effect of this filter. The threshold value of the conflict probability filter[10] (which implements a tradeoff between missed and false alerts) was set to a nominal value of 50% in postprocessing. The value of the conflict probability time filter[10] (a filter that disables the conflict probability filter for low values of time to conflict) was set at 7 min. Finally, it is noted that the CTAS conflict probe tool did not probe arrival aircraft for conflicts that began inside the TRACON of the primary airport (Denver in this case); the observed conflicts set was adjusted accordingly.

Example Results

Some results are presented here as examples; they do not represent the actual performance of the CTAS conflict probe tool under operational conditions. Missed and false alert rates, computed for overall and filtered categories at time-to-conflict values of 5, 10, 15, and 20 min, are presented in Fig. 9. It was determined that the mean conflict warning time for correct alerts was 12 min. Conflict prediction errors for minimum horizontal separation and conflict start time, computed for time-to-conflict values of 5, 10, 15, and 20 min, are presented in Fig. 10. From Figs. 9 and 10 it can be seen that conflict probe performance is strongly influenced by the value of time to conflict. It is evident that performance improves as the value of time to conflict decreases, as expected.

Discussion on Conflict Probe Performance

Any conflict probe will have a less than perfect performance when evaluated in a real-world environment. This can be attributed to flight

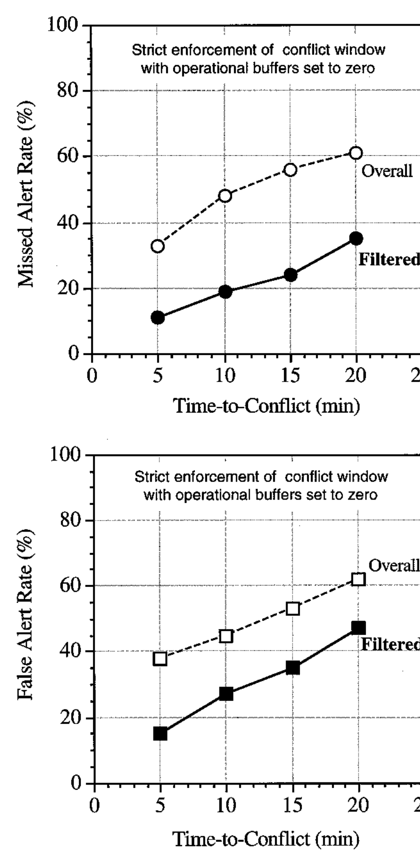


Fig. 9 Example of missed and false alert rates.

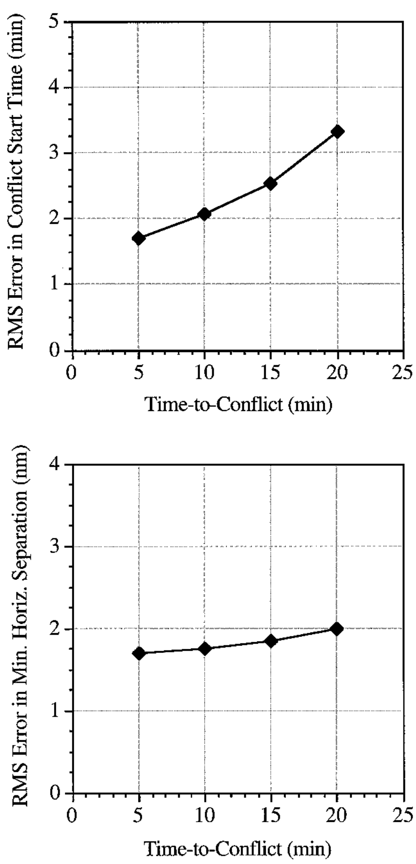


Fig. 10 Example of errors in conflict prediction parameters.

intent errors (lateral, vertical, and speed), wind model errors, aircraft (thrust, drag, weight) model errors, speed and heading errors due to radar tracker noise, and navigation errors. All of these contribute to the overall missed and false alert rates shown in Fig. 9. The errors in conflict prediction parameters for correct alerts, shown in Fig. 10, arise in part from bad flight intent information. Note that correct alerts are not partitioned by quality of flight intent information; a conflict probe may be able to predict conflicts correctly even with bad flight intent information, by the use of heuristic rules.

Although the lateral intent errors are relatively small for missed and false alerts in the filtered category, all of the other error sources identified are present (including speed intent errors); the filtered missed and false alert rates shown in Fig. 9 arise from a combination of all of these errors. Another factor, contributing to both the overall and filtered missed alert rates shown in Fig. 9, is that data on temporary altitude messages were not available in the flight plan data files used for this exercise; this means that some information on flight intent in the vertical dimension was unavailable to the conflict probe.

There are some additional factors that contribute significantly to the missed alert rates (both overall and filtered) shown in Fig. 9. The most important factor is that these rates were determined from a strict enforcement of the conflict window parameters. Consider, for example, an observed conflict with a minimum horizontal separation of 4.9 n mile. Suppose the conflict probe were to predict a minimum horizontal separation of 5.1 n mile due to trajectory prediction errors; in this case the probe would not declare a conflict (resulting in a missed alert) if the horizontal separation parameter was set to exactly 5 n mile. This would not have occurred if a small buffer had been added to the horizontal separation parameter of the conflict window. The missed alert rates shown in Fig. 9 are, therefore, indicative of worst-case performance. During operational field tests of the CTAS conflict probe tool at the Denver Center,[2] it was observed that air traffic controllers always use a horizontal separation buffer, typically setting the horizontal separation parameter that triggers conflict display to values between 8 and 10 n mile; vertical separation buffers were sometimes used for climbing/descending aircraft. A conflict probe evaluation that includes the effects of these controller-selected operational buffers would result in significantly lower rates of missed alerts, albeit at the cost of higher rates of false alerts relative to the operational 5-n mile separation standard.

The CTAS conflict probe tool uses a color-coding scheme to indicate the level of conflict probability.[10] The false alert rates shown in Fig. 9 are for all conflicts that would have been displayed to the air traffic controller. False alert rates associated with higher probability conflicts would be significantly lower.

Note that, although the example results presented simply used a conflict probability threshold value of 50%, changing this parameter could reduce the missed alert rate at the expense of the false alert rate, or vice versa. This tradeoff factor between missed and false alerts represents an important design parameter for a conflict probe, and fine-tuning of this parameter (and other parameters discussed earlier) could yield significant improvements in overall performance. For example, at small look-ahead times, a low missed alert rate is desired, even it is achieved at the expense of a higher false alert rate. On the other hand, at large look-ahead times, a low false alert rate may be desirable, even at the expense of a higher missed alert rate.

Achieving a large reduction in both missed and false alert rates would require a substantial improvement in the quality of information input into the probe. It is believed that the primary sources of conflict probe inaccuracies are real-world errors in flight intent and wind forecasts. A secondary source of inaccuracy (that primarily affects predictions of conflicts involving climbing/descending aircraft) is the presence of errors in the aircraft (thrust, drag, weight) models.

Conclusions

A comprehensive methodology has been developed to quantitatively evaluate the performance of a conflict probe. This methodology is universal in nature and can be applied to any conflict probe, thereby providing a framework for a comparative study of conflict probes. Real traffic data is utilized to preserve all real-world errors that degrade conflict probe performance. Because operational conflicts are generally not found in real traffic data, expanded separation criteria are utilized. A key result of this work is the determination that the characteristic properties of a vertically offset pseudoconflicts set generally match those of a corresponding operational conflicts set; therefore, the performance of a conflict probe using a vertically offset pseudoconflict window is indicative of its performance using an operational conflict window. A procedure is presented for partitioning conflict probe performance by quality of flight intent information.

A demonstration of the evaluation methodology was conducted by applying it to the CTAS conflict probe tool, using real traffic data from the Denver Center; example results from this numerical study are presented. Conflict probe performance improves as the time to conflict decreases, as expected. It was observed that lack of good quality flight intent information can significantly degrade conflict probe reliability. Although a tradeoff can be made between the rate of missed and false alerts, a significant improvement in the overall performance of a conflict probe would require a corresponding improvement in the quality of input data such as flight intent information and wind forecasts.

Acknowledgments

The software code for conflict probe performance evaluation was written by Hilda Lee. The contributions of Anupa Bajwa, Shon Grabbe, and Kapil Sheth are acknowledged. The author would like to thank Ralph Bach, Gano Chatterji, Heinz Erzberger, Michelle Eshow, Steven Green, Douglas Isaacson, B. David McNally, Russell Paielli, and Banavar Sridhar for valuable discussions over the course of this work.

References

[1]"Final Report of RTCA Task Force 3: Free Flight Implementation," RTCA, Washington, DC, Oct. 1995.

[2]McNally, B. D., Bach, R., and Chan, W., "Field Test of the Center-TRACON Automation System Conflict Prediction and Trial Planning Tool," AIAA Paper 98-4480, Aug. 1998.

[3]Brudnicki, D. J., Arthur, W. C., and Lindsay, K., "URET Scenario-Based Functional Performance Requirements Document," MITRE Center for Advanced Aviation System Development, CAASD Rept. MP97W0000044, McLean, VA, April 1998.

[4]Brudnicki, D. J., and McFarland, A. L., "User Request Evaluation Tool (URET) Conflict Probe Performance and Benefits Assessment," MITRE Center for Advanced Aviation System Development, CAASD Rept. MP98W0000112, McLean, VA, June 1997.

[5]Cale, M. L., Paglione, M., Ryan, H., Timoteo, D., and Oaks, R., "User Request Evaluation Tool (URET) Conflict Prediction Accuracy Report," Dept. of Transportation/Federal Aviation Administration, Rept. DOT/FAA/CT-TN98/8, April 1998.

[6]Paglione, M., Ryan, H. F., Kazunas, S., and Cale, M. L., "Generic Metrics for Conflict Probe Tools Developed for Free Flight," *Proceedings of the 42nd Annual Air Traffic Control Association Conference*, Air Traffic Control Association, Arlington, VA, 1997, pp. 77–81.

[7]Erzberger, H., Davis, T. J., and Green, S., "Design of the Center-TRACON Automation System," AGARD Guidance and Control Symposium on Machine Intelligence in Air Traffic Management, Berlin, May 1993, pp. 11-1–11-12.

[8]Denery, D. G., and Erzberger, H., "The Center-TRACON Automation System: Simulation and Field Testing," NASA TM 110366, Aug. 1995.

[9]Isaacson, D. R., and Erzberger, H., "Design of a Conflict Detection Algorithm for the Center-TRACON Automation System," *Proceedings of the 16th AIAA/IEEE Digital Avionics Systems Conference*, Irvine, CA, Oct. 1997.

[10]Erzberger, H., Paielli, R. A., Isaacson, D. R., and Eshow, M. M., "Conflict Detection and Resolution in the Presence of Prediction Error," *Proceedings of the 1st USA/Europe Air Traffic Management R&D Seminar*, Saclay, France, June 1997; URL: http://atm-seminar-97.eurocontrol.fr/erzberge.htm [cited 11 April 2001].

[11]Benjamin, S. G., Brewster, K. A., Brummer, R., Jewett, B. F., Schlatter, T. W., Smith, T. L., and Stamus, P. A., "An Isentropic Three-Hourly Data Assimilation System Using ACARS Aircraft Observations," *Monthly Weather Review*, Vol. 119, No. 4, 1991, pp. 888–906.