CHAPTER 1

INTRODUCTION

A modified version of this chapter has been published in the journal QUEST, Vol. 1, No. 4, 2005.

1.1 INTRODUCTION

A likely impossibility is always preferable to an unconvincing possibility. (Aristotle [1])

1.1.1 A General Description of the proposed Research

Congestion is ***ubiquitous*** in all domains of human endeavour. From a static point of view this implies the presence of congestion everywhere or in several places simultaneously in the broadest sense. In the dynamic sense ***ubiquitous*** implies variable pervasiveness of congestion with the passage of time.

A good understanding of the relationship between congestion and delay is essential in the design of mathematical congestion control models. In this respect Queueing Theory provides many tools needed for the analysis of Systems of Congestion.

Mathematically speaking, Systems of Congestion appear in many diverse and complicated guises which vary in extent and complexity. They often defy modelling efforts via discrete and continuous variables especially where the dynamics of a system must be *adequately* described, manipulated and controlled.

Congruently the term congestion also suggests that chaotic and disorderly (tumultuous) system conditions can be regarded as synonymous with congestion.

Over the past century a great number of publications have appeared which deal with the evolving field of Queueing Theory (Gross and Harris [2]). It is often so that many of the mathematical models fall short of being useful in applications-oriented practice as a result of mathematical complexity and the inability to deal with the dynamic (transient) operation of complex Systems of Congestion.

This thesis demonstrates the creation of an **eclectic collection** of models of system congestion and their efficacy in dealing with **the static and dynamic operation** of selected systems. These models are basically applications of probability theory and stochastic processes. The difficulty of using queueing models in practice are closely linked to:

- creating a representation of the queueing system by a mathematical model, and
- the flexibility of the mathematical model (Taha [3])

The thesis attempts to narrow the gap between theorists and practitioners by studying closed form functional representations and numerical approximations of the statics and dynamics of Systems of Congestion.

This goal is based on the premise that no other study in the field of Operations Research has displayed greater divergence between theoretical developments and applications. Therefore effort is needed to demonstrate the robustness of simple models which provide credible approximate solutions to complex design and operation problems. In this context a robust model is one which provides useful results even though the system being analysed may disregard the natural assumptions which are made when constructing the model. This may not mean that a new research frontier is being explored from the viewpoint of a theoretical mathematician and probabalist. It however remains imperative that the analysis of realistic Systems of Congestion be carried out by focussing on real physical problems (Taha [3]).

1.1.2 Exploring novel approaches to the modelling of Congestion.

Significant advances in the establishment of Chaos Theory over the past two decades suggest that it could be considered as a source of mathematical assistance in the modelling of Congestion.

The difficulty of applying queueing theory in practice is inter alia related to:

- modelling Systems of Congestion which are populated by intelligent entities,
- obtaining useful analytic results for certain mathematical models, and
- estimating system parameters.

Consequently the conjecture may be put forward that application of the fundamentals of Chaos Theory to congested systems via

- the use of computational approximations, and
- the use of model approximations based on the testing of model flexibility, holds promising potential.

Analysis in adapting prototype models to novel situations requires skills which are problem oriented in respect of transient operation and steady state operation. The worth of employing methods based on Chaos Theory will be measured by their usefulness in solving real Systems of Congestion rather than by way of mathematical elegance (Grosh [4]).

1.2 LITERATURE STUDY

1.2.1 Queues

The literature study on queues contains references to the definition and historical perspective of the modelling approach. Queueing theory has been a well-researched topic for many years and much published information is available. The literature study will give an overview sample of Queueing Theory. An important source of information is by Giffin [5].

1.2.1.1 Description of Queues

Queues are not an unfamiliar phenomenon. To define a queue requires specification of certain characteristics which describe the system:

An input process: This may be the arrival of an entity at a service location. The process may involve a degree of uncertainty concerning the exact arrival times and the number of entities arriving. To describe such a process the important attributes are the source of the arrivals, the size of each arrival, the grouping of such an arrival and the inter-arrival times.

A service mechanism: This may be any kind of service operation which processes arriving entities. The major features which must be specified are the number of servers and the duration of the service.

The queue discipline: It defines the rules of how the arrivals behave before service occurs.

The queue capacity: Finite or infinite

Examples of input and output processes which are as follows:

Table 1. Examples of Queueing systems

Situation	Input Process	Output Process
Bank	Entities arriving	Tellers serve entities
Toll Plaza	Vehicles arriving	Toll money is paid
Call Centre	Incoming Call	Call dealt with
Ferris wheel	Tourists arrive	Tourists are served
Intermittent Service Channel	Entities arrive	Entities are served intermittently
Naval Harbour	Ships that must unload	Unloading of ships

The presence of uncertainty makes these systems challenging in respect of analysis and design. The input rate/arrival rate together with the output rate/service rate mostly determine whether there are entities in the queue or not. These factors also determine the length of the queue.

In practice the arrival rate may be measured as the number of arrivals during a given period. The service rate can be measured in the same way. This is usually done for a system that has progressed from a transient state to a steady state. Most of these systems are described by arrival and service rates. It is however important to also focus on the transient characteristics of the system.

1.2.1.1 Historical perspective

The ground work for many of the earliest techniques of analysis in queueing theory was laid by A K Erlang, father of queueing theory, between 1909 and 1929. He is given credit for introducing the Poisson process to congestion theory, for the method of creating balance state equations (Chapman-Kolmogorov equations) to mathematically represent the notion of statistical equilibrium. Pollaczek [6] began studying non-equilibrium queueing systems by looking at finite intervals. However, the first truly time dependant solutions were not offered until Bailey [7] using generating functions and Lederman and Reuter[8], using spectral theory and Champernowne [9], using the combinatorial method found such solutions. Kendall [10] introduced his method of imbedded Markov Chains in analyzing non-Markovian queues. The important technique, known as the supplementary variable technique was introduced by Cox [11] and this method has been extensively used in the thesis.

Most of the pioneers of Queueing Theory were engineers seeking solutions to practical real world problems. The worth of queueing analysis was judged on model usefulness in solving problems rather than on the theoretical elegance of the proofs used to establish their logical consistency.

The trend toward the analytical investigation of the basic stochastic processes associated with queueing systems has continued up to the present time. Others associated with time dependant solutions and Markovian analysis are Bailey [7], Bhat [12], Cox [11], Kendall [10], Keilson and Kooharian [13], and Takacs [14].

<u>The focus of the non-research oriented engineer in this expansive theoretical development was on techniques which demonstrated applications of the results of the theory</u>. In Operations Research the only field that has few theoretical models with any useful applications is Queueing Theory. One may speculate why this has occurred. The commonly mentioned reason is that the equations resulting from many theoretical investigations are simply too overly complex to apply. The practitioner then often has to resort to simulation methods for analysis.

In practice common simple queues are scarce. Arrival and service rates may be constantly shifting over time so it is important to describe the distributions as functions of time. These systems are contrasted with steady-state solutions in which the arrival and service patterns are usually such that the state probability distribution is stationary. Dynamic systems require robust modelling that can provide useful results even though the analysis may violate assumptions used in constructing the model.

Most of the above discussion relates to what Bhat [12] refers to as behaviour problems of the system. The focus is to use mathematical models to seek understanding of a particular process. Other problems are statistical and operational. "Statistical" refers to analysis of empirical data, estimation of system characteristics and tests of hypotheses regarding queueing processes. "Operational" refers to design, testing and control of real life problems. All such problems have been addressed in this thesis.

1.2.1.3 Review of Queueing Models and their Modelling Approaches

The dynamics of queues has been analysed by using steady-state mathematics. Such queueing processes are described by using the Kendall-Lee notation which uses mnemonic characters that specify the queueing system:

A/B/C/D/E/F

A. Specifies the nature of the arrival process.

B. Specifies the nature of the service times.

C. Specifies the number of parallel servers

D. Specifies the queue discipline.

E. Specifies the maximum number of entities in the system.

F. Specifies the size of the population from which entities are drawn.

This notation is commonly used when deriving expressions for the average system length, number of entities in the queue, the average waiting time, and many other features.

For queueing models, entity arrivals and service times are summarized in terms of probability distributions normally referred to as arrival and service time distributions. These distributions may represent situations where entities arrive and are served individually (e.g. banks, supermarkets). In other situations, entities may arrive and/or be served in groups. (e.g. restaurants). The latter case is normally referred to as a bulk queue. A Poisson stream of entities arriving in groups is served at a counter in batches of varying size under the general rule for bulk service in which the server remains idle until the queue size reaches or exceeds a fixed number whereupon they are served. This system has been discussed by Borthakur [15].

Continuing in this vein example of several systems which differ widely may be described. In a queueing system the server can also offer two kinds of service in which entities arrive in batches of variable size. Just before service starts an entity chooses only one of the two service types. Such an M/M/1 queueing system has been studied by Madan [16]. In the same system, the first service is essential and the second optional service is offered in batches. This has been studied by Madan [17, 18] has also analysed an M/G/1 queueing system in which two services, one essential and the other optional, are offered where there is no waiting capacity. Sapna [19] has discussed an M/G/1 queueing system with non-perfect servers where there is no waiting capacity. Once the system becomes empty, the service is discontinued for a random length of time. When the service facility becomes ready to continue providing service and entities are waiting the service starts by serving the first entity in the queue. Otherwise service is again suspended and so on.

Queues with service interruption are found to exhibit a very interesting property called the Stochastic decomposition property (Fuhrmann [20]) i.e. the stationary number of entities present in the system at a random point in time is distributed as the sum of two or more independent random variables one of which is the stationary number of entities present in the standard M/G/1 queue (i.e. the server is always available) at a random point in time.

Doshi [21] has given a survey of queueing systems with interruptions and Levi and Yechiali [22] have discussed an M/M/s system with interruptions. Fuhrmann [20] considered an M/G/1 queueing system in which the server undergoes an interruption of random length each time the system becomes empty. He gives an intuitive explanation for these results, while simultaneously providing a more simple and elegant method of solution to show that the number of entities present in the system at a random point in time is distributed as the sum of two independent variables: (i) the number of Poisson arrivals during a time interval that is distributed as the forward recurrence time of an interruption, and (ii) the number of entities present in the corresponding standard M/G/1 system.

For the same system Fuhrmann and Cooper [23] have obtained two results that can lead to remarkable simplification when solving complex M/G/1 models. Shanthikumar [24] gives mechanistic (analytic) proof of the result which is more general than discussed by Fuhrmann and Cooper [23]. Keilson and Servi [25] have discussed the distributional form of Little's Law and the Fuhrmann and Cooper [23] decomposition. Keilson and Ramaswamy [26], Latouche and Ramaswami [27] have studied an M/G/1 queueing system in which the server attends interactively to secondary tasks upon primary service completion epochs when the primary queue is exhausted. Using the state space methods and simple renewal-theoretic tool they have obtained the ergodic distribution of the depletion time.

Levy and Kleinrock [28] have analysed both a queueing system that incurs a start up delay whenever an idle period ends and one in which the server undergoes interruption periods. They have seen that the delay distribution in the queue with a start up delay is composed of the direct sum of two independent variables. Leung [29] has shown that the customer waiting time in the system is distributed as the sum of the waiting time in a regular M/G/1 queue with interruptions and the additional delay due to interruptions which is a stochastic decomposition property. He has also derived a general formula for the additional delay.

A further example is when a single station provides service to customers who arrive in a Poisson stream with a constant intensity. All the customers require an equal and fixed amount of service but the service rate of the station varies randomly. The service station itself is subject to random breakdowns rendering it inoperative for random periods of time during which repair takes place. Maintenance centre allow a queue of breakdowns. The breakdowns are cleared, or a queue of breakdowns is not permissible, and the units will not be served unless the failure is repaired. In these two types White and Christie [30] have obtained queue length generating functions when the arrival processes are exponential. Using the inclusion of the supplementary variable technique Jaiswal [31] has obtained a solution for the first type of breakdown in which the repair and service time follow a general distribution. Heathcote [32] has obtained a solution if the arrival of breakdowns is restricted by imposing the condition that a breakdown cannot occur if there is no unit the system. Thiruvengadam [33] has obtained the time dependant and steady state queue length generating functions for a single server queueing process in which the service facility is subject to breakdowns as a pre-emptive resume priority.

To sum up, many conventional and classic models in queueing theory form the backdrop to this thesis. The foregoing examples of queueing systems have been selected to display typical complexities of systems to be modelled and the attendant difficulties of finding expressions for transient state and steady state operation.

It is clear that steady-state analysis is suited to certain design problems but only gives averages and no indication of how the queue characteristics vary with the passage of time.

The goal is to develop an expression/model which gives a measure of the number of entities present in the queueing system at certain instants in time and some measure of the delay experienced by entities passing through the system. It is the uncertainty which is present in most real systems which makes model building a challenging task.

Moving one step beyond the description of simple independent trials, the most widely researched and easily manipulated stochastic models are associated with Markov processes. A Markov process is one which exhibits one stage dependence; the probability distribution for future systems states can be developed from knowledge of the existing state distribution without regard to any other past history.

The study of continuous-time processes begins with the development of the general birth-and-death equations. These equations follow directly from a Markov chain discussion. In a queue a birth is normally seen as an arrival and a death as a departure from the system. The result will be a set of extremely versatile differential-difference equations which serve as the basis for many models by simply varying the state-dependant birth-and-death process. The chapters dealing with service interruption will use this process.

1.2.1.4 Confidence limits

To obtain confidence limits for the waiting time in the steady state (see chapter 2), one needs to use the following Multivariate Central Limit theorem (Rao [42]).

Suppose

$Y_1^*, Y_2^*, ..., Y_n^*$ are independent and identically distributed K-dimensional random variables such that,

$Y_n^* = (Y_{1n}, Y_{2n}, ..., Y_{Kn})$; n = 1,2,3,…

Having first and second order moments

$E(Y_n) = \mu$; $D(Y_n) = \sum$,

Define the sequence of random variables

$\overline{Y_n} = (\overline{Y_{1n}}, \overline{Y_{2n}}, ..., \overline{Y_{Kn}})$; n = 1,2,3,…

Where

$\overline{Y_n} = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$; i=1,2,…,K and j=1,2,…,n

Then

$\sqrt{n}\left[\overline{Y_n} - \mu\right] \xrightarrow{d} N_K(0, \sum\ )$ as $n \longrightarrow \infty$.

1.2.2 Chaos theory

A literature study on Chaos gives an introduction and serves to define the terminology of Chaos Theory. It is dealt with via definition, historical perspective and modelling approach. This study gives an overview of most important aspects of the theory used in this thesis.

1.2.2.1 Historical Perspective

The main focus of Chaos Theory is on Dynamical Systems, the branch of mathematics that attempts to understand the time behaviour of processes. This occurs in many fields such as the motion of the stars and galaxies which constitute a vast and incomprehensible dynamical system, the vagaries of the stock market, the changes that chemicals undergo, the rise and fall of populations, the motion of a simple pendulum and certain queue behaviour.

One of the remarkable discoveries of twentieth-century mathematics is that very simple systems, even systems depending on one variable, may behave just as unpredictably as the stock market, a turbulent waterfall, or a violent hurricane. Mathematicians have called the reason for this unpredictable behaviour, “CHAOS”.

Isaac Newton was one of the pioneers of dynamical systems when differential equations became the principal mathematical technique for describing processes that evolve continuously in time. In the 18th and 19th centuries, mathematicians devised numerous techniques to solve differential equations explicitly such as Laplace transforms, power series, variation parameters, linear algebra and many others. These techniques seldom succeed in solving nonlinear functions. Unfortunately many of the most important congestion processes are nonlinear.

There were four major landmarks in the past centuries that have revolutionised the way dynamical systems are viewed (Devaney [34]):

1. Henri Poincaré’s research in 1890 came close to solving the n-body problem that deals with the stability of the solar system. It dealt with the possible behaviour of the system and this was more important than an exact solution. It eventually concluded that stable and unstable manifolds might not match. When finally admitting this possibility, Poincaré saw that this would cause solutions to behave in a more complicated fashion than previously imagined. He had discovered Chaos Theory.
2. There were two notable exceptions that added to Poincarés work and results. They were the French mathematicians Pierre Fatou and Gaston Julia. In the 1920’s they found that the Julia set maps the dynamics of complex analytics. But the lack of computer power attenuated their work.
3. In the 1960’s Stephen Smale reconsidered Poincaré crossing stable and unstable manifolds with the aid of iteration. This meant that he could prove that the chaos, which his predecessors had uncovered, was real and could be analysed. The technique he used is named Symbolic Dynamics. The American meteorologist E.N. Lorenz used a very crude computer and

discovered that very simple differential equations could exhibit the type of chaos Poincaré had observed. He also realized that sensitive dependence on initial conditions was of paramount importance. A flurry of activity in the 1970's was led by contributions by Robert May, Mitchell Feigenbaum, Harry Swinney, Jerry Gollub, John Guckenheimer and Robert Williams.

4. The availability and speed of the modern computer made it possible to obtain a better understanding of a dynamical system. The foremost was Mandelbrot's discoveries of 1980. He discovered graphics that sparked renewed interest in the Julia set.

1.2.2.2 Modelling Approach

This research attempts to offer a set of tools from the field of DYNAMICAL SYSTEMS theory, which may be considered **as an alternative way of providing time-varying solutions** to flow problems encountered in Systems of Congestion.

The use of complex non-linear differential equations is the main mathematical technique for describing processes that evolve continuously in time. They describe processes that change smoothly over time; in the main they are analytically intractable.

Simpler types of equations — "difference equations", discrete in time, — may be used for processes that iteratively evolve from state to state (Gleick [35]).

A naïve approach to population growth was postulated as the classic Malthusian model of "unrestrained" growth (Malthus [36]). The Malthusian model evolved to what is known today as the *Logistic Model of Population Growth* (Verhulst [37]). Of parabolic nature, it is suited to modelling population flow systems in real life. It affords ease of computation, can be readily manipulated mathematically, and is suited to the iterative nature of step-by-step computation required by "difference equations".

Although it is deterministic in nature, it may be inferred that it is suitable for modelling dynamical systems which could exhibit chaotic characteristics. Stated in another way, even if it is of simple mathematical nature it has the ability to generate complex population dynamics that appear to be random, dynamics called ***CHAOS***.

The dictionary definitions of chaos are as follows:

(i) "The disordered formless matter supposed to have existed before the ordered universe."

(ii) "Complete disorder, utter confusion."

The Royal Society proposed *the following definition of chaos* in 1986:

(iii) "Stochastic behaviour occurring in a deterministic system."

This definition may be interpreted as "lawless behaviour governed entirely by law" (Stewart [38]).

To recapitulate, **an aim of the research is to study the transient behaviour of a dynamical system using mathematical features of *CHAOS theory***. Should the research lead to a fruitful result, to then advocate the use of the chosen chaos based model to support classical Queueing Theory models.

The research considers application of the Verhulst [37] *Logistic Model of Population Growth* (also known as the logistic parabola or *logistic mapping*) and other models to the chosen problem. The unadulterated version of the Verhulst [37] model in its "iteration" version (Schroeder [39]) is as follows:

$$x_{n+1} = F(x_n) = rx_n(1+ x_n) \qquad (1.1)$$

where: r is a constant between 0 and 4, and

x_n is the logistic map value at iteration *n.*

Fig. 1.1 shows a generated orbit. The range of the parameter *r* for values between 0 and 3 represent the *steady state regime* and is relatively uninteresting from the point of view of modelling a dynamical system. Values for the parameter *r* which lie in the region $3 < r < 3.5699$ are known to logistically map interesting orbiting dynamic characteristics where the population system being modelled may cycle between orbits of period length 2, 4, 6, 8, 16, 32, 64 and so forth (Schroeder [39]). As soon as a range of $3.5699 \leq r \leq 4$ is used, the logistics mapping of a system exists in the *"region of chaos"*.

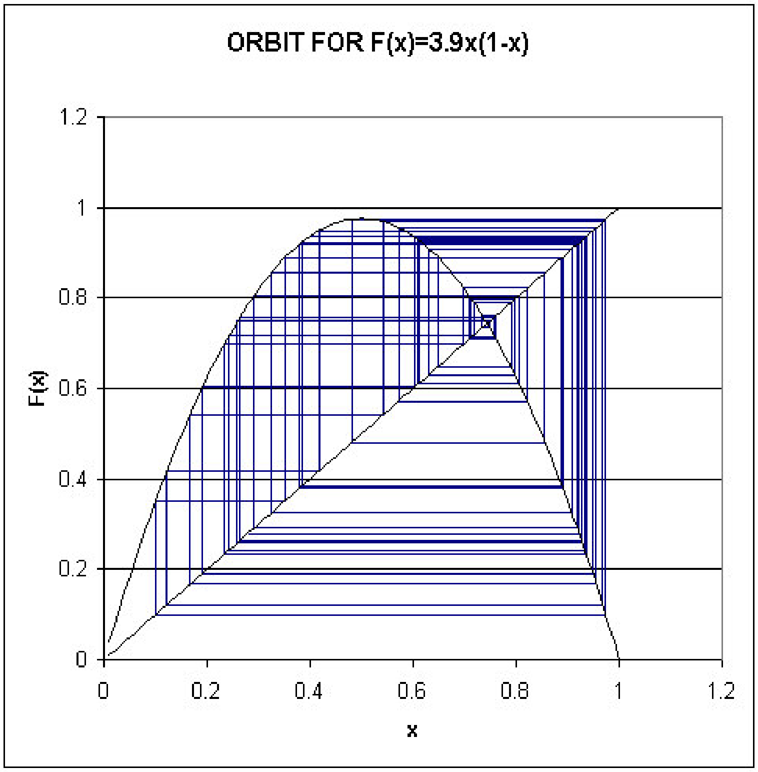

Fig. 1.1 Chaos generated orbit

When the question of why an equation such as (1.1) should be considered for purposes of logistic mapping, the reply is contained in statements such as "the details of the equation are beside the point. What matters is that the function should have a *hump*" (Stewart [38]).

Feigenbaum [40] proved that no matter which type of mapping is used such as logistic, polynomial, or trigonometric, — as long as the function is unimodal in the range of interest, simplistic iterative modelling methods are adequate.

1.2.3 The need for a new theory

It is clear that there may be an opportunity for developing a theory or applying current theories to achieve a better end result. In the study it is shown that Chaos Theory based models may be applied to describe the transient behaviour of a System of Congestion by using some or other form of logistic mapping. But the need exists to build a model that manages the system, by changing the resources to increase service levels. Such a model is shown in Fig. 1.2.2.

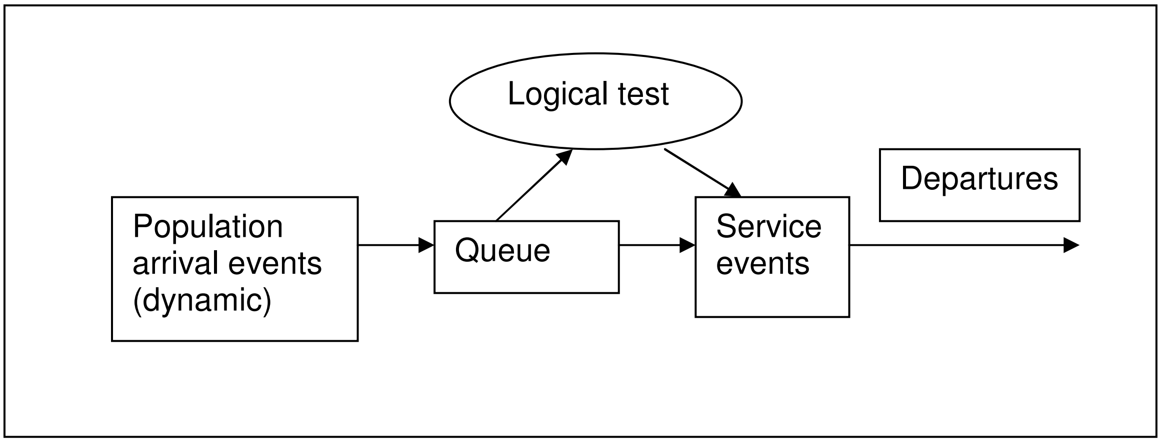

Fig 1.2.2 Proposed model for dynamical system

The proposed model firstly shows that arrivals are dynamic. The logistic mapping of Chaos will be considered to generate arrivals at the system. The r-value and a scaling factor will determine the effectiveness of generation of the arrivals. The service rate will be modelled in the same way.