Agent-based simulation of a multi-queue emergency services call centre to evaluate resource allocation

Bruce G. Lewis[1,2], **Ric D. Herbert**[1], **Peter F. Summons**[1], **and William J. Chivers**[1]

[1]Faculty of Science and Information Technology, The University of Newcastle, Ourimbah, NSW 2259, Australia.
[2]NSW Police Assistance Line, PO Box 3427, Tuggerah NSW 2259, Australia.
E-Mail: lewi1bru@police.nsw.gov,au

Keywords: *Emergency services, call centre, Police Assistance Line, agent-based modelling*

ABSTRACT

Emergency services call centres operate to stringent parameters based on providing rapid support to their communities. It is this that differentiates the emergency service call centre from other call centre categories.

The focus of this paper is to model the resource allocation within the NSW Police Assistance Line (NSWPAL), which is a multi-site, multi-queue, multi-server call centre that takes urgent and non-urgent calls from the community of New South Wales, Australia. The major resource for NSWPAL is its Customer Service Representatives (CSRs) who answer the calls and, although multi-skilled, may not be equally skilled across the queues.

The business case for the model is simple. Staff costs are a major consideration for the call centre with unscheduled incoming calls on the different queues leading to additional staff costs. Although rostering of staff for the queues is based on historical forecasts, one of the challenges in running a multi-queue centre is to staff the queues to meet the unexpected demands, especially during emergencies. This leads to the shuffling of staff to meet the load and invariably affects the service level of one of the queues.

One such event occurred in July 2007 during a normally busy period. The inbound call rate was such that, although all calls were answered, service level fluctuated and could not be maintained at the target level for a period of time. It is these situations that we aim to model to examine staffing policies and real-time management regimes.

The literature suggests that traditional modelling techniques will not provide the flexibility required to model the call centre. In this paper we propose an agent-based model of the call centre, with the main agents being callers, call distribution manager, Team Leader, CSRs and the queues. The objective of the model is to improve the business through generating practices that lead to better customer service level metrics at reduced costs through more efficient resource allocation. The agent based model provides the flexibility to model the resource allocation without having to commit considerable computing resources and undertaking a rigorous mathematical approach. The model uses the extensive data base of the call centre to provide information for modelling the agents. In this paper we describe the model at the systems level and present the agent-based model we will be using.

1 INTRODUCTION

In this paper we propose a model to examine the allocation of limited server resources in a multi-queue Emergency Services call centre. Staff costs are a major consideration for the call centre with unexpected incoming calls on the different queues leading to additional staff costs. This leads to the shuffling of staff between queues to meet the load and this invariably affects the service level of one of the queues. Although rostering of Customer Service Representatives (CSRs) for the queues is based on historical forecasts, one of the challenges in running a multi-queue centre is to staff the queues to meet the sudden unexpected demands, especially during emergencies. The NSW Police Assistance Line (NSWPAL), which is an Emergency Services call centre, will be used as a source of rich data and as a case study for the agent-based model. PAL is a multi-site, multi-queue, multi-server call centre that takes urgent and non-urgent calls from the community of New South Wales, Australia.

In the following sections we examine previous call centre studies, looking at the need for cost minimisation while trying to meet service level. We also explain how an emergency call centre can be differentiated from a commercial centre. We describe the operation of NSWPAL and an situation the affected the performance of the call centre, we identify a modelling technique that gives us the freedom to test a number of management regimes to deal with unexpected increases in call volume and finally, describe the model we will be building.

2 PREVIOUS STUDIES

The literature mainly defines a call centre in commercial terms (Mehrotra & Fama 2003, Omari & Al-Zubaidy 2005, Gilmore 2001). The things that the definitions have in common are that they are based on telephone systems, they are staffed by dedicated people known as agents, there may be one or more queues and that there is a high variability and uncertainty in the call arrival rate for inbound centres.

Broadly speaking, the Emergency Services call centre fits this mould but is differentiated by two important aspects; timeliness and caller state. The callers may be in a state of high anxiety due to the situation in which they are involved and, irrespective of this, the agent must extract the information required to facilitate the timely dispatch of resources. Another differentiator is the call arrival rate. In the context of an inbound emergency services call centre, the organisation has no control over the arrival rate which is contingent on natural and human phenomena. Chen & Henderson (2001) identify this an important factor in determining the staffing levels at a New Zealand police communications centre. As we will discuss later, this occurs at the NSWPAL and we will describe a particular incident. Much of the research around call centres is on finding the optimum way to staff them Erdem & Gedikoglu (2006).

Bennington et al. (2000) explain that the advantages to customers who deal with a call centre include a reduced cost to make the call, greater convenience due to reduced travel, quick service once connected due to newer technologies, and the ability to communicate in the language of their choice. This is the case for both urgent and non-urgent calls at NSWPAL. Customers dial NSWPAL on a toll-free number available from fixed telephone services, are connected quickly due to accurate staffing and technical systems and can request an interpreter if so required. They state that the advantage to the organisation that has a call centre include the ability to service many customers at the one time, improved call streaming with customers making selections and less management costs due to reduced office space for customers. To this we add the use of common technical systems, common technical infrastructure and in-house support services located with the call centre; this is the case at NSWPAL. The disadvantages relate to lack of face-to-face contact and unknown customer perceptions (Bennington, Cummane & Conn 2000).

One of the major drivers for call centre efficiency is cost. There are different types of costs including staff and telecommunications. We divide these into tangible and intangible costs. The tangible include staff and telecommunications and the intangible include political fallout, reputation and media reporting. Hughes (1995) notes that since "incoming call centers are often viewed as cost rather than profit centers accurate staffing is a primary concern" and that staffing to meet service levels with tight cost constraints is a challenge. Consequently, much of the driving force associated with call centre management and research is centred around improving efficiencies due to the high cost of labour (Bennington et al. 2000). Chassioti & Worthington (2004) identify this as being in the order of 70% of the a call centre's total running cost and there are significant management challenges in the areas of human resource being recruitment, absenteeism, emotional support, burnout. Duder & Rosenwein (2001) identify that the cost to provide trained operators exceeds half of the operational costs of a call centre. In addition, increased delays in servicing customers translate to increased system and telecommunications costs due to customers abandoning and retrying or not retrying. We identify another operational cost that is associated with an emergency service; the cost to call in staff on overtime or to shorten meal breaks within the business rules to meet the service level for urgent calls.

Call centre performance is judged by a variety of metrics. Of these, the most common presented in the literature are the Average Speed of Answer (ASA), Service Level or Telephone Service factor (TSF) being the percentage of calls answered within a specified time and, percentage of abandoned calls (Duder & Rosenwein 2001, Mehrotra & Fama 2003), Gans et al. (2003 cited Robbins, Medeiros & Dunn (2006)). Pichitlamken et al.(2003) note that since call centres aggregate their measurement into intervals of, say, 30 minutes, lack of call by call information complicates analysis of the data since standard parameter estimation techniques do not usually apply. At NSWPAL, the granularity is 15 minutes and this is used for forecasting the call rate and scheduling the CSRs.

One of the parameters that underpins the performance of an inbound call centre is the call arrival rate and it is this that is driving our research. Betts, Meadows & Walley (2000) note that call centres have seasonal demand patterns and many experience short term spikes in demand, for example due to weather conditions, with capacity management being a trade-off between operator boredom and high service levels. Consequently, the nature of peaks and troughs in demand can have a significant effect on operation and so need to be understood. In a previous study (Lewis, Herbert & Bell 2003) of the NSWPAL non-urgent inbound queue, it was shown that there were strong daily and weekly patterns together with unexpected spikes in the incoming call rate. This then is the dilemma for the Team Leaders who monitor the call centre operations and performance; (how) can the service level be maintained during an unexpected increase in incoming calls?

Mehrotra & Fama (2003) note a number of reasons as to why call centres are interesting as being the worldwide size of the industry, the operational and mathematical complexity, multi-queue nature, the inbound calls arrival times randomness, the random call durations, routing technologies complexity and the agent skill sets differences to handle a variety of call types. For these reasons the authors state that decision makers have difficulty understanding the system dynamics without effective modelling and explain the challenges faced by call centre managers including balancing the main areas of service quality, cost and employee satisfaction. These, they say, lead to important questions for which decision support models are valuable; number of agents; their skill sets; how agents should be scheduled including shifts, breaks, training, meetings and other activities; call types, quantities and when they will arrive; inbound calls' response time; agent multiple queue training, "Call centre performance against forecast, schedule and routing design" and the performance of the centre. In their research model, the authors find that "call centre managers can have only a small influence on short-term performance".

3 THE NSW POLICE ASSISTANCE LINE (NSWPAL)

NSWPAL is an emergency services multi-queue, multi-server, inbound call centre. It has been in existence as a NSW statewide operation since December 1999. It is a 24-hour inbound telephone call centre available to police and the community. NSWPAL operates as a single virtual call centre by virtue of its technology base. It is a niche facility in the call centre industry, being used for the reporting of both urgent and non-urgent crime and incidents, for providing police-related information to the community and for providing an intelligence source for the NSW Police.

NSWPAL currently runs five significant business streams and caters for planned and unplanned emergency and special operations. The main business streams are Triple Zero (000), Crime Stoppers, 131-444, Customer Assistance Unit and the police telephone switchboard. NSWPAL also provides call centre services for special operations and other NSW State Government agencies. Depending on the number of calls in a queue CSRs change queues to service the priorities and maintain service level under the management of their Team Leaders.

NSWPAL uses modern information technology and systems throughout all of its businesses. A Computer-Telephony Interface (CTI) system monitors the incoming calls and ensures that all calls are routed to the suitably trained agent. The same system maintains a comprehensive database that is used to show the centre's performance in real time and to prepare a variety of management reports. A Workforce Management System is used to forecast the various incoming call queues and to schedule staff in accordance with their and PAL's business needs to meet the forecast demand.

In Australia, Triple Zero (000) is the number dialled when people require urgent assistance for an emergency service organisation. When Triple Zero (000) is dialled, the call goes to the telecommunications carrier (Telstra) Triple Zero (000) operator who asks whether the caller wants police, fire or ambulance. If police is requested, the Telstra operator transfers the call to a police operator and the Telstra operator stays on the line until the call is taken by a police agent. At the time of writing, about 75% of police Triple Zero (000) calls in NSW are taken by the Police Assistance Line. A PAL operator takes the call and immediately starts entering the details into the police despatch system. The job is received by the radio operator who dispatches sworn officers. Underpinning this is a series of systems and redundancies to ensure the call is serviced expeditiously.

PAL's performance is measured in terms of its service level, being the percentage of call answered in a given time. This metric is different for different queues. Daily performance for the 131-444 non-urgent queue are published on the home page of the NSW Police website at http://www.police.nsw.gov.au/. This overall performance is directly related to the performance on the staff including the CSRs and the Team Leaders who continually manage the queues.

4 PROBLEM DEFINITION

While cost is an important factors in effectively managing an emergency call centre, it is the risk to the community that needs to be minimised. At PAL, the CSR four-weekly rosters are prepared six-weeks in advance. These are based on the forecast call volume, service level and the CSR preferences. Each of the queues has a different service level. For the urgent queue the target is to answer 90% of call in 10 seconds while the non-urgent is 80% in 27 seconds.

Friday nights and Saturday mornings are usually a busy period for the urgent queue. In July 2007 unexpected events caused a drop in performance of the urgent queue. Both urgent and non-urgent queues were functional. In Figure 1, can be seen the forecast calls and the received calls and the fact that the forecast was very good. However, the urgent queue tells a different story. With reference to Figure 2, for the period 10 pm to 6 am the number of calls received far exceeded the forecast volume and in particular, the period 11 pm to 1 am service level was such that it was not possible to answer the calls within 10 seconds. Since for each queue, the number of CSRs is related to the expected call rate, the capability to meet the extended peak was not available. The problem for the Team Leader on duty was how best to optimally manage the situation with the available resources. The options available to the team leaders are limited and include moving CSRs between queues, changes in CSR meal breaks, calling in additional staff and extending the shifts of CSRs. Changing CSRs between queues can only occur if the other queues are able to maintain their performance. The the latter two options are at a financial cost to the NSWPAL and so requires management approval.

5 MODELLING AS A SOLUTION

Over the past 20 years, social science modelling has undergone a significant change. Gilbert & Abbott (2005) explain that it has evolved from "durational models of the 1980s to the logistical regressions

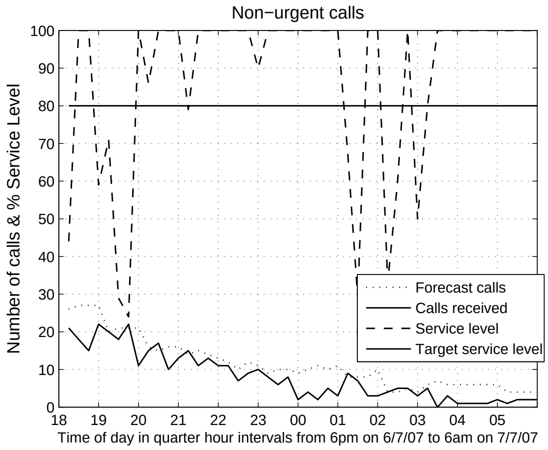

Figure 1. Non-urgent calls for the period 6pm on 6 July 2007 to 6am on 7 July 2007

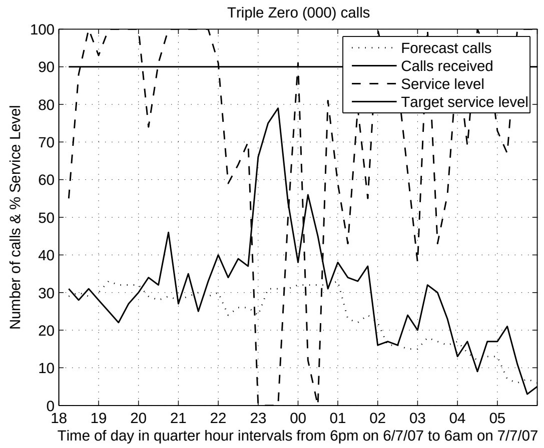

Figure 2. Urgent calls for the period 6pm on 6 July 2007 to 6am on 7 July 2007

of the 1990s and the count-data regressions of the most recent decade". Furthermore, the increase of computational power has enabled a shift towards simulation and maximum likelihood data approach compared to the previously used estimation and optimisation approaches. The authors identify the different ways computers can be used to examine social structures and highlight agent-based modelling as an important change in social science since the implications of human actions in social structures can be examined. They also note that large amounts of data based on social structure evolution can be rendered visually to allow patterns to be easily detected.

Gilbert & Terna (2000) state that although statistical and mathematical models provide a formal base, the number of equations required to "represent real social phenomena", particularly where non-linear relationships are present, and the often implausible simplifying assumptions can lead to misleading theories. They say that with computer simulation, there is no difficulty in representing non-linear relationships although these can produce some methodological problems. They note that “major growth” of computer simulation use in the social sciences is due to the development of agent-based models (ABM).

Gilbert (1999) explains that simulation allows emergence and development of macro-level responses from individual action. Hitchins (1992) notes that emergence occurs where “whole entities exhibit properties which are meaningful only when attributed to the whole, not its parts” and that “every system exhibits emergent properties which derive from its component activities and structure but cannot be reduced to them.”

In the case of NSWPAL, the main parameter of interest in the model is the service level. This is an emergent property derived from the speed of answering the calls which in turn is dependent on the number of calls waiting, the number of agents available and the individual agents’ performance parameters.

There are a number of techniques and computer-based programs that have been used to model call centres. Abdulmalek (2006) notes that operational research modelling can be used to staff call centres efficiently. Duder & Rosenwein (2001) use an intuitive modelling approach to influence management strategy for ”a large scale call centre operation that provides a service function”. Omari & Al-Zubaidy (2005) use a modelling tool called OPNET to simulate their call centre under study. They evaluate quality of service, the average wait time, the agent’s utilisation and cost based on salaries. Gilbert & Terna (2000) identify a number of suitable tools that implement object-oriented programming systems including Lisp-Stat, Swarm, MAML (Multi-Agent Modelling Language), SDML (Strictly Declarative Modelling Language), LSD (Library for Simulation Development), SIM_AGENT, SimBioSys and StarLogo. They explain their use and the situations in which they are suitable. Buist & L’Ecuyer (2005) describe a Java library called ContactCentres that is used to write contact centre simulators. The authors note that specialised software such as ccProphet and the Arena Contact Center Edition from Rockwell significantly ease modelling.

Object Oriented Design (OOD)is used for computer-based modelling. In his book, Roff (2003) discusses the attributes and advantages of OOD. He says that one advantage of OOD in comparison to procedural programming and says that by hiding functionality with the use of encapsulation makes software easier to read and prevents problems. Gilbert & Terna (2000) state that object-oriented techniques avoid memory management problems and high level tools such as Swarm avoid both memory management and time synchronisation problems.

Macal & North (2006) explain that ABM is different from object-oriented simulation and that object-oriented simulation is used as a basis for ABM. They also note that ABM toolkits are generally object-oriented. Bonabeau (2001 cited Macal & North (2006)) states that ABM is grounded in the modelling of “human social behaviour and individual decision making.” Macal & North (2006) state that ABM is becoming popular because systems and interdependencies have been and are becoming more complex and, data is being stored in databases at finer granularity.

In our model, we want to simulate the interaction between the rate of incoming calls, the CSR’s ability to service this and the management of the queues by the Team Leaders under different management regimes. In this section we describe briefly describe the way computers are used and conclude that because of the non-linear parallel processes involving human interaction, agent-based modelling (ABM) based on Object Oriented programming techniques is the most suitable approach for our model

6 THE MODEL

Under the management of the Team Leaders, the CSRs work autonomously in processing the calls for the queues into which they are logged. The CTI system works independently of human intervention to ensure the calls for the various queues are routed correctly. Associated with each queue is a complex decision process as to how to allocate CSRs in dynamic call environment. For this, the team leaders rely solely on the presented CTI statistics. One of the goals of our model will be to allow team leaders to assess the impact of sudden or unexpected changes in the call rate and assess the impact on performance of different queue management scenarios.

The daily operation of the NSWPAL involves parallel processes, as noted in Gilbert (1999), being a number of CSRs on concurrent calls and calls arriving concurrently into the queues. When answering and dealing with the calls, the CSRs act autonomously and make decisions (Macal & North 2006) within the bounds of their training and the NSWPAL business rules. This is similar for the Team Leaders who monitor the business queues and allocate staff the queues based on the queue and CSR performance. For these reasons, and since we wish to simulate the complex processes underpinning the operation of the call centre, we have chosen to use ABM to model and simulate the NSWPAL call centre facility. We also see the appeal of the ABM approach as allowing the performance parameters of individual CSRs to be included in the model to allow better tuning of the queue management process.

The model we propose is a simplified version of the NSWPAL and captures the essence of the operation to maintain an optimum service level at minimum cost. We examine the model on two levels. First we take a systems approach and then we use this to identify the agents together with their actions and attributes as an ABM. Figure 3 shows the model as a system with feedback to optimise the performance. The PABX/ACD/CTI subsystem routes an incoming call to the relevant queue. This is determined by the number dialled by the caller. Calls wait in one of the queues until answered by a CSR. Once a call is finished, the CSR waits for the next call. Calls take a finite time to complete. This is called the handle time and consists of the talk time with the caller and any after call work time without the caller. In our model we measure the handle time as a single entity. There are two feedback paths. In one of the feedback paths, the individual performance of the CSRs emerges to the centre's performance. In the other path, the Team Leaders monitor the call queue metrics and determine what needs to be done to optimise the service level and ensure callers do not wait an excessive time. Particular attention is given to the urgent queue over the non-urgent.

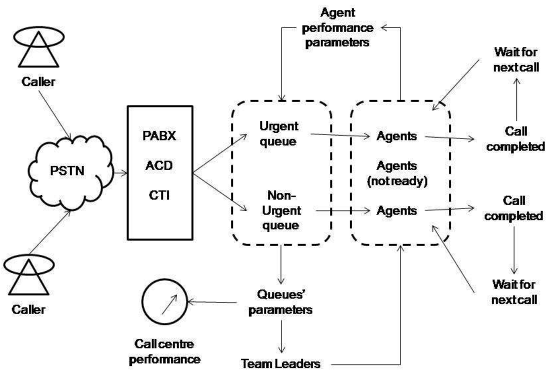

Figure 3. The call centre model showing the main blocks that will be modelled

The next level is the ABM and this is best seen in the UML class diagram hown in Figure 4. The model has both agents and an operating environment.

We see the agents in the model as being callers, a call management system which is the subsystem that routes the call to a waiting queue or a CSR, CSRs who answer and processes the calls in each of the urgent and non-urgent queues and a Team Leader who manages the queues in real time. The environment in which the model will operate consists of a clock to time the events, the call centre statistic files, and the reports. We assume that the focus is on the operation of the call centre after a call enters the call management system, that there are more incoming lines than callers and so there is no call blocking, that all customers are answered from a FIFO queue, and that no calls are abandoned from the queue.

7 CONCLUSION

By examining the operation of an emergency services call centre during times of an unexpected increase in incoming call volume and by examining the various computer-based methods of solving the associated resource problem, we have determined that an agent-based model will be the most suitable approach to modelling the NSWPAL. We have distilled the essence of the multi-queue operation of the NSW Police Assistance Line and have translated the systems approach to an agent-based model with agent classes, their attributes and their methods.

Our next step will be to build the model and apply it to previous events while looking for emergent properties and theories Further research may see the theories being applied to other emergency services organisations.

8 ACKNOWLEDGEMENT

The authors wish to acknowledge the helpful comments and suggestions of the referees, the permission granted by the NSW Police Force to undertake the research and the support of the NSWPAL Director.

9 REFERENCES

Abdulmalek, F. (2006), Improving a call center performance at a telecommunication company, *in* A. Genco, A. Gentile & S. Soce, eds, '4th International Industrial Simulation Conference', EUROSIS-ETI, Palermo, Italy, pp. 93–96.

Bennington, L., Cummane, J. & Conn, P. (2000), 'Customer satisfacion and call centers: an australian study', *International Journal of Service Industry Management* **11**(2), 162–173.

Betts, A., Meadows, M. & Walley, P. (2000), 'Call centre capacity management', *International Journal of Service Industry Management* **11**(2), 185–196.

Buist, E. & L'Ecuyer, P. (2005), 'A java library for simulating contact centers', *Proceedings of the 2005 Winter Simulation Conference* pp. 556–565.

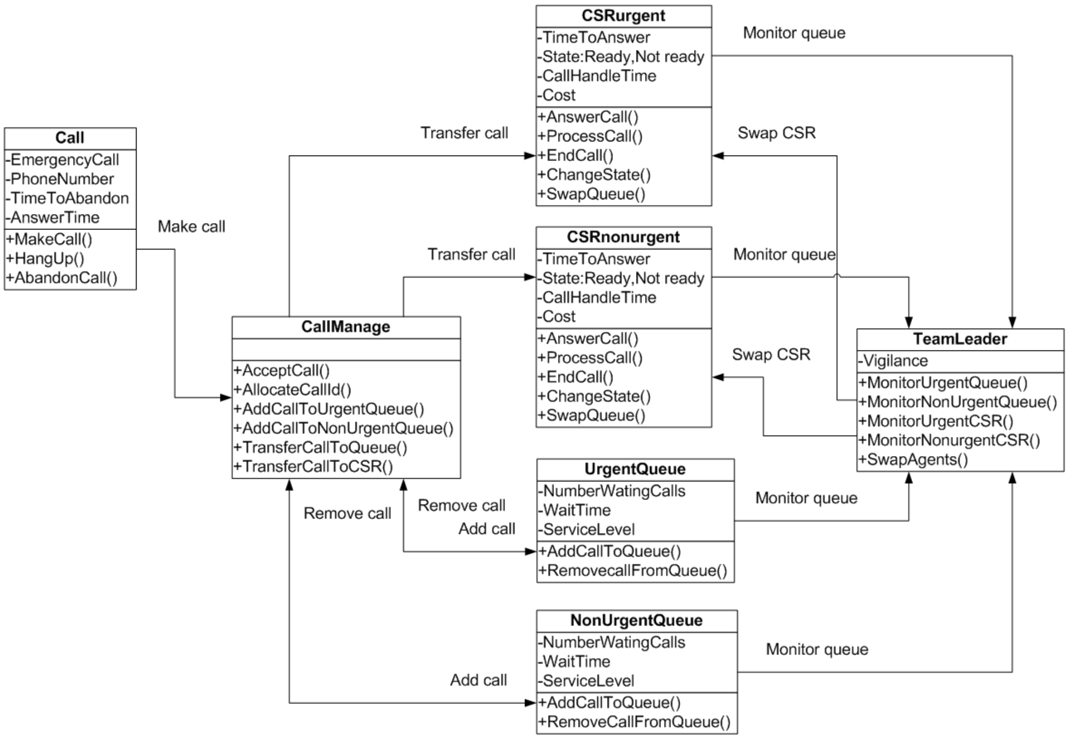

Figure 4. UML representation of the call centre model

Chassioti, E. & Worthington, D. (2004), 'A new model for call centre queue management', *Journal of the Operational Research Society* **55**, 1352–1357.

Chen, B. P. & Henderson, S. G. (2001), 'Two issues in setting call centre staffing levels', *Annals of Operations Research* **108**, 175–192.

Duder, J. C. & Rosenwein, M. B. (2001), 'Towards "zero abandonments" in call center performance', *European Journal of Operations Research.* **135**, 50–56.

Erdem, A. S. & Gedikoglu, B. (2006), 'A dss for shift design and workforce allocation in a call center', *PICMET 2006 Proceedings* pp. 1279–1289. Istanbul, Turkey.

Gilbert, N. (1999), 'Simulation: A new way of doing social science', *The American Behavioural Scientist* **42**(10), 1485–1487.

Gilbert, N. & Abbott, A. (2005), 'Introduction', *The American Journal of Sociology* **110**(4), 859.

Gilbert, N. & Terna, P. (2000), 'How to build and use agent-based models in social science', *Mind and Society* **1**(1), 57–72.

Gilmore, A. (2001), 'Call centre management: is service quality a priority?', *Managing Service Quality* **11**(3), 153–159.

Hitchins, D. K. (1992), *Putting Systems to Work*, Wiley, England.

Hughes, C. (1995), 'Four steps for accurate call-center staffing', *HRMagazine* pp. 87–89.

Lewis, B. G., Herbert, R. D. & Bell, R. D. (2003), 'The application of fourier analysis to forecasting the inbound call time series of a call centre', *MODSIM 2003, International Congress on Modelling and Simulation Proceedings* pp. 1281–1286.

Macal, C. M. & North, M. J. (2006), Tutorial on agent-based modelling and simulation part 2: How to model with agents, *in* L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol & R. M. Fujimoto, eds, 'Proceedings of the 2005 Winter Simulation Conference', IEEE, pp. 73–83.

Mehrotra, V. & Fama, J. (2003), Call center simulation modelling: Methods, challenges and opportunities, *in* S. Chick, P. J. Sanchez, D. Ferrin & D. J. Morrice, eds, 'Proceedings of the 2003 Winter Simulation Conference', IEEE, pp. 135–143.

Omari, T. & Al-Zubaidy, H. (2005), Call center performance evaluation, *in* 'Canadian Conference on Electrical and Computer Engineering', IEEE, pp. 1805 – 1808.

Robbins, T. R., Medeiros, D. J. & Dunn, P. (2006), Evaluating arrival rate uncertainty in call centers, *in* L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol & R. M. Fujimoto, eds, 'Proceedings of the 2006 Winter Simulation Conference', IEEE, pp. 2180–2186.

Roff, J. T. (2003), *UML A Beginner's Guide*, McGraw-Hill/Osborne, New York.