The Computational Air Traffic Control Brain: Computational Red Teaming and Big Data for Real-time Seamless Brain-Traffic Integration

Hussein A. Abbass*, Jiagnjun Tang*, Rubai Amin*, Mohamed Ellejmi**, and Stephen Kirby**

*University of New South Wales, Canberra Campus, Australia.
**Eurocontrol Experimental Centre, Brétigny, France.

ABSTRACT

As SESAR, NextGen and the world move towards more automation of Air Traffic Control (ATC) tasks to safely accommodate future growth in air traffic movements, it is clear that the Air Traffic Controllers (ATCOs) role will need to change significantly as more tasks are automated. We believe that controller workload will remain a key factor in future systems, but the focus will change from detailed repetitive activities to higher level monitoring and assurance of system-level safety. Therefore new controller workload indices will need to be developed or adapted to truly be effective.

To address this need, our research models the future division of ATC roles between the human and automation on the human brain’s two hemispheres. The left hemisphere is responsible for arithmetic, logic, and looks at the details; the right hemisphere is responsible for human qualities such as emotion, spatial awareness, and the capacity to judge holistically.

We call our integrated model of decision making within the future ATC system, the “computational ATC brain” (CAB); it consists of automation (left) and controller (right) components that will deliver ATC services in the future. For example, calculating time to closest point of approach of two aircraft to estimate time of conflict is a typical systematic brain’s left hemisphere problem. As a classical problem for automation, we assign it to the left-hemisphere of the CAB. The brain’s right hemisphere qualities—such as intuition and spatial awareness—that make an airspace safe are not well suited for automation and are assigned to the controller.

Up to now most controller workload indices developed have focused on:

1. Static, not adaptive, models that work for many operators and in many traffic situations resulting in ‘fixed’ indices. Although dynamic data can be used in an index, for example, dynamic density, a static index does not adapt to changes in the environment. Little research has been done on adaptive workload indices that adapt the impact of a variable, such as the number of aircraft in a sector, from one situation to another. As in much of life, in ATC assessments one size does not fit all.

2. Providing a situational awareness only based on history without factoring in potential future actions. This works well for analyzing what happen but they are of limited use for helping to develop systems that can avert potential overload situations. That is once a situation is assessed to be of high workload, it is too late to do anything about it without a major disruption of traffic.

This paper will address part of the wider research we have been doing in evaluating the true effectiveness of workload models, and provide an overview of a prototype environment for evaluating adaptive and predictive workload models in the real time management of complexity. We will present

1 Corresponding author: Email: h.abbass@adfa.edu.au
results from a November 2013 experiment performed by the authors at Eurocontrol Brétigny in France. This experiment aimed at evaluating a proof of concept for comparing mental workload assessed directly from electroencephalograph (EEG) data obtained from the human brain and a classical “fixed” workload model while closing the automation loop.

**BACKGROUND**

Since our paper uses some terminologies not usually found in the ATC lexicon, we provide some background information below:

**The Big Data Challenge**

Big data is an opportunity that we cannot miss out in this century. The classical definition of big data relies on the 5V model: volume, velocity, variety, veracity and value. The air traffic domain and the work in this paper are exemplars of the 5Vs.

Volume is about the size of the data. Repeating the analysis presented in this paper in a single command and control centre with 20 operators will generate close to 10 terabytes (TB) of psychophysiological data alone every day. In our Australian ATC research Lab based at the University of New South Wales, Canberra Campus, we developed the TOP-LAT (Trajectory Optimization and Prediction of Live Air Traffic) system; a passive ground watchdog system of air traffic. The system operated from 2008 to 2013 receiving in real-time the traffic of the two Australian flight information regions, processing trajectories, estimating different emissions, calculating different task complexity metrics, estimating risk and analyzing the traffic network, to name a few. In this system, a few TB of data can be easily generated on a daily basis. We only saved raw data that enables us to recalculate what we may need in the future and system-level statistics as we did not have the storage capacity to save all the data being generated.

The above examples capture velocity, whereby the situation is highly dynamic (for example, human mental load), causing continuous changes in the data. It is clear that the data is heterogeneous (variety); in this paper, traffic and psychophysiological data needed to be aligned and fused together. The data contains noise (veracity) that need to be dealt with, from artifacts in the EEG signal, noise in radar and possible inaccuracy of ADS-B in certain geographic regions. With the complexity of the future air traffic system, it is hard to deny the value that will be generated from properly utilizing big data within the air traffic domain. The era of big data is an era of opportunities for improving ATC worldwide.

**The Computational ATC Brain (CAB)**

A brain with two hemispheres: the left hemisphere is made of silicon automating complex ATC tasks including tactical controller tools, mid-term conflict detection and resolution, complex real-time adaptive tactical flow management tools, dynamic airspace sectorization tools, data links, and satellite navigation capabilities. The right hemisphere of the CAB is made of humans including ATCOs, pilots, supervisors, and different stakeholders.

The human’s right hemisphere brain functionalities represent those qualities that no machine will replace, at least for years to come. The right hemisphere of the human brain needs to continue to operate as the right hemisphere of the CAB in the future. The role of ATCOs will change to emphasize right hemisphere functions. The integration of the computations done in Silico in the left-hemisphere of the CAB and the computations done by humans in the right hemisphere is a complex Brain-Traffic
Integration problem [1]. This problem sees the controller’s brain as an influential component ‘inside’ and an integral part of the automation loop rather than an external recipient of those decisions made by the automation loop.

But how to achieve seamless integration between the left and right hemispheres—the silicon occupying the left hemisphere and the humans occupying the right hemisphere—of the CAB remains to be a scientific challenge. The benefits from NextGen and SESAR will not reach their maximum potentials until we think anew of this interface to achieve seamless integration between the human brain and the silicon brain that will manage the complexity of the future air traffic environment.

This Brain-Traffic integration problem requires out-of-the-box thinking. The massive amount of data (the Big Data challenge) that will exist at the interface of the human-brain and the air traffic environment require efficient mechanisms to analyze these data, extract the right cues at the right time for the right context and seamlessly exchange cues between the left and right hemispheres of the computational ATC brain. Computational red teaming (CRT) is possibly one scientific endeavor to study this interface in the form of a challenge from a risk lens to disentangle the uncertainties occurring at the interface and their impact on the environment and different objectives of the participants.

**Computational Red Teaming**

Red teaming (RT) is an ancient military concept to play ‘devil advocates’ against our own plans, concepts, and systems to discover vulnerabilities, risk assess our thinking and systems, and design innovative mitigation strategies to secure these systems. The difference between RT and other test and evaluation approaches lies in its reliance on two concepts: “deliberate challenge” where the process assesses system boundaries and “deliberately” pushes a situation beyond these boundaries to stress-test the system, and “risk thinking” which emphasizes (1) continuous innovation, (2) uncertainty analysis in time and space, and (3) analyses focused on systems’ objectives.

‘Computational Red Teaming’ (CRT), an innovation of this research group [3], attempts to design artificial intelligence systems to assume the role of, or support, the human red team. CRT relies on an integrated architecture with optimization, data mining and simulation components to design an intelligent system that is able to ‘challenge’ and risk-assess a concept, technology or a human.

We emphasize the concept of ‘challenge’ in CRT as opposed to a classical decision support system which aims to “aid” or “support” humans. A challenge is a pro-active process whereby the situation is continuously evaluated, risk assessed, and actions are dynamically introduced to counteract a user, system or environmental effects. To design a system that is able to challenge something, it needs to be able to (1) continuously receive information, extract patterns and trends, and assess the risk associated with the situation; (2) have available models or mechanisms to mimic the behavior of the system it is trying to challenge so that it is able to evaluate the impact of a potential challenge; (3) to exhibit the ability to search for optimal actions that will influence the system it is trying to challenge.

CRT systems are designed to either discover vulnerabilities by proactively discovering holes (exposure of critical elements in a system to hazards or threats), or support humans’ blind-spots by proactively and continuously anticipating threats in the environment and providing counter-actions to help the human operator overcoming these challenges. In this paper, an undesired change in the complexity of the traffic in the environment is the hazard that needs to be continuously anticipated, assessed, and mitigated.
CRT has been applied to different application domains. Within ATC, it has been used to evaluate mid-term conflict detection tools [4], ground-air interaction in the terminal maneuvering area [8], correction of air traffic events [5], to name a few. In our previous work on CRT, we used it to discover vulnerabilities in systems. In this paper, we demonstrate, for the first time, its use to support humans.

In this paper, CRT is used to demonstrate the two workload challenges discussed in the abstract, and to close the loop between measuring workload and intervening within the environment to manage workload. We share the first experiment of its kind to demonstrate that seamless integration of human brain data and air traffic data is possible, and that humans and machines can work in harmony.

**Real-Time Seamless Integration**

The objective of this work is twofold. From an academic perspective, it demonstrates that real-time, dynamic and seamless integration of cues from human mental processes within automation improves the effectiveness of the decision support environment. From an applied ATC perspective, it reveals the advantages of using CRT to assess workload metrics.

First, by acting in real-time on the workload information obtained through different metrics, we can assess the suitability of the metric for a real-time environment. For example, the dynamic density metric we used in the experiment reflected nicely the complexity of the traffic situation based on subjective metrics and informal discussions we had with the ATCOs. However, by the time the load is detected to be high, it was too late to do anything about it. In fact, we found that the interference of automation at this late stage increased ATCOs workload. Therefore, this metric is not suitable as an anticipatory metric for ATC complexity.

Second, by assessing psychophysiological metrics obtained from sensors such as EEG, we can calibrate whether or not the traffic-dependent metric can truly assess mental load. By closing the automation loop through feeding EEG cues back into the decision support environment, CRT provides an objective and effective test for any workload metric. CRT couples the development of the metric with the `act` on the metric. If a metric measures or even anticipates workload, identifies that the system is outside its safe operating envelope, but we have no way of knowing what maneuvers and actions required to steer the system back to a safe operating envelop, it is then a clear indication that the solution is incomplete. To put it succinctly, it is no longer sufficient to measure; we need to understand causes and controls that enable us to influence the system under consideration when a measure indicates a state of risk.

**The Study**

The hypothesis of the experiment was that it is possible to continuously, simultaneously and seamlessly monitor and analyze data from the traffic and human mental/brain activities to control the complexity of the traffic in real-time in order to (1) seamlessly manage workload; (2) and compare task and mental load indicators in real-time scenarios.
Figure 1: A schematic diagram of the experimental environment.

The system (Figure 1) is designed to extract cues in real time from both traffic and mental data to trigger CRT to balance complexity. Subjective assessment of traffic complexity through the ATC Workload Input Technique (ATWIT) developed by FAA [7] is performed every two minutes. Brain signals are measured continuously from the ATCO in real-time using a Nexus-32™ EEG cab with 19 channels (Figure 2). These signals are analyzed and cognitive cues on the mental states of the ATCO are extracted. Simultaneously, traffic information is analyzed in real-time and task cues are extracted to capture the complexity of the traffic in play. One or both type of cues is used to make a decision based on the scenario being used.

Figure 2: The EEG Cap during a pre-experiment testing on one of the Authors. The diagram also shows the ATWIT system displayed on the right screen, the traffic scenario in the middle screen, and a communication panel on the left screen.

Changes in these two sources of complexity are continuously being assessed. When an undesired situation arises, the optimization component of CRT gets triggered to assess which maneuver or action
in the environment is best to counter-act the undesired state. It uses the simulation environment for impact and look-ahead analysis.

Once this action is identified through the CRT process, it gets displayed on the advisory screen and the process continues. The tagging position has software that serves four purposes: (1) keeping track of the experimental protocol ensuring that every step has been completed; (2) recording subjective observations during the session; (3) sending event-markers to other software components, such as the start of session event; (4) time-stamped synchronization messages to all other software components.

The experiments provided us with wealth of data that we can use to improve the models we built. For example, Figure 3 shows a picture of the areas activated in the brain for one of the controllers during one of the tasks. Red color indicates a high level of mental processing. This type of analysis provides us with deeper insight when evaluating workload.

![Figure 3: An example of brain activities for one of the controllers during a task. The two lines in the north symbolises the nose, while the two curves in the east and west depict ears.](image)

The four participants were previous ATCOs who are currently working within the Eurocontrol experimental centre. Each air traffic scenario lasted for 50 minutes. During the first 25 minutes, the aim for CRT was to identify maneuvers that will increase complexity, while in the last 25 minutes, the aim was to identify maneuvers to decrease complexity. Four different trials were conducted for each subject: one with CRT turned off, one with CRT relying on task complexity (using the index in [6]) alone to make decisions, one with CRT relying on cognitive complexity alone to make decisions, and one with CRT relying on both task and cognitive (based on brain data) complexity working together. CRT worked best when both task and cognitive complexity were used. Subjective assessments using ATWIT was used to assess ATCO’s perception of workload every two minutes. It revealed that the reliance on task complexity alone increased complexity in the last 25-minutes instead of decreasing it. The reason is that classical workload metrics detect complexity after they happen. The belated automated advice increased the complexity of the situation when the operator was already attempting to manage the complexity in the space. Cognitive complexity metrics acted, however, as lead indicators; thus, the advices were given on time for controllers to act.

**DISCUSSIONS AND CONCLUSION**

A recent human-based experiment conducted at the Eurocontrol Experimental Centre at Brétigny, France, used Computational red teaming (CRT) as a proof of concept for a test and evaluation environment for cognitive load and workload metrics.
The results demonstrate the feasibility of using cognitive and task indicators to help ATCOs to manage complexity in airspace. Cues extracted from EEG brain signals acted as lead indicators that allowed the system to manage complexity. The situation reversed with workload indicators. Better anticipatory cognitive complexity metrics are needed to support air traffic controllers in real-time environments.

Future research of this work is twofold. First, we will explore the concept of anticipatory adaptive workload indices where the metric for workload changes during one session based on circumstances. Secondly, we will continue to enhance our Computational Red Teaming (CRT) environment to better test and evaluate the ability of a proposed metric to adapt to human mental processes and traffic situations, and potentially be used in real-time to inform better decision making.

**ACKNOWLEDGEMENT**

The authors are obliged to Mr Edward Stevens for his thorough comments that tremendously helped in improving the readability of this paper, and to Mr Steven Carver and Ms Marion Brophy for their professionalism and hard work in handling this submission. We thank Miss Shen Ren for her support during data collection, Dr Jon Hegg for training the team on the hygienic and fast fitting of EEG caps and Dr Anastasia Lampropoulou for support with Nexus 32. A special thank you goes to all participants who were very professional, patient and supportive; and for Eurocontrol for being a forward looking organization. This work was done with human ethics approval HC13251 under the title “Computational Red Teaming for Air Traffic Control”.

**REFERENCES**


