Assessing the Capacity of Novel ATM Systems

A J Hudgell (QinetiQ) & R M Gingell (CGP)
alison.hudgell@atc.QinetiQ.com *rodg@cgp.co.uk*

Abstract

This paper describes a methodology for assessing the capacity of future Air Traffic Management systems that may have a radically different concept of operation from today's. The methodology was developed by the EUROCONTROL CARE–INTEGRA *Metrics and Methodologies* Project. Designed to be generic with regard to operational concept, it involves the objective counting of observable events in a simulation of the system. It assesses sustainable throughput for a given maximum acceptable delay. It also addresses capacity at a whole-system level, providing the framework to ensure that the capacity of a sector is assessed within the appropriate network environment. Its major innovative elements are a new surrogate measure for capacity – information processing load (IPL) – and the use of defined traffic samples to ensure that capacity is assessed within the appropriate context. The detailed assumptions of the IPL model for particular simulation scenarios are parameterised, facilitating comparison of best and worse cases and of studies carried out by different simulation teams. An initial exercise to apply the methodology to real simulation data has given promising results, but further validation is required.

1. Introduction

The objective of the EUROCONTROL CARE–INTEGRA *Metrics and Methodologies* Project is to develop methodologies for assessing capacity, safety, efficiency and environmental impact, that can be applied to any novel Air Traffic Management (ATM) system. This paper describes the capacity assessment methodology developed by that project. The aim is to develop a methodology that is generic with regard to operational concept and can be used to assess the capacity of future ATM systems by simulation. The project views the capacity assessment of current ATM systems, based on controller workload, as a special case. From there, we developed a generic model of ATM system capacity, which aims to be applicable to (at least) all currently proposed future operational concepts.

In considering the needs that a capacity assessment methodology for novel ATM systems must fulfil, the authors drew on their experience of shortcomings that have become evident when current methods have been employed for assessing future systems' capacity in studies such as PHARE. In particular, we focussed on the need for an objective measure on which to base assessment of future systems' capacity. Many current, workload-based methods rely on controller experience (e.g. to provide a precise value for how much work something will cause), but for future systems this is not available. The proposed new methodology parameterises the assumptions upon which capacity assessment is based, so that, for example, likely best and worse cases can easily be compared. This parameterisation should also facilitate comparison of studies carried out by different teams using different simulation tools. The close coupling of many current capacity assessment methods with the simulation system used, makes this virtually impossible today.

A second and related point is the need to decouple the results of the capacity assessment from the performance of the system model that is being used for the simulation. The proposed new methodology looks at the loadings of different system components in isolation from one another, so that the critical components can be objectively identified.

Thirdly, we address the need to look at capacity at a whole-system level. This is something that is very poorly addressed by today's methods. The introduction of advanced tools (e.g. arrival/departure managers, 4D trajectory predictors, etc.) and more air/ground co-operation in ATC decision making is likely to lead to a higher degree of coupling between different control areas: TMAs and en-route sectors. This results in the need, potentially, to consider a large part of the airspace around the region of operation when assessing capacity; to consider how each sector fits into the overall airspace and Air Traffic Management environment.

2. Capacity in novel ATM systems

2.1 What is capacity?

The capacity of a part of the ATM system is traditionally measured in terms of traffic throughput – the number of flights that can pass through it, being safely and efficiently handled (of which more later), in a given time period. This notion is simple enough at first sight, but as we look deeper, we uncover a number of important aspects that must be understood if we are to assess capacity successfully.

Firstly, it is important to recognize that the achievable throughput depends on the distribution of the flights over the airspace – that is, on the pattern of traffic that the ATM system element has to deal with. For example, the achievable throughput of an airspace sector might be very high if flights only ever went through from north to south, flying in formation. But this bears no relation to the throughput that is achievable with typical sector traffic – travelling along a variety of intersecting paths, at different speeds, with some flights climbing and descending through the levels of others. So in order to assess the capacity of a part of the ATM system, we first have to define the traffic pattern for which we are providing capacity.

The fundamental reason for providing capacity is satisfaction of demand. We therefore propose that ATM system capacity should be assessed for the traffic pattern arising from ***expected demand***.

The second important aspect of ATM capacity is its stochastic nature. Even for a given traffic pattern, the flights will not be exactly the same every day; real operations lead to innumerable, random variations. Each flight will start at a slightly different time, will fly slightly differently (because it is late, early, heavy, light, etc.). Some flights will be cancelled and there will be some "unscheduled" flights. Furthermore there are variations in the pattern of traffic during the day. In order to assess capacity we have to look at an average over time and traffic sample; it is not sufficient simply to note one particular throughput that was achieved in one particular hour of operation. Capacity must be considered as the ***long-term average sustainable throughput***. At times, the system will be able to manage more traffic than this, but in other periods it will be able to manage less, depending on the particular traffic that is present.

Thirdly, queuing theory tells us that there is an inescapable link between the *sustainable throughput* of a system and the maximum acceptable waiting time. This is true for any system (hospital beds, comms networks, ATC systems, …) and again arises from the stochastic nature of the traffic. Considering traffic arriving at an ATM sector, inter-arrival times are not uniform and hence instantaneous demand varies. If we allow no delay in "being served", the system must at all times be able to cope with the maximum instantaneous demand that can be envisaged, otherwise it will sometimes be overloaded. This means that most of the time the system will be operating at significantly less than its potentially available capacity. Potential capacity is being wasted, which makes the service provided inefficient.

The existence of a "queue" allows the peaks in input demand to be smoothed out. Flow Management provides the queuing mechanism for en-route airspace. As a result a system that would have been overloaded by the short-term demand peaks, can now cope – the traffic arriving in peaks waits in the queue and is served in turn. Because the input stream is smoothed, less of the potentially available capacity is wasted. Given the requirement to provide an efficient ATM service and not a wasteful one, some kind of queuing mechanism is essential, and we can assume that Flow Management is a fundamental part of the ATM system. Hence, in defining the traffic pattern used to assess *sustainable throughput*, the INTEGRA methodology applies flow management-type smoothing to the raw traffic demand.

The longer waiting time we allow in the queue, the more the input stream is smoothed, and hence the greater is the *sustainable throughput* of a given system. In practice there must be a limit on the maximum acceptable waiting time, and this will in turn limit the sustainable throughput. The maximum acceptable delay is a parameter in the INTEGRA methodology.

So, we consider ATM system capacity to be the long-term average *sustainable throughput*, for the traffic pattern arising from demand, with Flow Management smoothing applied up to a maximum acceptable delay. These aspects are fundamental to ATM system capacity – they apply to today's system just as much as to future systems.

2.2 What limits capacity?

In order to assess ATM capacity for any specific system, it is necessary to know, for that system, what limits the *sustainable throughput*. Possibilities include:

- Physical limits:
  - physical space for flying and manoeuvring the given traffic safely (e.g. wake vortex separation minima limit capacity within an approach traffic stream);

- Workload limits:
  - controller or pilot workload,
  - capacity of controller or pilot to maintain situational awareness,
  - capacity of comms networks (with assumptions about necessary slack to provide required performance, speed and reliability),
  - computer system workload (given available memory, processor speed, etc.).

Under operational concepts different from today's, different factors will influence the *sustainable throughput* of the ATM system. The tactical controller is no longer necessarily the system component that limits throughput. We cannot tell in advance which factors will be critical. Hence a generic method needs to address all of them, and to assess which factors are limiting *sustainable throughput*.

In fact this is not the end of the story, for on top of *sustainable throughput* there may be additional, indirect limits on capacity. The most important of these is safety. In the same way that there is a trade-off between queuing delay and *sustainable throughput*, there is a dependency between *sustainable throughput* and the available margin for error – the amount of "slack" in the system. Considering examples from the current ATM system: controllers must always be able to find some "spare" effort to cope with exceptions such as loss of radio contact, and the airspace structure must have sufficient space to accommodate a flight that loses pressurisation and must descend rapidly.

In designing and implementing a real ATM system, safety must have the highest priority. We define the safety level that we are prepared to accept, and capacity follows from that. Although we *could* pack more flights in, this would result in more accidents. So we limit capacity to ensure safety. Hence the physical limits noted above are in fact set by safety considerations; they define the required balance between safety and capacity.

Other factors that we must balance with capacity include:

- the required efficiency of the system – we have already discussed one major aspect of this: maximum acceptable delay;
- environmental impact – for example, noise quotas limit approach capacity at many major airports.

It is important to recognise that these factors are inter-dependent with capacity. They may be measured at the same time, with changes to the modelled operational concept to improve one at the expense of another[1]. Alternatively, requirements on safety, efficiency and environmental impact may be built into the simulation scenario, placing pre-defined limits on capacity. The experimenter should be aware of the distinction!

None of the above contradicts current methods of assessing capacity. Current, workload-based methods can be thought of as a special case of this generic argument.

3. Proposed new method

3.1 Information Processing Load

At the most fundamental level, what the ATM system has to do to control air traffic is *process information*. Hence, **information processing load (IPL)** provides a surrogate measure for capacity that works for any element of a system. It provides a way of assessing all possible capacity-limiting factors that are not related to physical space. This new surrogate measure is the first key element of the INTEGRA methodology.

To assess the information processing load in a system, the first step is to consider which elements of the system have to process information. We term these system elements **actors.** For example, there may be two or more **controllers**, and **computer tools** (e.g. trajectory

[1] INTEGRA is also developing metrics and methodologies for assessing safety, efficiency and environmental impact.

predictor, conformance monitor, conflict probe). In a free-flight system, the information processing would be distributed among many **pilots** and their in-cockpit separation tools, as well as the controllers and their ground-based tools. In addition, information is handled by **communications systems**, and by the controllers' and pilots' **HMIs**.

The second step is to determine how much information processing each actor has to do in each unit time, during the simulation. This step is exactly analogous with the way that existing capacity methods assess controller workload as a function of time. It produces a graph like Figure 1 (below) for each actor.

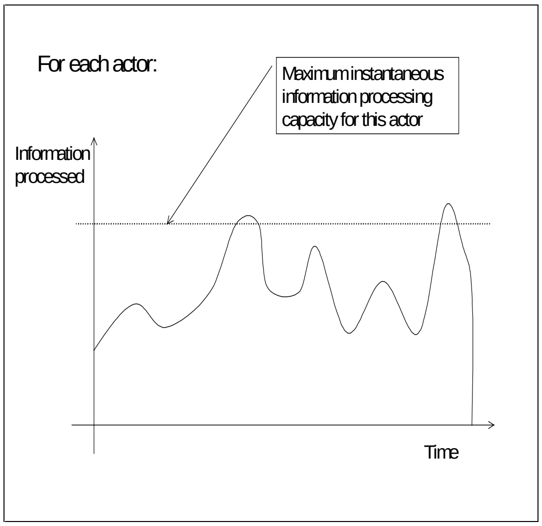

Figure 1: Graph of IPL vs. time for an actor

The third step is to use these graphs of IPL against time to assess capacity. Again, the approach is analogous to that used in current methods. There are two options:

- The graphs for a given actor from different simulation scenarios can be compared to indicate the *relative* loading of that actor in the two scenarios. This relative assessment may be all that is needed.
- The graphs can be compared with an estimate of the maximum information processing capacity of each actor, to assess which actors came near their IPL capacity at times in the simulation. This indicates which actors are critical, limiting capacity in the modelled system. (Subsequent development of the operational concept can focus on reducing the information processing load of these actors).

The paper will now look at each of these three steps in more detail, to show how the INTEGRA capacity assessment methodology can be applied in practice.

3.2 Identification of actors

In using this capacity methodology, the first task is to determine which system components need to be represented as information-processing *actors* in the simulation. As just noted, these could include:

- Human
  - Tactical
  - Planner
  - First Pilot
  - Co-pilot

- Computer assistance tools
  - Conflict probe
  - Trajectory predictor
  - Flight Path monitor
  - In-cockpit separation tools

- Communications
  - Human–system communications, i.e. the HMI
  - System data network
  - Ground–ground comms network
  - Air–ground communications
    - datalink
    - r/t (voice)

The information processed by all these actors needs to be considered, as any one may be the limiting factor. However, the precise set that is relevant depends on the operational concepts of the systems being assessed and compared.

In future systems involving human controller roles, it is very likely that one of these will be the capacity-limiting actor, because it is easier to increase the processing power of a computer assistance tool or a comms link than a human actor. However, there are instances in existing systems where other actors (e.g. the comms network) may become overloaded before the controllers. Moreover, this is more likely to be the case in simulation systems, and unless explicitly identified it may cause a misleading assessment of the capacity of the future system.

Note that automated tools may process information at different times from a controller performing the analogous tasks. For example the processing required to identify a conflict may be completed by the tool as soon as the initial trajectories are in the system, whereas a radar controller may not be aware of a conflict until much later. Therefore the peaks of loading in an automated system will be different from those in a mainly manual system.

3.3 Allocation of information processing load to actors, as a function of time

So, having identified the relevant actors, the next step is to assess how much IPL each actor has at each moment. How can we assess this? As with controller workload, it is not something that can always be counted directly. We first identify generic *causes of information processing* in an ATM system, and then (in line with the requirement for an objective measure) identify observable simulation events that can be used to mark their occurrence. We also outline, in general terms, the timing of the information processing load in relation to the observable events.

In this way, we define a model that represents information processing load in terms of observable simulation events where:

- the actors are labelled *i* where i = 1..<number of actors>;
- the causes of processing are labelled α where $\alpha = a..g$;
- the amount of processing carried out by each actor *i*, in response to each cause α, in each time increment is[2] $p_{i,\alpha}(t)$.

The amount of processing that each actor must do in response to each cause, and when, depends on the operational concept being simulated, and must therefore be defined by the experiment analyst. In order that different approaches and assumptions can easily be compared and discussed, the rule set for allocating processing from each of the sources to the appropriate actor is parameterised by the set of weighting factors $\{\lambda_{i,\alpha}\}$, a two-dimensional array of values λ.

Employing a set of parameters has the additional benefit that the algorithms can be re-run with different parameter values, using the same simulation outputs, a number of times. This allows:

- comparison of best and worst case assumptions;
- analysis of sensitivity of the capacity assessment to the parameter values used.

As an initial assumption, we can take λ=1 for any actor that must process information as a result of cause α, λ=0 otherwise. Depending on the fidelity of the simulation, and what things are changing between the simulation scenarios, the experiment analyst may wish to assign more precise values for λ. [1] gives further discussion of this. To obtain an initial indication of the potential capacity gain from introduction of a new tool, for example, detailed setting of λs is not expected to be necessary. However, inclusion of these weighting factors means that the proposed methodology is readily extensible to simulations of higher fidelity.

The following paragraphs deal in turn with each of the identified *causes of information processing*, showing how the overall IPL model is built up.

a) Flight "arrival" – not necessarily in a geographical area, but in the *list of flights that are of concern to the actor*. Processing is carried out to:

- Add the flight to the actor's "database" (i.e. mental picture in the case of a human actor);
- Co-ordinate acquisition with the pilot/aircraft (e.g. sign on).

Processing is done for each flight that "arrives". It is done at the time the flight first "appears in the list".

Hence: $p_{i,a}(t) = \lambda_{i,a}\, a(t)$

where: a(t) = number of flights "arriving in the list" in each unit time.
$\lambda_{i,a}$ = weighting factor for actor i for flight arrival.

b) Interaction detection: i.e. determining the forecast interactions of a flight with others in the simulation. Processing is carried out to compare each "new" predicted trajectory with trajectories of all other flights "on the list". A new trajectory may arise because a flight arrives, or because of a change in plan for a flight that is already "on the list".

[2] The notation p(t) indicates that *p* is a function of time *t*. In other words, we are not measuring a single value for *p*, but measuring it at each time increment, and noting how it changes with time.

Processing is done for each new or changed trajectory "multiplied by" each flight in the "list" at that time. It is done at a time dependent on the operational concept – e.g. interaction search could be done by a controller when a flight arrives in the area of interest to him, or it may be done in advance of that time by a tool.

Hence: $p_{i,b}(t) = \lambda_{i,b}(a(t) + c(t))(n(t)-1)$

where: a(t) = number of flights "arriving in the list" in each unit time.
c(t) = number of changes to predicted trajectories generated in each unit time.
n(t) = number of flights "on the list" at that time.
$\lambda_{i,b}$ = weighting factor for actor i for interaction search.

c) **Resolution planning:** i.e. planning a resolution for each forecast interaction[3]. This resolution may consist of manoeuvring an aircraft or doing nothing except careful monitoring to ensure that the forecast interaction is not "worsening".

Processing is done for each forecast interaction. Processing is carried out when (or shortly after) the actor concerned "realises" there is an interaction (e.g. when the conflict probe highlights the interaction to the controller).

Depending on the fidelity of the throughput comparisons required, it may be appropriate to weight the interactions according to their "difficulty". The simplest assumption is that all interactions require the same amount of information processing to plan a resolution. This assumption may be adequate for simulation experiments aiming to obtain an initial indication of the potential capacity benefit from the introduction of a new tool, for example a conflict probe, where the *nature* of interactions is not expected to change significantly between two simulation scenarios. However, where the nature of interactions *is* expected to change significantly (e.g. free routing vs. structured routes), additional detail may be required. The following approach is recommended initially: the processing due to resolution planning should be weighted according to the number of constraints present on the resolution. This means that:

- if the predicted interaction occurs surrounded by plenty of free space, the weighting factor, *w*, is set to 1;
- *w* is incremented for each other aircraft that is within interaction thresholds *d* (horizontally) and *h* (vertically) of either of the interacting aircraft;
- *w* may also be incremented for each airspace boundary (sector boundary, edge of restricted airspace, etc.) that is within interaction thresholds *d* and *h* of either of the interacting aircraft if such boundaries are significant to the operation of the aircraft.

Hence: $p_{i,c}(t) = \lambda_{i,c}\, wf(t)$

where: f(t) = number of forecast interactions identified in each unit time.
w = weighting factor for "difficulty" of resolution of each interaction, as defined above.
$\lambda_{i,c}$ = weighting factor for actor i for resolution planning.

d) **Resolution implementation:** i.e. implementing the planned resolution manoeuvres. It is assumed that a similar amount of work is required to implement any resolution.

Processing is done for each resolution manoeuvre. It is done at a time dependent on the control paradigm. For example, in today's system it is done shortly before the manoeuvre is carried out, while in a future system an instruction to manoeuvre may be sent in advance by air-ground datalink.

[3] **Definition of an interaction.** Predicted interactions may be identified within the simulation or by post-process comparison of the predicted trajectories recorded during the simulation. If two aircraft are predicted to come within threshold distance *d* and altitude *h* of one another, that is an interaction. The values of d and h are concept-dependent and should be set by the experiment analyst. For example, in today's system under a radar separation minimum of 5NM, a controller may consider a pair of aircraft to be interacting if their predicted trajectories (as predicted in the controller's head) come within, say, d = 10NM of one another. In a future system with a more accurate trajectory predictor, that threshold may be reduce to, say, d = 7NM. In addition, these threshold values may vary depending on the navigation capability of the aircraft involved.

Hence: $p_{i,d}(t) = \lambda_{i,d}\, r(t)$

where: $r(t)$ = number of resolution manoeuvres implemented in each unit time.
$\lambda_{i,d}$ = weighting factor for actor i for resolution implementation.

e) Monitoring conformance to plan. Monitoring is done for every flight that is of concern to the actor. Hence, processing is carried out for each flight in the “list”. However, additional monitoring may be done to ensure that a forecast interaction, for which no resolution manoeuvre has been planned, is not “worsening”. In order to allow for this, flights that are involved in an interaction for which no resolution manoeuvre is planned are weighted by parameter *v*. The value of that parameter is concept-dependent. For example, in today’s system, the controller may do twice as much monitoring for flights that are interacting but not conflicting; however a conformance monitoring tool may treat all flights equally.

Hence: $p_{i,e}(t) = \lambda_{i,e}\, vn(t)$

where: $n(t)$ = number of flights “on the list” at time t.
v = weighting factor for each flight as discussed above.
$\lambda_{i,e}$ = weighting factor for actor i for monitoring.

Monitoring processing can be assumed to be iterated, with each iteration a parameter time, *m*, apart. Hence, vn(t) should be sampled at time intervals of *m* seconds. Parameter *m* is concept-dependent and should be set by the experiment analyst. The resulting processing, $p_{i,e}$, can be taken to have occurred in that unit time, or spread evenly over the interval to the next sample (i.e. over *m* seconds), whichever is deemed to be more representative of the simulated system.

f) Other trajectory changes. For example instructions to allow a flight to comply with a planned exit level.

Processing is done for each such change made. As for resolution manoeuvre implementation, processing is done at a time dependent on the control paradigm. For example, in today’s system it is done shortly before the manoeuvre is carried out, while in a future system an instruction to manoeuvre may be sent in advance by air-ground datalink.

Hence: $p_{i,f}(t) = \lambda_{i,f}\, o(t)$

where: $o(t)$ = number of other trajectory changes required in each unit time.
$\lambda_{i,f}$ = weighting factor for actor i for other trajectory changes.

g) Co-ordination with other control agencies (e.g. neighbouring sectors or centres, or between “free-flight” pilot and overseeing controller).

Processing is done for each deviation from plan (where plans would include e.g. standing agreements). It is done at the time that the deviating manoeuvre is planned.

Hence: $p_{i,g}(t) = \lambda_{i,g}\, c(t)$

where: $c(t)$ = number of changes to trajectories generated in each unit time (same factor as in (b) above).
$\lambda_{i,g}$ = weighting factor for actor i for co-ordination.

Summing over causes: For each actor, the total information processing carried out in each unit time[4] is the sum of the above contributions (remembering that some of them may be zero):

$$P_i(t) = \Sigma_{\alpha=a}^{g}\, p_{i,\alpha}(t)$$

This expands to:

$$P_i(t) = \lambda_{i,a}a(t) + \lambda_{i,b}(a(t)+c(t))(n(t)-1) + \lambda_{i,c}wf(t) + \lambda_{i,d}r(t) + \lambda_{i,e}vn(t) + \lambda_{i,f}o(t) + \lambda_{i,g}c(t)$$

[4] Obviously, the IPL “in each unit time” will depend on the time step used. Since the information processing load is assigned to the time step in which observable events occur, the step needs to be:

- small compared with the average length of time a flight spends in the simulation area,
- large enough that the assignment of processing within that time step forms reasonable model of reality.

Initially a time step in the region of 1 to 5 minutes is suggested, but further work is needed to check the sensitivity of the proposed approach to the time step used.

Each of the terms in this expression can be evaluated by post-processing simulation data – for fast-time or real-time simulations. Hence, P(t) can be plotted for each actor, to give a graph like the example in figure 1.

3.4 Assessing sustainable throughput

The third step in using IPL as a surrogate measure for capacity involves using these graphs to assess capacity. So how do we proceed from these graphs to an indication of *sustainable throughput*?

The first point to note is that it may be sufficient simply to compare the graphs from the different simulation scenarios. Comparing the $\{P_i(t)\}$ from different simulation scenarios allows comparison of the amount of processing that must be done by actor *i* in each scenario. If actor *i* is known to be at or near its maximum processing capacity in one scenario, then a second scenario can be said to increase capacity if P_i is reduced[5]. On the other hand, if actor *i* has spare processing capacity, then P_i can be increased (within limits) without affecting system capacity.

So comparing the graphs for a given actor, to indicate the relative loading of that actor in the two scenarios, allows the analyst to answer such questions as:

- Would the introduction of a conflict probe reduce loading on the planning controller?
- Would the use of ASAS increase the aircrew's load at times when they are already very busy?

However, if the simulations seek to answer wider questions, for example to compare the capacities of ATM systems with radically different task distributions, then we need to go on to assess the *sustainable throughput* of each actor.

Our graphs show how much information each actor must process in each unit time to control the traffic sample. *Sustainable throughput* is determined by how much capacity the actors have to process that information. The experiment analyst needs to assess the "cut-off" for each actor – the maximum information the actor can process in unit time (e.g. figure 1). The maximum instantaneous IPL a given actor can achieve should be independent of operational concept.

- For controllers and pilots the cut-off can be assessed by calibration simulations in which the actor is known (by other means) to be at capacity. Real-time simulations may be necessary, using controllers or pilots who well understand the operational roles being simulated.
- For machines and networks we can take a straightforward engineering approach.

It is not a simple cut-off. There is likely (both in human and machine actors) to be some "elasticity", whereby the actor can accept an overload of information for a short period, provided this is followed by a slack period in which it can process the "backlog". This is analogous to current human workload models. Hence, "actor capacity" is likely to be defined by a rule of the form: "not more than 100% loaded for not more than x% of the time and never more than y% overloaded" (i.e. temporal distribution rules). This determines the highest and widest peaks in processing load that the actor can manage, which in turn determines *sustainable throughput*.

In this way, the IPL graphs can be used to assess which actors were modelled as near capacity in the simulation. This indicates which actors are critical, limiting capacity in the modelled system. (This information can be used to focus subsequent developments of the operational concept on reducing the IPL of the critical actors).

Finally, the amount of traffic in the simulation traffic samples can be increased until the critical actor(s) are modelled as being at their maximum sustainable processing level. The throughput then being achieved is the *sustainable throughput*. As was noted earlier, the stochastic nature of capacity means that the sustainable throughput needs to be measured over many hours, and an average taken.

In common with the majority of today's (workload-based) capacity assessment methods, this method provides *relative assessments* of sustainable throughput, rather than absolute values. These are calibrated by comparison with a scenario/system of known capacity (e.g. as determined by observation of current sector operations). Until now, the practical success of this approach has been limited to operational concepts that are very similar to today's, so that real-life experience can be relied upon. The

[5] Either the average over time is reduced, or the height/width of peaks are reduced, or both.

INTEGRA methodology's more generic view, using IPL, aims to extend the useful scope of the approach.

This section of the paper has explained what we mean by information processing load, and how we can go about measuring it in a simulation. We now return to the second important aspect of the INTEGRA capacity method: the context in which IPL is measured.

4. Measurement context

The INTEGRA capacity assessment methodology uses defined traffic samples. The idea is to use a pre-simulation process which involves developing a set of traffic samples that represents the demand for the relevant region of airspace for a given time period (e.g. 2005). The development of the samples involves a degree of simulated strategic and executive control to ensure that the samples provide a realistic representation of expected traffic flows in that region at that time.

First, the INTEGRA Traffic Sample Generator [2] is used to generate a set of traffic samples, all complying with the same forecast demand. Forecast demand is taken as input, and is represented through a traffic pattern description in the form of flows between pairs of regions or airports. The airport arrival and departure capacities are also taken as inputs, and applied as flow management constraints to act as a queuing mechanism, smoothing out the demand into a manageable flow. The "flow management" process takes a parameter maximum acceptable delay. If the demand cannot be met within that acceptable delay then flights are removed from the sample.

Each of these traffic samples is then "pre-simulated" to ensure that it is in principle controllable, again within the specified maximum acceptable Flow Management delay. This is achieved through the following steps:

1. Calculate trajectories for all aircraft in the sample.
2. Probe for conflicts between the trajectories.
3. Analyse the most complex conflicts (e.g. those involving large numbers of aircraft) and attempt simplistic resolution by trajectory manipulation, including delay up to the maximum acceptable delay.
4. Remove from the sample any flights that cannot be de-conflicted within the acceptable delay criterion after all acceptable trajectory variations have been exhausted. (What is acceptable will be dependent on aircraft performance operating procedures and airline preferences, etc.).

The resulting traffic samples are the baseline samples, which can be used as the starting point for simulation of any ATM system and airspace structure.

Where structured airspace is to be simulated (for example, defined routes and airways), these baseline traffic samples need to be reviewed to ensure they are controllable in principle *within that airspace structure*. This is achieved by a second iteration of the pre-simulation steps 1 to 4 above, this time taking into account the required airspace structure. A record must be kept of the flights removed by this second iteration, if the structured airspace simulation is to be compared with simulations of different airspace structures (or unstructured airspace).

The baseline traffic samples provide commonality between different simulations using different simulation tools and carried out by different teams. It is important to note that, while current simulations often require a pre-processing of the traffic sample to provide something that is realistic in the current operating environment, the INTEGRA traffic samples should not be doctored in this way because that would destroy the aimed-for commonality. The pre-simulation processes outlined above ensure that the traffic samples are sufficiently realistic for the purpose of simulation. If it is found necessary to doctor the sample (e.g. for reasons of controller acceptability) then, as with the iteration to accommodate airspace structure, the changes made must be recorded for subsequent use.

Finally, we turn to the requirement to assess the capacity of ATM system elements not just in isolation, but in relation to the overall ATM system. The difficulty comes in how to bound what one has to simulate and yet still make the simulation realistic. Obviously to model a large region in great detail is impracticable, but to model only a small area is limiting. The INTEGRA methodology intends that generic baseline traffic samples will be developed to represent the traffic demand over a wide area, e.g. Europe. Traffic samples for the simulation of any individual region of airspace will be provided by “scissoring” out the corresponding geographic region from the generic traffic samples. This will include the required feed sectors.

This process essentially involves modelling the complete airspace as a network, at the highest level of abstraction, and then isolating specific areas, or nodes, for individual detailed fast- or real-time simulations. The effects of throughput in these individual nodes can then be fed back to the network model to ascertain the effects on overall system capacity.

The assessment of capacity needs to consider demand *and* delay. It is difficult to derive from a simulation of one or two sectors, the effect of any delay on the progress of the complete flight. The pre-simulation process can be repeated post simulation, to assess the effect on overall flight delay of delays introduced due to the control strategy adopted during a simulation. This is achieved by recording from the simulation the times that the aircraft entered and left the simulation. The pre-simulation process can be re-run with these times as constraints, so that the effects on the aircraft destination times can be assessed. The more parts of the overall airspace that have been simulated using “scissored” out portions of the overall demand sample, the greater the significance of this overall delay assessment.

5. Conclusion

The INTEGRA Capacity Assessment Methodology provides a generalisation of today’s workload-based methods, designed to be generic with regard to operational concept. It involves post-facto observation of simulated systems, not a priori forecasting, being based on the objective counting of observable simulation events.

It also addresses capacity at a whole-system level, providing the framework to ensure that the capacity of a sector (or, more generally, an ATM system element) is assessed within the appropriate network environment.

Its major innovative elements are:

- its new surrogate measure for capacity: information processing load (IPL);
- its use of defined traffic samples ensures that the capacity of a system element is assessed within the appropriate context, or environment.

The methodology provides a *framework* for capacity assessment – the detailed assumptions about relative weightings of different IPL elements must be set by the user to complete the IPL model for particular simulation scenarios. However, these detailed assumptions are made transparent through the use of a clearly specified set of weighting factors. This parameterisation of the model allows best and worst cases to be established and compared, and facilitates comparison of studies carried out by different teams using different simulation tools.

The INTEGRA Capacity Assessment Methodology assesses *sustainable throughput* for a given maximum acceptable delay. We have noted that capacity may in fact be constrained to less than the sustainable throughput by safety, efficiency or environmental concerns. It may therefore be necessary to review capacity assessments made, in the light of assessments of safety, efficiency and environmental impact.

The proposed methodology is not yet validated. More experimental use is required, and will probably lead to further refinement of the work presented here. However, software tools that implement the INTEGRA Capacity Assessment Methodology have been developed within the INTEGRA project. These tools have been used to apply the methodology to real simulation data from EUROCONTROL’s EACAC simulations, as a post-simulation process. The findings of this initial exercise were promising [3]:

- The software tool ran successfully in terms of performance on a Pentium 3 with 256 Mbytes of RAM, with a typical traffic sample size of 140 aircraft in two hours.
- The capacity metric algorithm has been demonstrated to be responsive to the to the data and has shown the expected differences between simulation scenarios.

This work was intended just as an initial exercise of the metric – the data recorded in the simulations was not sufficient to exercise all aspects of the methodology – and hence this only a very first step toward its validation.

References

[1] *INTEGRA Metrics & Methodologies, Detailed Specification of Capacity Metric, Inputs – Processing – Outputs*, ed.1.0 (proposed issue), EUROCONTROL 14/8/2000

[2] *INTEGRA Metrics & Methodologies, Detailed Specification of Traffic Sample Generator, Inputs – Processing – Outputs*, ed.3.0 (proposed issue), EUROCONTROL 20/9/2000

[3] *INTEGRA Metric in EACAC – Final Report*, ed. 0.A. EUROCONTROL 9/2/2001

Biographical notes

Alison Hudgell was a technical consultant to the INTEGRA Metrics and Methodologies Capacity Project carried out by DERA (now QinetiQ) for EUROCONTROL. She has 14 years' experience of research in the field of civil Air Traffic Management, characterised by work on the definition and assessment of new operational concepts, including ATC system capacity assessment, and supported by expertise in systems modelling and simulation. Her most recent work has been focussed on realising the Collaborative Decision Making (CDM) concept for civil air traffic operations in Europe.

Rod Gingell was project manager for the INTEGRA Metrics and Methodologies Capacity and Safety Projects carried out by DERA (now QinetiQ) for EUROCONTROL. He has 12 years' experience in technical consultancy and project management in the defence and Air Traffic Management domains, latterly as a Director of CGP. He has particular expertise in the area of ATM concept evaluation and integration, gained through work for CAA/NATS, EUROCONTROL and the European Commission, including PHARE Demonstration 1.