ELECTRONIC WORKSHOPS IN COMPUTING

Series edited by Professor C.J. van Rijsbergen

C.R. Roast and J.I. Siddiqi, Sheffield Hallam University, UK (Eds)

BCS-FACS Workshop on Formal Aspects of the Human Computer Interface

Proceedings of the BCS-FACS Workshop on Formal Aspects of the Human Computer Interface, Sheffield Hallam University, 10-12 September 1996

Modelling Interactions as Queues

C. Karagiannidis, A. Koumpis, C. Stephanidis and A. C. Georgiou

Published in Collaboration with the
British Computer Society

©Copyright in this paper belongs to the author(s)

ISBN 3-540-76105-5

Modelling Interactions as Queues

Charalampos Karagiannidis

Adamantios Koumpis

Constantine Stephanidis

Institute of Computer Science, Foundation for Research and Technology - Hellas (ICS-FORTH)
Science and Technology Park of Crete, P.O. Box 1385, GR-71110 Heraklion, Crete, Greece
Tel: +30-81-391741, Fax: +30-81-391740, E-mail: {karagian, koumpis, cs}@ics.forth.gr

Andreas C. Georgiou

Department of Business Administration, University of Macedonia
GR-54006 Thessaloniki, Greece
Tel: +30-31-891569, Fax: +30-31-891282, E-mail: acg@macedonia.uom.gr

Abstract

Queueing Modelling is proposed in the domain of Intelligent User Interfaces and, more generally, in the field of Human-Computer Interaction. It is argued that Queueing Modelling can facilitate assessment of the current state and prediction of the future behaviour of user-computer interaction. This assessment can be used as the basis for adaptations either at design-time or at run-time, and can benefit the development and evaluation of adaptable or adaptive user interfaces. In order to illustrate our approach, we present specific queueing systems which model the sensory load of the user in the context of Intelligent Multimedia User Interfaces. Indicative adaptation policies that build on this analysis are also exemplified.

1 Introduction and Background

The proliferation of computer-based systems and applications in every walk of life and the anticipated widespread use of emerging telematic services has introduced new dimensions to the issue of human-computer interaction, necessitating the design of user interfaces accessible and usable by a diverse user population with different abilities, requirements and preferences. Thus, it has become increasingly important to design human-computer interfaces, which not only support more efficient and effective user interaction, but also address the individual end user expectations, while exhibiting intelligent and co-operative behaviour [1].

In this context, Intelligent User Interfaces (IUIs) are expected to play a crucial role. The development of such interfaces can provide solutions to all users by supporting user-adaptability, and can address the need for systems usable by users with diverse requirements, users with changing requirements, or in the context of changing environments, being intuitive to use, supporting nomadic uses, being flexible and interoperable, and empowering for the citizen [2].

Although several approaches have been reported in the relevant literature, e.g. [3, 4], currently there is no general consensus about the characteristics, behaviour and essential components of IUIs [3, 5]. The relevant literature acknowledges a number of issues concerning IUIs, such as recording and analysis of user interactions, drawing assumptions about user experiences, tasks, abilities, preferences and cultural backgrounds, representing these assumptions in an appropriate formal system, and suggesting or performing different kinds of adaptations to user requirements [4, 6].

In this paper, we take the view that the user interface needs to make several ‘communication decisions’ about the concepts it communicates. These decisions include *what* to communicate, *when* to communicate, and *how* to convey information effectively and efficiently. A user interface, in this context, is called *intelligent* in the measure to which it *adapts* itself, and makes communication decisions dynamically, at run-time, conditioned on an analysis of information regarding the current context of the interaction, which is defined by the state of the user profile, the task being performed, the application nature, etc [7]. Thus, adaptivity is considered as a major aspect of IUIs.

The issue of adaptivity can be addressed at different levels. At the highest level, two *independent* phases can be identified [8]:

- deciding on the need for adaptations, based on some sort of assessment of the current state of user-computer interaction and prediction of its future behaviour;
- performing adaptations through a certain ‘adaptation strategy’.

In this paper, we focus on the *assessment* of certain aspects that provide *evidence* for the *quality* of user-computer interaction.

Assessment of user-computer interaction is currently achieved through the use of models that capture the semantics of several fundamental aspects of the interaction, such as the user model, task model, application model, etc [5, 9]. Developing and maintaining such models is, however, challenging. These models are usually complicated, employing complex knowledge representation and inference techniques. They are not easily modifiable by the user interface designer. Moreover, most of these models are essentially application or domain dependent (e.g. task model depends strictly on the application), making their re-use to different applications impractical. An *application and domain independent* approach, that can be used *complementary* with existing ones, would, therefore, be very useful.

In the following, we are proposing such an approach, based on Queueing Modelling. It is argued that queueing systems can model user-computer interaction, thus provide evidence for the quality of interaction. After a brief introduction to Queueing Modelling, *indicative* models are presented, which model the user sensory load in the context of a multimedia / multimodal user interface. Also, indications are provided concerning the modifications of the adaptation strategy that can be based on these data. Similar models can be developed in order to describe other aspects of user-computer interaction, such as the utilisation of interface objects, I/O devices, etc.

2 Queueing Modelling in User-Computer Interaction

Several efforts have been reported in the literature, towards the employment of applied mathematical techniques in the domain of Software Engineering. For example, the use of Time Series Analysis in the analysis of user behaviour has been addressed in [10], and Markov Chain Analysis has been employed for the analysis of software specifications in [11]. Also, the analysis of user interfaces based on repetition of user sessions has been reported in [12].

Queueing Theory is widely used in many scientific disciplines, including several fields of Computer Science, in order to describe dynamic systems, assess their current state and predict their future behaviour [13, 14, 15]. In the following, we provide a brief introduction to basic notions of queueing theory concerning only the aspects that are of interest in the context of this paper. More detailed material can be found in [15].

Generally, a queueing system is described by the customer and server, the input process, the service mechanism, the discipline and the capacity of the queue. The main objective of analysing real systems through queueing modelling, is the *development of performance measures, in order to evaluate a system or predict its future behaviour, and adjust optimally various aspects concerning its functional abilities*.

Typical metrics that are provided by queueing modelling include the probability p_n of n customers being in the system, the probability of a customer finding at least k other customers waiting in the queue upon arrival, the average number of customers waiting in the queue to be served L_q , the average number of time spend in the queue W_q , etc.

The work reported in this paper is based on the notion that user-computer interaction can be modelled as a stochastic process. More specifically, the system is considered in a time interval, in which it can be described by a random variable $n(t)$ with realisations $t=0, 1, 2, \dots$. In the domain of user-computer interaction, this variable may denote the number of ‘information messages’ or ‘signals’ waiting or being served by the ‘human component’ of the system.

We shall make some assumptions relevant to the behaviour of this process that will lead to useful modelling situations. In particular, we make the convention that several ‘chunks’ of information, thereof called ‘messages’, exchanged to both directions (user-computer) are served by specific combinations of human sensory channels. Generally, these combinations may include all the human sensors, but according to the current state of User Interface Technology, they usually include:

- human eyes (e.g. eye gaze for input, VDU for output)
- human ears (e.g. microphone for input, speakers for output)
- human hands (e.g. mouse and keyboard for input, Braille display for output)

Therefore, it is assumed that information coming to the user is served *exclusively* by one of the following combinations of human sensory channels (Figure 1):

- eyes	queue ₁
- eyes and ears	queue ₂
- eyes and hands	queue ₃
- ears	queue ₄
- ears and hands	queue ₅
- hands	queue ₆
- eyes and ears and hands	queue ₇

Figure 1: User-computer interaction in a multimedia user interface modelled as a queueing system.

In the following, we shall attempt to accommodate a queueing modelling framework to the interaction between user and computer in a Multimedia User Interface (MMI), following the conventions stated above. The user is considered to be a single server, acting under single or multiple serving rates, i.e. processing times, according to customer types, i.e. messages requesting service from combinations of sensory channels. The utilisation of seven parallel single channels may seem, at first glance, inaccurate, given the fact that the user can not serve simultaneously by his ‘eyes’ and ‘ears’ combination and single ‘eyes’ combination. Perhaps, a service forbidding model could be employed further in this direction. On the other hand, we accept the postulation that the ‘eyes’ channel can be used with ‘ears’ or ‘hands’ simultaneously, to form combinations and therefore permit the ‘server’, i.e. sensory channel to operate with different rates.

Based on the above introductory analysis, we argue that an IMUI can monitor the interaction, identify/estimate the values of input and service parameters pertaining to each model employed, and calculate the metrics that are considered critical for every specific application. These metrics provide *evidence* for the quality of the user-computer interaction¹. In the light of these metrics, the user interface can adapt its media / modalities allocation policy, in order to meet the specific requirements of the application. Alternatively, these metrics can be calculated at design-time based on the specification of the user interface, and can be used for the evaluation of the user interface before it is actually implemented.

¹ Of course, there are several other aspects affecting user-computer interaction, such as the content of the messages, user’s physical abilities, etc. The metrics provided by queueing modelling, however, may provide evidence of the quality of interaction.

3 Example Models

In the following, we present indicative queueing models that can be employed in order to assess the load posed on user's sensory channels, in the context of a multimedia user interface. The first model assumes that all messages are of the same type; the second model assumes that each message belongs to one of seven different queues, which are served by the user with equal rates; the third model assumes that the seven different queues of messages are served by the user with different rates. It is noted that each of the models presented in this paper can be seen as a different model of the human as an information processor, intending to capture the load posed on user's sensory channels under different assumptions.

The calculation of the metrics of each model require only the estimation of the input and service rates of the messages. These rates can be determined from the dialogue specification of the user interface, or at run-time, based on monitoring of user-computer interaction. Thus, the results may be used for the assessment of the user interface either before or during its use. For each model, we do not present the underlying theory, but only the metrics that can be calculated from the model.

3.1 The M/M/1/GD model

The most fundamental queueing system is the one with a single service point, where inter-arrival and service times are exponential (that is, input and service rates are Poisson). In this case, the mean rates of input and service are $\tilde{e}_n = \tilde{e}$ and $\tilde{i}_n = \tilde{i}$ for all n . The discipline for this queueing system is usually FCFS, but the metrics provided here are valid for any discipline. This model can be seen as a model of the human as an information processor, where all the messages are treated in the same way. This is the simplest modelling framework of an IMUI under the queueing context (Figure 2).

Figure 2: The single channel, single service point model.

Based on Queueing Theory [15], several metrics can be calculated from the above model. Let us assume, for example, that for a given application, the input and output rates are: $\tilde{e}_n = \tilde{e} = 20$ and $\tilde{i}_n = \tilde{i} = 30$, with time unit one minute. This means that, on the average, 20 messages of any sensory combination appear in one minute, and the human component can respond adequately to 30 messages on average, per minute. In the light of the above model, the following metrics can be calculated (in steady state condition):

- *user's idle time:* $p_0 = 1/3$; that is, on average, 33% of time, all combinations of the user's sensory channels are idle; that is, the probability that a message of any kind will find the user idle is 33%;
- *probability that a message will find one message in the system:* $p_1 = 2/9$;
- *probability that an incoming message will find at least 1 message being in the system:* $\tilde{\pi}^1 = 2/3$;
- *probability that an incoming message will find a message being served and at least one in the queue:* $\tilde{\pi}^2 = 4/9$;
- *expected number of messages in the system:* $L_s = 2$; that is, on average, one message is being processed by the human component and one waits;

- *expected time that an information stays in the system: $W_s = 0.1$* ; thus, a message spends about six seconds in the system, from which $W_q = 2/30$, that is four seconds are spent in the ‘waiting line’;
- *expected length of a busy period: 6 seconds.*

3.2 The M/G/1/GD model

In this case, we assume that messages arrive according to a Poisson process with overall arrival rate $\tilde{e}$. The user serves the seven different types of information with a general service distribution. Let us assume that S denotes the random variable of service time, with mean $E\{S\}$ and variance σ_s^2 . In the case considered here, seven, not necessarily equal, rates of service are involved, namely $\hat{i}_i$, $i = 1, 2, \dots, 7$. These rates correspond to different types of messages or types of sensory channels. Thus, this may be considered as a model of the human as an information processor, where the messages are categorised into seven types, according to the sensory channel utilised (Figure 3).

The type of a message is not predictable in advance, but the system may determine by sampling that there exist probabilities q_i , $i = 1, 2, \dots, 7$ (e.g. limits of relative frequencies for each class of information), that an entering message belongs to a class i . It is clear that the partial rate of arrivals for each class is $\tilde{e}_i = q_i \tilde{e}$ (still Poisson processes). A special, but similar case, can be analysed, in which a single service rate $\hat{i}$ is utilised but the seven types of customers appear through the aforementioned probabilities.

To illustrate this approach, consider the following situation. Assume that the mean arrival rate is 5 messages per minute. Furthermore assume that $q_1 = 0.2$, $q_2 = 0.1$, $q_3 = 0.1$, $q_4 = 0.2$, $q_5 = 0.2$, $q_6 = 0.1$ and $q_7 = 0.1$. Also, assume that the corresponding service rates are $\hat{i}_1 = 10$, $\hat{i}_2 = 5$, $\hat{i}_3 = 5$, $\hat{i}_4 = 10$, $\hat{i}_5 = 10$, $\hat{i}_6 = 5$, $\hat{i}_7 = 5$ and therefore the expected partial times of service are 6, 12, 12, 6, 6, 12 and 12 seconds respectively (again, this may be determined either at design or at run-time.). Then the expected service time is $E\{S\} = 8.4$ seconds, which gives *0.14 minutes*, that is a service rate of approximately *7.14 messages per minute*. From the above, we conclude that the overall utilisation factor for the user is $\tilde{n} = 0.7$. Also, *the probability that an information of any type will find the user idle is 0.3 that is 30%. The expected length of a busy period of the user is approximately 28 seconds.*

Figure 3: A model with Poisson arrival and general service with distribution probabilities.

3.3 The Mi/Gi/1/NPRP model

This approach assumes that seven parallel queues are formed by seven different types of customers-messages. The messages arrive according to a Poisson process to the single general distribution channel. Upon arrival, each message is considered to be a member of one of the seven priority classes $i = 1, 2, \dots, 7$ (the number refers to combination of sensory channels), according to the adaptation strategy employed by the IMUI, and therefore each queue includes the messages belonging to a certain priority class. The priority classes are numbered, so that the smaller the value of the associate index the higher the priority. The messages from group *1* are preferred over messages from all other groups and so on. We assume that the Poisson arrivals of the *i-th* priority have mean rate $\tilde{e}_i$.

Priority service may follow either the preemptive or the non-preemptive rule. In the first case, the service (i.e. processing) of a lower priority message may be interrupted to commence service of an arriving information with higher priority. In the second case discussed here, if a message is being served and another message of higher priority arrives, the latter leaves the system only after the completion of its service. It is assumed that messages within a priority class are served on a FCFS basis. The service distributions for each class are general with mean service rate $E_i\{S\}$ and variance σ_{is}^2 .

The priority favouring of one class will definitely cause a decrease of mean service times of other classes. As it is pointed out in [17] *the classes with the faster service rates should be given the highest priorities if the optimality criterion for such an interaction system is the reduction of messages waiting for processing* (the shortest processing time rule, SPT).

In fact, if always higher priority classes of messages have faster service (processing) rates, then the average wait over all classes given previously is less than that of a model with a single no priority channel, with a mean service rate emerging as the weighted average among the mean service rates of the classes. Therefore, a simple optimisation rule for the user interface regulated by this model is to *force a priority discipline being applied by the user, estimate mean rates for each priority class, and reconsider the priorities according to these rates*. In this case, the model may be time-dependent, in the sense that increases or decreases in service rates of specific sensory combinations, may affect accordingly the priority rule in order to preserve low average queue lengths, and therefore forward greater amount of information towards the user. A graphical representation of this model appears in Figure 4.

Figure 4: Poisson arrivals, general or exponential service, with priorities.

To illustrate the above modelling framework let us consider an example. Assume that information arrives according to seven classes emerging from the combinations of sensory channels. The user interface monitors the system, estimates mean rates of service, and designates priorities to the classes according to these rates.

Class ‘eyes’, being the first one, is favoured with the higher priority. Although ‘eyes’ messages are processed prior to all other classes, ears are processed prior to hands, and so on, once a message of any priority enters service, the process must be completed before a new information of any kind begins service.

Policy 1

Let the mean arrival rates for each class be

$$\bar{e}_1 = 3, \bar{e}_2 = 3, \bar{e}_3 = 6, \bar{e}_4 = 3, \bar{e}_5 = 3, \bar{e}_6 = 1, \bar{e}_7 = 1$$

where all rates are messages per minute. The overall arrival rate is the sum of the class rates resulting to $\bar{e} = 20$. For simplicity assume that the user processes messages exponentially and the mean rates are

$$i_1 = 20, i_2 = 30, i_3 = 60, i_4 = 15, i_5 = 12, i_6 = 10, i_7 = 20$$

where all rates are mean number of messages processed per minute. Therefore, a message of the class ‘eyes’ needs 3 seconds of service, a message of class ‘ears’ needs 2 seconds, a message of the class ‘hands’ needs 1 second, and so on.

The highest priority class, that is a message of type ‘eyes’, stays in the system (user-computer) for about 7 seconds, while the lowest priority messages remain on the average more than 11 minutes. The overall length of queue is $L_q = 18.585$ whereas the weighted average waiting time is 55.75 sec.

Policy 2

If the IMUI recognises that priorities are no longer in accordance with the faster service rates, then it may ‘adjust’ its adaptation strategy. Assume that rates are rearranged, so that the highest priorities are designated to the fastest service rates. That is

$$\check{e}_1 = 6, \check{e}_2 = 3, \check{e}_3 = 3, \check{e}_4 = 1, \check{e}_5 = 3, \check{e}_6 = 3, \check{e}_7 = 1$$

and the corresponding service rates are

$$\check{i}_1 = 60, \check{i}_2 = 30, \check{i}_3 = 20, \check{i}_4 = 20, \check{i}_5 = 15, \check{i}_6 = 12, \check{i}_7 = 10$$

Then the overall length of queue is 12.72, whereas the weighted average waiting time is 38.16 sec. As we can see *the adjustment of the adaptation strategy with respect to the priority classes, decreased the overall queue length more than 31%, while the partial waiting mean times as well as the overall weighted time were reduced. The overall mean weighted time was also reduced according to the same ratio.*

If we employ a single queue, single server model with a weighted rate of service taken over all priority classes, the shortest processing time rule will still provide better performances of waiting times and sizes than this model. Mixed priorities and mean processing times by the human component will provide results that heavily depend on the actual values of the rates (for detailed discussion on priorities see [17]).

The discussion of the model assumes nothing on the service rates or times distribution (which for simplicity were taken to be exponential). We can investigate further various configurations using different distributions, or apply simulation techniques to estimate reasonable empirical distributions for the service rates (as well as the arrival rates).

This model could be considered as a first approach to a server vacation case. If some sensory channels are blocked or generally interrupted, then the service rate falls near zero. Then the user interface could continue sending information messages to the blocked priority queue for some time, but as soon as a recent estimation of the rates takes place, the priority policy would be reconsidered, and other types of information would be given higher priorities. For example, if the user stops processing the ‘eyes’ channel, then the user interface could designate low priority to these messages and consider increasing the priority of a class which seems to be still functional. Such modelling situations which take advantage of the concept of vacations are discussed in [16].

4 Discussion and Conclusions

We have proposed the employment of Queueing Modelling for the description, analysis, assessment and prediction of certain issues affecting the quality of user-computer interaction. We have presented concrete queueing systems that model the load posed on user’s sensory channels, in the context of a multimedia user interface. The proposed approach may be used for adaptations either at design-time or at run-time, and address the development and evaluation of adaptable or adaptive user interface development.

One critical factor affecting the proposed approach is the estimation and validation of the data pertaining to each model, namely the arrival and service rates of messages. This estimation may take place either at design-time or at run-time. In the first case, the rates may be estimated by the specification of the interface, either through an ‘uninformed’ approach, i.e. the probability of each event in the interaction is considered identical, or through an ‘informed’ approach, i.e. through actual user trials, or simulation [11]. In the case that the data are estimated at run-time, it is presumed that the user interface has the capability of monitoring certain aspects of the interaction, and calculate the required data through statistical sampling. Also, the user interface is assumed to have the capability of identifying when a message has ‘arrived’ to or has been ‘served’ by a user’s sensory channel. For example, we can assume that a visual message has been served when it has been discarded by the user, or when it is put at the background; alternatively, the service time can be estimated based on the data contained in it, e.g. the number of characters of a text message.

Another important factor affecting the proposed approach, especially if it is used at run-time, is the computational effort required for the calculation of the metrics for each model. Like any other modelling situation,

there is a trade-off between the accuracy of the model, and the computational effort required for the calculation of the corresponding metrics. For example, the models presented in this paper could be enhanced, by taking into account the content of each message. That is, assuming that each server, i.e. sensory channel, is serving several queues of messages, each one containing messages of the same content type. This would make the model more accurate, but also more computationally expensive. Thus, the complexity of the model to be employed depends on the nature of the specific application.

5 References

1. Stephanidis C. Towards User Interfaces for All: Some Critical Issues. In HCI International '95, Special Session on "User Interfaces for All: Everybody, Everywhere and Anytime", 1995.
2. Dieterich H, Malinowski U, Kuehme T, Schneider-Hufschmidt M. State of the Art in Adaptive User Interfaces. In Schneider-Hufschmidt M, Kuehme T, Malinowski U (eds) Adaptive User Interfaces. Elsevier, 1993, pp 13-48.
3. Sullivan J, Tyler S (eds). Intelligent User Interfaces. ACM Press, 1991.
4. Maybury M (ed). Intelligent Multimedia Interfaces. AAAI Press, 1993.
5. Puerta A. The Study of Models of Intelligent Interfaces. In 1993 International Workshop on Intelligent User Interfaces, pp 71-78.
6. Edmonds E. The future of Intelligent Interfaces: Not just 'How?', but 'What?' and 'Why?'. In 1993 International Workshop on Intelligent User Interfaces, pp 13-17.
7. Szekely P. Structuring Programs to support Intelligent Interfaces. In Sullivan J, Tyler S (eds). Intelligent User Interfaces. ACM Press, 1991, pp 445-464.
8. Karagiannidis C, Koumpis A, Stephanidis C. Intelligent User Interfaces: Deciding 'What', 'When', 'Why' and 'How' to Adapt. In review.
9. Arens Y, Hovy E, Vossers M. On the Knowledge Underlying Multimedia Presentations. In Maybury M (ed). Intelligent Multimedia Interfaces. AAAI Press, 1993, pp 280-306.
10. Dix A, Finlay J, Beale R. Analysis of User Behaviour as Time Series. In People and Computers VII, Cambridge University Press, 1992, pp 429-444.
11. Whittaker J. Markov Chain Techniques for Software Testing and Reliability Analysis. PhD Thesis, University of Tennessee at Knoxville, 1992.
12. Siochi A, Ehrich R. Computer Analysis of User Interfaces Based on Repetition in Transcripts of User Sessions. In ACM Transactions on Information Systems, 9(4), 1991, pp 309-335.
13. Eager D, Lazowska E, Zahorjan J. Adaptive Load Sharing in Homogeneous Distributed Systems. In IEEE Transactions on Software Engineering, 12(5), 1986, pp 662-675.
14. Mitrani I, Wright P. Routing in the presence of breakdowns. In International Journal of Performance Evaluation, 20, 1994, pp 151-164.
15. Kleinrock L. Queueing Systems, Volume 1: Theory, Volume 2: Computer Applications. John Wiley & Sons, 1975-1976.
16. Karagiannidis C, Koumpis A, Stephanidis C, Georgiou A.C. Assessing User-Computer Interaction through Queueing Modelling. In review.
17. Gross D, Harris C. Fundamentals of Queueing Theory. John Wiley & Sons, 1985.
18. Moray N (ed). Mental Workload: Its Theory and Measurement, Part III: Mathematical Models and Mental Workload. Plenum Press, pp 253-286, 1979.
19. Life MA, Narborough-Hall CS, Hamilton WI. (eds). Simulation and the User Interface. Taylor & Francis, 1990.