

**Stochastic models of steady state and dynamic operation
of systems of congestion**

By

Gert Botha Erasmus

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor

(Industrial Engineering)

In the

Faculty of Engineering, Built Environment and Information Technology

At the

UNIVERSITY OF PRETORIA

Pretoria

June 2005

Acknowledgements

I would like to express my sincere thanks and appreciation to the following persons:

- a) Prof. Kris Adendorff and Prof. Sarma Yadavalli that acted as my supervisors and mentors, not only on academic level, but especially personally.
- b) My father and mother, I thank God each day for the good parents that I received. Thank you for providing the resources and motivation to complete my studies.
- c) Marnella, Andreas, H.P., Marelee, Martinus and Nicolaas for your devoted time and love that you have given me during my studies.
- d) All the personnel of the Department of Industrial and System Engineering, especially Anne-Marie and Christa for their motherly care and patience.
- e) To all my friends that have supported me mentally and physically during these studies.
- f) The University of Pretoria and all other personnel for providing me with the resources and research environment to complete my studies.

“Soli deo Gloria”

To the best two Mentors, Supervisors and friends in the whole world,
my Family that stood by me and all my colleagues.

Stochastic models of steady state and dynamic operation of systems of congestion

Gert Botha Erasmus

Supervisor: Prof. K. Adendorff

Co-supervisor: Prof. V.S.S. Yadavalli

Department of Industrial and Systems Engineering

Philosophiae Doctor (Industrial Engineering)

ABSTRACT

Key terms: Systems of Congestion, Queueing Theory, Chaos Theory, Steady state, Transient state, System Dynamics, Waiting time, Bulk system with interruption.

- (i) The thesis sets out to address the problematic phenomenon of Systems of Congestion via Basic Queueing Theory. The theory, and its application in practice, appears to be a field of study which is the common domain of “theorists” and “practitioners”.
- (ii) This professional dichotomy has come about due to diverging interests in that one group is mainly interested in the purity of mathematical modelling, and the other group is motivated to use modelling, which conveniently employs applications oriented solutions.
- (iii) The schism between the groups has been accentuated by the “practitioners” who in addition to having an interest in steady state system behaviour make use of methods of modelling of the transient operation of complex Systems of Congestion.
- (iv) At the outset the thesis demonstrates how closed form solutions are obtained for steady state and transient state operation of a selection of Systems of Congestion. The attendant mathematical derivations are elegant and intricate.

- (v) Having revealed the limited utility of closed-form solutions the thesis proceeds to investigate the feasibility of using dynamical systems theory to study the transient behaviour of complex Systems of Congestion.
- (vi) The creation of Chaos Theory in recent decades suggests that it may be employed as a useful tool in analysing Systems of Congestion. Iterative Chaos Theory methods of orbit generation for complete Systems of Congestion are therefore examined. The use of such orbit generation methods is found to be satisfactory for simple Systems of Congestion. More than a perfunctory knowledge of chaos mapping is however required. The simplicity of modelling is emphasized.
- (vii) Based on the results of benchmarking the creation of dynamic system orbits against an existing simulation method, the research advances to modelling of the transient operation of complex systems. Once again the iterative method of orbit generation displays the ease of modelling while simultaneously unfolding system dynamics graphically.
- (viii) One may hopefully contend that a tool of eminent utility has been developed to aid practitioners in studying and optimizing Systems of Congestion.

TABLE OF CONTENTS

CHAPTER 1	5
INTRODUCTION	5
1.1 INTRODUCTION	6
1.1.1 <i>A General Description of the proposed Research</i>	6
1.1.2 <i>Exploring novel approaches to the modelling of Congestion</i>	8
1.2 LITERATURE STUDY.....	9
1.2.1 <i>Queues</i>	9
1.2.1.1 <i>Description of Queues</i>	9
1.2.1.1.1 <i>Historical perspective</i>	10
1.2.1.1.3 <i>Review of Queueing Models and their Modelling Approaches</i>	12
1.2.1.1.4 <i>Confidence limits</i>	16
1.2.2 <i>Chaos theory</i>	17
1.2.2.1 <i>Historical Perspective</i>	17
1.2.2.2 <i>Modelling Approach</i>	19
1.2.3 <i>The need for a new theory</i>	22
CHAPTER 2	24
CONFIDENCE LIMITS FOR EXPECTED WAITING TIME OF TWO QUEUEING MODELS	24
2.1.1 <i>Introduction</i>	25
2.1.3 <i>The ML and CAN estimators for expected waiting time</i>	29
2.1.3.1 <i>The ML Estimator</i>	29
2.1.3.2 <i>An application of the multivariate central limit theorem</i>	30
2.1.3.3 <i>The CAN Estimator</i>	31
2.1.4 <i>Confidence limits for the expected waiting time</i>	32
2.2 STATISTICAL ANALYSIS FOR A TANDEM QUEUE WITH BLOCKING	36
2.2.1 <i>Introduction</i>	36
2.2.2 <i>System description and assumptions</i>	38
2.2.3 <i>Analysis of the system</i>	38
2.2.3.1 <i>Transient Solution</i>	40
2.2.3.2 <i>The Steady state solution</i>	41
2.2.3.3 <i>Expected service time per entity in the system</i>	42
2.2.4 <i>MLE and CAN estimator for the expected service time per entity in the system</i>	42
2.2.4.1 <i>The ML estimator</i>	42
2.2.4.2 <i>An application of the multivariate central limit theorem</i>	43
2.2.4.3 <i>The CAN Estimator</i>	44
2.2.4.4 <i>Confidence limits for the expected waiting time</i>	45

CHAPTER 3	47
A SINGLE CHANNEL QUEUEING MODEL WITH OPTIONAL SERVICE AND SERVICE INTERRUPTION.....	47
3.1 <i>Introduction.....</i>	48
3.2 <i>Model description.....</i>	48
3.3 <i>Assumptions and notation.....</i>	49
3.4 <i>Time dependant solution.....</i>	51
3.5 <i>Some special Cases.....</i>	55
3.6 <i>The steady state solution.....</i>	57
3.7 <i>Some special Cases.....</i>	59
3.8 <i>Concluding remark.....</i>	61
CHAPTER 4	62
AN M/M/1 QUEUEING SYSTEM WITH BATCH ARRIVALS OF VARYING SIZE, SERVICE OF FIXED BATCH SIZE AND TWO MODES OF FAILURE OF SERVICE FACILITY.....	62
4.1 <i>Introduction.....</i>	63
4.2 <i>Model description.....</i>	63
4.3 <i>Assumptions and notation.....</i>	64
4.4 <i>Equations describing the system.....</i>	65
4.5 <i>Time dependant solution.....</i>	66
4.6 <i>The steady state solution.....</i>	71
4.7 <i>Some special cases.....</i>	72
4.8 <i>Concluding remark.....</i>	75
CHAPTER 5	76
AN M/G/1 QUEUEING SYSTEM WITH TWO MODES OF FAILURE....	76
5.1 <i>Introduction.....</i>	77
5.2 <i>Model description.....</i>	77
5.3 <i>System description.....</i>	78
5.4 <i>Equations governing the system.....</i>	79
5.5 <i>Time dependant solution.....</i>	81
5.6 <i>Steady state solution</i>	87
5.7 <i>Some special cases.....</i>	88
5.8 <i>Concluding remarks.....</i>	89
CHAPTER 6	90
CHAOS THEORY BASED MODELS OF SIMPLE SYSTEMS OF CONGESTION.....	90
6.1 INTRODUCTION.....	91
6.1.1 <i>The classical Poisson arrival system.....</i>	91
6.1.1.1 <i>The general modelling approach.....</i>	91

6.1.2	<i>The classical exponential service system</i>	93
6.1.2.1	<i>The general modelling approach</i>	93
6.1.3	<i>The classical M/M/1 queue</i>	95
6.1.3.1	<i>The general modelling approach</i>	95
6.2	INTRODUCTION TO CHAOS GENERATION	99
6.2.1	<i>The Verhulst generated arrival system</i>	99
6.2.1.1	<i>The general modelling approach</i>	99
6.2.	<i>The Verhulst generated service system</i>	101
6.2.2.1	<i>The general modelling approach</i>	101
6.2.3	<i>The Verhulst generated single channel queue</i>	103
6.2.3.1	<i>The general modelling approach</i>	103
6.2.4	<i>Benchmarking the Verhulst generated single channel queue model</i>	105
6.2.5	<i>Extending the Verhulst generated single channel queue model to deal with variable traffic intensity</i>	108
6.3	FURTHER EXAMPLES OF CHAOS GENERATION	111
6.4	CONCLUDING REMARKS ON SINGLE CHANNEL ORBITS RESULTING FROM A MENU OF METHODS OF GENERATION.	119
CHAPTER 7		120
ANALYSIS OF THE DYNAMIC CHARACTERISTICS OF PRACTICAL SYSTEMS OF CONGESTION USING CHAOS GENERATION METHODS.....		120
7.1	INTRODUCTION	121
7.2	SYSTEM NO. 1	122
7.2.1	<i>System scenario</i>	122
7.2.2	<i>The system model</i>	122
7.2.3	<i>Diagnosis of the model results</i>	125
7.2.4	<i>Using realtime feedback to improve system performance</i>	132
7.2.5	<i>The effect of the size of system waiting area on system performance</i>	132
7.2.6	<i>Concluding comments on system no. 1</i>	135
7.3	SYSTEM NO. 2	136
7.3.1	<i>System scenario</i>	136
7.3.2	<i>The system model</i>	136
7.3.3	<i>Diagnosis of the model results</i>	142
7.3.4	<i>Using realtime feedback to improve system performance</i>	142
7.3.6	<i>Concluding comments on system no. 2</i>	144
7.4	SYSTEM NO. 3	145
7.4.1	<i>System scenario</i>	145
7.4.2	<i>The system model</i>	145
7.4.3	<i>Diagnosis of the model results</i>	146
7.4.4	<i>Using realtime feedback to improve system performance</i>	149
7.4.5	<i>Concluding comments on system no. 3</i>	151
7.5	SYSTEM NO. 4	152
7.5.1	<i>System scenario</i>	152
7.5.2	<i>The system model</i>	152

7.5.3	<i>Diagnosis of the model results</i>	157
7.5.4	<i>Using realtime feedback to improve system performance</i>	157
7.5.5	<i>Concluding comments on System No. 4</i>	159
7.6	EVALUATION OF THE MODELLING METHODS AND ACHIEVEMENT OF DYNAMIC OPERATION RESULTS OF COMPLEX SYSTEMS OF CONGESTION.....	160
CHAPTER 8	161
CONCLUSION	161
APPENDIX A	165
	FLOW DIAGRAM FOR THE DESIGN OF AN ARRIVALS/SERVICE ORBIT GENERATING FUNCTION	165
APPENDIX B	166
	FLOW DIAGRAM FOR THE DESIGN OF A SYSTEM ORBIT GENERATING FUNCTION	166
REFERENCES	167

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 probabilist. 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 Champenowne [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].

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. 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

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

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

Having first and second order moments

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

Define the sequence of random variables

$$\bar{Y}_n = (\bar{Y}_{1n}, \bar{Y}_{2n}, \dots, \bar{Y}_{Kn}); n = 1, 2, 3, \dots$$

Where

$$\bar{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_{ij}; i=1, 2, \dots, K \text{ and } j=1, 2, \dots, n$$

Then

$$\sqrt{n}[\bar{Y}_n - \mu] \xrightarrow{d} N_K(0, \sum) \text{ as } n \longrightarrow \infty .$$

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 Poincaré’s 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:

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

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"*.

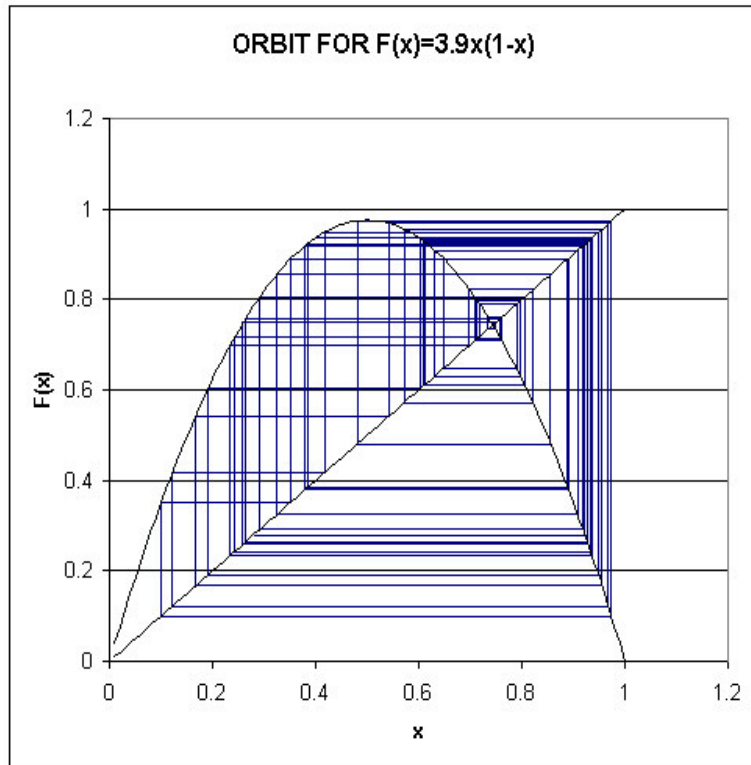


Fig. 1.1 Chaos generated orbit

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.

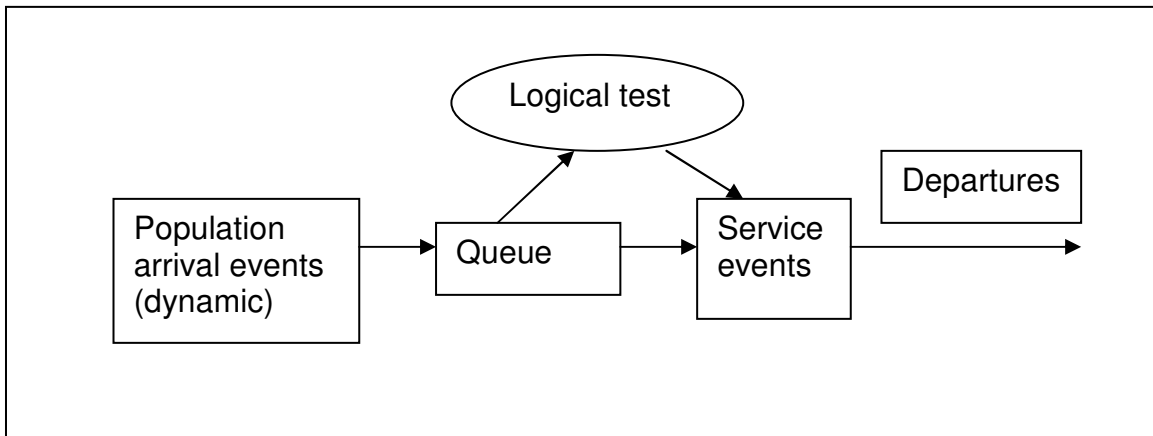


Fig 1.2.2 Proposed model for dynamical system

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.

CHAPTER 2

CONFIDENCE LIMITS FOR EXPECTED WAITING TIME OF TWO QUEUEING MODELS

A modified version of this chapter has been published in ORION, Vol. 20, No. 1, 2004

2.1.1 Introduction

Once one is armed with a queueing model of a system, one which is described by equations which emulate the relevant birth-death process, parametric estimation is one of the essential tools to understand the random phenomena using stochastic models. Whenever systems are fully observable in terms of their basic random components such as inter arrival times and service times, standard parametric estimation techniques of statistical theory are quite appropriate. Most of the studies of several queueing models are confined to only obtaining expressions for transient or stationary (steady state) solutions and do not consider the associated inference problems. Recently, Bhat [41] has provided an overview of methods available for estimation, when the information is restricted to the number of entities in the system at certain discrete points in time. Narayan Bhat has also described how maximum likelihood estimation (MLE) is applied directly to the underlying Markov chain in the queue length process in $M|G|1$ and $GI|M|1$. An attempt is made in this chapter to obtain MLE, a consistent asymptotically normal estimator (CAN) and asymptotic confidence limits for the expected waiting time per entity in $M|M|1|_{\infty}$ and $M|M|1|N$ queues. These two models and the expected waiting time per entity for each model are explained briefly.

2.1.2 Description of Systems

Model I The $(M|M|I):(FCFS|\infty|\infty)$ queue

It can be readily seen that (Taha [3]) the difference-differential equations governing $M|M|1$ are given by

$$p'_n(t) = \lambda p_{n-1}(t) - (\lambda + \mu)p_n(t) + \mu p_{n+1}(t), \quad n = 1, 2, 3, \dots \quad (2.1.1)$$

$$p'_0(t) = -\lambda p_0(t) + \mu p_1(t), \quad n = 0 \quad (2.1.2)$$

As $t \rightarrow \infty$, the steady state solution can be proved to exist, when $\lambda < \mu$. Assuming that $p'_n(t) \rightarrow 0$ and $p_n(t) \rightarrow p_n$ as $t \rightarrow \infty$, for $n = 0, 1, 2, \dots$, it is clear that

$$-\lambda p_0 + \mu p_1 = 0, \quad n = 0 \quad (2.1.3)$$

$$\lambda p_{n-1} - (\lambda + \mu)p_n + \mu p_{n+1} = 0, \quad n = 1, 2, 3, \dots \quad (2.1.4)$$

Solving these difference-differential equations,

$$p_n = (1 - \rho)\rho^n, \quad n = 0, 1, 2, \dots \quad (2.1.5)$$

where $\rho = \frac{\lambda}{\mu} < 1$.

Clearly (2.1.5) corresponds to the probability mass function of the Geometric distribution. The expected waiting time per entity in the queue is given by

$${}_1W_\rho = \frac{\lambda}{\mu(\mu - \lambda)}. \quad (2.1.6)$$

Model II The $(M|M|1):(GD|N|\infty)$ queue

The model is essentially the same as Model I, except that the maximum number of entities in the system is limited to N (maximum queue length is $N-1$) (Taha [3]).

The steady state equations for the model are given by

$$-\rho p_0 + p_1 = 0, \quad n = 0 \quad (2.1.7)$$

$$\rho p_{n-1} - (\rho + 1)p_n + p_{n+1} = 0, \quad n = 1, 2, 3, \dots, N - 1 \quad (2.1.8)$$

$$\rho p_{N-1} - p_N = 0, \quad n = N \quad (2.1.9)$$

The solution of the above difference-differential equations is given by

$$p_n = \frac{(1 - \rho)}{(1 - \rho^{N+1})} \rho^n, \quad n = 0, 1, 2, \dots, N \quad (2.1.10)$$

The expected number of entities in the system is given by

$$L_s = \frac{\rho \{1 - (N + 1)\rho^N + N\rho^{N+1}\}}{(1 - \rho)(1 - \rho^{N+1})}, \quad \rho \neq 1 \quad (2.1.11)$$

Since the queue length is limited and some entities are lost, it is necessary to compute the effective arrival rate λ_{eff} , which is given by

$$\lambda_{eff} = \lambda(1 - p_N).$$

The expected number of entities in the queue L_Q is

$$\begin{aligned} L_Q &= L_s - \frac{\lambda_{eff}}{\mu} \\ &= \frac{\rho^2 [1 - N\rho^{N-1} + (N-1)\rho^N]}{(1-\rho)(1-\rho^{N+1})} \end{aligned} \quad (2.1.12)$$

Hence the expected waiting time per entity in the queue is given by

$$\begin{aligned} {}_2W_Q &= \frac{L_Q}{\lambda_{eff}} \\ &= \frac{\lambda [(\mu^N - \lambda^N) - N\lambda^{N-1}(\mu - \lambda)]}{\mu(\mu - \lambda)(\mu^N - \lambda^N)} \end{aligned} \quad (2.1.13)$$

2.1.3 The ML and CAN estimators for expected waiting time

2.1.3.1 The ML Estimator

Considering $X_{i1}, X_{i2}, \dots, X_{in}$ (with $i = 1, 2$ representing Models I and II) to be random samples of size n , each randomly drawn from different exponential inter arrival time populations with the parameter λ . and letting $Y_{i1}, Y_{i2}, \dots, Y_{in}$ (with $i = 1, 2$ representing Models I and II) be random samples of size n , each drawn from different exponential service time populations with the parameter μ , it follows that $E(\bar{X}_i) = \frac{1}{\lambda}$ and $E(\bar{Y}_i) = \frac{1}{\mu}$, where \bar{X}_i and \bar{Y}_i , $i = 1, 2$, are the sample means of inter arrival times and service times respectively corresponding to Models I and II. Further \bar{X}_i and \bar{Y}_i (with $i = 1, 2$ representing Models I and II) are the MLEs of $\frac{1}{\lambda}$ and $\frac{1}{\mu}$ respectively. Let $\theta_1 = \frac{1}{\lambda}$ and $\theta_2 = \frac{1}{\mu}$ respectively.

Model I

The average waiting time per entity in the queue given in (2.1.6) reduces to

$${}_1W_Q = \frac{\theta_2^2}{(\theta_1 - \theta_2)} \tag{2.1.14}$$

and hence the MLE of W_Q is given by

$${}_1\hat{W}_Q = \frac{\bar{Y}_1^2}{(\bar{X}_1 - \bar{Y}_1)} \tag{2.1.15}$$

Model II

The average waiting time per entity in the queue given in (2.1.13) reduces to

$${}_2W_Q = \frac{\theta_2^2[(\theta_1^N - \theta_2^N) + N\theta_2^{N-1}(\theta_2 - \theta_1)]}{(\theta_2 - \theta_1)(\theta_2^N - \theta_1^N)} \quad (2.1.16)$$

and hence the MLE of W_Q is given by

$${}_2\hat{W}_Q = \frac{\bar{Y}_2^2[(\bar{X}_2^N - \bar{Y}_2^N) + N\bar{Y}_2^{N-1}(\bar{Y}_2 - \bar{X}_2)]}{(\bar{Y}_2 - \bar{X}_2)(\bar{Y}_2^N - \bar{X}_2^N)} \quad (2.1.17)$$

It may be noted that ${}_i\hat{W}_Q$ given in (2.1.15) and (2.1.17) are real valued functions in \bar{X}_i and \bar{Y}_i , $i=1,2$, which are also differentiable. The following application of the multivariate central limit theorem may be considered (Rao [42]).

2.1.3.2 An application of the multivariate central limit theorem

Suppose T'_1, T'_2, T'_3, \dots are independent and identically distributed k -dimensional random variables such that

$$T'_n = (T'_{1n}, T'_{2n}, T'_{3n}, \dots, T'_{kn}), \quad n = 1, 2, 3, \dots$$

having the first and second order moments $E(T'_n) = \mu$ and $Var(T'_n) = \Sigma$. The sequence of random variables may be defined as

$$\bar{T}'_n = (\bar{T}'_{1n}, \bar{T}'_{2n}, \bar{T}'_{3n}, \dots, \bar{T}'_{kn}), \quad n = 1, 2, 3, \dots$$

$$\text{where } \bar{T}_{in} = \frac{\sum_{j=1}^n T_{ij}}{n}, \quad i = 1, 2, \dots, k \text{ and } j = 1, 2, \dots, n$$

Then, $\sqrt{n}(\bar{T}_n - \mu) \xrightarrow{d} N(0, \Sigma)$ as $n \rightarrow \infty$

2.1.3.3 The CAN Estimator

Model I

By applying the multivariate central limit theorem given to (2.1.15), it readily follows that

$$\sqrt{n}[(\bar{X}_1, \bar{Y}_1) - (\theta_1, \theta_2)] \xrightarrow{d} N(0, \Sigma)$$

as $n \rightarrow \infty$, where the dispersion matrix $\Sigma = ((\sigma_{ij}))$ is given by

$$\Sigma = \text{diag}(\theta_1^2, \theta_2^2)$$

From (Rao [42]), it follows that

$$\sqrt{n}(\hat{W}_Q - W_Q) \xrightarrow{d} N(0, \sigma^2(\theta)), \text{ as } n \rightarrow \infty, \text{ where } \theta = (\theta_1, \theta_2) \text{ and}$$

$$\begin{aligned} \sigma^2(\theta) &= \sum_{i=1}^2 \left(\frac{\partial W_Q}{\partial \theta_i} \right)^2 \cdot \sigma_{ii} \\ &= \frac{\theta_2^2 [\theta_1^2 + \theta_2^2 (2\theta_1 - \theta_2)^2]}{(\theta_1 - \theta_2)^4} \end{aligned} \quad (2.1.18)$$

Hence, ${}_1\hat{W}_\varrho$ is a CAN estimator of ${}_1W_\varrho$. There are several methods for generating CAN estimators and the Method of Moments and the Method of Maximum likelihood are commonly used to generate such estimators (Sinha [43]).

Model II

As in Model I,

$\sqrt{n}({}_2\hat{W}_\varrho - {}_2W_\varrho) \xrightarrow{d} N(0, {}_2\sigma^2(\theta))$, as $n \rightarrow \infty$, where $\theta = (\theta_1, \theta_2)$, ${}_2W_\varrho$ and ${}_2\hat{W}_\varrho$ are given by (4.16) and (4.17) respectively. Further, ${}_2\sigma^2(\theta)$ is computed from the partial derivatives $\left(\frac{\partial {}_2W_\varrho}{\partial \theta_i}\right)$, $i = 1, 2$ as in Model I. Thus ${}_2\hat{W}_\varrho$ is a CAN estimator of ${}_2W_\varrho$.

2.1.4 Confidence limits for the expected waiting time

Let ${}_i\sigma^2(\hat{\theta})$ be the estimator of ${}_i\sigma^2(\theta)$ (with $i = 1, 2$ representing Models I and II) obtained by replacing θ by a consistent estimator ${}_i\hat{\theta}$ namely ${}_i\hat{\theta} = (\bar{X}_i, \bar{Y}_i)$, $i = 1, 2$. Let ${}_i\hat{\sigma}^2 = {}_i\sigma^2(\hat{\theta})$. Since ${}_i\sigma^2(\theta)$ is a continuous function of ϑ , ${}_i\hat{\sigma}^2$ is a consistent estimator of ${}_i\sigma^2(\theta)$, i.e., ${}_i\hat{\sigma}^2 \xrightarrow{p} {}_i\sigma^2(\theta)$ as $n \rightarrow \infty$, $i = 1, 2$. By the Slutsky theorem

$$\frac{\sqrt{n}({}_i\hat{W}_\varrho - W_\varrho)}{{}_i\hat{\sigma}} \xrightarrow{d} N(0, 1)$$

$$\text{i.e., } \Pr \left[-k_{\frac{\alpha}{2}} < \frac{\sqrt{n}({}_i\hat{W}_Q - {}_iW_Q)}{{}_i\hat{\sigma}} < k_{\frac{\alpha}{2}} \right] = (1 - \alpha)$$

where $k_{\frac{\alpha}{2}}$ is obtained from Normal tables. Hence, a $100(1 - \alpha)\%$ asymptotic confidence interval for ${}_iW_Q$ is given by

$${}_i\hat{W}_Q \pm k_{\frac{\alpha}{2}} \cdot \frac{{}_i\hat{\sigma}}{\sqrt{n}}, \quad i = 1, 2 \tag{2.1.19}$$

Numerical Results

Table 2

Confidence limits for M/M/1/∞ : FCFS with 99% confidence interval and sample size of 20

$\lambda \backslash \mu$	0.04	0.06	0.08	0.1
0.01	(8.289842877:8.376823789)	(3.316926722:3.349739944)	(1.777150118:1.794278453)	(1.105858754:1.116363468)
0.02	(24.83223528:25.16776472)	(8.286718407:8.37994826)	(4.144904413:4.18842892)	(2.487402722:2.512597278)
0.03	(74.16104529:75.83895471)	(16.55477386:16.77855947)	(7.456343866:7.543656134)	(4.262400695:4.309027877)

Table 3

Confidence limits for M/M/1/N: FCFS with 99% confidence interval and sample size of 20

	$\lambda \backslash \mu$	0.04	0.06	0.08	0.1
N=10	0.01	(8.331269275:8.334920554)	(3.331916366:3.334744788)	(1.784518941:1.786909398)	(1.110057009:1.112165194)
	0.02	(24.75248955:24.7587519)	(8.328275966:8.332745576)	(4.164721818:4.168373097)	(2.498408635:2.501570885)
	0.03	(60.07542472:60.08759256)	(16.50061597:16.50687832)	(7.490679322:7.495571505)	(4.283054105:4.287193479)
	$\lambda \backslash \mu$	0.04	0.06	0.08	0.1
N=20	0.01	(8.331507591:8.335159075)	(3.33191912:3.334747547)	(1.784519057:1.786909514)	(1.110057019:1.112165204)
	0.02	(24.99636102:25.0026853)	(8.33109717:8.335569306)	(4.164840925:4.168492408)	(2.498418861:2.501581139)
	0.03	(73.40387298:73.41482651)	(16.66318663:16.66951092)	(7.497549755:7.502448733)	(4.283644082:4.287784475)
	$\lambda \backslash \mu$	0.04	0.06	0.08	0.1
N=40	0.01	(8.331507591:8.335159075)	(3.33191912:3.334747547)	(1.784519057:1.786909514)	(1.110057019:1.112165204)
	0.02	(24.99683772:25.00316228)	(8.331097265:8.335569401)	(4.164840925:4.168492409)	(2.498418861:2.501581139)
	0.03	(74.98446701:74.99541962)	(16.66350439:16.66982894)	(7.49755051:7.50244949)	(4.283644089:4.287784482)

As is to be expected, W_q is an increasing function of λ , and a decreasing function of μ , for both M/M/1/ ∞ and M/M/1/N queueing systems [See Tables 2&3].

2.2 Statistical analysis for a tandem queue with blocking

A maximum likelihood estimator (MLE), a consistent asymptotically normal (CAN) estimator and asymptotic confidence limits for the expected service time per customer in the system in a two station tandem queue with zero queue capacity and with blocking are obtained.

2.2.1 Introduction

Many studies of queueing models are confined to obtaining expressions for transient or stationary (steady state) solutions and do not consider the associated statistical inference problems. Parametric estimation is one of the essential tools to understand random phenomena using stochastic models. Analysis of queueing systems in this context has not received due attention. Whenever the systems are fully observable in terms of their basic random components such as inter-arrival times and service times, standard parametric techniques of statistical theory are quite appropriate. Recently Bhat [41] has provided an overview of methods available for estimation, when the information is restricted to the number of entities in the system at some discrete point in time. Bhat has also described how maximum likelihood estimation is applied directly to the underlying Markov chain in the queue length process in M/G/1 and GI/M/1 queues. Yadavalli *et al* [44] have obtained asymptotic confidence limits for the expected waiting time per customer in the queues of M/M/1/ ∞ and M/M/1/N. Further, Yadavalli *et al* [45] have extended the same results to c parallel servers ($c \geq 1$).

Generally speaking, the queueing models assume that each service channel consists of only one station. Situations do exist, where each service channel may consist of several stations in series. In this situation, an entity must successively pass through all the stations before completing service. Such situations are known as queues in series or tandem queues. Examples of such situations are as follows:

- a) In a manufacturing process, units must pass through a series of service channels (work stations), where each service channel performs a given task or job.
- b) In a University registration process, each registrant must pass through a series of counters such as advisor, departmental chairman (Head of the Department), Cashier etc.
- c) In a clinical physical examination procedure, a patient goes through a series of stages such as laboratory tests, Electro Cardio Graph, Chest X-ray etc.

In all these model structures, it is not only sufficient to know how many persons are in the system but also where they are.

An attempt is made in this paper to study a two station tandem queue with blocking in detail, Taha [3]. An MLE, CAN and asymptotic confidence limits are obtained for the expected service time per entity in the system.

2.2.2 System description and assumptions

Consider a simplified single channel queueing system consisting of two series stations as below:

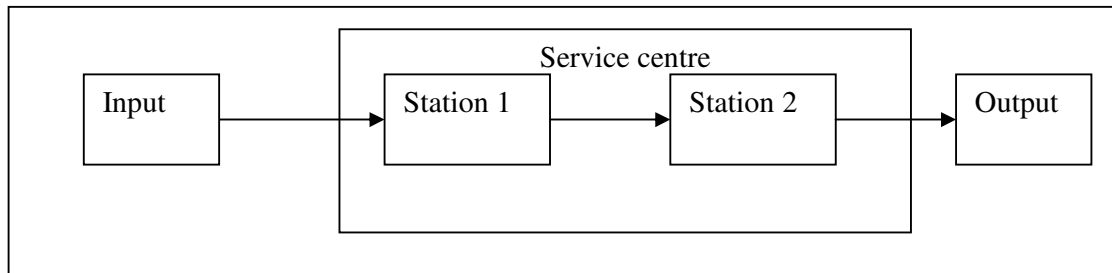


Fig. 2.2.1 System configuration

An entity arriving for service must pass through station 1 and station 2 before completion of service. The precise assumptions of the model are as follows:

- (i) Arrivals occur according to a Poisson distribution with a mean rate λ .
- (ii) Service times at each station are exponentially distributed with a service rate μ .
- (iii) Queues are not permitted ahead of station 1 or station 2.
- (iv) Each station is either free or busy.
- (v) Station 1 is said to be blocked when the entity in station 1 completes service before station 2 becomes free. In such a case the entity cannot wait between the stations, since this is not allowed.

2.2.3 Analysis of the system

Let the symbols 0,1 and b represent free, busy or blocked states of a station. Let $X(t)$ and $Y(t)$ respectively denote the states of station 1 and station 2 and the vector process $Z(t) = \{(X(t), Y(t)), t \geq 0\}$ with state space

$$E = \{(0,0), (0,1), (1,0), (1,1), (b,1)\}, \quad (2.2.3.1)$$

the state of the system at time t . Since the inter-arrival and service times are exponential, it follows that the process $Z(t)$ is a Markov process with the infinitesimal generator given by

$$\begin{array}{c}
 E \\
 \mathbf{(0,0)} \\
 \mathbf{(0,1)} \\
 \mathbf{(1,0)} \\
 \mathbf{(1,1)} \\
 \mathbf{(b,1)}
 \end{array}
 \begin{array}{c}
 \mathbf{(0,0)} \\
 \mathbf{(0,1)} \\
 \mathbf{(1,0)} \\
 \mathbf{(1,1)} \\
 \mathbf{(b,1)}
 \end{array}
 \begin{array}{c}
 \mathbf{(0,1)} \\
 \mathbf{(1,0)} \\
 \mathbf{(1,1)} \\
 \mathbf{(b,1)}
 \end{array}
 \begin{array}{c}
 \mathbf{(1,0)} \\
 \mathbf{(1,1)} \\
 \mathbf{(b,1)}
 \end{array}
 \begin{array}{c}
 \mathbf{(1,1)} \\
 \mathbf{(b,1)}
 \end{array}
 \begin{array}{c}
 \mathbf{(b,1)}
 \end{array}
 \left[\begin{array}{ccccc}
 -\lambda & 0 & \lambda & 0 & 0 \\
 \mu & -(\lambda + \mu) & 0 & \lambda & 0 \\
 0 & \mu & -\mu & 0 & 0 \\
 0 & 0 & \mu & -2\mu & \mu \\
 0 & \mu & 0 & 0 & -\mu
 \end{array} \right]$$

(2.2.3.2)

Let $p_{ij}(t) = p[Z(t) = (i, j)], \forall (i, j) \in E$ represent the probability that the system is in state (i, j) at time t with the initial condition $p_{00}(0) = 1$. From the infinitesimal generator given in (2.2.3.2), the following system of differential-difference equations is obtained:

$$\frac{dp_{00}(t)}{dt} = -\lambda p_{00}(t) + \mu p_{01}(t)$$

(2.2.3.3)

$$\frac{dp_{01}(t)}{dt} = -(\lambda + \mu) p_{01}(t) + \mu p_{10}(t) + \mu p_{b1}(t)$$

(2.2.3.4)

$$\frac{dp_{10}(t)}{dt} = \lambda p_{00}(t) - \mu p_{10}(t) + \mu p_{11}(t)$$

(2.2.3.5)

$$\frac{dp_{11}(t)}{dt} = \lambda p_{01}(t) - 2\mu p_{11}(t)$$

(2.2.3.6)

$$\frac{dp_{b1}(t)}{dt} = \mu p_{11}(t) - \mu p_{b1}(t)$$

(2.2.3.7)

2.2.3.1 Transient Solution

Solving the system of equation (2.2.3.3)-(2.2.3.7) along with the equation

$\sum_{(i,j) \in E} p_{ij}(t) = 1$ and using Laplace transforms, it is evident that:

$$p_{00}(t) = \frac{2\mu^2}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} + \lambda\mu^2 \sum_{i=1}^3 \frac{(\alpha_i + 2\mu)}{\alpha_i(\alpha_i + \lambda) \prod_{\substack{j=i \\ j \neq 1}}^3 (\alpha_i - \alpha_j)} e^{\alpha_i t} \quad (2.2.3.8)$$

$$p_{01}(t) = \frac{2\lambda\mu}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} + \lambda\mu \sum_{i=1}^3 \frac{(\alpha_i + 2\mu)}{\alpha_i \prod_{\substack{j=i \\ j \neq 1}}^3 (\alpha_i - \alpha_j)} e^{\alpha_i t} \quad (2.2.3.9)$$

$$p_{10}(t) = \frac{\lambda(\lambda + 2\mu)}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} + \lambda^2 \mu^2 \sum_{i=1}^3 \frac{(2\alpha_i + \lambda + 2\mu)}{\alpha_i(\alpha_i + \lambda)(\alpha_i + \mu)} * \frac{e^{\alpha_i t}}{\prod_{\substack{j=i \\ j \neq 1}}^3 (\alpha_i - \alpha_j)} + \frac{\lambda}{(2\lambda - \mu)} e^{\mu t} \quad (2.2.3.10)$$

$$p_{11}(t) = \frac{\lambda^2}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} + \lambda^2 \mu \sum_{i=1}^3 \frac{1}{\alpha_i \prod_{\substack{j=i \\ j \neq 1}}^3 (\alpha_i - \alpha_j)} e^{\alpha_i t} \quad (2.2.3.11)$$

$$p_{b1}(t) = \frac{\lambda^2}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} + \lambda^2 \mu^2 \sum_{i=1}^3 \frac{1}{\alpha_i(\alpha_i + \mu) \prod_{\substack{j=i \\ j \neq 1}}^3 (\alpha_i - \alpha_j)} e^{\mu t} + \frac{\lambda}{(\mu - 2\lambda)} e^{-\mu t} \quad (2.2.3.12)$$

where α_1, α_2 and α_3 are the roots of

$$s^3 + (2\lambda + 4\mu)s^2 + (\lambda^2 + 7\lambda\mu + 5\mu^2)s + \mu(3\lambda^2 + 4\lambda\mu + 2\mu^2) = 0$$

2.2.3.2 The Steady state solution

Since the stationary behaviour of the system is to be modelled, let $\lim_{t \rightarrow \infty} p_{ij}(t) = p_{ij}$. Let $\underline{p} = (p_{00}, p_{01}, p_{10}, p_{11}, p_{bi})$ be the stationary distribution corresponding to the Markov process $z(t)$. It readily follows from (2.2.3.8)-(2.2.3.12) that

$$p_{00}(t) = \frac{2\mu^2}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \quad (2.2.3.13)$$

$$p_{01}(t) = \frac{2\lambda\mu}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \quad (2.2.3.14)$$

$$p_{10}(t) = \frac{\lambda(\lambda + 2\mu)}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \quad (2.2.3.15)$$

$$p_{11}(t) = \frac{\lambda^2}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \quad (2.2.3.16)$$

$$p_{bi}(t) = \frac{\lambda^2}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)} \quad (2.2.3.17)$$

It may be noted that the solution given in (2.2.3.13)-(2.2.3.17) is in agreement

with Taha [3] with $\rho = \frac{\lambda}{\mu}$

2.2.3.3 Expected service time per entity in the system

The expected number of entities in the system is given by

$$\begin{aligned}
 L_s &= \sum_{n=0}^{\infty} np_n \\
 &= (p_{01} + p_{10}) + 2(p_{11} + p_{b1}) \\
 &= \frac{\lambda(5\lambda + 4\mu)}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)}
 \end{aligned}
 \tag{2.2.3.18}$$

The probability that an entity will enter station 1 is

$$\begin{aligned}
 &(p_{00} + p_{01}) \\
 &= \frac{2\mu(\lambda + \mu)}{(3\lambda^2 + 4\lambda\mu + 2\mu^2)}
 \end{aligned}
 \tag{2.2.3.19}$$

W_s represents the expected service time per entity in the system since queues are allowed and is given by

$$W_s = \frac{L_s}{\lambda_{eff}} = \frac{L_s}{\lambda(p_{00} + p_{01})} = \frac{(5\lambda + 4\mu)}{2\mu(\lambda + \mu)}
 \tag{2.2.3.20}$$

In the next section, the maximum likelihood and consistent asymptotically normal estimators for the expected service time per entity in the system are obtained.

2.2.4 MLE and CAN estimator for the expected service time per entity in the system

2.2.4.1 The ML estimator

Let X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_n be random samples of size n , each drawn from exponential inter-arrival time and exponential service time populations with parameters λ and μ respectively. It is clear that $E(\bar{X}) = \frac{1}{\lambda}$ and $E(Y) = \frac{1}{\mu}$, where \bar{X}_i and \bar{Y}_i are the sample means of inter-arrival times and service time respectively.

It can be shown that \bar{X} and \bar{Y} are MLEs of $\frac{1}{\lambda}$ and $\frac{1}{\mu}$ respectively.

Let $\theta_1 = \frac{1}{\lambda}$ and $\theta_2 = \frac{1}{\mu}$. The average service time per customer in the system given in (2.2.3.20) reduces to

$$W_s = \frac{\theta_2 (4\theta_1 + 5\theta_2)}{2(\theta_1 + \theta_2)} \quad (2.2.4.1)$$

and hence the MLE of W_s is given by

$$\hat{W}_s = \frac{\bar{Y}(4\bar{X} + 5\bar{Y})}{2(\bar{Y} + \bar{X})} \quad (2.2.4.2)$$

It may be noted that \hat{W}_s given in (2.2.4.2) is a real valued function in \bar{X} and \bar{Y} , which are also differentiable. Consider the following application of the multivariate central limit theorem. See Rao [42].

2.2.4.2 An application of the multivariate central limit theorem

Suppose T'_1, T'_2, T'_3, \dots are independent and identically distributed k -dimensional random variables such that

$$T'_n = (T_{1n}, T_{2n}, T_{3n}, \dots, T_{kn}), \quad n = 1, 2, 3, \dots$$

having the first and second order moments $E(T_n) = \mu$ and $Var(T_n) = \Sigma$. The sequence of random variables may be defined as

$$\bar{T}'_n = (\bar{T}_{1n}, \bar{T}_{2n}, \bar{T}_{3n}, \dots, \bar{T}_{kn}), \quad n = 1, 2, 3, \dots$$

where $\bar{T}_{in} = \frac{\sum_{j=1}^n T_{ij}}{n}$, $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, n$

Then, $\sqrt{n}(\bar{T}_n - \mu) \xrightarrow{d} N(0, \Sigma)$ as $n \rightarrow \infty$

2.2.4.3 The CAN Estimator

By applying the multivariate central limit theorem to (2.2.4.2), it readily follows that

$$\sqrt{n}[(\bar{X}, \bar{Y}) - (\theta_1, \theta_2)] \xrightarrow{d} N(0, \Sigma)$$

as $n \rightarrow \infty$, where the dispersion matrix $\Sigma = ((\sigma_{ij}))$ is given by

$$\Sigma = \text{diag}(\theta_1^2, \theta_2^2)$$

Again from Rao [42] it follows that

$$\sqrt{n}(\hat{W}_s - W_s) \xrightarrow{d} N(0, \sigma^2(\theta)), \text{ as } n \rightarrow \infty, \text{ where } \theta = (\theta_1, \theta_2) \text{ and}$$

$$\begin{aligned} \sigma^2(\theta) &= \sum_{i=1}^2 \left(\frac{\partial W_s}{\partial \theta_i} \right)^2 \cdot \sigma_{ii} \\ &= \frac{\vartheta_2^2 [\theta_1^2 \theta_2^2 + (4\vartheta_1^2 + 10\theta_1 \theta_2 + 5\vartheta_2^2)^2]}{4(\vartheta_1 + \vartheta_2)^4} \end{aligned}$$

Thus, \hat{W}_s is a CAN estimator of W_s . There are several methods for generation of CAN estimators and the Method of Moments and the Method of Maximum likelihood are commonly used to generate such estimators. See Sinha [43].

2.2.4.4 Confidence limits for the expected waiting time

Let $\sigma^2(\hat{\theta})$ be the estimator of $\sigma^2(\theta)$ obtained by replacing θ by a consistent estimator $\hat{\theta}$ namely. Let $\hat{\sigma}^2 = \sigma^2(\hat{\theta})$. Since $\sigma^2(\theta)$ is a continuous function of θ , $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2(\theta)$, i.e., $\hat{\sigma}^2 \xrightarrow{P} \sigma^2(\theta)$ as $n \rightarrow \infty$, $i = 1, 2$. By the Slutsky theorem

$$\sqrt{n}(\hat{W}_s - W_s) \xrightarrow{d} N(0,1)$$

$$\text{i.e., } \Pr \left[-k_{\frac{\alpha}{2}} < \frac{\sqrt{n}(\hat{W}_s - W_s)}{\hat{\sigma}} < k_{\frac{\alpha}{2}} \right] = (1 - \alpha)$$

where $k_{\frac{\alpha}{2}}$ is obtained from Normal tables. Hence, a $100(1 - \alpha)\%$ asymptotic confidence interval for W_s is given by

$$\hat{W}_s \pm k_{\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}} \tag{2.2.5.1}$$

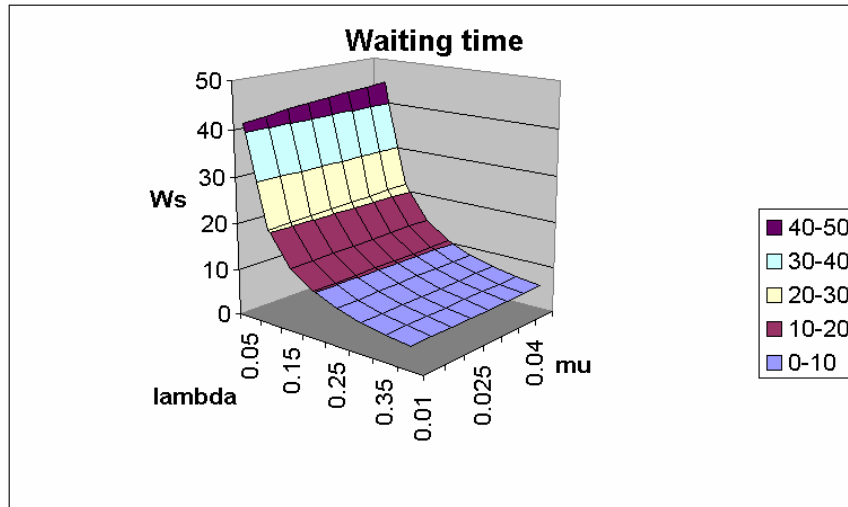
As is to be expected, W_q is an increasing function of λ , and a decreasing function of μ , for a tandem queue with blocking. The numerical illustration of the confidence interval of this model (tandem queues) is shown in Table 4.

Numerical Results

Table 4

Confidence limits for a tandem queue with blocking: 99% confidence interval and sample size of 20

λ	0.05		0.1		0.15		0.2		0.25		0.3		0.35		0.4	
	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL
0.01	41.57	41.76	20.41	20.50	13.51	13.57	10.10	10.14	8.06	8.10	6.71	6.74	5.74	5.77	5.02	5.04
0.015	42.21	42.41	20.60	20.70	13.61	13.67	10.15	10.20	8.09	8.13	6.73	6.76	5.76	5.79	5.03	5.06
0.02	42.76	42.96	20.79	20.88	13.69	13.76	10.20	10.25	8.13	8.17	6.76	6.79	5.78	5.80	5.05	5.07
0.025	43.23	43.44	20.95	21.05	13.78	13.84	10.25	10.30	8.16	8.20	6.78	6.81	5.80	5.82	5.06	5.09
0.03	43.65	43.85	21.10	21.20	13.86	13.92	10.30	10.35	8.20	8.23	6.80	6.83	5.81	5.84	5.08	5.10
0.035	44.01	44.22	21.25	21.35	13.93	14.00	10.35	10.40	8.23	8.26	6.83	6.86	5.83	5.86	5.09	5.11
0.04	44.34	44.55	21.38	21.48	14.00	14.07	10.39	10.44	8.26	8.29	6.85	6.88	5.85	5.87	5.10	5.13
0.045	44.63	44.84	21.50	21.60	14.07	14.14	10.43	10.48	8.29	8.32	6.87	6.90	5.86	5.89	5.11	5.14



Graph illustrating W_q as a function of λ and μ .

CHAPTER 3

A SINGLE CHANNEL QUEUEING MODEL WITH OPTIONAL SERVICE AND SERVICE INTERRUPTION

3.1 Introduction

One of the important characteristics of a queueing system is the service process. Entities in the system may be served individually or in batches. An arriving entity may not get satisfactory service rendered by the server. An intelligent entity may think of better service from the same server or may seek some other server (i.e. leaving the system unsatisfied). If the service given to an intelligent entity is not satisfactory and if it needs a second service, it has to join the end of the queue and wait for its turn of service. Either it may join the queue for the second service, or balk since it has waited too long. To satisfy such intelligent entities, the server can offer two kinds of service. Either an arriving entity can choose one of two servers before service starts. For example, a patient decides to undergo ordinary surgery or laparoscopic surgery; or a vehicle uses the existing road or the by-pass road. Or if an arriving entity is not satisfied by the first essential service, it can opt for the second optional service immediately. Else it can opt for the second optional service immediately. The former kind of service has been studied by Madan [16] and later by Madan [17] where there is no waiting capacity. However, in queueing systems where the server offers two services, one essential and the other optional, and interruption may take place; the system implies queueing models with service interruption which have been extensively studied in the past by Takagi [46]. Service interruption models with more than one service offered by a single server have not been considered so far. To fill the gap this chapter presents a Markov queueing model where the server offers two services, one essential and the other optional.

3.2 Model description

The server offers two services, one essential and the other optional. The essential service follows an exponential distribution and the optional service follows an arbitrary distribution. The server offers only one service at a time. The first service is essential for all entities while the second service is optional. In addition the service is interrupted for a random period whenever

the system becomes empty. It is assumed that the duration of interruptions is independent and, identically randomly distributed and is independent of the arrival process and the service time. Therefore the system has three states, namely

- (i) The operating state providing the first essential service.
- (ii) The operating state providing the second optional service.
- (iii) The state of interruption.

3.3 Assumptions and notation

λ : average arrival rate of entities.

μ : the average service rate of the server when offering essential service.

$W_n(t)$: the joint probability that at time t , there are $n > 0$ entities in the system and the server is providing essential service for the entities.

$S_n(x, t)$: the joint probability that at time t , there are $n > 0$ entities in the system with elapsed service time between R and $R+dx$ and the server is offering optional service to the entities.

$V_n(x, t)$: the joint probability that at time t there are $n > 0$ entities in the system with elapsed interruption time lying between R and $R+dx$ and the server is interrupted.

On completion of the regular (essential) service, an entity leaves the system with probability p and desires to have the second optional service with probability q ; $p+q=1$.

$\mu_1(x)dx$: the first order probability that the optional service will be completed in the time interval x and $x+dx$ given that the same was not completed

before time x and is related to the density function $B_1(x)$ by the hazard function relation.

$$B_1(\mu) = \mu_1(x)e^{-\int_0^x \mu_1(x)dx}$$

$\alpha(x)dx$: The first order probability that the elapsed interruption will take place in time x and $x+dx$ given that the same was not complete until time x and is related to the density function $V(x)$ by the relation

$$V(x) = \alpha(x)e^{-\int_0^x \alpha(t)dt}$$

Equations governing the system are as follows:

$$\begin{aligned} \frac{d}{dt}W_n(t) = & -(\lambda + \mu)W_n(t) + \lambda W_{n-1}(t) + p\mu W_{n+1}(t) \\ & + \int_0^\infty V_n(x,t)\alpha(x)dx + \int_0^\infty S_{n+1}(x,t)\mu(x)dx \quad ,n > 1 \end{aligned} \quad (3.4.1)$$

$$\frac{d}{dt}W_1(t) = -(\lambda + \mu)W_1(t) + \lambda W_2(t) + \int_0^\infty V_1(x,t)\alpha(x)dx + \int_0^\infty S_2(x,t)\mu(x)dx \quad (3.4.2)$$

$$\frac{\partial}{\partial x}S_n(x,t) + \frac{\partial}{\partial t}S_n(x,t) = -(\lambda + \mu_1(x))S_n(x,t) + \lambda S_{n-1}(x,t) \quad ,n > 1 \quad (3.4.3)$$

$$\frac{\partial}{\partial x}S_1(x,t) + \frac{\partial}{\partial t}S_1(x,t) = -(\lambda + \mu_1(x))S_1(x,t) \quad (3.4.4)$$

$$\frac{\partial}{\partial x}V_n(x,t) + \frac{\partial}{\partial t}V_n(x,t) = -(\lambda + \alpha(x))V_n(x,t) + \lambda V_{n-1}(x,t) \quad ,n > 0 \quad (3.4.5)$$

$$\frac{\partial}{\partial x}V_0(x,t) + \frac{\partial}{\partial t}V_0(x,t) = -(\lambda + \alpha(x))V_0(x,t)$$

(3.4.6)

Subject to boundary conditions

$$S_n(0, t) = \nu \mu W_n(t); \quad n > 0$$

(3.4.7)

$$V_n(0, t) = 0; \quad n > 0$$

(3.4.8)

$$V_n(0, t) = \rho \mu W_1(t) + \int_0^{\infty} S_1(x, t) \mu(x) dx$$

(3.4.9)

Further, it is assumed that the system starts initially when there are k units in the system so that the initial conditions are:

$$W_n(0) = \delta_{nk} = \begin{cases} 1 & \text{if } n = k \\ 0 & \text{if } n \neq k \end{cases}$$

where δ_{nk} is Kronecker's delta function

$$S_n(0) = 0 \text{ for all } n > 0$$

$$V_n = 0 \text{ for all } n \geq 0$$

(3.4.10)

3.4 Time dependant solution

The Laplace transform of a function $f(t)$ is defined as

$$f^*(s) = L\{f(t)\} = \int_0^{\infty} e^{-st} f(t) dt, \quad \text{Re}(s) > 0$$

(3.5.1)

The Laplace transform of $\frac{d}{dt} f(t)$ is

$$L\left[\frac{d}{dt} f(t)\right] = s f^*(s) - f(0)$$

(3.5.2)

Using the Laplace transform of equations (3.4.1) to (3.4.9) and equations (3.4.10), (3.5.1) and (3.5.2) result in:

$$(s + \lambda + \mu)W_n^*(s) = \delta_{nk} + \lambda W_{n-1}^*(s) + p\mu W_{n+1}^*(s) + \int_0^\infty V_n^*(x, s)\alpha(x)dx + \int_0^\infty S_{n+1}^*(x, s)\mu_1(x)dx$$

$n > 1$

(3.5.3)

$$(s + \lambda + \mu)W_1^*(s) = \delta_{1k} + p\mu W_2^*(s) + \int_0^\infty V_1^*(x, s)\alpha(x)dx + \int_0^\infty S_2^*(x, s)\mu_1(x)dx$$

(3.5.4)

$$\frac{\partial}{\partial x} S_n^*(x, s) + (s + \lambda + \mu_1(x))S_n^*(x, s) = \lambda S_{n-1}^*(x, s) \quad , n > 1$$

(3.5.5)

$$\frac{\partial}{\partial x} S_1^*(x, s) + (s + \lambda + \mu_1(x))S_1^*(x, s) = 0$$

(3.5.6)

$$\frac{\partial}{\partial x} V_n^*(x, s) + (s + \lambda + \alpha(x))V_n^*(x, s) = \lambda V_{n-1}^*(x, s) \quad , n > 0$$

(3.5.7)

$$\frac{\partial}{\partial x} V_0^*(x, s) + (s + \lambda + \alpha(x))V_0^*(x, s) = 0$$

(3.5.8)

Subject to boundary conditions:

$$S_n^*(0, s) = q\mu W_n^*(s) ; \quad , n > 0$$

(3.5.9)

$$V_n^*(0, s) = 0 ; \quad , n > 0$$

(3.5.10)

$$V_0^*(0, s) = p\mu W_1^*(s) + \int_0^x S_1^*(x, s)\mu_1(x)dx$$

(3.5.11)

The following generating functions are defined;

$$W^*(s, z) = \sum_{n=1}^{\infty} W_n^*(s) z^n$$

$$S^*(x, s, z) = \sum_{n=1}^{\infty} S_n^*(x, s) z^n$$

$$S^*(0, s, z) = \sum_{n=1}^{\infty} S_n^*(0, s) z^n$$

$$S^*(s, z) = \int_0^{\infty} S^*(x, s, z) dx$$

$$V^*(x, s, z) = \sum_{n=1}^{\infty} V_n^*(x, s) z^n$$

$$V^*(0, s, z) = \sum_{n=1}^{\infty} V_n^*(0, s) z^n$$

$$V^*(s, z) = \int_0^{\infty} V^*(x, s, z) dx \quad (3.5.12)$$

$\sum_{n=2}^{\infty} z^{n+1} * (3.5.3) + z^2 * (3.5.4)$ and using (3.5.12):

$$\begin{aligned} [(s + \lambda - \lambda z + \mu)z - p\mu]W^*(s, z) &= z^{k+1} - z[p\mu W_1^*(s) + \int_0^{\infty} S_1^*(x, s)\mu_1(x)dx] \\ &+ z \int_0^{\infty} V^*(x, s, z)\alpha(x)dx + \int_0^{\infty} S^*(x, s, z)\mu_1(x)dx \end{aligned} \quad (3.5.13)$$

$\sum_{n=2}^{\infty} z^{n+1} * (3.5.5) + z * (3.5.6)$ and using (3.5.12):

$$\frac{\partial}{\partial x} S^*(x, s, z) - (s + \lambda - \lambda z + \mu_1(x))S^*(x, s, z) = 0 \quad (3.5.14)$$

$\sum_{n=2}^{\infty} z^{n+1} * (3.5.7) + (3.5.8)$ and using (3.5.12):

$$\frac{\partial}{\partial} V^*(x, s, z) + (s + \lambda - \lambda z + \alpha(x))V^*(x, s, z) = 0 \quad (3.5.15)$$

$\sum_{n=1}^{\infty} z^n$ * (3.5.9) gives

$$S^*(0, s, z) = q\mu W^*(s, z) \quad (3.5.16)$$

$\sum_{n=1}^{\infty} z^n$ * (3.5.10) + (3.5.11) gives

$$V^*(0, s, z) = p\mu W_1^*(s) + \int_0^{\infty} S_1^*(x, s)\mu_1(x)dx \quad (3.5.17)$$

Integrating the equations (3.5.14) and (3.5.15) from 0 to x, gives

$$S^*(x, s, z) = S^*(0, s, z) \exp\left[-(s + \lambda - \lambda z)x - \int_0^x \mu_1(t)dt\right] \quad (3.5.18)$$

$$V^*(x, s, z) = V^*(0, s, z) \exp\left[-(s + \lambda - \lambda z)x - \int_0^x \alpha(t)dt\right] \quad (3.5.19)$$

Integrating (3.5.8) from 0 to x, gives

$$V_0^*(x, s) = V_0^*(0, s) \exp\left[-(s + \lambda)x - \int_0^x \alpha(t)dt\right] \quad (3.5.20)$$

Integration of (3.5.20) from 0 to ∞ yields

$$V_0^*(s) = V_0^*(0, s) \frac{1 - V^*(s + \lambda)}{s + \lambda} \quad (3.5.21)$$

From (3.5.11), (3.5.17) and (3.5.21):

$$V^*(0, s, z) = V_0^*(0, s) = \frac{(s + \lambda)V_0^*(s)}{1 - V^*(s + \lambda)} \quad (3.5.22)$$

Using (3.5.16) in (3.5.18):

$$S^*(x, s, z) = q\mu W^*(s, z) \exp[-(s + \lambda - \lambda z)x - \int_0^x \mu_1(t) dt]$$

Integrating this from 0 to ∞ , gives

$$S^*(s, z) = q\mu W^*(s, z) \frac{1 - B_1^*(s + \lambda - \lambda z)}{s + \lambda - \lambda z} \quad (3.5.23)$$

using (3.5.22) in (3.5.19) and integrating from 0 to ∞

$$V^*(s, z) = \frac{(s + \lambda)V_0^*(s)}{1 - V^*(s + \lambda)} * \frac{1 - V^*(s + \lambda - \lambda z)}{s + \lambda - \lambda z} \quad (3.5.24)$$

using (3.5.17), (3.5.18), (3.5.19) and (3.5.22) in (3.5.13) and simplifying yields

$$\begin{aligned} & [(s + \lambda - \lambda z + \mu)z - \mu(p + qB_1^*(s + \lambda - \lambda z))]W^*(s, z) \\ &= z^{k+1} + \frac{(s + \lambda)V_0^*(s)}{1 - V^*(s + \lambda)} [V^*(s + \lambda - \lambda z) - 1] \end{aligned}$$

Thus

$$W^*(s, z) = \frac{z^{k+1} + \frac{z(s + \lambda)V_0^*(s)}{1 - V^*(s + \lambda)} [V^*(s + \lambda - \lambda z) - 1]}{[(s + \lambda - \lambda z + \mu)z - \mu(p + qB_1^*(s + \lambda - \lambda z))]} \quad (3.5.25)$$

Since $W^*(s, z)$ is a regular function and the denominator of the right hand side vanishes for some z in $|z| < 1$, the number at 1 also vanishes for the same value of z . Applying Rouché's theorem the only unknown $V_0^*(s)$ can be determined. Hence $S^*(s, z)$, $W^*(s, z)$ and $V^*(s, z)$ can be completely determined.

3.5 Some special Cases

Case 1:

If the optional service is exponential then:

$$B_1^*(s + \lambda - \lambda z) = \frac{\mu_1}{c + \lambda - \lambda z + \mu_1}$$

Therefore (3.5.23) to (3.5.25) become

$$W^*(s, z) = \frac{(s + \lambda - \lambda z + \mu_1) \left[z^{k+1} + \frac{z(s + \lambda)V_0^*(s)}{1 - V^*(s + \lambda)} [V^*(s + \lambda - \lambda z) - 1] \right]}{[(s + \lambda - \lambda z + \mu)z - p\mu](s + \lambda - \lambda z + \mu_1) - q\mu\mu_1} \quad (3.5.26)$$

$$S^*(s, z) = \frac{q\mu W^*(s, z)}{s + \lambda - \lambda z + \mu_1} \quad (3.5.27)$$

$$V^*(s, z) = \frac{(s + \lambda)V_0^*(s)}{1 - V^*(s + \lambda)} * \frac{1 - V^*(s + \lambda - \lambda z)}{s + \lambda - \lambda z} \quad (3.5.28)$$

Case 2:

In addition to the condition of case 1, if there is no optional service and the server is offering the essential service only, then $p = 1, q = 0$. (3.5.26) to (3.5.28) become

$$W^*(s, z) = \frac{\left[z^{k+1} + \frac{z(s + \lambda)V_0^*(s)}{1 - V^*(s + \lambda)} [V^*(s + \lambda - \lambda z) - 1] \right]}{[(s - \lambda - \lambda z + \mu)z - \mu]}$$

$$S^*(s, z) = 0$$

$$V^*(s, z) = \frac{(s + \lambda)V_0^*(s)}{1 - V^*(s + \lambda)} * \frac{1 - V^*(s + \lambda - \lambda z)}{s + \lambda - \lambda z}$$

3.6 The steady state solution

Taking W_n , S_n and V_n as the respective steady state probabilities corresponding to $W_n(t)$, $S_n(t)$ and $V_n(t)$ and correspondingly $W(z)$, $S(z)$ and $V(z)$ as the probability generating functions, then the steady state solution can be obtained by using the Tauberian theorem Widder [47].

$$\lim_{s \rightarrow 0} sf^*(s) = \lim_{t \rightarrow \infty} f(t)$$

If the limit on the right exists the equations (3.5.23) to (3.5.25) become

$$W(z) = \frac{\frac{\lambda z}{1-V(\lambda)} [V(\lambda - \lambda z) - 1] V_0}{(\lambda - \lambda z + \mu)z - \mu(p + qB_1(\lambda - \lambda z))} \quad (3.6.1)$$

$$S(z) = q\mu W(z) \frac{1 - B_1(\lambda - \lambda z)}{\lambda - \lambda z} \quad (3.6.2)$$

$$V(z) = \frac{\lambda V_0}{1-V(\lambda)} * \frac{1 - V(\lambda - \lambda z)}{\lambda - \lambda z} \quad (3.6.3)$$

If $P(z)$ is the probability generating function of the number of entities in the system irrespective of the state, then $P(z) = W(z) + S(z) + V(z)$,

Thus :

$$P(z) = \frac{[-V(\lambda - \lambda z)][p + qB_1(\lambda - \lambda z)]}{[p + qB_1(\lambda - \lambda z)]\mu - z(\lambda - \lambda z + \mu)} + \frac{V_0}{1-V(\lambda)} \quad (3.6.4)$$

Using the normalization condition $P(1) = 1$,

$$V_0 = \frac{\mu - \lambda[1 + q\mu E(B_1)]}{\lambda E(V)} [1 - V(\lambda)]$$

where $E(B_1)$ is the expected optional service time and $E(V)$ is the expected interruption time

Therefore (3.6.4) becomes

$$P(z) = \frac{[\mu - \lambda[1 + q\mu E(B_1)]] * [1 - V(\lambda - \lambda z)][\rho + qB_1(\lambda - \lambda z)]}{\lambda E(V) \mu(\rho + qB_1(\lambda - \lambda z) - z(\lambda - \lambda z + \mu))}$$

or

$$P(z) = \frac{\mu - \lambda[1 + q\mu E(B_1)][\rho + qB_1(\lambda - \lambda z)](1 - z) * \frac{1 - V(\lambda - \lambda z)}{(\lambda - \lambda z)E(V)}}{\mu(\rho + qB_1(\lambda - \lambda z) - z(\lambda - \lambda z + \mu))}$$

(3.6.5)

$$= \alpha_-(z) * P_{M/M/1}(z)$$

Where

$$\alpha_-(z) = \frac{1 - V(\lambda - \lambda z)}{(\lambda - \lambda z)E(V)}$$

$$P_{M/M/1}(z) = \frac{[\mu - \lambda[1 + q\mu E(B_1)][\rho + qB_1(\lambda - \lambda z)](1 - z)}{\mu(\rho + qB_1(\lambda - \lambda z) - z(\lambda - \lambda z + \mu))}$$

$\alpha_-(z)$ is the probability generating function of the number of entities which arrive before an arbitrary entity during an interruption period in which the arbitrary entity arrives (Fuhrmann [20]) and $P_{M/M/1}(z)$ is the probability generating function of the number of entities in the M/M/1 queueing system with additional optional service and

$$\delta = \frac{\lambda(1 + q\mu E(B_1))}{\mu}$$

3.7 Some special Cases

Case 1:

If there is no essential service (the server is offering only the optional service),

then $p = 0, q = 1$ and $\frac{1}{\mu} \rightarrow 0$ (3.6.5) becomes

$$P(z) = \frac{[1 - \lambda E(B_1)]B_1(\lambda - \lambda z)(1 - z)}{B_1(\lambda - \lambda z) - z} * \frac{1 - V(\lambda - \lambda z)}{(\lambda - \lambda z)E(V)} \quad (3.6.6)$$

$$= \alpha_-(z) * P_{M/G/1}(z)$$

which is the stochastic decomposition for the M/G/1 queueing system Takagi [46]

Case 2:

Suppose there is no additional optional service so that $\rho = 1, q = 0$ Then

(3.6.5) becomes:

$$P(z) = \frac{\mu(1 - \delta)(1 - z)}{\mu - z(\lambda - \lambda z + \mu)} * \frac{1 - V(\lambda - \lambda z)}{(\lambda - \lambda z)E(V)} \quad (3.6.7)$$

where $\delta = \frac{1}{\mu}$

$$P(z) = \alpha_-(z) * P_{M/M/1}(z)$$

which is the stochastic decomposition for the M/M/1 queueing system by Takagi [46]

Case 3:

Suppose the additional service follows an exponential distribution. Then

$$B_1(\lambda - \lambda z) = \frac{\mu_1}{\lambda - \lambda z + \mu_1}$$

Thus (3.6.5) becomes

$$P(z) = \frac{[\mu - \lambda(1 + q\mu\mu_1)][\rho(\lambda - \lambda z + \mu_1) + q\mu_1]}{\mu[\rho(\lambda - \lambda z + \mu_1) + q\mu_1] - z(\lambda - \lambda z + \mu)(\lambda - \lambda z + \mu_1)} * \frac{1 - V(\lambda - \lambda z)}{\lambda E(V)}$$

(3.6.8)

Further if there is no additional optional service, then equation (3.6.8) leads to

(3.6.7)

Case 4:

If $E(N)$ denotes the expected number of entities in the system, then

$$E(N) = \frac{d}{dz} P(z) \text{ at } z=1$$

Thus

$$E(N) = \frac{\lambda E(V^2)}{qE(V)} + \frac{\lambda}{\mu - \lambda(1 + q\mu E(B_1))} + \frac{\lambda^2 \mu q E(B_1^2)}{2[\mu - \lambda(1 + q\mu E(B_1))]} + \lambda q E(B_1) \tag{3.6.9}$$

where $E(B_1) = (-1)^i B_1^{(i)}(0)$

$$= (-1)^i \frac{d^i B_1^*(s)}{ds^i} \text{ at } s=0$$

$$E(V^i) = (-1)^i V^{*i}(0) = (-1)^i \frac{d^i V^*(s)}{ds^i} \text{ at } s=0$$

By Little's [46] formula, the mean entity response time is given by

$$E(T) = \frac{E(N)}{\lambda}$$

Therefore,

$$E(I) = \frac{E(V^2)}{2E(V)} + \frac{1}{\mu - \lambda(1 + q\mu E(B_1))} + \frac{\lambda\mu q E(B_1^2)}{2[\mu - \lambda(1 + q\mu E(B_1))]} + qE(B_1) \quad (3.6.10)$$

Further if there is no additional optional service so that $\rho = 1, q = 0$ then equation (3.6.9) and (3.6.10) become:

$$E(N) = \frac{\lambda E(V^2)}{\lambda E(V)} + \frac{\lambda}{\mu - \lambda} \text{ and}$$

$$E(T) = \frac{E(V^2)}{2E(V)} + \frac{1}{\mu - \lambda}$$

which are the results given in (Takagi [46]) for the M/M/1 queueing system with interruption. On the other hand, if the server offers only the optional service so that $\rho = \zeta, q = 1$ and $\frac{1}{\mu} \rightarrow 0$ then equations (3.6.9) and (3.6.10)

become

$$E(N) = \frac{\lambda E(V^2)}{2E(V)} + \frac{\lambda^2 E(B_1^2)}{2[1 - \lambda E(B_1)]} + \lambda E(B_1)$$

$$E(T) = \frac{E(V^2)}{2E(V)} + \frac{\lambda E(B_1^2)}{2[1 - \lambda E(B_1)]} + E(B_1)$$

which are the results given in Takagi [46] where

$$\alpha(z) = V(\lambda - \lambda z)$$

3.8 Concluding remark

The description of the queueing system given in the introduction of this chapter leads the reader to believe that the low degree of system complexity would result in ease of mathematical modelling. The eventual mathematical manipulations required to create the model are far from insignificant, rather they are extensive and involved, and demand treatment by a highly proficient practitioner.

CHAPTER 4

AN M/M/1 QUEUEING SYSTEM WITH BATCH ARRIVALS OF VARYING SIZE, SERVICE OF FIXED BATCH SIZE AND TWO MODES OF FAILURE OF SERVICE FACILITY

4.1 Introduction

In many industrial processes, the service is interrupted because of the occurrence of breakdown in the facility that provides the service. The entities will not be serviced unless the facility is repaired. The server if human, may be in need of rest from time to time (Yadavalli *et al* [44]) or if non-human may be subject to two modes of failure, partial or total. That is, when the service facility is in partial failure mode, it gives service with a lower rate than in normal operating conditions. Various authors have analysed queueing systems where the service facility is subject to two modes of failure (Madan [49], Jain and Sharma [50], Reddy [51], and Sridharan and Jayashree [52]). Queueing systems with two modes of failure and arrivals and services in batches have not been considered so far. Such types of service interruptions are common in industry, factories, telephone booths and in operation of mechanical devices such as electronic computers, etc. In this chapter an M/M/1 queueing system is considered where the service facility is subject to two modes of failure, arrivals are in batches of varying size and service is rendered for batches of fixed size.

4.2 Model description

In this model units arrive at the system in batches of varying size and batches are pre-ordered for service purposes. The service of units is rendered in batches of fixed size and the service times of successive batches are distributed exponentially by a single server with rate μ_1 in normal working condition and at a slower rate μ_2 , ($\mu_2 < \mu_1$) in case of partial failure of the service channel.

One of the underlying assumptions about the repair process is that it starts instantaneously. If the service channel repair in the partial failure mode is complete, the unit enters the normal working mode; otherwise it goes to the failure mode.

After the repair of the service channel in total failure is complete, the unit goes directly to the normal working mode without passing through the partial failure mode. The repair times of failure modes, and the failure times are exponentially distributed with different derivatives.

4.3 Assumptions and notation

The system may be described as follows:

1. Entities arrive in batches in varying size. Let $\lambda c_i dt (i = 1, 2, 3, \dots, k)$ denote the probability that a batch of i entities arrives in a small interval of time dt , where $0 \leq c_i < 1$ and $\sum_{i=1}^k c_i = 1$ and $\lambda > 0$ is the mean arrival rate of batches which are pre-ordered for service purposes.
2. The service of entities is rendered in batches of fixed size b , ($b \geq 1$) and the service times of successive batches are distributed exponentially with mean service time $\frac{1}{\mu_1}$, ($\mu_1 > 0$) and $\frac{1}{\mu_2}$, ($\mu_2 > 0$) when the service channel is in the normal and partial failure mode respectively.
3. $\alpha_1(\alpha_2)dt$ is the first order probability that a total (partial) failure occurs during a short interval of time dt .
4. $\beta_1(\beta_2)dt$ is the first order probability of completion of a repair of total (partial) failure during a short interval of time dt .
5. $W_n(t)$ is the joint probability that at time t , there are ($n \geq 0$) entities in the queue when the service channel is in the normal working mode (i.e. excluding the batch of entities in service if any).
6. $S_n(t)$ is the joint probability that at time t , there are n entities ($n \geq 0$) in the queue when the service channel is in partial failure mode (i.e. excluding the batch of entities in service if any).
7. $R_n(t)$ is the probability that at time t , there are ($n \geq 0$) entities in the queue when the service channel is in total failure mode (i.e. excluding the batch of entities in service if any)

8. $Q(t)$ is the probability that at time t , there are no entities either in service or in the queue and the service channel is in normal working mode, and though operative, is idle
9. $F(t)$ is the probability that at time t , there are no entities either in service or in the queue, and the service channel is in partial failure mode, and though partially operative, is idle.
10. $P_n(t)$ is the probability that there are ($n \geq 0$) entities in the queue irrespective of the state of the service channel and that $P_n(t) = W_n(t) + S_n(t) + R_n(t)$.
11. If repair in the partial failure mode is in the process of being completed, the system will not enter the total failure mode.

4.4 Equations describing the system

Using probability arguments, the following difference-differential equations are obtained:

$$\frac{d}{dt}W_n(t) + (\lambda + \mu_1 + \alpha_2)W_n(t) = \sum_{i=1}^n \lambda c_i W_{n-i}(t) + \mu_1 W_{n+b}(t) + \beta_1 R_n(t) + \beta_2 S_n(t); n > 0, \quad (4.4.1)$$

$$\frac{d}{dt}W_0(t) + (\lambda + \mu_1 + \alpha_2)W_0(t) = \lambda Q(t) + \mu_1 \sum_{k=1}^b W_k(t) + \beta_1 R_0(t) + \beta_2 S_0(t), \quad (4.4.2)$$

$$\frac{d}{dt}S_n(t) + (\lambda + \mu_2 + \alpha_1 + \beta_2)S_n(t) = \sum_{i=1}^n \lambda c_i S_{n-i}(t) + \mu_2 S_{n+b}(t) + \alpha_2 W_n(t); n > 0, \quad (4.4.3)$$

$$\frac{d}{dt}S_0(t) + (\lambda + \mu_2 + \alpha_1 + \beta_2)S_0(t) = \mu_2 \sum_{k=1}^b S_k(t) + \alpha_2 W_0(t) + \lambda F(t), \quad (4.4.4)$$

$$\frac{d}{dt}R_n(t) + (\lambda + \beta_1)R_n(t) = \sum_{i=1}^n \lambda c_i R_{n-i}(t) + \alpha_1 S_n(t) \quad (4.4.5)$$

$$\frac{d}{dt}R_0(t) + (\lambda + \beta_1)R_0(t) = \alpha_1 S_0(t)$$

(4.4.6)

$$\frac{d}{dt}Q(t) + \lambda Q(t) = \mu_1 W_0(t)$$

(4.4.7)

$$\frac{d}{dt}F(t) + \lambda F(t) = \mu_2 S_0(t)$$

(4.4.8)

It is assumed that the system initially starts when there are m entities in the queue and the service channel is in the normal working condition so that the initial conditions are

$$W_n(0) = \delta_{n,m} \text{ where}$$

$$\delta_{n,m} = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}$$

$$S_n(0) = 0, \forall n \geq 0$$

$$R_n(0) = 0, \forall n \geq 0$$

(4.4.9)

4.5 Time dependant solution

Let $f^*(s)$ be the Laplace transform of $f(t)$. Taking the Laplace transform of equations (4.4.1)-(4.4.8) and using (4.4.9), it follows that:

$$(s + \lambda + \mu_1 + \alpha_2)W_n^*(s) = \delta_{n,m} + \sum_{i=1}^n \lambda c_i W_{n-i}^*(s) + \mu_1 W_{n+b}^*(s) + \beta_1 R_n^*(s) + \beta_2 S_n^*(s), n > 0$$

(4.5.1)

$$(s + \lambda + \mu_1 + \alpha_2)W_0^*(s) = \delta_{0,m} + \lambda Q^*(s) + \mu_1 \sum_{k=1}^b W_k^*(s) + \beta_1 R_0^*(s) + \beta_2 S_0^*(s)$$

(4.5.2)

$$(s + \lambda + \mu_2 + \alpha_2 + \beta_2)S_n^*(s) = \sum_{i=1}^n \lambda c_i S_{n-i}^*(s) + \mu_2 S_{n+b}^*(s) + \alpha_2 W_n^*(s)$$

(4.5.3)

$$(s + \lambda + \mu_2 + \alpha_1 + \beta_2)S_0^*(s) = \mu_2 \sum_{k=1}^b S_k^*(s) + \alpha_2 W_0^*(s) + \lambda F^*(s) \quad (4.5.4)$$

$$(s + \lambda + \beta_1)R_n^*(s) = \sum_{i=1}^n \lambda c_i R_{n-i}^*(s) + \alpha_1 S_n^*(s) \quad (4.5.5)$$

$$(s + \lambda + \beta_1)R_0^*(s) = \alpha_1 S_0^*(s) \quad (4.5.6)$$

$$(s + \lambda)Q^*(s) = \mu_1 W_0^*(s) \quad (4.5.7)$$

$$(s + \lambda)F^*(s) = \mu_2 S_0^*(s) \quad (4.5.8)$$

The following probability generating functions are defined as follows:

$$W^*(s, z) = \sum_{n=0}^{\infty} W_n^*(s) z^n$$

$$S^*(s, z) = \sum_{n=0}^{\infty} S_n^*(s) z^n$$

$$R^*(s, z) = \sum_{n=0}^{\infty} R_n^*(s) z^n$$

$$C(z) = \sum_{i=1}^{\infty} c_i z^i \quad (4.5.9)$$

$\sum_{z=1}^{\infty} z^{n+b} * (4.5.1) + z^b * (4.5.2)$ and using (4.5.9), it follows that

$$\begin{aligned} [(\eta + \mu_1 + \alpha_2)z^b - \mu_1]W^*(s, z) &= z^{m+b} + \mu_1 \sum_{i=1}^b (z^b - z^i)W_i^*(s) - \mu_1 W_0^*(s) \\ &\quad + \beta_1 z^b R^*(s, z) + \beta_2 z^b S^*(s, z) + \lambda z^b Q^*(s) \end{aligned} \quad (4.5.10)$$

$\sum_{z=1}^{\infty} z^{n+b} * (4.5.3) + z^b * (4.5.4)$ and using (4.5.9), the result is

$$\begin{aligned} [(\eta + \alpha_1 + \beta_2 + \mu_2)z^b - \mu_2]S^*(s, z) &= \mu_2 \sum_{i=1}^b (z^b - z^i)S_i^*(s) - \mu_2 S_0^*(s) \\ &\quad + \alpha_2 z^b W^*(s, z) + \lambda z^b F^*(s) \end{aligned} \quad (4.5.11)$$

$\sum_{z=1}^{\infty} z^n * (4.5.5) + (4.5.6)$ and using (4.5.9), it follows that

$$(\eta + \beta_1)R^*(s, z) = \alpha_1 S^*(s, z) \quad (4.5.12)$$

where $\eta = s + \lambda - \lambda C(z)$

Simplification of (4.5.10), (4.5.11) and (4.5.12) yields

$$\begin{aligned} K(s, z)W^*(s, z) &= (\eta + \beta_1)[(\eta + \mu_2 + \alpha_1 + \beta_2)z^b - \mu_2]^* \\ &\quad \left[z^{m+b} + \mu_1 \sum_{i=1}^b (z^b - z^i)W_i^*(s) + \lambda Q^*(s)z^b - \mu_1 W_0^*(s) \right] + \\ &\quad z^b [(n + \beta_1)\beta_2 + \alpha_1 p_1]^* \\ &\quad \left[\mu_2 \sum_{i=1}^b (z^b - z^i)S_i^*(s) + \lambda F^*(s)z^b - \mu_2 S_0^*(s) \right] \end{aligned} \quad (4.5.13)$$

$$K(s, z)S^*(s, z) = (\eta + \beta_1) \left[\begin{aligned} &\alpha_2 z^b \left(z^{m+b} + \mu_1 \sum_{i=1}^b (z^b - z^i)W_i^*(s) + \lambda Q^*(s)z^b - \mu_1 W_0^*(s) \right) \\ &+ ((\eta + \mu_1 + \alpha_2)z^b - \mu_1) \left(\mu_2 \sum_{i=1}^b (z^b - z^i)S_i^*(s) + \lambda F^*(s)z^b - \mu_2 S_0^*(s) \right) \end{aligned} \right] \quad (4.5.14)$$

$$\begin{aligned} K(s, z)R^*(s, z) &= \alpha_1 \left[\alpha_2 z^b \left(z^{m+b} + \mu_1 \sum_{i=1}^b (z^b - z^i)W_i^*(s) + \lambda Q^*(s)z^b - \mu_1 W_0^*(s) \right) \right] \\ &\quad + \left[((\eta + \mu_1 + \alpha_2)z^b - \mu_1) \left(\mu_2 \sum_{i=1}^b (z^b - z^i)S_i^*(s) + \lambda F^*(s)z^b - \mu_2 S_0^*(s) \right) \right] \end{aligned} \quad (4.5.15)$$

$$(s + \lambda)Q^*(s) = \mu_1 W_0^*(s) \quad (4.5.16)$$

$$(s + \lambda)F^*(s) = \mu_2 S_0^*(s) \quad (4.5.17)$$

where

$$K(s, z) = (\eta + \beta_1) \left[(\eta + \mu_1 + \alpha_2)z^b - \mu_1 \right] \left[(\eta + \mu_2 + \alpha_1 + \beta_2)z^b + \mu_2 \right] - \alpha_2 \beta_2 z^{2b} - \alpha_1 \alpha_2 \beta_1 z^{2b}$$

$$f(z) = (s + \lambda - \lambda C(z) + \beta_1) \left[\frac{((s + \lambda - \lambda C(z) + \mu_1 + \alpha_2)z^b - \mu_1)^*}{((s + \lambda - \lambda C(z) + \mu_2 + \alpha_1 + \beta_2)z^b - \mu_2) - \alpha_2 \beta_2 z^{2b}} \right]$$

$$g(z) = \alpha_1 \alpha_2 \beta_1 z^{2b}$$

For $|z| = 1$

$$\begin{aligned} |f(z)| &= \left| \frac{(s + \lambda - \lambda C(z) + \beta_1) \left[((s + \lambda - \lambda C(z) + \mu_1 + \alpha_2)z^b - \mu_1) \right]}{((s + \lambda - \lambda C(z) + \mu_2 + \alpha_1 + \beta_2)z^b - \mu_2) - \alpha_2 \beta_2 z^{2b}} \right| \\ &= |(s + \lambda - \lambda C(z) + \beta_1)| \\ &\quad \left| \frac{((s + \lambda - \lambda C(z) + \mu_1 + \alpha_2)z^b - \mu_1)}{((s + \lambda - \lambda C(z) + \mu_2 + \alpha_1 + \beta_2)z^b - \mu_2) - \alpha_2 \beta_2 z^{2b}} \right| \\ &\geq |s + \lambda + \lambda C(z) + \beta_1| \left| \frac{((s + \lambda - \lambda C(z) + \mu_1 + \alpha_2)|z^b| - \mu_1)}{((s + \lambda - \lambda C(z) + \mu_2 + \alpha_1 + \beta_2)|z^b| - \mu_2) - \alpha_2 \beta_2 |z^{2b}|} \right| \\ &= |s + \lambda - \lambda C(z) + \beta_1| \left| \frac{(s + \lambda - \lambda C(z) + \mu_1 + \alpha_2) - \mu_1}{((s + \lambda - \lambda C(z) + \mu_2 + \alpha_1 + \beta_2) - \mu_2) - \alpha_2 \beta_2} \right| \\ &\geq |s + \lambda + \beta_1 - \lambda C(z)| \left| \frac{((s + \lambda + \mu_1 + \alpha_2 - \lambda C(z)) - \mu_1)}{((s + \lambda + \mu_2 + \alpha_1 + \beta_2 - \lambda C(z)) - \mu_2) - \alpha_2 \beta_2} \right| \end{aligned}$$

$$\geq |s + \lambda + \beta_1 - \lambda| \left| \frac{(s + \lambda + \mu_1 + \alpha_2 - \lambda - \mu_1)}{(s + \lambda + \mu_2 + \alpha_1 + \beta_2 - \lambda - \alpha_2) - \alpha_2 \beta_2} \right|$$

$$= |s + \beta_1| |(s + \alpha_2)(s + \alpha_1 + \beta_2) - \alpha_2 \beta_2|$$

($\because \operatorname{Re}(s) > 0$)

$$\geq |\beta_1| |\alpha_2(\alpha_1 + \beta_2) - \alpha_2 \beta_2|$$

$$\geq |\beta_1| |\alpha_1 \alpha_2|$$

$$= \alpha_1 \alpha_2 \beta_1$$

For $|z| = 1$

$$|g(z)| = \alpha_1 \alpha_2 \beta_1 |z^{2b}| = \alpha_1 \alpha_2 \beta_1$$

$\therefore |f(z)| > |g(z)|$ on $|z| = 1$

Since $f(z)$ and $g(z)$ are differentiable inside and on the contour $|z| = 1$ and $|f(z)| > |g(z)|$ on $|z| = 1$, $f(z) - g(z)$, i.e. the denominator of equations (4.5.13), (4.5.14) and (4.5.15) have the same number of zeros inside $|z| = 1$ as that of $f(z)$ by Rouché's theorem. The zeros are given by the equations

$$s + \lambda - \lambda C(z) + \beta_1 = 0 \text{ and}$$

$$[(s + \lambda - \lambda C(z) + \mu_1 + \alpha_2)z^b - \mu_1]^* [(s + \lambda - \lambda C(z) + \mu_2 + \alpha_1 + \beta_2)z^b - \mu_2] - \alpha_2 \beta_2 z^{2b} = 0$$

The later equation has $2b$ zeros inside $|z| = 1$. Thus the denominator $K(s, z)$ has $2b$ zeros inside $|z| = 1$. Since $\overline{W}(s, z)$ is regular inside the contour $|z| = 1$, the numerator must vanish from the zeros of the denominator, as such there are $2b$ linear equations in $2b$ unknowns, $\overline{W}_r^*(s), r = 1, 2, \dots, b$, $\overline{S}_r^*(s), r = 1, 2, \dots, b$. These together with (4.5.16) and (4.5.17) are sufficient to determine all the

$2b+2$ unknowns. $\overline{W}^*(s, z)$ and hence $\overline{S}^*(s, z)$ and $\overline{R}^*(s, z)$ can be completely determined.

4.6 The steady state solution

If W_n, S_n, R_n, Q and F represent the respective steady state probabilities corresponding to $W_n(t), S_n(t), R_n(t), Q(t)$ and $F(t)$, and correspondingly $W(z), S(z)$ and $R(z)$ are the probability generating functions of W_n, S_n, R_n , then the steady state solution can be obtained by using the Tauberian property (see Widder [47]).

$$\lim_{s \rightarrow 0} sf^*(s) = \lim_{t \rightarrow \infty} f(t) \text{ if the limit on the right hand side exists.}$$

Thus equations (4.5.13), (4.5.14) and (4.5.15) yield

$$W(z) = \frac{N_1}{D_1} \quad (4.6.1)$$

$$S(z) = \frac{N_2}{D_1} \quad (4.6.2)$$

$$R(z) = \frac{N_3}{D_1} \quad (4.6.3)$$

$$\lambda Q = \mu_1 W_0 \quad (4.6.4)$$

$$\lambda F = \mu_2 S_0 \quad (4.6.5)$$

where

$$N_1 = (\lambda - \lambda C(z) + \beta_1)((\lambda - \lambda C(z) + \mu_2 + \alpha_1 + \beta_2)z^b - \mu_2) \left[\mu_1 \sum_{i=1}^b (z^b - z^i) W_i + \lambda z^b Q - \mu_1 W_0 \right] \\ + z^b [(\lambda - \lambda C(z) + \beta_1)\beta_2 + \alpha_1 \beta_1]^* \left[\mu_2 \sum_{i=1}^b (z^b - z^i) S_i + \lambda z^b F - \mu_2 S_0 \right]$$

$$N_2 = (\lambda - \lambda C(z) + \beta_1) \left[\alpha_2 z^b (\mu_1 \sum_{i=1}^b (z^b - z^i) W_i + \lambda z^b Q - \mu_1 W_0) + ((\lambda - \lambda C(z) + \mu_1 + \alpha_2)z^b - \mu_1)^* \left(\mu_2 \sum_{i=1}^b (z^b - z^i) S_i + \lambda F z^b - \mu_2 S_0 \right) \right]$$

$$N_3 = \alpha_1 \left[\begin{array}{l} \alpha_2 z^b (\mu_1 \sum_{i=1}^b (z^b - z^i) W_i + \lambda z^b Q - \mu_1 W_0) + ((\lambda - \lambda C(z) + \mu_1 + \alpha_2) z^b - \mu_1)^* \\ \left(\mu_2 \sum_{i=1}^b (z^b - z^i) S_i + \lambda F z^b - \mu_2 S_0 \right) \end{array} \right]$$

and

$$D_1 = (\lambda - \lambda C(z) + \beta_1) \left[\begin{array}{l} ((\lambda - \lambda C(z) + \mu_1 + \alpha_2) z^b - \mu_1) z^b - \mu_1)^* \\ ((\lambda - \lambda C(z) + \mu_2 + \alpha_1 + \beta_2) z^b - \mu_2) \end{array} \right] - \alpha_1 \alpha_2 \beta_1 z^{2b}$$

As earlier the unknowns can be determined by applying Rouché's theorem.

4.7 Some special cases

Case 1

If there are single arrivals and single departures, then $C_1=1$, $C_i=0$ for $i \neq 1$, $C(z)=z$ and $b=1$. Substituting these values in equations (4.6.1),(4.6.2) and (4.6.3) and using (4.6.4), it follows that

$$W(z) = \frac{(z-1)[((\lambda - \lambda z + \mu_2 + \alpha_1 + \beta_2)z - \mu_2)(\lambda - \lambda z + \beta_1)^* \mu_1 W_0 + \alpha_1 \beta_1 z \mu_2 S_0]}{D_2} \quad (4.6.6)$$

$$S(z) = \frac{(z-1)(\lambda - \lambda z + \beta_1)[\alpha_2 z \mu_1 W_0 + ((\lambda - \lambda z + \mu_1 + \alpha_2)z - \mu_1)^* \mu_2 S_0]}{D_2} \quad (4.6.7)$$

$$R(z) = \frac{\alpha_1 (z-1)[\alpha_2 z \mu_1 W_0 + ((\lambda - \lambda z + \mu_1 + \alpha_2)z - \mu_1)^* \mu_2 S_0]}{D_2} \quad (4.6.8)$$

where

$$D_2 = (\lambda - \lambda z + \beta_1)[((\lambda - \lambda z + \mu_1 + \alpha_2)z - \mu_1)((\lambda - \lambda z + \mu_2 + \alpha_1 + \beta_2)z - \mu_2) - \alpha_2 \beta_2 z^2] - \alpha_1 \alpha_2 \beta_1 z^2$$

As earlier, the two unknowns can be determined by applying Rouché's theorem.

To determine the unknown Q+F, one may use the normalising condition

$$W(1)+S(1)+R(1)+Q+F=1$$

For $z=1$, the equations (4.6.6),(4.6.7) and (4.5.8) are indeterminate.

Hence using L'Hospital rule one obtains

$$W(1) = \frac{\lambda\beta_2(\alpha_1 + \beta_2)(Q + F)}{D} \quad (4.6.9)$$

$$S(1) = \frac{\lambda\alpha_2\beta_1(Q + F)}{D} \quad (4.6.10)$$

$$R(1) = \frac{\lambda\alpha_1\alpha_2(Q + F)}{D} \quad (4.6.11)$$

where

$$D = \beta_1(\mu_1(\alpha_1 + \beta_2) + \alpha_2\mu_2) + \lambda(\alpha_2\beta_1 + \alpha_2\beta_1 + \beta_1\beta_2 + \alpha_1\alpha_2)$$

and using (4.6.9)-(4.6.11), and simplifying

$$Q + F = 1 - \frac{\lambda(\alpha_1\beta_1 + \alpha_2\beta_1 + \beta_1\beta_2 + \alpha_1\alpha_2)}{\beta_1(\mu_1(\alpha_1 + \beta_2) + \mu_2\alpha_2)} \quad (4.6.12)$$

This is the steady state probability that the system is in a working condition, but idle. Therefore, the utilization factor is:

$$\rho = \frac{\lambda(\alpha_1\beta_1 + \alpha_2\beta_1 + \beta_1\beta_2 + \alpha_1\alpha_2)}{\beta_1(\mu_1(\alpha_1 + \beta_2) + \mu_2\alpha_2)} \quad (4.6.13)$$

and the steady state condition is given by $\rho < 1$.

Further, if there is no failure in the service channel, then setting $\alpha_1 = \alpha_2 = 0$ in (4.6.12) and (4.6.13), the result is

$$Q + F = 1 - \frac{\lambda}{\mu}$$

$$\rho = \frac{\lambda}{\mu_1} \text{ and } \rho < 1.$$

The system state probabilities are

$$W(1) = \frac{\lambda\beta_1(\alpha_1 + \beta_2)}{\beta_1(\mu_1(\alpha_1 + \beta_2) + \mu_2\alpha_2)}$$

$$S(1) = \frac{\lambda\alpha_2\beta_1}{\beta_1(\mu_1(\alpha_1 + \beta_2) + \mu_2\alpha_2)}$$

$$R(1) = \frac{\lambda\alpha_1\alpha_2}{\beta_1(\mu_1(\alpha_1 + \beta_2) + \mu_2\alpha_2)}$$

Case 2

In addition to the condition of Case 1, if the repair rates, service rates and failure rates are identical, then setting $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$ and $\mu_1 = \mu_2 = \mu$ it follows that

$$\rho = \frac{\lambda(\alpha + \beta)^2}{\beta\mu(\beta + 2\alpha)}$$

and the steady state condition is

$$\frac{\lambda}{\mu} < \frac{\beta(\beta + 2\alpha)}{(\alpha + \beta)^2}$$

If $P(z)$ denotes the queue length probability generating function irrespective of the state of system, then $P(z) = W(z) + S(z) + R(z)$

Further, if $E(N_q)$ denotes the expected number of entities in the queue, then

$$E(N_q) = \frac{d}{dz} P(z) \Big|_{z=1}$$

Denoting

$\beta_1(\mu_1\alpha_1 + \mu_1\beta_2 + \mu_2\alpha_2) - \lambda(\alpha_1\alpha_2 + \alpha_1\beta_1 + \alpha_2\beta_1 + \beta_1\beta_2)$ by **X** and

$$\lambda^2(\alpha_1 + \alpha_2 + \beta_1 + \beta_2) - \lambda(\mu_1\alpha_1 + \mu_2\alpha_2 + 2\alpha_1\alpha_2 + \mu_1\beta_2 + \mu_1\beta_1 + \mu_2\beta_1 + 2\alpha_1\beta_1 + 2\alpha_2\beta_1 + 2\beta_1\beta_2) + \beta_1\mu_1\mu_2 + \beta_1\mu_1\alpha_1 + \mu_1\beta_1\beta_2 + \beta_1\mu_2\alpha_2$$

by **Y**.

Then

$$E(N_q) = [X[Q(-\lambda^2(\alpha_1 + \alpha_2 + \beta_1 + \beta_2) + \lambda\alpha_1\alpha_2 + \lambda\mu_2\beta_1 + \lambda\alpha_1\beta_1 + \lambda\alpha_2\beta_1 + \lambda\beta_1\beta_2) +$$

$$F(-\lambda^2(\alpha_1 + \alpha_2 + \beta_1 + \beta_2) + \lambda\alpha_1\alpha_2 + \lambda\mu_1\beta_1 + \lambda\alpha_1\beta_1 + \lambda\alpha_2\beta_1 + \lambda\beta_1\beta_2 + \lambda\mu_1\alpha_1) -$$

$$(Q + F)(\lambda\alpha_1\alpha_2 + \lambda\alpha_1\beta_1 + \lambda\alpha_2\beta_1 + \lambda\beta_1\beta_2)Y] / X^2$$

(4.6.14)

Further, if there are no failures at all, setting $\alpha_1 = \alpha_2 = 0$ and $\mu_1 = \mu_2 = \mu$ in (4.6.14), the end result is

$$E(N_q) = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\rho^2}{1 - \rho}$$

which is the expected number of entities in the queue for the M/M/1 queueing system (Saaty [53]).

4.8 Concluding remark

The model of the system considered in this chapter of the thesis emphasizes that the model is constructed mathematically in an advanced and elegant fashion. However it is suspected that its utility would be limited if it were to be used in practical applications as a result of complexity of the time dependant solution model of the system.

CHAPTER 5

AN M/G/1 QUEUEING SYSTEM WITH TWO MODES OF FAILURE

5.1 Introduction

In the previous chapter, queueing systems with an exponential service distribution were considered. However in real life situations the service distribution need not be exponential. It may be of phase type or k-Erlang as in the case of buying cosmetics and provisions in a default mental state, taking X-rays, blood test etc. in a hospital; receiving cash from a bank. Besides, exponential service is found in industry or in production or in mechanical devices. Hence the study of such systems is absolutely essential. In this chapter, an M/G/1 queueing system where the service facility is subject to failure in two modes is considered, partial and total.

5.2 Model description

In this model the inter-arrival time of entities follows a negative exponential distribution i.e. the arrival process is Poisson. The service time X_n of the n th entities follows a general distribution by with an average service rate of μ_1 when the service channel is in the normal working condition. The service rate of the n th customer follows a general distribution and the average service time is denoted by $\mu_2 (< \mu_1)$ when the service channel is in a partial failure mode. After completion of the repair of the total failure mode the channel directly changes to the normal mode without passing through the partial failure mode. If the service channel repair in the partial failure mode is completed, the system enters the normal working mode; otherwise it goes into the total failure mode. The repair times of the partial failure mode and the total failure mode are exponentially distributed with different densities. The failure times from normal to partial, and partial to total failure mode are also exponentially distributed with different densities. If the repairs in the partial failure mode are in the process of being completed, the system will not enter the total failure mode. Further, it is assumed that the repair process starts instantaneously after the completion of repair.

5.3 System description

λ : Arrival rate of entities ($\lambda > 0$).

$W_n(n,t)$: The joint probability that at time t , there are n ($n > 0$) entities in the system with elapsed service time lying between x and $x+dx$ and the service channel is in the normal working condition.

$S_n(x,t)$: The joint probability that at time t , there are n ($n > 0$) entities in the system with elapsed service time lying x and $x+dx$ and the service channel is in the partial failure mode.

$R_n(t)$: The probability that at time t , there are n ($n > 0$) entities in the system and the service channel is in the total failure mode.

$\alpha_1(\alpha_2)dt$: The first order probability that a total (partial) failure occurs during the short interval of time dt .

$\beta_1(\beta_2)dt$: The first order probability that the repair of the service channel is in the total (partial) failure mode will be complete during the short interval of time dt . As soon as the service channel is subject to total failure, it ceases to provide service instantaneously.

$\mu_1(x)dx$: The first order probability that the service will be complete in time x and $x+dx$, when the service channel is in normal working condition, given that the same was not complete till time x and is related to the density function $B_1(x)$ by the relation.

$$B_1(x) = \mu_1(x) \exp \left[- \int_0^x \mu_1(y) dy \right]$$

$\mu_2(x)dx$: The first order probability that the service will be complete in time x and $x+dx$, when the service channel is in the partial failure mode, given that the same was not complete till time x and is related to the density function $B_2(x)$ by the relation.

$$B_2(x) = \mu_2(x) \exp \left[- \int_0^x \mu_2(y) dy \right]$$

5.4 Equations governing the system

Using probability arguments, the following partial difference-differential equations are obtained.

$$\frac{\partial}{\partial x} W_n(x,t) + \frac{\partial}{\partial t} W_n(x,t) + [\lambda + \mu_1(x) + \alpha_2] W_n(x,t) = \lambda W_{n-1}(x,t) \quad ; n > 1 \quad (5.4.1)$$

$$\frac{\partial}{\partial x} W_1(x,t) + \frac{\partial}{\partial t} W_1(x,t) + [\lambda + \mu_1(x) + \alpha_2] W_0(x,t) = 0 \quad (5.4.2)$$

$$\frac{d}{dt} W_0(t) + [\lambda + \alpha_2] W_0(t) = \int_0^{\infty} W_1(x,t) \mu_1(x) dx + \beta_1 R_0(t) + \beta_2 S_0(t) \quad (5.4.3)$$

$$\frac{\partial}{\partial x} S_n(x,t) + \frac{\partial}{\partial t} S_n(x,t) + [\lambda + \mu_2(x) + \alpha_1 + \beta_2] S_n(x,t) = \lambda S_{n-1}(x,t) \quad ; n > 1 \quad (5.4.4)$$

$$\frac{\partial}{\partial x} S_1(x,t) + \frac{\partial}{\partial t} S_1(x,t) + [\lambda + \mu_2(x) + \alpha_1 + \beta_2] S_1(x,t) = 0 \quad (5.4.5)$$

$$\frac{d}{dt} S_0(t) + [\lambda + \alpha_1 + \beta_2] S_0(t) = \int_0^{\infty} S_1(x,t) \mu_2(x) dx + \alpha_2 W_0(t) \quad (5.4.6)$$

$$\frac{d}{dt} R_n(t) + [\lambda + \beta_1] R_n(t) = \lambda R_{n-1}(t) + \alpha_1 S_n(t) \quad ; n > 0 \quad (5.4.7)$$

$$\frac{d}{dt} R_0(t) + [\lambda + \beta_1] R_0(t) = \alpha_1 S_0(t) \quad (5.4.8)$$

subject to boundary conditions

$$W_n(0,t) = \int_0^{\infty} W_{n+1}(x,t)\mu_1(x)dx + \beta_2 S_n(t) + \beta_1 R_n(t) \quad ; n > 1 \quad (5.4.9)$$

$$W_1(0,t) = \int_0^{\infty} W_2(x,t)\mu_1(x)dx + \beta_2 S_1(t) + \lambda W_0(t) \quad (5.4.10)$$

$$S_n(0,t) = \int_0^{\infty} S_{n+1}(x,t)\mu_1(x)dx + \alpha_2 W_n(t) \quad ; n > 1 \quad (5.4.11)$$

$$S_1(0,t) = \int_0^{\infty} S_2(x,t)\mu_2(x)dx + \alpha_2 W_1(t) + \lambda S_0(t) \quad (5.4.12)$$

Without loss of generality it may be assumed that the system is initially empty and the server is in an idle period when the service channel is in the normal working condition $W_0(0) = 1$ and all other initial probabilities are zero. i.e.

$W_n(0) = \delta_{n,0}$ where $\delta_{n,0}$ is Kronecker's delta function

$S_n(0) = 0, \text{ for } n \geq 0$

$R_n(0) = 0, \text{ for } n \geq 0 \quad (5.4.13)$

5.5 Time dependant solution

Denoting the Laplace transform of a function $f(t)$ by $f^*(s)$, and taking the Laplace transform of equations (5.4.1) to (5.4.12) and using (5.4.13), it follows that

$$\frac{\partial}{\partial x} W_n^*(x, s) + [s + \lambda + \mu_1(x) + \alpha_2] W_n^*(x, s) = \lambda W_{n-1}^*(x, s) \quad , n > 1 \quad (5.5.1)$$

$$\frac{\partial}{\partial x} W_1^*(x, s) + [s + \lambda + \mu_1(x) + \alpha_2] W_1^*(x, s) = 0 \quad (5.5.2)$$

$$(s + \lambda + \alpha_2) W_0^*(s) = \int_0^{\infty} W_1^*(x, s) \mu_1(x) dx + \beta_1 R_0^*(s) + \beta_2 S_0^*(s) + 1 \quad (5.5.3)$$

$$\frac{\partial}{\partial x} S_n^*(x, s) + [s + \lambda + \mu_2(x) + \alpha_1 + \beta_2] S_n^*(x, s) = \lambda S_n^*(x, s) \quad , n > 1 \quad (5.5.4)$$

$$\frac{\partial}{\partial x} S_1^*(x, s) + [s + \lambda + \mu_2(x) + \alpha_2 + \beta_2] S_1^*(x, s) = 0 \quad (5.5.5)$$

$$(s + \lambda + \alpha_1 + \beta_2) S_0^*(s) = \int_0^{\infty} S_1^*(x, s) \mu_2(x) dx + \alpha_2 W_0^*(s) \quad (5.5.6)$$

$$(s + \lambda + \beta_1) R_n^*(s) = \lambda R_{n-1}^*(s) + \alpha_1 S_n^*(s) \quad , n > 0 \quad (5.5.7)$$

$$(s + \lambda + \beta_1) R_0^*(s) = \alpha_1 S_0^*(s) \quad , n > 0 \quad (5.5.8)$$

subject to boundary conditions

$$W_n^*(0, s) = \int_0^{\infty} W_{n+1}^*(x, s) \mu_1(x) dx + \beta_2 S_n^*(s) + \beta_1 R_n^*(s) \quad ; n > 1$$

(5.5.9)

$$W_1^*(0, s) = \int_0^{\infty} W_2^*(x, s) \mu_1(x) dx + \beta_2 S_1^*(s) + \beta_1 R_1^*(s) + \lambda W_0^*(s)$$

(5.5.10)

$$S_n^*(0, s) = \int_0^{\infty} S_{n+1}^*(x, s) \mu_2(x) dx + \alpha_2 W_n^*(s) \quad ; n > 1$$

(5.5.11)

$$S_1^*(0, s) = \int_0^{\infty} S_2^*(x, s) \mu_2(x) dx + \alpha_2 W_1^*(s) + \lambda S_0^*(s)$$

(5.5.12)

Defining the following probability generating functions:

$$W^*(x, s, z) = \sum_{n=1}^{\infty} W_n^*(x, s) z^n$$

$$W^*(0, s, z) = \sum_{n=1}^{\infty} W_n^*(0, s) z^n$$

$$W^*(s, z) = W_0^*(s) + \int_0^{\infty} W^*(x, s, z) dx$$

$$S^*(x, s, z) = \sum_{n=1}^{\infty} S_n^*(x, s) z^n$$

$$S^*(0, s, z) = \sum_{n=1}^{\infty} S_n^*(0, s) z^n$$

$$S^*(s, z) = S_0^*(s) + \int_0^{\infty} S^*(x, s, z) dx$$

$$R^*(s, z) = \sum_{n=0}^{\infty} R_n^*(s) z^n$$

(5.5.13)

$\sum_{n=2}^{\infty} z^n$ * (5.5.1) + z * (5.5.1) and using (5.5.13), it follows that

$$\frac{\partial}{\partial x} W^*(x, s, z) + [s + \lambda + \lambda z + \mu_1(x) + \alpha_2] W^*(x, s, z) = 0$$

Integrating this from 0 to x gives

$$W^*(x, s, z) = W^*(0, s, z) \exp \left[-(s + \lambda + \lambda z + \alpha_2)x - \int_0^x \mu_1(x) dx \right]$$

(5.5.14)

$\sum_{n=2}^{\infty} z^n$ * (5.5.4) + z * (5.5.5) and using (5.5.13), it follows that

$$\frac{\partial}{\partial x} S^*(x, s, z) + [s + \lambda - \lambda z + \mu_2(x) + \alpha_1 + \beta_2] S^*(x, s, z) = 0$$

Integrating this from 0 to x

$$S^*(x, s, z) = S^*(0, s, z) \exp \left[-(s + \lambda + \lambda z + \alpha_1 + \beta_2)x - \int_0^x \mu_2(x) dx \right] \quad (5.5.15)$$

$\sum_{n=1}^{\infty} z^n$ * (5.5.7) + z^* (5.5.8) and using (5.5.13), it follows that

$$(s + \lambda + \lambda z + \beta_1)R^*(s, z) = \alpha_1 S^*(s, z) \quad (5.5.16)$$

$\sum_{n=2}^{\infty} z^{n+1}$ * (5.5.9) + z^2 * (5.5.10) + z^* (5.5.3) and using (5.5.13), it further follows that

$$\begin{aligned} zW^*(0, s, z) + (s + \lambda + \alpha_2)zW_0^*(s) &= \int_0^{\infty} W^*(x, s, z)\mu_1(x)dx + \beta_2 zS^*(s, z) \\ &\quad + \beta_1 zR^*(s, z) + \lambda^2 zW_0^*(s) + z \end{aligned} \quad (5.5.17)$$

$\sum_{n=2}^{\infty} z^{n+1}$ * (5.5.11) + z^2 * (5.5.12) + z^* (5.5.6) and using (5.5.13), it results in

$$zS^*(0, s, z) + (s + \lambda + \alpha_1 + \beta_2)zS_0^*(s) = \int_0^{\infty} S^*(x, s, z)\mu_2(x)dx + \alpha_2 zW^*(s, z) + \lambda z^2 S_0^*(s) \quad (5.5.18)$$

Denoting $s + \lambda - \lambda z$ by η and using (5.5.14) in (5.5.17), and then integrating from 0 to ∞ , it is clear that

$$\begin{aligned} \frac{(\eta + \alpha_2)[z - \beta_1^*(\eta + \alpha_2)]}{1 - \beta_1^*(\eta + \alpha_2)} W^*(s, z) &= \frac{(\eta + \alpha_2)(z - 1)\beta_1^*(\eta + \alpha_2)}{1 - \beta_1^*(\eta + \alpha_2)} W_0^*(s) \\ &\quad + \beta_2 zS^*(s, z) + \beta_1 zR^*(s, z) + z \end{aligned} \quad (5.5.19)$$

Using (5.5.15) in (5.5.18), and then integrating from 0 to ∞ , it follows that

$$\frac{(\eta + \alpha_1 + \beta_2)[z - \beta_2^*(\eta + \alpha_2 + \beta_2)]}{1 - \beta_2^*(\eta + \alpha_2 + \beta_2)} S^*(s, z) = \frac{(\eta + \alpha_1 + \beta_2)(z - 1)\beta_2^*(\eta + \alpha_2 + \beta_2)}{1 - \beta_2^*(\eta + \alpha_2 + \beta_2)} S_0^*(s) + \alpha_2 z W^*(s, z) \quad (5.5.20)$$

$$(\eta + \beta_1)R^*(s, z) = \alpha_1 S^*(s, z) \quad (5.5.21)$$

Using (5.4.20) and (5.5.21) and simplifying

$$\begin{aligned} K(s, z)W^*(s, z) &= \frac{(\eta + \alpha_2)(z - 1)\beta_1^*(\eta + \alpha_2)}{1 - \beta_1^*(\eta + \alpha_2)} * \frac{(\eta + \alpha_1 + \beta_2)[z - \beta_2^*(\eta + \alpha_1 + \beta_2)]}{1 - \beta_2^*(\eta + \alpha_2 + \beta_2)} * (\eta + \beta_1)W_0^*(s) \\ &+ \frac{z(z - 1)(\eta + \alpha_1 + \beta_2)}{1 - \beta_2^*(\eta + \alpha_1 + \beta_2)} * [\beta_2(\eta + \beta_1) + \alpha_1\beta_1][\beta_2^*(\eta + \alpha_1 + \beta_2)]S_0^*(s) \\ &+ \frac{z(\eta + \beta_1)(\eta + \alpha_1 + \beta_2)[z - \beta_2^*(\eta + \alpha_1 + \beta_2)]}{1 - \beta_2^*(\eta + \alpha_1 + \beta_2)} \end{aligned} \quad (5.5.22)$$

$$\begin{aligned} K(s, z)S^*(s, z) &= \frac{(\eta + \alpha_2)[z - \beta_2^*(\eta + \alpha_2)]}{1 - \beta_1^*(\eta + \alpha_2)} * \frac{(\eta + \beta_1)(\eta + \alpha_1 + \beta_2)(z - 1)\beta_2^*(\eta + \alpha_1 + \beta_2)}{1 - \beta_2^*(\eta + \alpha_1 + \beta_2)} * S_0^*(s) \\ &+ \alpha_2 z (\eta + \beta_1) \left[z + \frac{(\eta + \alpha_2)(z - 1)\beta_2^*(\eta + \alpha_1 + \beta_2)}{1 - \beta_1^*(\eta + \alpha_2)} * W_0^*(s) \right] \end{aligned} \quad (5.5.23)$$

$$\begin{aligned} K(s, z)R^*(s, z) &= \frac{\alpha_1(\eta + \alpha_2)[z - \beta_1^*(\eta + \alpha_2)]}{1 - \beta_1^*(\eta + \alpha_2)} * \frac{(\eta + \alpha_1 + \beta_2)(z - 1)\beta_2^*(\eta + \alpha_1 + \beta_2)}{1 - \beta_2^*(\eta + \alpha_1 + \beta_2)} * S_0^*(s) \\ &+ \alpha_1 \alpha_2 z \left[z + \frac{(\eta + \alpha_2)(z - 1)\beta_1^*(\eta + \alpha_2)}{1 - \beta_1^*(\eta + \alpha_2)} * W_0^*(s) \right] \end{aligned} \quad (5.5.24)$$

where

$$K(s, z) = (\eta + \beta_1) \left[\frac{(\eta + \alpha_2)(z - \beta_1^*(\eta + \alpha_2))}{1 - \beta_1^*(\eta + \alpha_2)} * \frac{(\eta + \alpha_1 + \beta_2)(z - \beta_2^*(\eta + \alpha_1 + \beta_2))}{1 - \beta_2^*(\eta + \alpha_1 + \beta_2)} - \alpha_2 \beta_2 z^2 \right] - \alpha_1 \alpha_2 \beta_1 z^2$$

Since $W^*(s, z)$ is a regular function and denominator of (5.5.22) i.e. $K(s, z)$ vanishes for some z in $|z| < 1$, the numerator also vanishes for the same value of z . Applying Rouché's theorem the unknown $W_0^*(s)$ and $S_0^*(s)$ can be determined. Hence $W^*(s, z), S^*(s, z)$ and $R^*(s, z)$ can be completely determined.

Special cases

If it is assumed that the service is in the normal working condition and that partial failure of the service channel is exponential, then

$$\beta_1^*(s + \lambda - \lambda z + \alpha_2) = \frac{\mu_1}{s + \lambda - \lambda z + \alpha_2 + \mu_1}$$

$$\beta_2^*(s + \lambda - \lambda z + \alpha_1 + \beta_2) = \frac{\mu_2}{s + \lambda - \lambda z + \alpha_1 + \beta_2 + \mu_2}$$

Consequently equations (5.5.22) to (5.5.24) would become

$$\begin{aligned} K_1(s, z)W^*(s, z) &= [(\eta + \alpha_1 + \beta_1 + \mu_1)z - \mu_2]^*(\eta + \beta_1)^*[z + (z - 1)\mu_1W_0^*(s)] \\ &\quad + z(z - 1)^*[\beta_2(\eta + \beta_1) + \alpha_1\beta_1]^*\mu_2S_0^*(s) \end{aligned} \quad (5.5.25)$$

$$\begin{aligned} K_1(s, z)S^*(s, z) &= \alpha_2 z(\eta + \beta_1)^*[z + \mu_1(z - 1)W_0^*(s)] \\ &\quad + [(\eta + \alpha_2 + \mu_1)z - \mu_1]^*(\eta + \beta_1)\mu_2(z - 1)S_0^*(s) \end{aligned} \quad (5.5.26)$$

$$K(s, z)R^*(s, z) = \alpha_1 \alpha_2 z[z + \mu_1(z - 1)W_0^*(s)] + \alpha_2 [(\eta + \alpha_2 + \mu_1)z - \mu_1]\mu_2(z - 1)S_0^*(s) \quad (5.5.27)$$

where

$$K_1(s, z) = (s + \lambda - \lambda z + \beta_1) * [(s + \lambda - \lambda z + \alpha_2) - \mu_1] *$$

$$[(s + \lambda - \lambda z + \alpha_1 + \beta_2 + \mu_2)z - \mu_2] - \alpha_2 \beta_2 z^2 - \alpha_1 \alpha_2 \beta_1 z^2$$

Since $W^*(s, z)$ is a regular function $W_0^*(s)$ and $S_0^*(s)$ can be determined as before.

5.6 Steady state solution

In taking the steady state probability corresponding to $W_n(t)$, $S_n(t)$ $R_n(t)$ as W_n , S_n , R_n , and the corresponding probability generations as $W(z)$, $S(z)$ and $R(z)$ by using the Tauberian property (see Widder [47]).

$$\lim_{s \rightarrow 0} s f^*(s) = \lim_{t \rightarrow \infty} f(t)$$

if the limit on the right exists.

The steady state solution can be obtained from (5.5.22), (5.5.23) and (5.5.24) as

$$K(z)W(z) = \frac{(\eta_1 + \alpha_2)(z-1)\beta_1(\eta_1 + \alpha_2)}{1 - \beta_1(\eta_1 + \alpha_2)} * \frac{(\eta_1 + \alpha_1 + \beta_2) * [z - \beta_2(\eta_1 + \alpha_1 + \beta_2)]}{1 - \beta_2(\eta_1 + \alpha_1 + \beta_2)} * (\eta_1 + \beta_1)W_0$$

$$+ \frac{z(z-1) * (\eta_1 + \alpha_1 + \beta_2)}{1 - \beta_1(\eta_1 + \alpha_1 + \beta_2)} * [\beta_2(\eta_1 + \beta_1) + \alpha_1 \beta_1] * S_0$$

(5.6.1)

$$K(z)S(z) = \alpha_2 z (\eta_1 + \beta_1) * \frac{(\eta_1 + \alpha_2)(z-1)\beta_1(\eta_1 + \alpha_2)}{1 - \beta_1(\eta_1 + \alpha_2)} * W_0 + \frac{(\eta_1 + \alpha_2)[z - \beta_1(\eta_1 + \alpha_2)]}{1 - \beta_1(\eta_1 + \alpha_2)}$$

$$* \frac{(\eta_1 + \alpha_1 + \beta_2)(z-1)\beta_2(\eta_1 + \alpha_1 + \beta_2)}{1 - \beta_2(\eta_1 + \alpha_1 + \beta_2)} * (\eta_1 + \beta_1)S_0$$

(5.6.2)

$$\begin{aligned}
 K(z)R(z) = & \alpha_1 * \frac{(\eta_1 + \alpha_2)[z - \beta_1(\eta_1 + \alpha_2)]}{1 - \beta_1(\eta_1 + \alpha_2)} * \frac{(\eta_1 + \alpha_1 + \beta_2)(z-1)\beta_2(\eta_1 + \alpha_1 + \beta_2)}{1 - \beta_2(\eta_1 + \alpha_1 + \beta_2)} * S_0 \\
 & + \alpha_1 \alpha_2 z * \frac{(\eta_1 + \alpha_2)(z-1)\beta_1(\eta_1 + \alpha_2)}{1 - \beta_1(\eta_1 + \alpha_2)} * W_0
 \end{aligned}
 \tag{5.6.3}$$

where $\eta_1 = \lambda - \lambda z$ and

$$K(z) = (\eta_1 + \beta_1) * \left[\frac{(\eta_1 + \alpha_2)[z - \beta_1(\eta_1 + \alpha_2)]}{1 - \beta_1(\eta_1 + \alpha_2)} * (\eta_1 + \alpha_1 + \beta_2) * \right. \\
 \left. \frac{[z - \beta_2(\eta_1 + \alpha_1 + \beta_2)]}{1 - \beta_2(\eta_1 + \alpha_1 + \beta_2)} - \alpha_2 \beta_2 z^2 \right] - \alpha_1 \alpha_2 \beta_1 z^2$$

As earlier, applying Rouche's theorem, the two unknowns W_0 and S_0 may be determined.

5.7 Some special cases

Case 1

If the service times in the normal working mode and the partial failure mode are exponential, then

$$\beta_1(\lambda - \lambda z + \alpha_2) = \frac{\mu_1}{\lambda - \lambda z + \alpha_2 + \mu_1}$$

$$\beta_2(\lambda - \lambda z + \alpha_1 + \beta_2) = \frac{\mu_2}{\lambda - \lambda z + \alpha_1 + \beta_2 + \mu_2}$$

Therefore equations (5.6.1), (5.6.2) and (5.6.3) become

$$\begin{aligned}
 K_1(z)W(z) = & (z-1) * [(\eta_1 + \alpha_1 + \beta_2 + \mu_2)z - \mu_2] * (\eta_1 + \beta_1)\mu_1 W_0 \\
 & + z(z-1) * [\beta_2(\eta_1 + \beta_1) + \alpha_1 \beta_1] * \mu_2 S_0
 \end{aligned}
 \tag{5.6.4}$$

$$K_1(z)S(z) = \alpha_2 z (\eta_1 + \beta_1)^* (z-1) \mu_1 W_0 + (\eta_1 + \beta_1)^* [(\eta_1 + \alpha_2 + \mu_1)z - \mu_1]^* (z-1) \mu_2 S_0$$

(5.6.5)

$$K_1(z)R(z) = \alpha_1 \alpha_2 z (z-1) \mu_1 W_0 + \alpha_1 [(\eta_1 + \alpha_2 + \mu_1)z - \mu_1]^* (z-1) \mu_2 S_0$$

(5.6.6)

where

$$K_1(z) = (\eta_1 + \beta_2)^* [(\eta_1 + \alpha_2 + \mu_1)z - \mu_1]^* ((\eta_1 + \alpha_1 + \beta_2 + \mu_2)z - \mu_2) - \alpha_2 \beta_2 z^2] - \alpha_1 \alpha_2 \beta_1 z^2$$

The two unknowns W_0 and S_0 can be determined as before

Case 2

If there is no failure at all, then $\alpha_1 = \alpha_2 = 0$. Using this in (5.6.1) to (5.6.3), it follows that

$$W(z) = \frac{(z-1)\beta_1(\lambda - \lambda z)}{z - \beta_1(\lambda - \lambda z)} W_0$$

$$S(z) = \frac{(z-1)\beta_2(\lambda - \lambda z)}{z - \beta_2(\lambda - \lambda z)} S_0$$

$$R(z) = 0$$

The identical forms of $W(z)$ and $S(z)$ confirm that the service time distributions in the normal working condition and the partial failure mode are the same when there is no failure at all.

5.8 Concluding remarks

At this juncture of the modelling process one may admire the modelling elegance achieved despite the attendant intricacy of applications in practice.

CHAPTER 6

CHAOS THEORY BASED MODELS OF SIMPLE SYSTEMS OF CONGESTION

A modified version of this Chapter was presented at a Southern African Institute for Industrial Engineering Conference, 2004.

6.1 Introduction

When embarking on the use of Chaos Theory in modelling simple Systems of Congestion it is considered prudent to provide a benchmark based on the **classical M/M/1 queue** to serve as the necessary introductory backdrop to the investigation:

6.1.1 The classical Poisson arrival system

6.1.1.1 The general modelling approach

Modelling a completely random arrival process traditionally involves using the Poisson distribution (negative exponentially distributed inter-arrival times) as the cornerstone of analysis in generating an ordered sequence of arrival events. This implies that the arrival system is treated as being Markovian.

If arrivals are considered to occur within a temporal sequence of equal time intervals, the cumulative Poisson distribution can adequately generate arrivals with the passage of time.

The Poisson distribution of arrivals is given by

$$P_n = \frac{\lambda^n e^{-\lambda}}{n!} \quad n=0,1,2,\dots \quad \text{and } \lambda > 0 \quad (6.1)$$

where n = no. of arrivals in a given time interval

λ = average no. of arrivals in the temporal sequence of time intervals

An example of the generation of a Poisson based arrival process for $\lambda = 8$ over 200 one minute time intervals is shown in Fig. 6.1.1.

The generation of the arrival process is driven by a random number generator. The adequacy of the generation process is demonstrated by the achieved results.

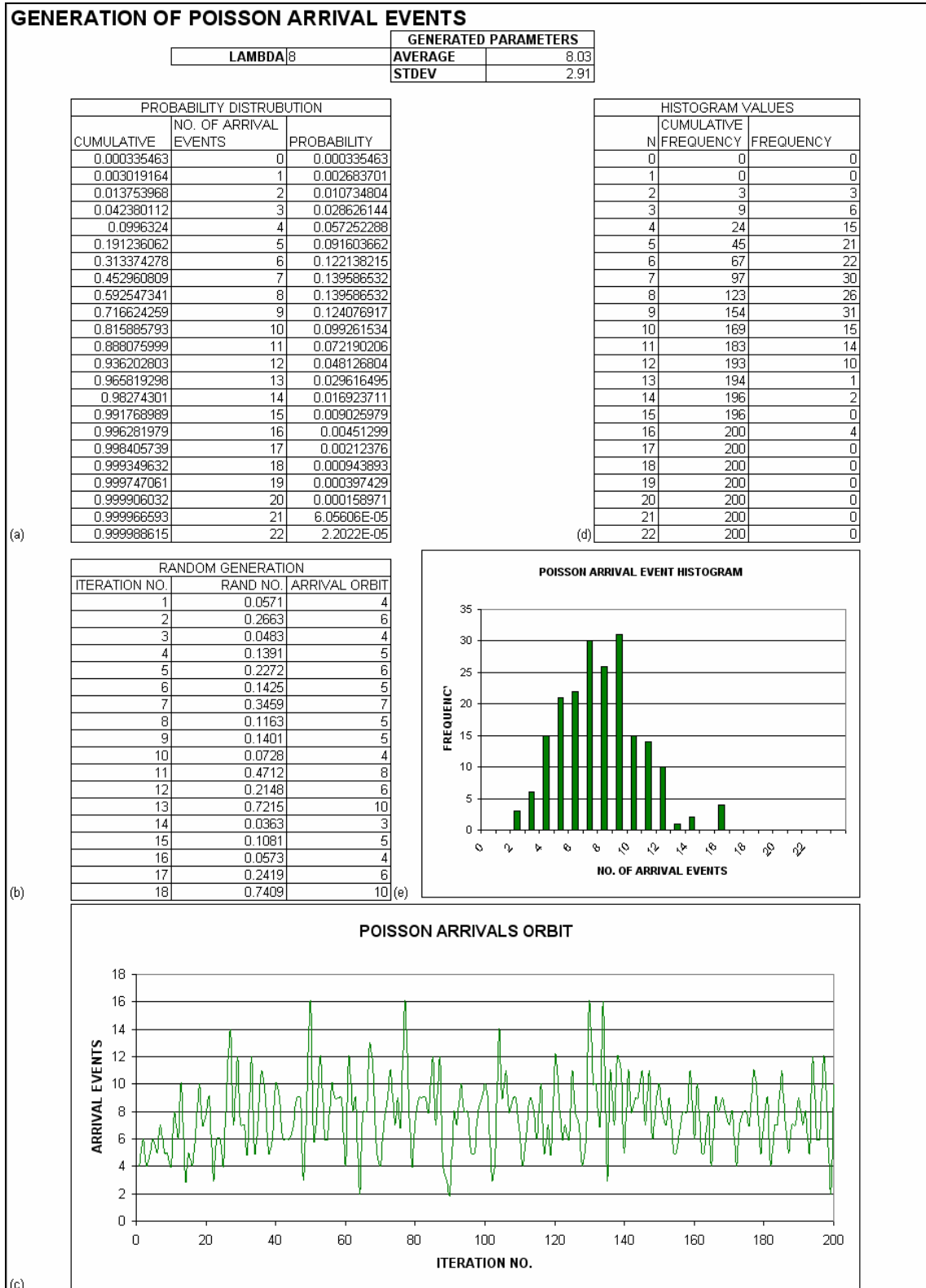


Fig. 6.1.1 GENERATION OF THE ORBIT OF POISSON ARRIVAL EVENTS

6.1.2 The classical exponential service system

6.1.2.1 The general modelling approach

In a similar fashion to the modelling of completely random arrivals (See par. 6.1.1), the modelling of a single completely random service process often involves the Poisson distribution (negative exponentially distributed service times) in generating an ordered sequence of service events. This implies that the service system is treated as being Markovian.

If consecutive service events are considered to occur within a temporal sequence of equal time intervals (synchronously identical to the arrival time intervals) the cumulative Poisson distribution can adequately generate service events with the passage of time.

The Poisson distribution of service events is given by

$$P_n = \frac{\mu^n e^{-\mu}}{n!} \quad n=0,1,2,\dots \text{ and } \mu > 0 \quad (6.2)$$

where n = no. of service events **offered** in a given time interval
 μ = average no. of service events **offered** in the temporal sequence of time intervals

An example of the generation of the service process for $\mu = 10$ over 200 one minute time intervals is shown in Fig. 6.1.2

The generation of the service process is driven by a random number generator. The adequacy of the generation process is demonstrated by the achieved results.

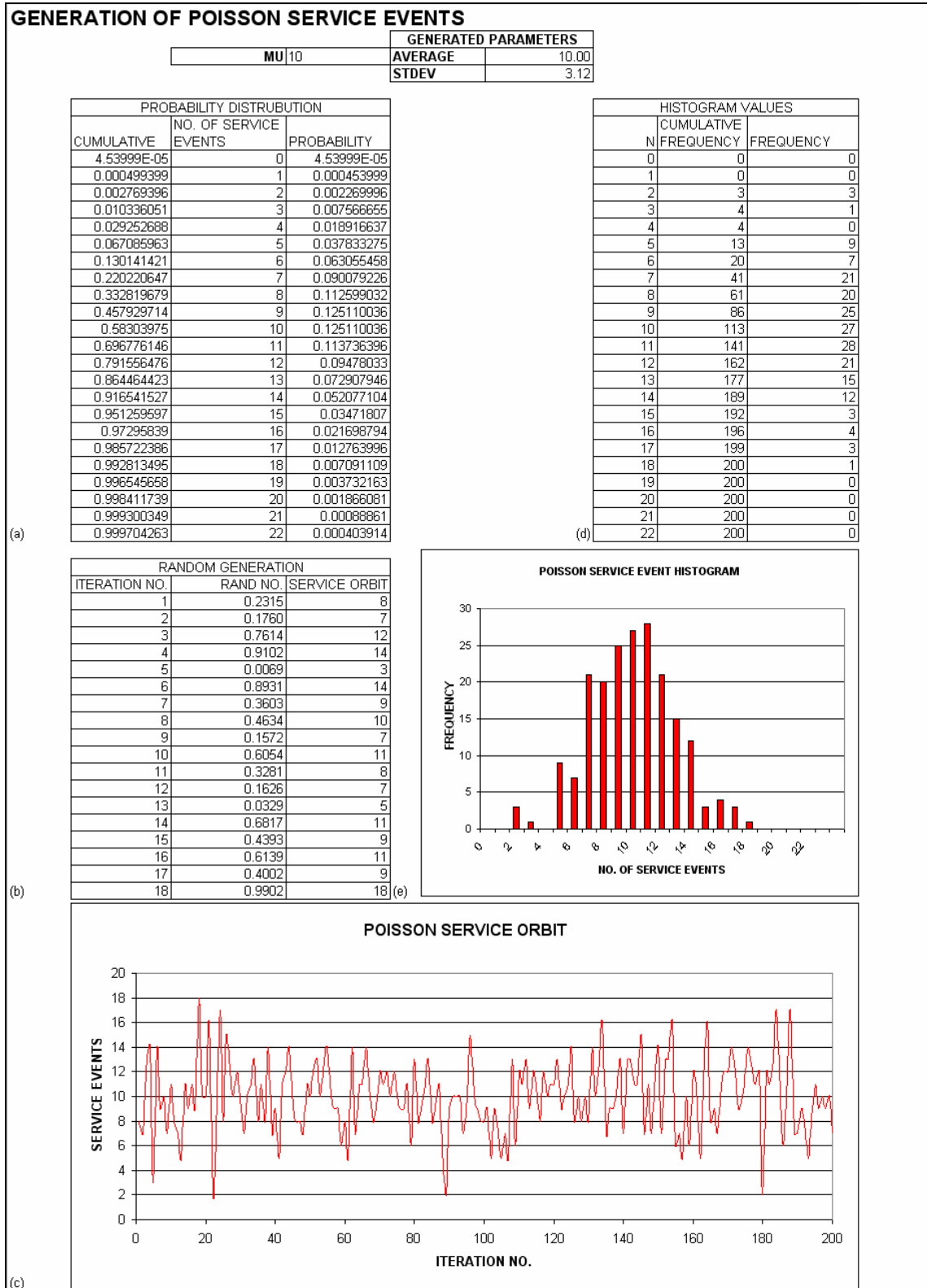


Fig. 6.1.2 GENERATION OF THE ORBIT OF POISSON SERVICE EVENTS

6.1.3 The classical M/M/1 queue

6.1.3.1 The general modelling approach

From the point of view of analyzing Systems of Congestion recent significant developments have addressed approximations and numerical techniques in manipulating steady-state and non steady-state systems. With this in mind and obeying the requirement of **model simplicity and robustness**, the concept of studying ***a temporal sequence of equal time intervals*** plays a central role in modelling the M/M/1 queueing system as it deals with arrival and service events. The novelty of the proposed system model is based on the flow of entities as follows: During a given time interval (t) the number of entities in the system at the end of time t equals the number of system entities at the beginning of time t plus the number of arrival events in time t minus the number of service events offered (available) in time t , i.e.

$$\begin{aligned} \text{No. in system at the end of } (t + \Delta t) = & \quad [\text{No. in system at the beginning of } t] + \\ & \quad [\text{No. of arrival events in } \Delta t] - \\ & \quad [\text{No. of service events offered in } \Delta t] \quad (6.3) \end{aligned}$$

The model calculates the average number in the system during the interval (Δt) as follows:

If the number of service events offered in Δt exceeds the sum of the number in the system at t plus the number of arrival events in Δt , the average number of units in the system during Δt is given by:

$$\begin{aligned} & \quad [(\text{No. at } t + \text{No. of arrival events in } \Delta t) / 2] \times \\ & \quad [(\text{No. at } t + \text{No. of arrival events in } \Delta t) / \\ & \quad (\text{No. of service events in } \Delta t)] \quad (6.4) \end{aligned}$$

If the sum of the number in the system at t plus the number of arrival events in Δt exceeds the number of service events offered in Δt the average number of units in the system during Δt is given by:

$$\frac{[(\text{No. at } t + \text{No. of arrival events in } \Delta t) + (\text{No. at } t + \text{No. of arrival events in } \Delta t - \text{No. of service events offered in } \Delta t)]}{2} \quad (6.5)$$

A model of the events which take place within a time interval is an example of a highly simplified model of a deterministic **instantaneous** replenishment inventory system which allows shortages to occur during the time interval i.e. when some service events are analogously on offer but not used within the interval as a result of insufficient arrivals.

One may speculate that such an elementary model does not meet the requirement of mathematical elegance, or that an attempt is being made to approach the modelling problem pragmatically to avoid immersion into higher mathematics. At this juncture of the modelling process one should await the results which follow, results which are based on further development of the system modelling approach before prematurely judging the merit of the model.

The resulting orbit of number of entities in the system which is obtained by merging the arrival and service processes used in sections 6.1.1 and 6.1.2 does not deliver the required theoretical mean number in the system for the temporal sequence of time intervals. To compensate for this state of affairs the data stream of system entities must be manipulated by means of a **designer equation**(Appendix B) The designer equation is a necessary adjunct to equations (6.3) and (6.4) to shape the data stream of system entities to reflect reality of system operation modelled via passing reference to interevent times (arrival and service).

The generation of the system state with the passage of time is driven by random number generation and is shown in Fig. 6.1.3. The adequacy of the generation process, which includes the use of a designer equation, is demonstrated by achieved results.

The model can now be used in spreadsheet form for the analysis of steady state and transient operation of an M/M/1 queue. Consequently it may also serve as a ***touchstone*** in evaluating the use of the Chaos based models which follow. One should however not lose sight of the fact that the Poisson/exponential assumption is a mathematical concept and that no real process can be expected to constantly be in agreement with it. It is however heartening to know that use of it as a benchmark will lead to a conservative evaluation of alternative modelling methods.

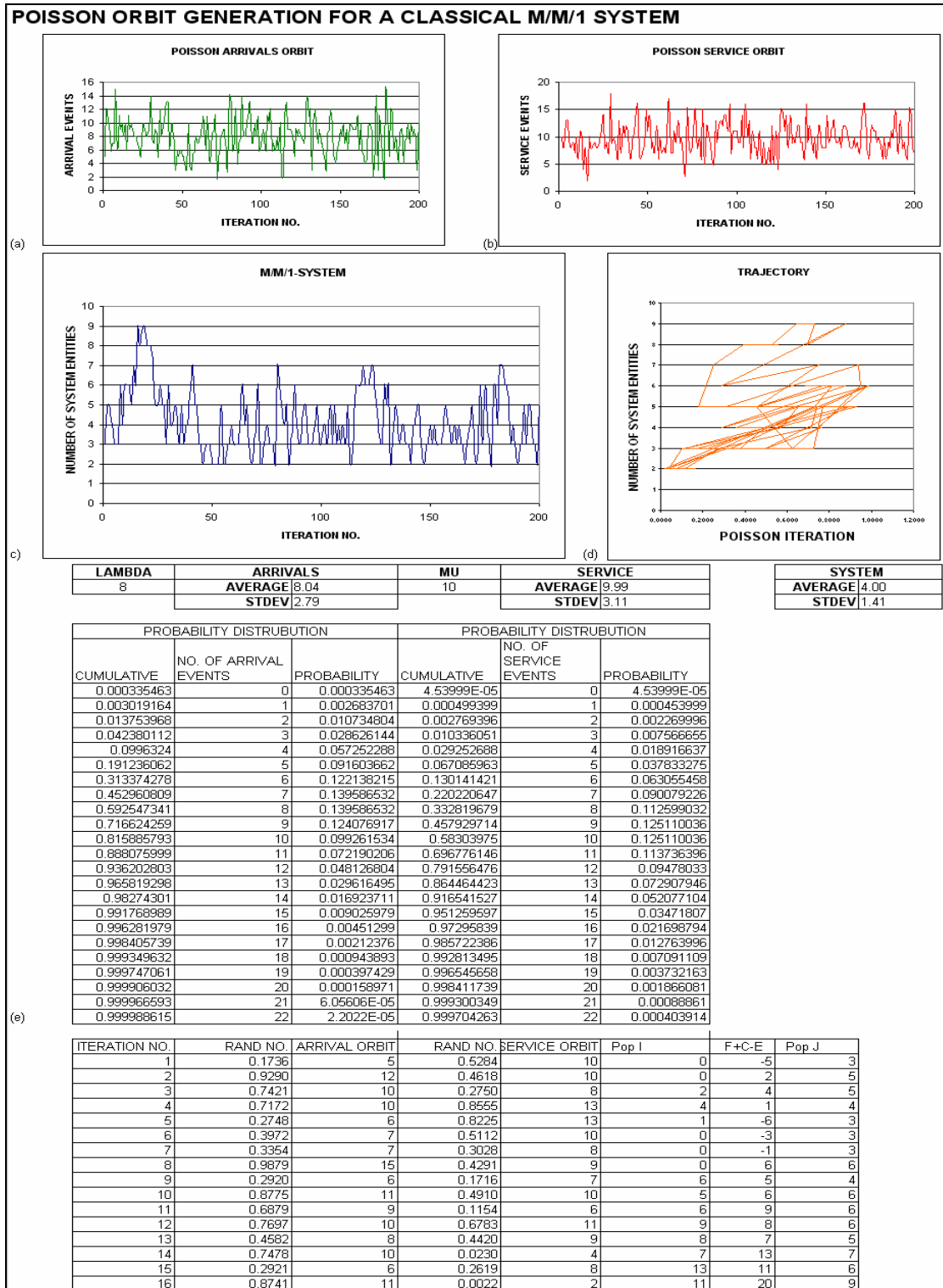


Fig. 6.1.3 GENERATION OF THE ORBIT OF A CLASSICAL POISSON M/M/1 SYSTEM

6.2 Introduction to Chaos generation

Having established the classical M/M/1 queue as the benchmark for the general use of chaos based models the research may progress to create the relevant method of analysis for a chaos driven single channel queue with an average arrival rate of $\lambda=8$ and an average service rate of $\mu=10$. The initial research efforts are based on:

- Verhulst logistic mapping
- Weibull based mapping
- Trigonometric mapping

Fig. 1.1 serves as an example which displays the nature of iterative mapping of the Verhulst type.

6.2.1 The Verhulst generated arrival system

6.2.1.1 The general modelling approach

In attempting to emulate arrival events of an M/M/1 system by using the Verhulst logistic generation method it is necessary to at least achieve “Poissonness” (Grosh [4]) by:

- selecting an appropriate logistic parameter to ensure that “chaotic” randomness is generated, and
- creating an emulated mean and standard deviation which are related as in a Poisson distribution.

At this juncture it must be emphasized that the use of a designer equation (Appendix A) becomes mandatory to fashion the data stream of generated arrivals effectively.

An example of the temporal sequence of the number of arrival events in equal time intervals for an average arrival rate of $\lambda=8$ as generated by a Verhulst logistic model over 200 one minute time intervals is shown in Fig. 6.2.1. The adequacy of the generation process is demonstrated by the achieved results.

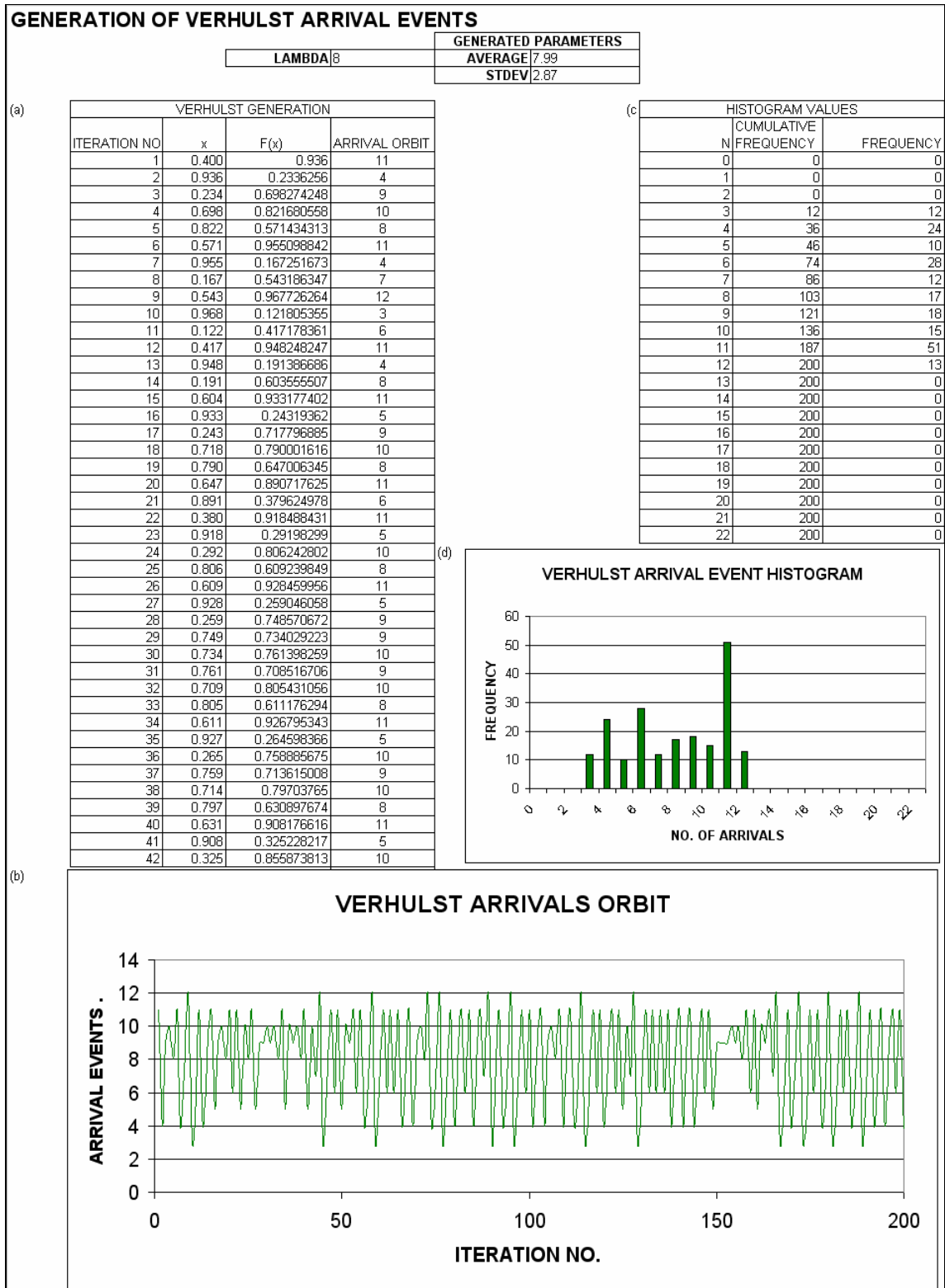


Fig. 6.2.1 GENERATION OF THE ORBIT OF VERHULST ARRIVAL EVENTS

6.2. The Verhulst generated service system

6.2.2.1 The general modelling approach

In attempting to emulate service events of an M/M/1 system by using the Verhulst generation method it is necessary as in the case of arrival events to at least achieve “Poissonness” (Grosh [4]) by:

- selecting an appropriate logistic parameter to ensure that “chaotic” randomness is generated, and
- creating an emulated mean and standard deviation which are related as in a Poisson distribution.

An example of the temporal sequence of the number of service events in 200 one minute equal time intervals for an average service rate of $\mu=10$ as generated by a Verhulst logistic model is shown in Fig. 6.2.2. The adequacy of the generation process is demonstrated by the achieved results.

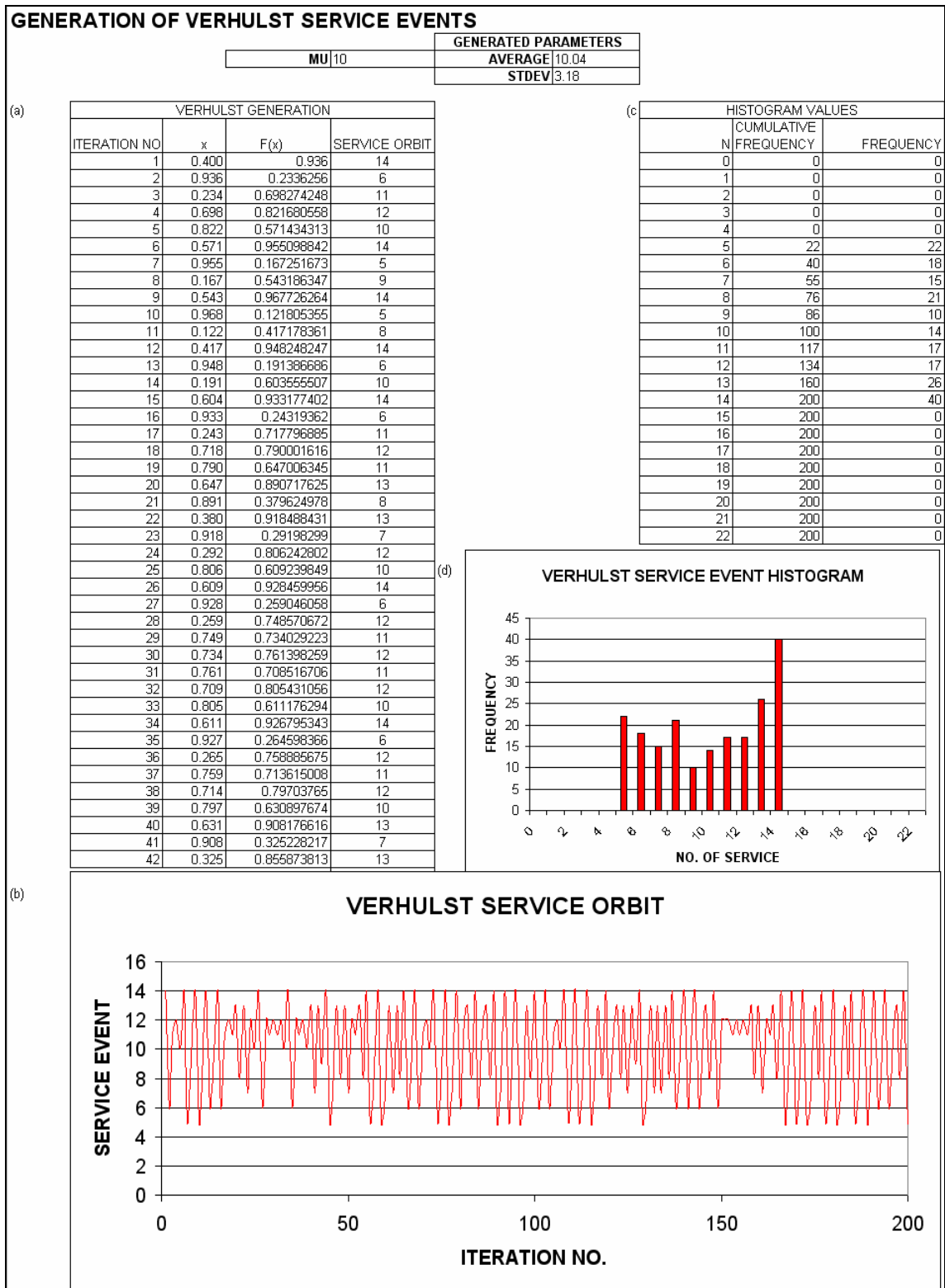


Fig. 6.2.2 GENERATION OF THE ORBIT OF VERHULST SERVICE EVENTS

6.2.3 The Verhulst generated single channel queue

6.2.3.1 The general modelling approach

If as at the outset of this chapter considering the use of chaos generation methods to model a single channel queueing system by means of approximations and numerical techniques is heeded, and robustness and simplicity of modelling is to be achieved, the concept of studying a temporal sequence of equal time intervals which accommodate arrival and service events is justified.

As in the case of the classical M/M/1 queue analysis of par. 6.1.3.1 the Verhulst system model makes use of the highly simplified model described in equation (6.3) which also requires manipulation of the generated data stream by ***designer equations***.

The generation of the system state with the passage of time is driven by chaos iterative generation and is shown in Fig 6.2.3. The adequacy of the generation process, which includes the use of a designer equation, is demonstrated by the achieved results.

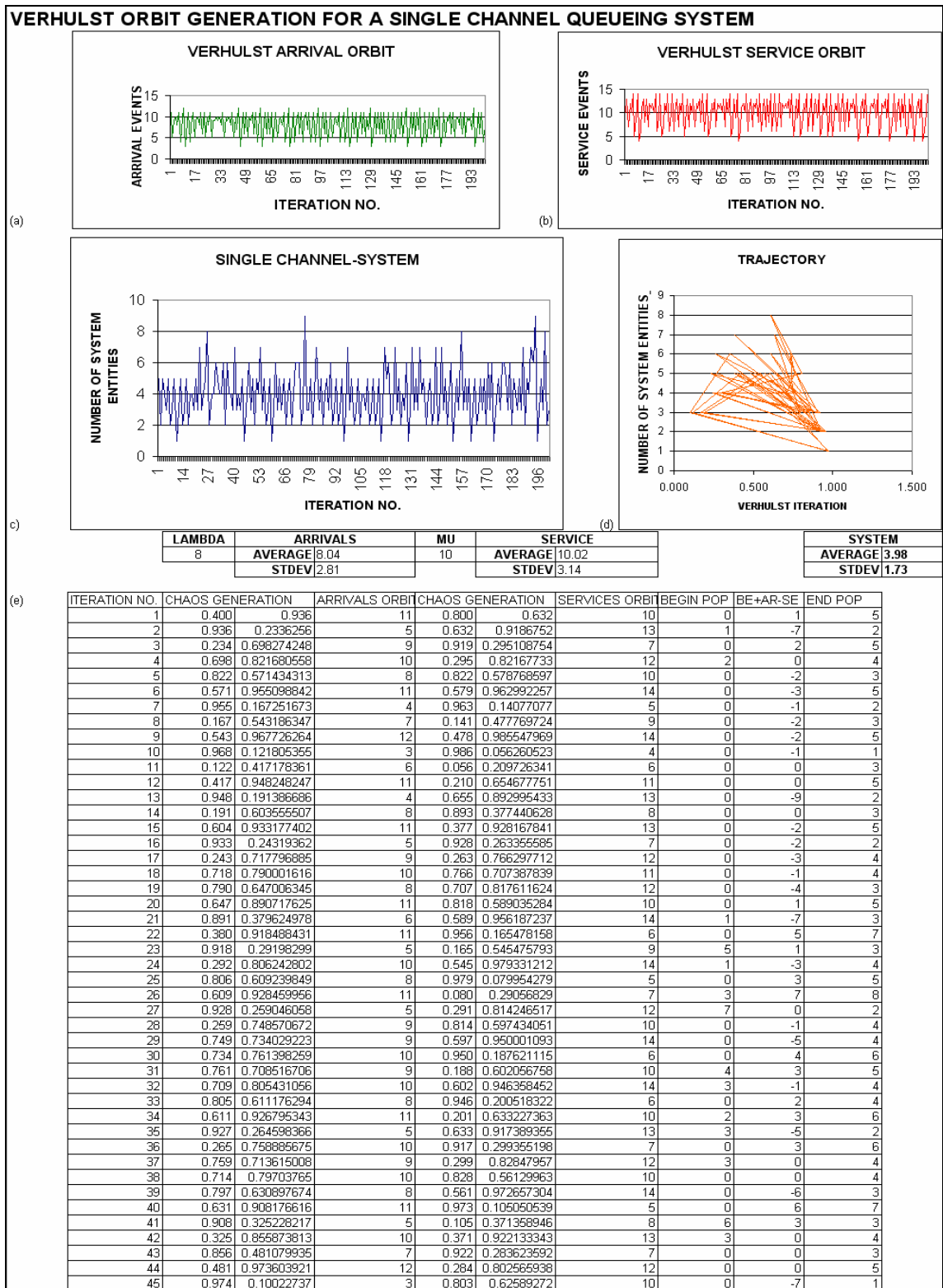


Fig. 6.2.3 GENERATION OF THE ORBIT OF A VERHULST SINGLE CHANNEL QUEUEING SYSTEM

6.2.4 Benchmarking the Verhulst generated single channel queue model

Comparison of the Poisson M/M/1 and Verhulst methods of generating system dynamics as depicted in Figs. 6.1.3 and 6.2.3 respectively results in

- achieving equivalence of mean and standard deviation values for the arrival and service processes,
- achieving graphical plausibility of system orbit likeness i.e. applying the TLAR criterion (“that looks about right”) in comparing the two system entity orbits.

No quantitative justification for “Poissonness” other than the foregoing parameter determination and application of the TLAR plausibility criterion has been carried out.

As a further matter of interest the Verhulst methods of generating system dynamics over 200 one minute intervals are shown in Fig. 6.2.4 for a **general service distribution** queueing system for $\lambda = 8, \mu = 10$ and $\sigma = 0.010$. The average number of entities in the system is given by:

$$L = L_q + \rho \quad (6.6)$$

$$\text{for } L_q = \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} \quad (6.7)$$

$$\text{and } \rho = \frac{\lambda}{\mu} \quad (6.8)$$

where:

L = the average number of entities in the system

L_q = the average number of entities in the queue

ρ = the traffic intensity

λ = the average number of arrivals entering the system per unit time

σ^2 = the variance of the service time

μ = the average number of services offered per unit time

The results indicate

- achieving equivalence of mean and standard deviation values for the arrival and service processes,
- achieving graphical plausibility of system orbit likeness i.e. applying the TLAR criterion in comparing the two system entity orbits.

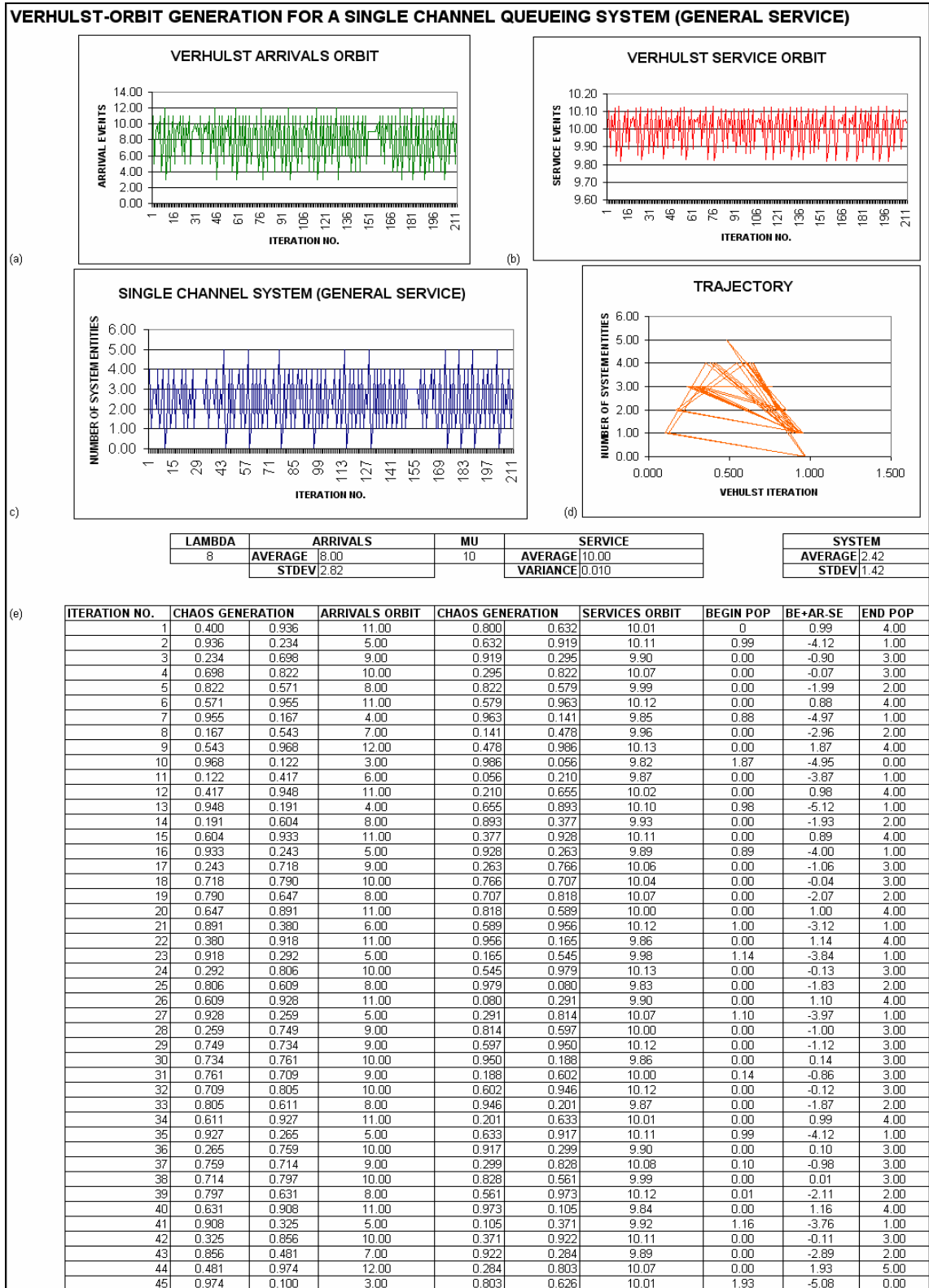


Fig. 6.2.4 GENERATION OF THE ORBIT OF A VERHULST QUEUEING SYSTEM (GENERAL SERVICE DISTRIBUTION)

6.2.5 Extending the Verhulst generated single channel queue model to deal with variable traffic intensity

Having achieved a degree of likeness greater than a scant semblance between the classical M/M/1 and Verhulst queueing system one may embark on extending the Verhulst model to include a range of traffic intensities which may prove to be beneficial in analysing the transient (dynamic) and steady state operation of a single channel queue.

Consequently the Verhulst queueing system has been extended to include a range of average arrival rates ($0.2 \leq \lambda < 1$) for an average service rate $\mu = 10$.

$$\text{Traffic intensity } \rho = \frac{\lambda}{10}$$

An example of a Verhulst generated single channel queue for a chronological sequence of values of λ of 9.8; 8.0; 9.5; and 7.0 over 200 one minute intervals is shown in Fig. 6.2.5.

Each of the chronological values of λ are employed for four consecutive epochs of 50 consecutive intervals.

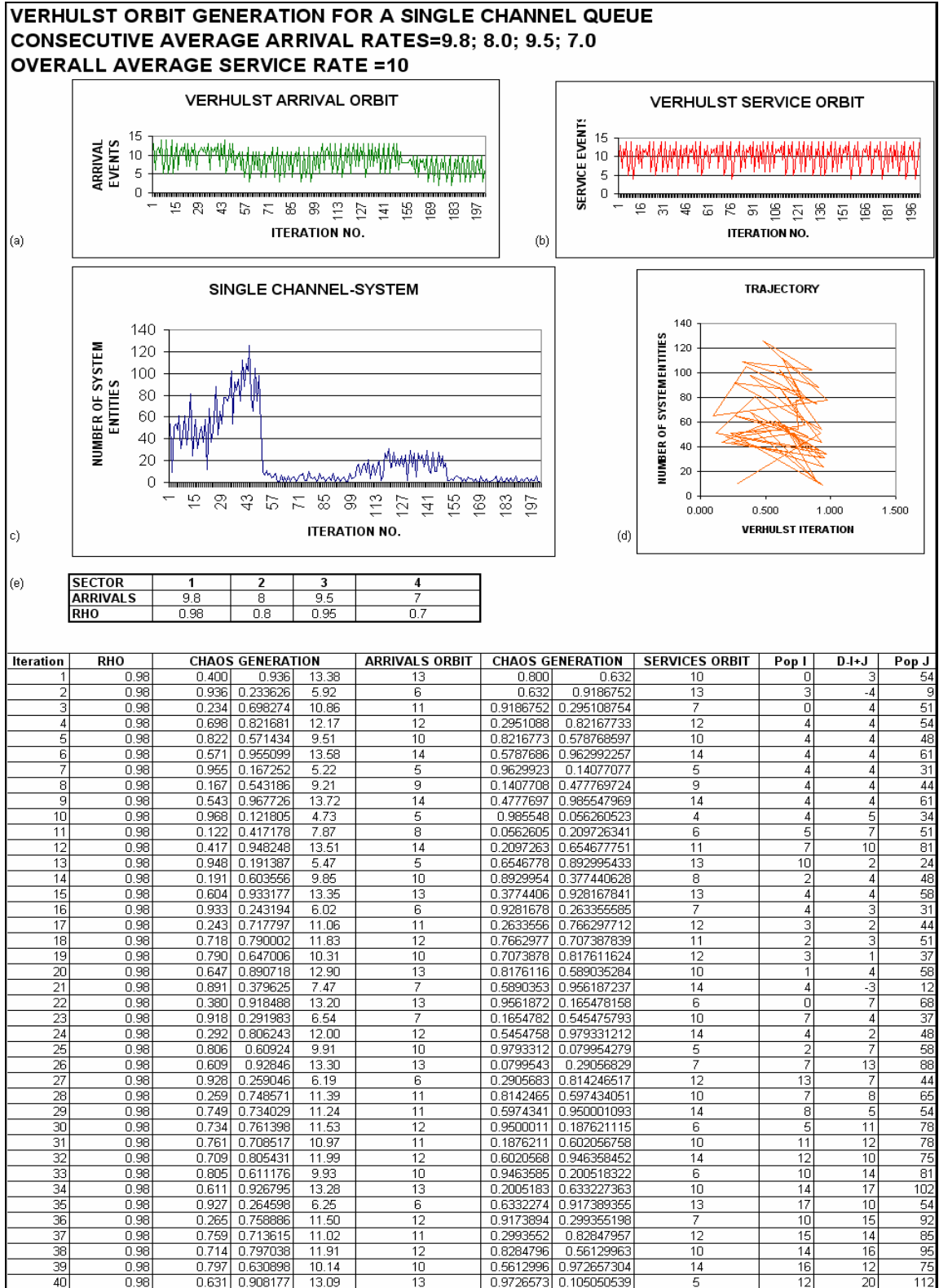


Fig. 6.2.5 GENERATION OF THE ORBIT OF AN EXTENDED VERHULST SINGLE CHANNEL QUEUE

The system orbit generated for a total of 200 one minute consecutive intervals unambiguously displays how the system behaves dynamically in a natural sense to being subjected to step functions in average arrival rate, albeit that the transitions from one steady state to a following steady state are ephemeral.

The extended model is versatile and amenable to use of many values of traffic intensity which may occur in practical situations. Such traffic intensity values may be selected a priori by external control or by automatically adjusting the arrival and service processes by means of internal system feedback mechanisms.

6.3 Further examples of Chaos generation

The introduction to Chaos generating methods described in par. 6.2 makes mention of other methods of mapping which may be considered as alternatives to Verhulst logistic mapping i.e.

- Weibull based mapping, and
- Trigonometric mapping. (Stewart [38])

The general modelling approach used for the generation of orbits for the two above- mentioned methods of mapping slavishly follows the underlying mathematical regimen employed in par.6.2.

The results which have been achieved are shown in :

- Fig. 6.3.1: Generation of the orbit of Weibull arrival events
- Fig. 6.3.2: Generation of the orbit of Weibull service events
- Fig. 6.3.3: Generation of the orbit of a Weibull single channel queueing system
- Fig. 6.3.4: Generation of the orbit of Sin arrival events
- Fig. 6.3.5: Generation of the orbit of Sin service events
- Fig. 6.3.6: Generation of the orbit of a Sin single channel queueing system.

The orbits shown have all been prepared for an average arrival rate of $\lambda = 8$ and an average service rate of $\mu = 10$.

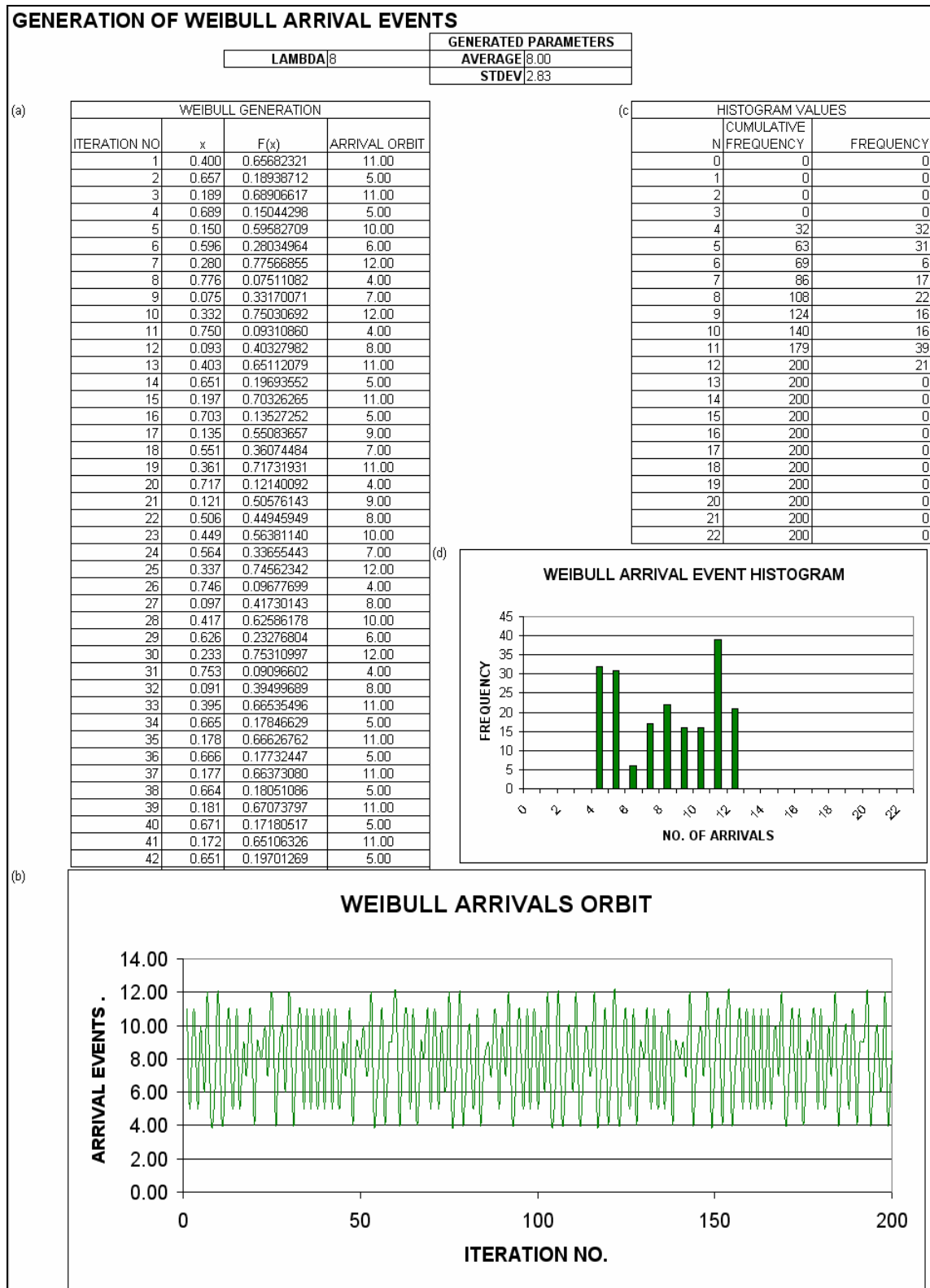


Fig. 6.3.1 GENERATION OF THE ORBIT OF WEIBULL ARRIVAL EVENTS

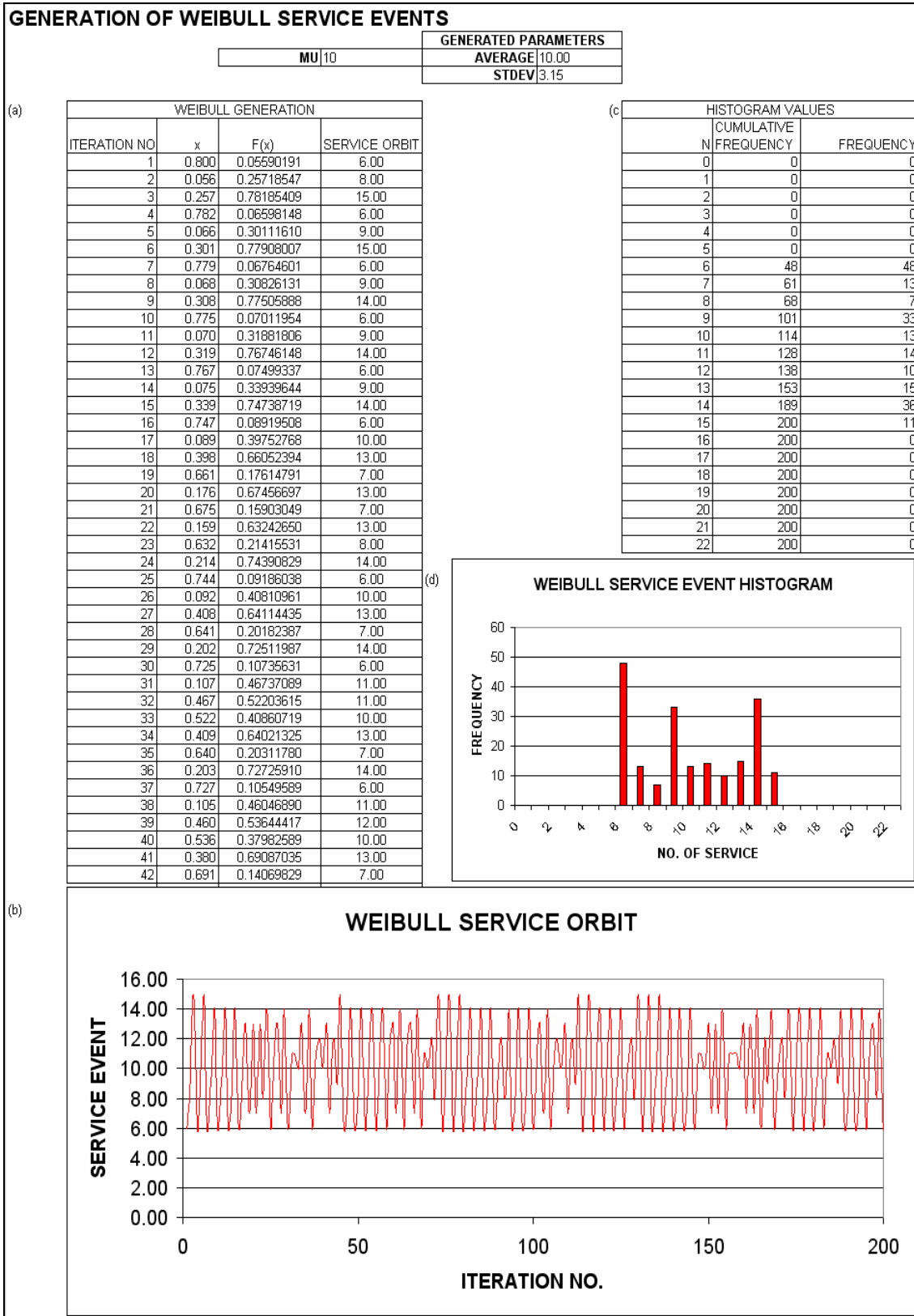


Fig. 6.3.2 GENERATION OF THE ORBIT OF WEIBULL SERVICE EVENTS

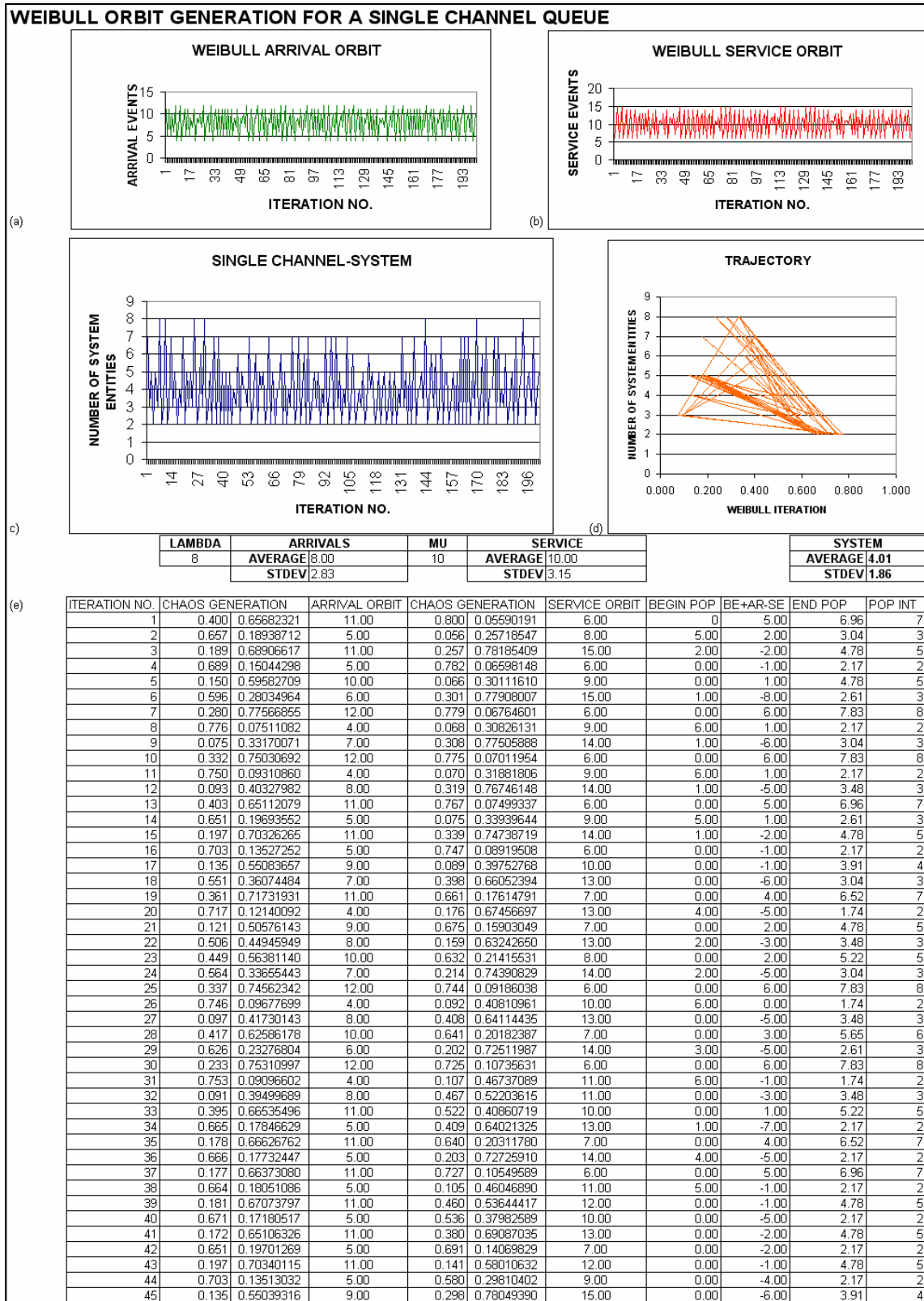


Fig. 6.3.3 GENERATION OF THE ORBIT OF A WEIBULL SINGLE CHANNEL QUEUEING SYSTEM

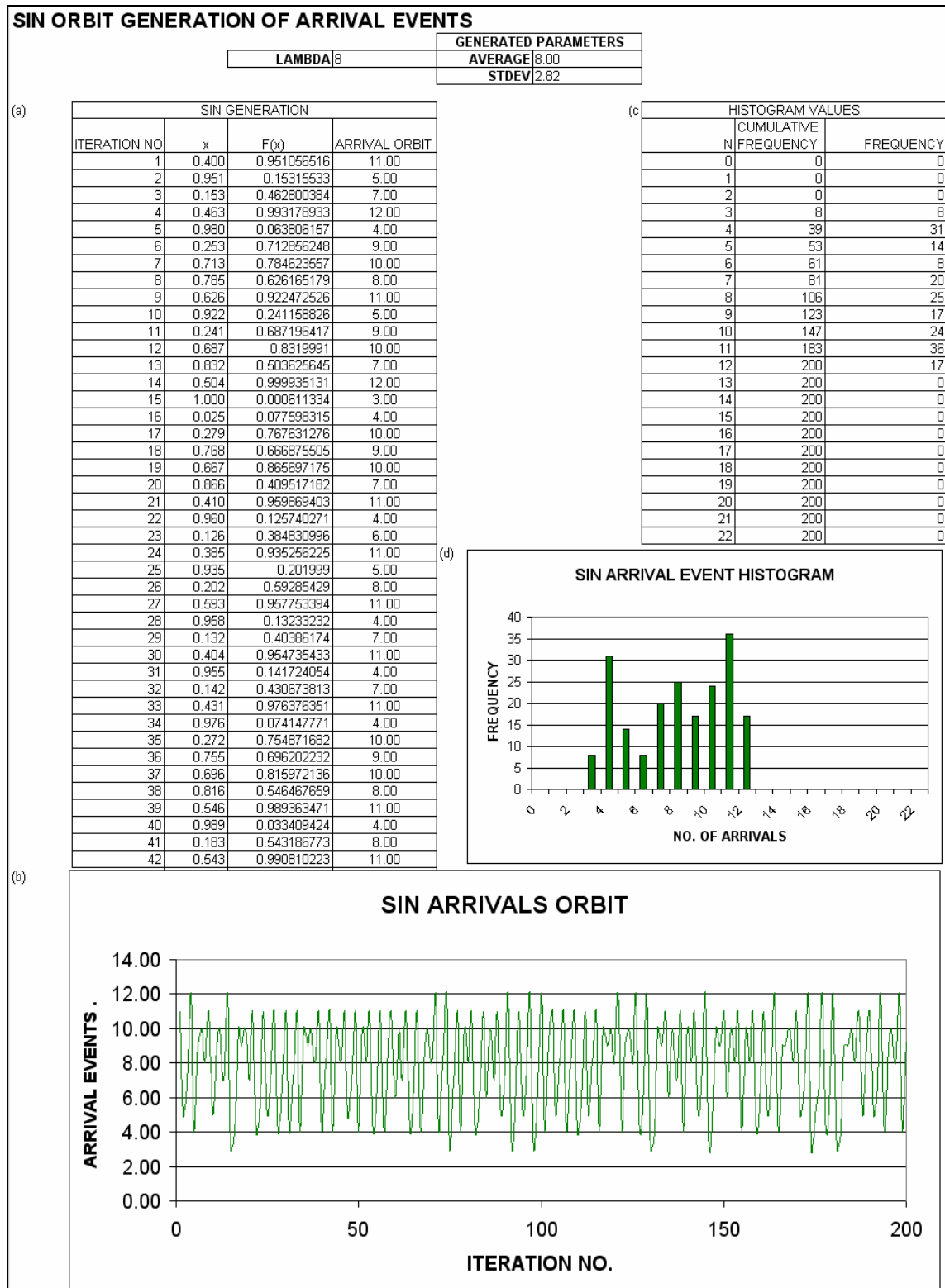


Fig. 6.3.4 GENERATION OF THE ORBIT OF SIN ARRIVAL EVENTS

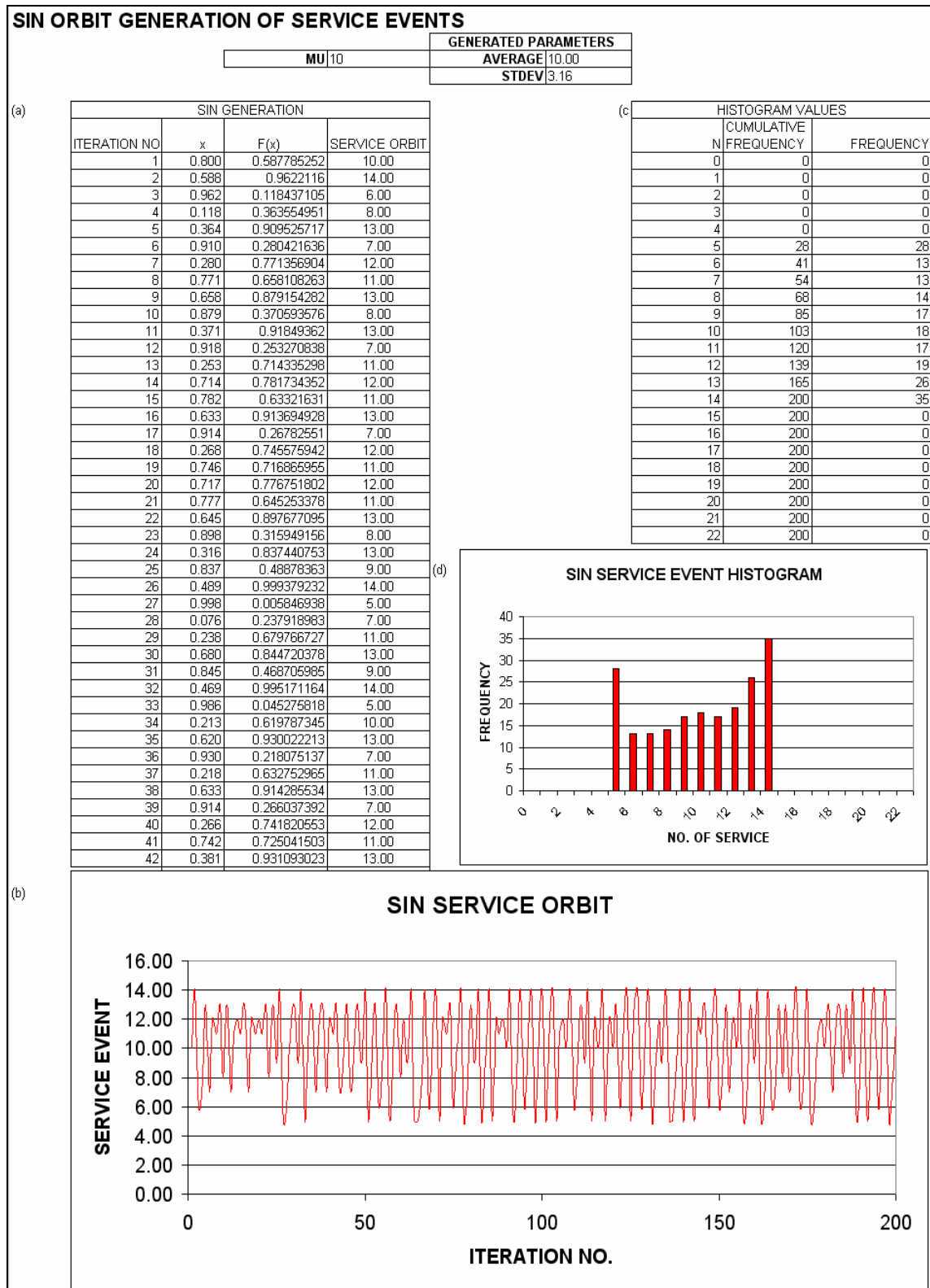


Fig. 6.3.5 GENERATION OF THE ORBIT OF SIN SERVICE EVENTS

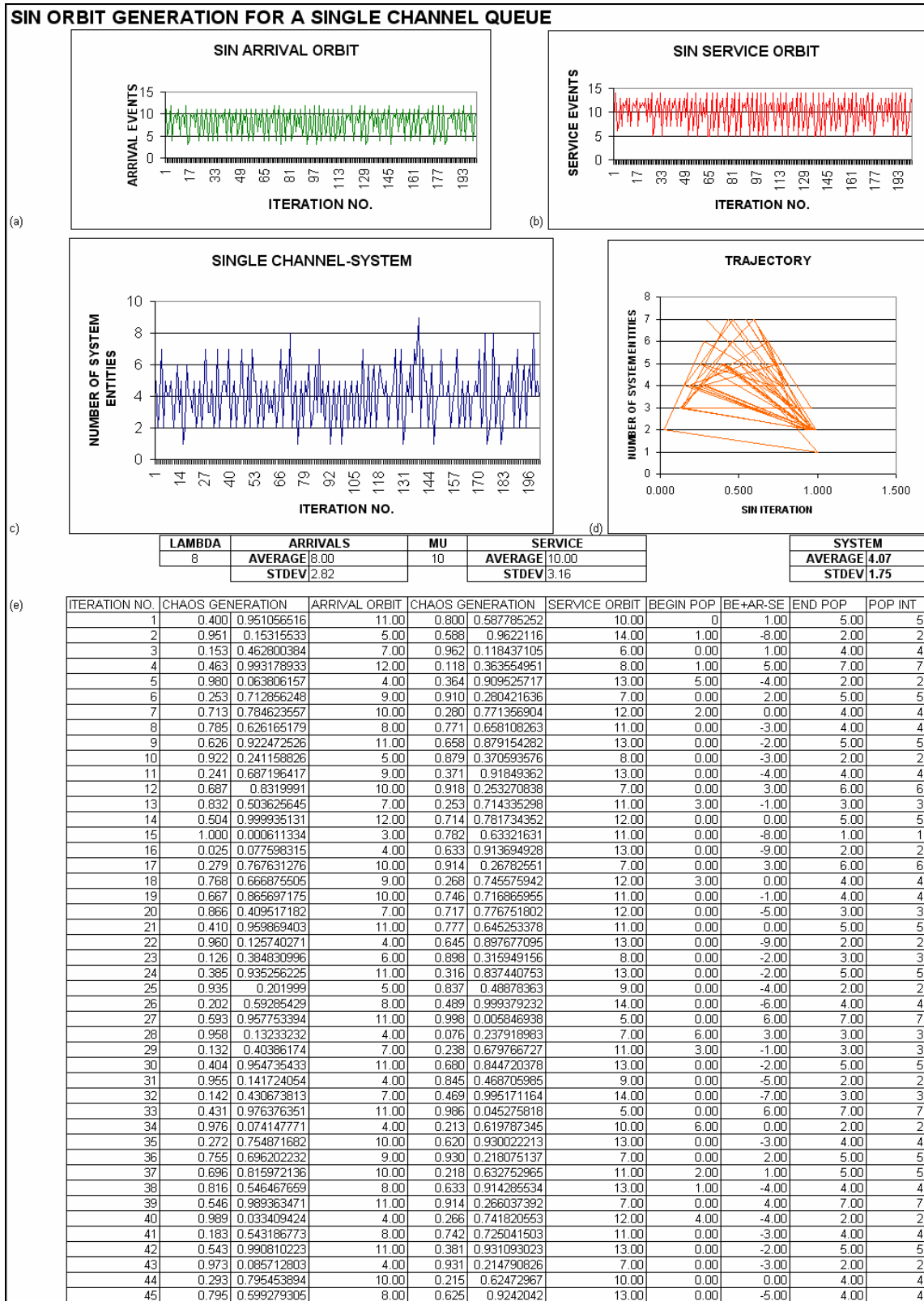


Fig. 6.3.6 GENERATION OF THE ORBIT OF A SIN SINGLE CHANNEL QUEUEING SYSTEM

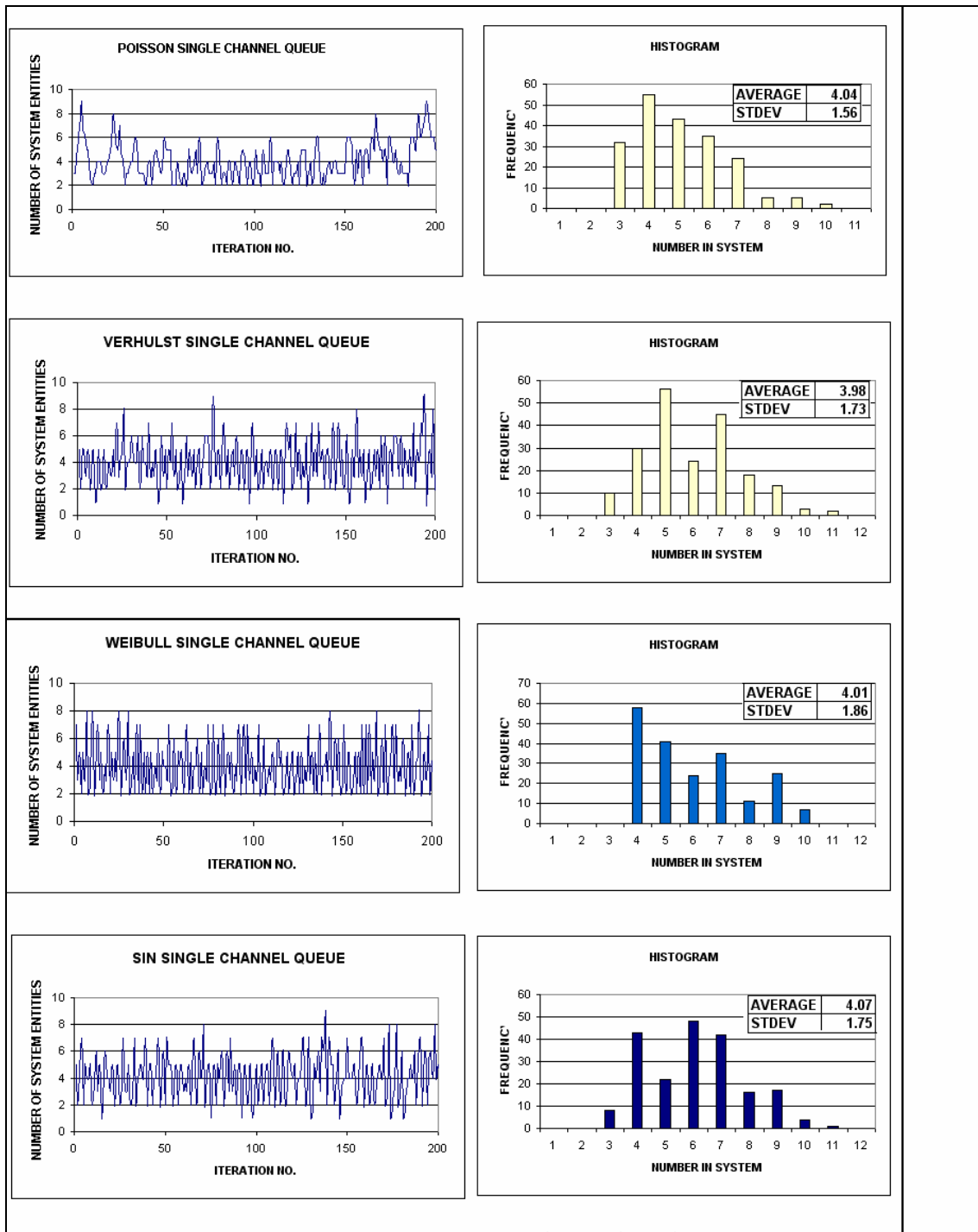


Fig. 6.3.7 PICTORIAL COMPARISON OF ORBITS OF SEVERAL SINGLE CHANNEL QUEUEING MODELS

6.4 Concluding remarks on single channel orbits resulting from a menu of methods of generation.

When viewing the various generated orbits shown in Fig. 6.3.7 one perceives that

- the various numerical values of average and standard deviation are virtually identical,
- one is inclined to believe that a measure of similarity exists in the histograms, and
- one consequently cautiously harbours the suspicion that **further extension and embellishment of the concept of chaos based system orbit generation** to match examples from the plethora of practical complex Systems of Congestion which exist, may be attempted.

The practical complex Systems of Congestion which are to be modelled in the following chapter are of a divergent nature and of necessity at least contain real time feedback rules to support decision making in achieving optimum transient and stable system operation.

CHAPTER 7

ANALYSIS OF THE DYNAMIC CHARACTERISTICS OF PRACTICAL SYSTEMS OF CONGESTION USING CHAOS GENERATION METHODS

A modified version of this Chapter will be presented at a Southern African Institute for Industrial Engineering Conference, 2005.

7.1 INTRODUCTION

In pursuing the search for an alternative way of providing time-varying solutions for Systems of Congestion the thesis proceeds to examine a number of practical systems of a divergent and complex nature. **It only attempts to depict the transient operation of each practical system via chaos based system orbit generation and in so doing endeavours to furnish modelling techniques for use in achieving optimum dynamic operation.**

To eventually achieve optimum dynamic operation depends on the nature of operation of the System of Congestion i.e. that an operational objective be formulated against a background of economic, physical, social and other constraints of the system.

The various systems to be considered from the point of view of dynamic operation are

- System No. 1: Two single channel queues which alternatively make use of a single server and are combined to form a single System of Congestion.
- System No. 2: A multi-channel queue which serves a population of entities which arrive in a pattern which varies daily in time by orders of magnitude.
- System No. 3: A multi-channel queue (30 channels) each with a constant service rate combined to form a single System of Congestion.
- System No. 4: A multi-channel queueing system which serves an extensive population by communication when emergency conditions occur.

The various system configurations are explained in the following paragraphs:

7.2 SYSTEM NO. 1

7.2.1 System scenario

A crossing point over a river for vehicles in a rural area (Hartebeespoort dam wall in South Africa) consists of a single vehicle width bridge. The flow of vehicles over the bridge consists of eastbound and westbound traffic. Eastbound traffic cannot use the bridge while westbound traffic is using it and vice versa. The traffic flow is controlled by an existing automated signalling system which allows sequential periods of two minutes for traffic flow in a given direction. If sufficient entities are waiting to use the channel the service rate is approximately constant at 10 entities per two minute interval.

On certain days the traffic arrival rate increases (eastbound and westbound) over the period from 11h00 to 14h00 and then decreases over the period from 14h00 to 17h00.

7.2.2 The system model

The Verhulst generation of eastbound and westbound arrivals are shown in Fig. 7.2.2.1 and Fig. 7.2.2.2 respectively and are based on fixed average arrival rates for sequential periods of 15 minutes. Each consecutive set of 15 minute temporally sequential periods of arrivals is further subdivided into 2 minute intervals for purposes of orbit generation. The arrivals generation process is based on actual observations on site.

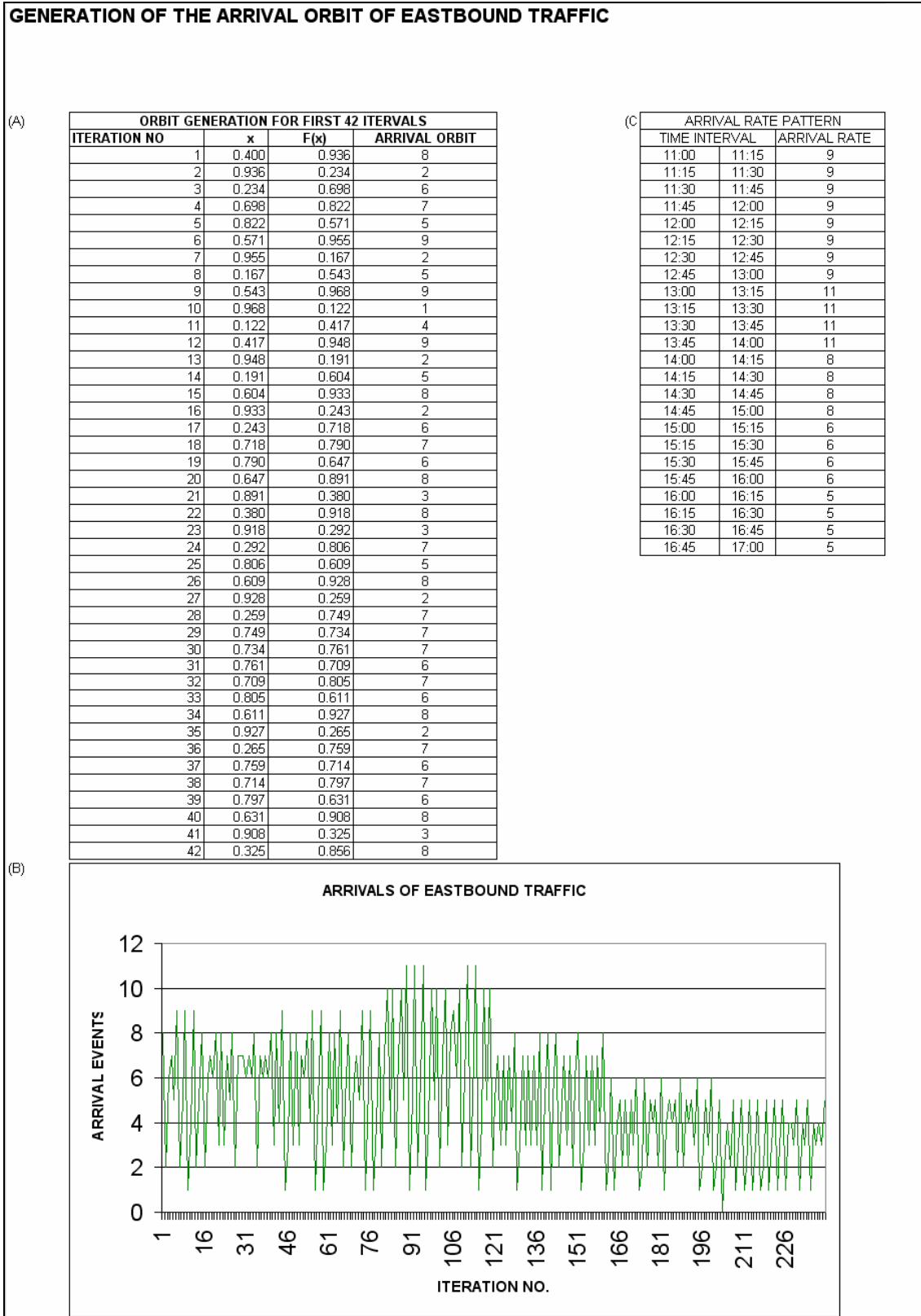


Fig. 7.2.2.1 GENERATION OF ARRIVALS OF EASTBOUND TRAFFIC

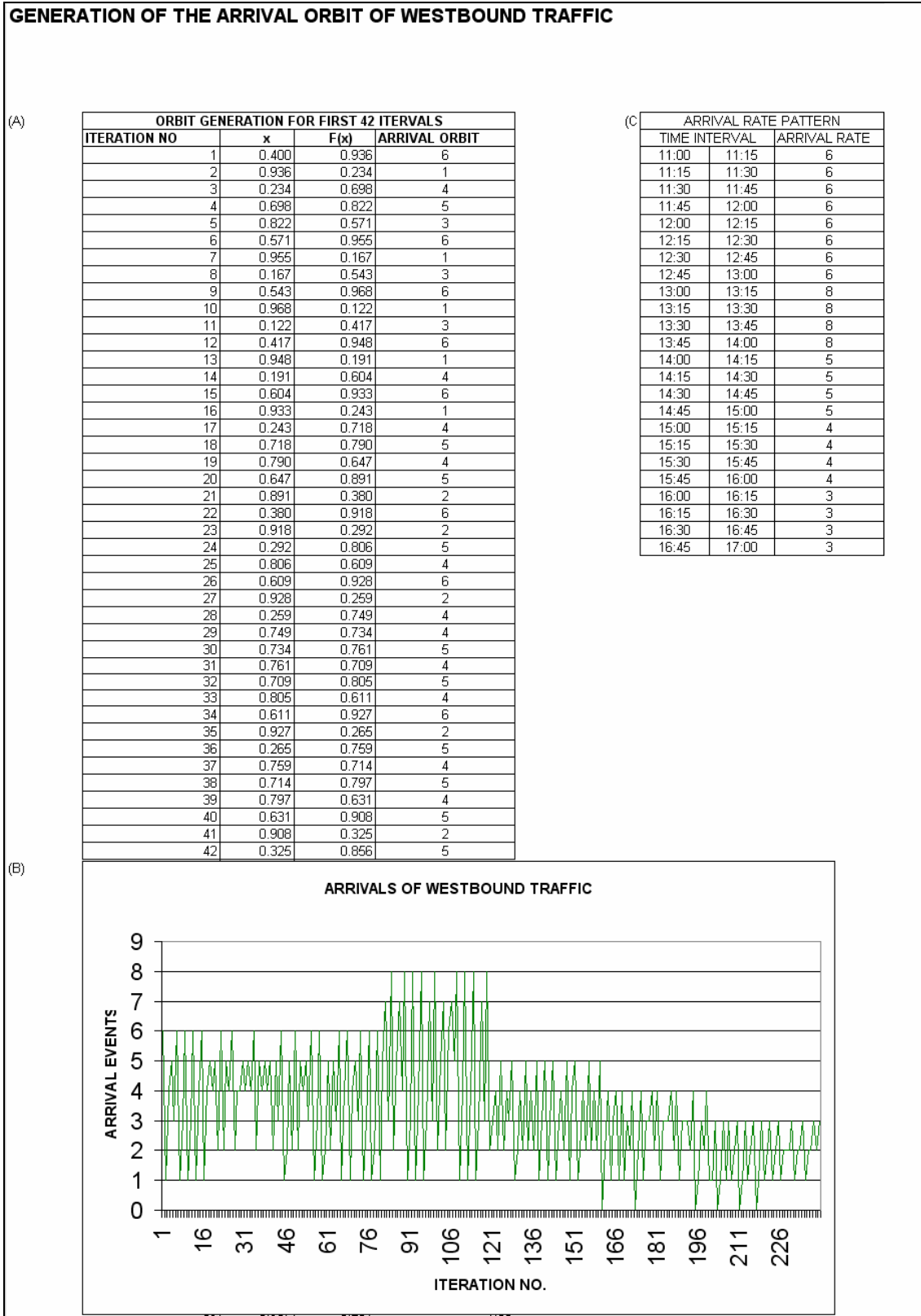


Fig. 7.2.2.2 GENERATION OF ARRIVALS OF WESTBOUND TRAFFIC

The Verhulst generation of service is shown in Fig. 7.2.2.3 and is based on an average service rate of 10 units per 2 minute interval with a standard deviation of 0,10 units. (i.e. an approximately constant service rate). The service event process is based on actual observations on site.

Combination of the eastbound traffic arrival orbit and interrupted service (each alternating 2 min. service time interval) results in a portrayal of the system event dynamics shown in Fig. 7.2.2.4 (C).

Pursuing the abovementioned combination method for the westbound traffic results in a portrayal of the system event dynamics shown in Fig. 7.2.2.5 (C).

Eventual combination (superposition) of the eastbound and westbound situations results in the portrayal of total system event dynamics as shown in Fig. 7.2.2.6 (D).

7.2.3 Diagnosis of the model results

When viewing the system population values of Fig. 7.2.2.6 (A), (B) and (C), which closely match actual site conditions on a particular day, one could consider altering the service cycle pattern in an effort to decrease the system population values thereby improving the state of congestion.

Consequently several service cycle patterns have been considered as alternative patterns to the service cycle pattern employed in par. 7.2.2. The five service cycle patterns considered are shown in Fig. 7.2.2.7. Each of the situations of the system were analysed in the same way as described in par. 7.2.1 and par. 7.2.2.

The results of the analysis are shown in Fig. 7.2.2.8. A portrayal of the total system event dynamics for service cycle pattern No. 3 is shown in Fig 7.2.2.9.

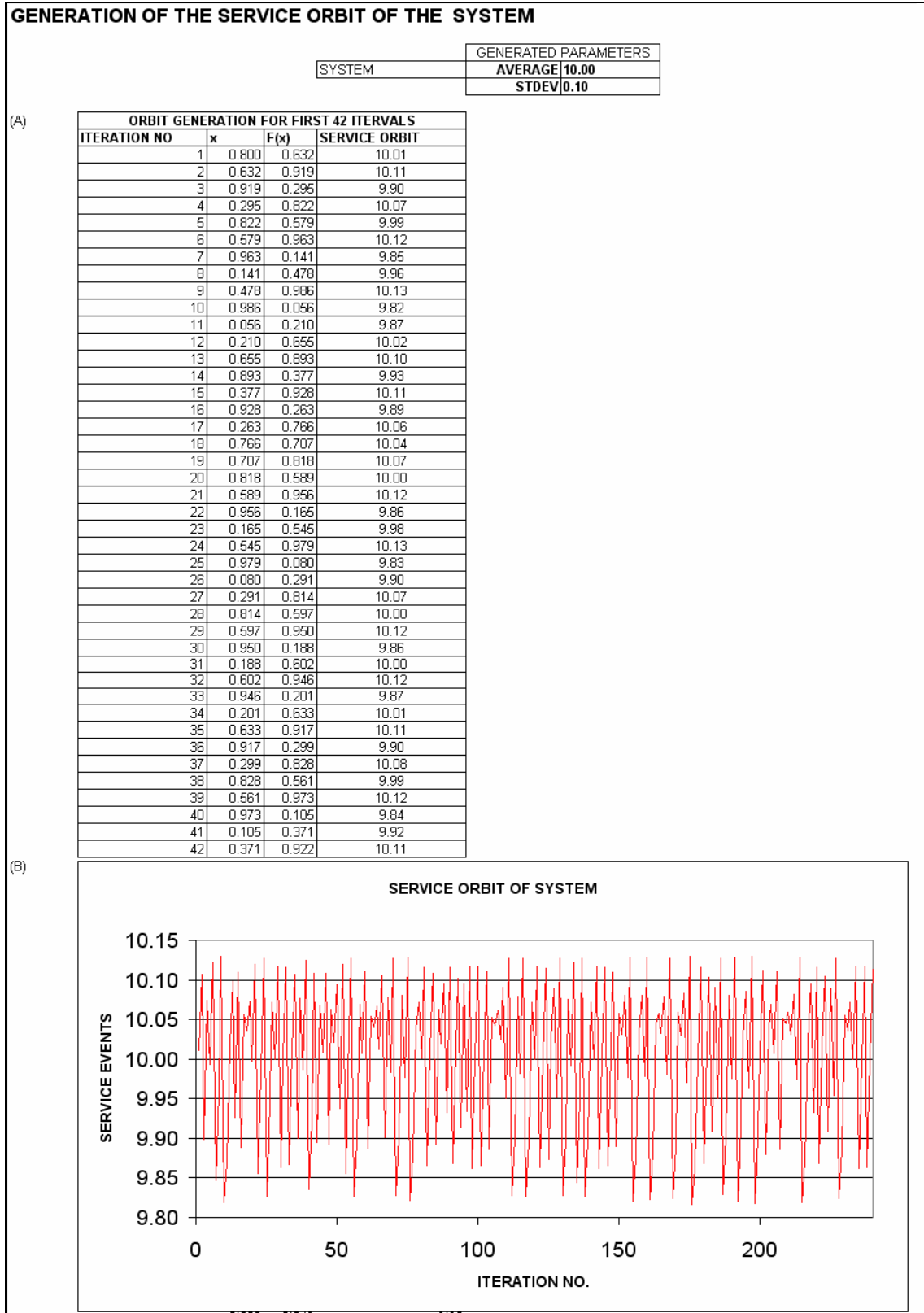


Fig. 7.2.2.3 GENERATION OF THE SERVICE ORBIT OF THE SYSTEM

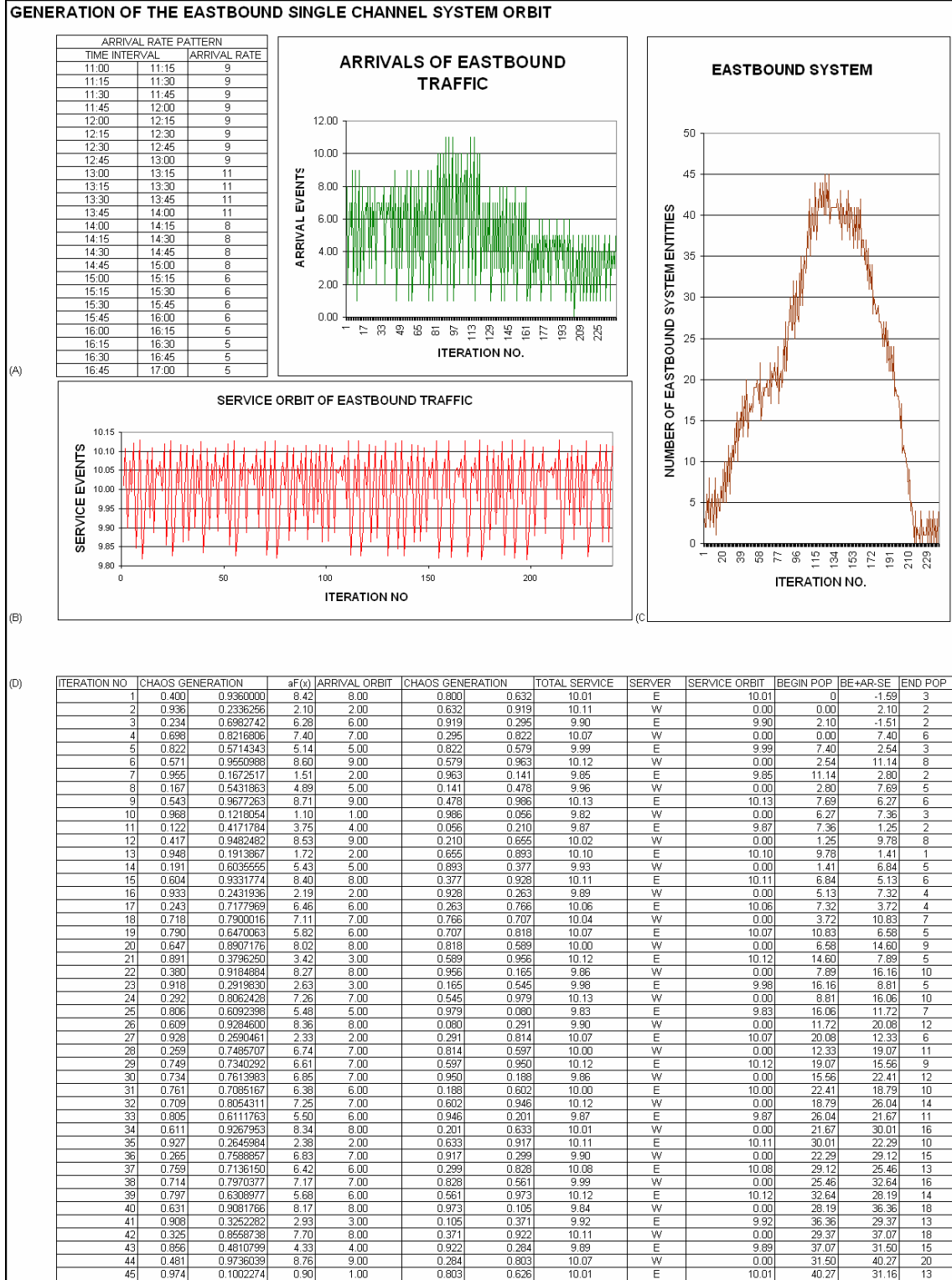


Fig. 7.2.2.4 GENERATION OF THE EASTBOUND SINGLE CHANNEL QUEUE ORBIT

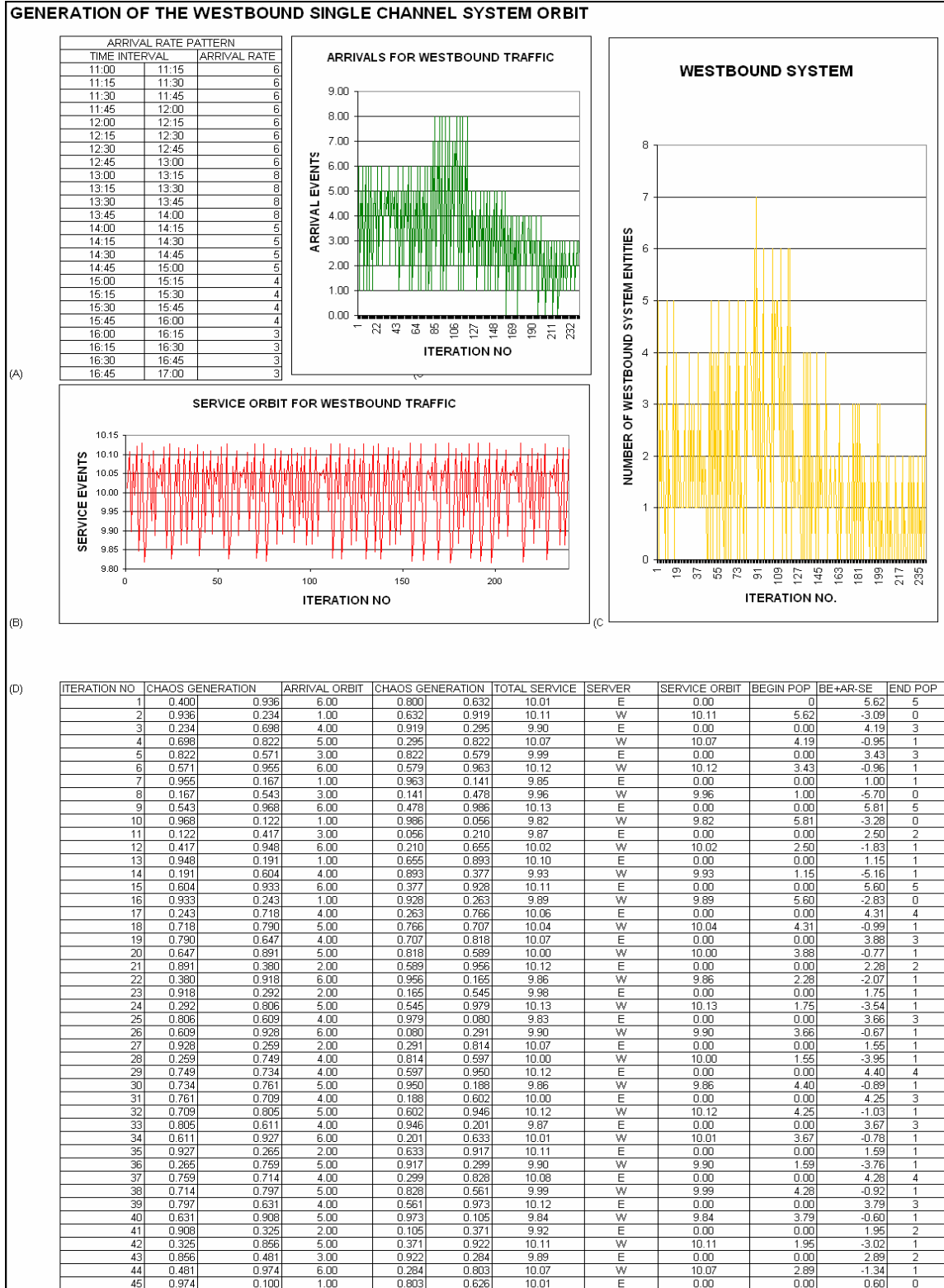


Fig. 7.2.2.5 GENERATION OF WESTBOUND TRAFFIC SINGLE CHANNEL QUEUE ORBIT

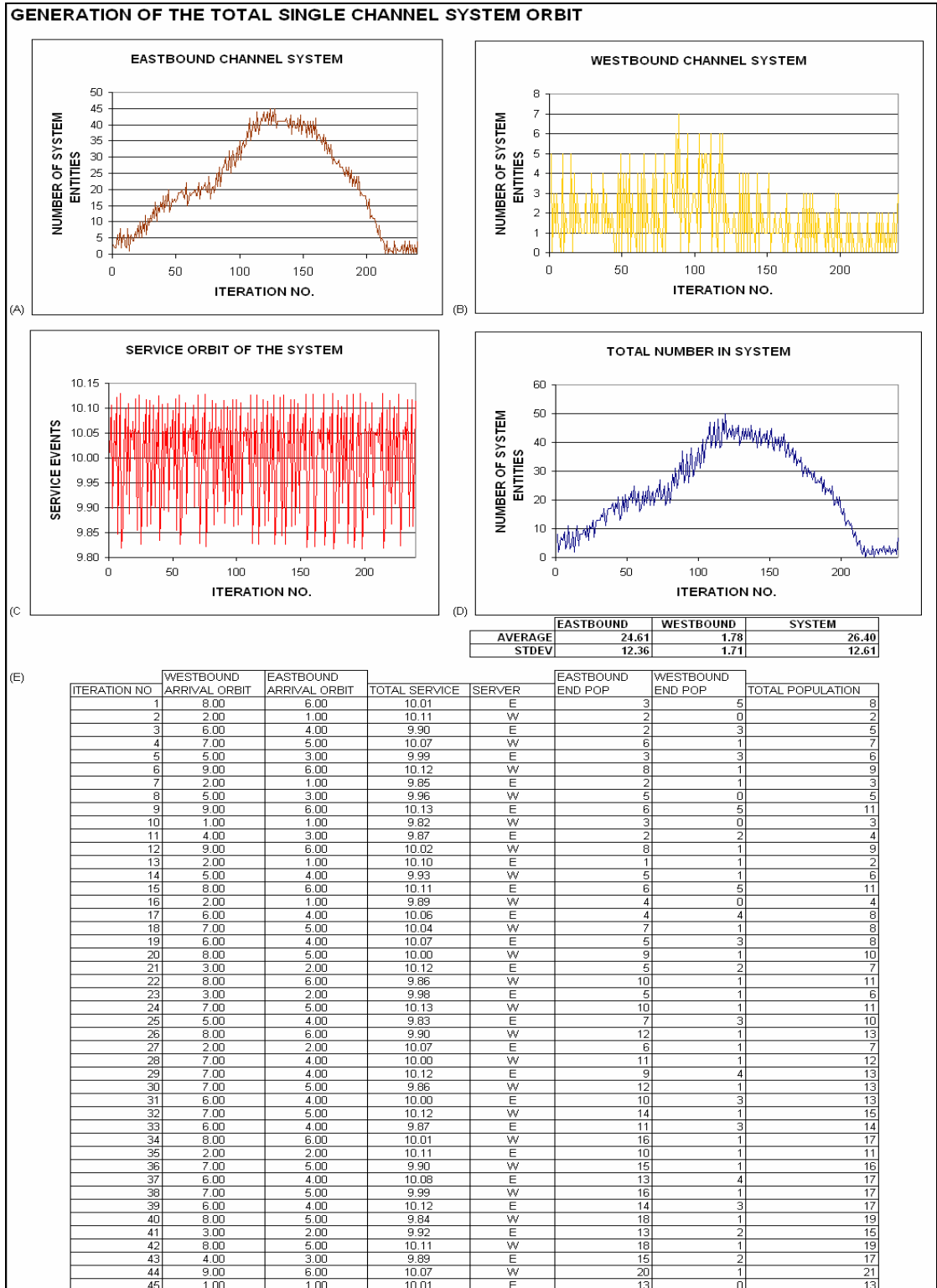


Fig. 7.2.2.6 GENERATION OF THE TOTAL SYSTEM ORBIT

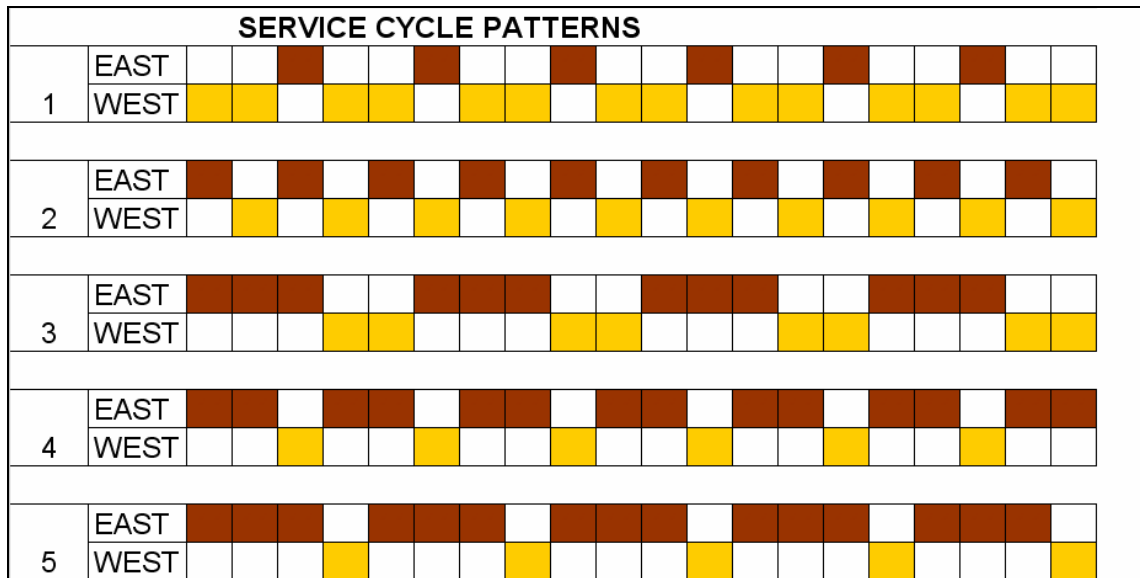


Fig. 7.2.2.7 ALTERNATIVE TEMPORAL SERVICE CYCLES FOR THE SYSTEM

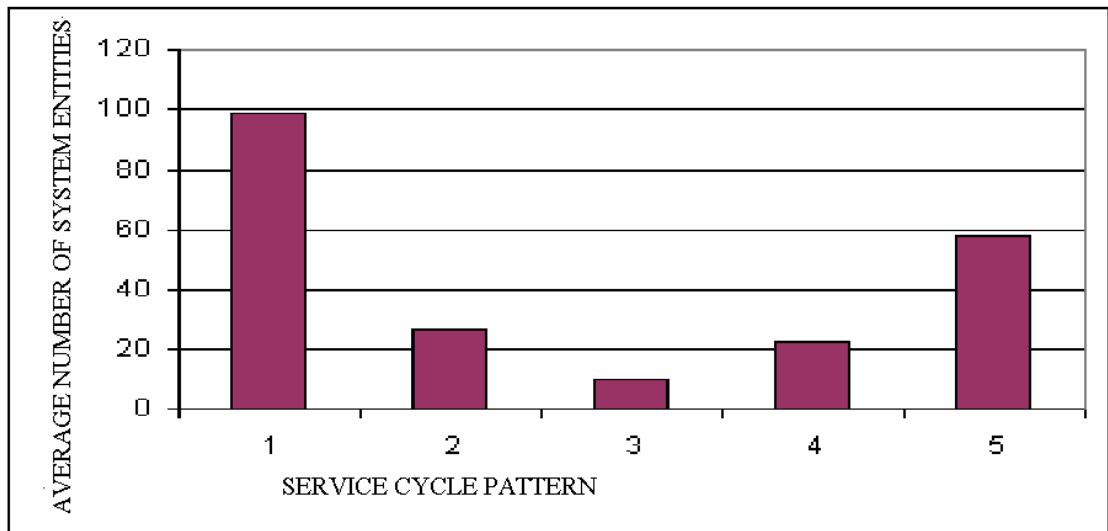


Fig. 7.2.2.8 AVERAGE SYSTEM POPULATION VALUES FOR SELECTED SERVICE CYCLE PATTERNS

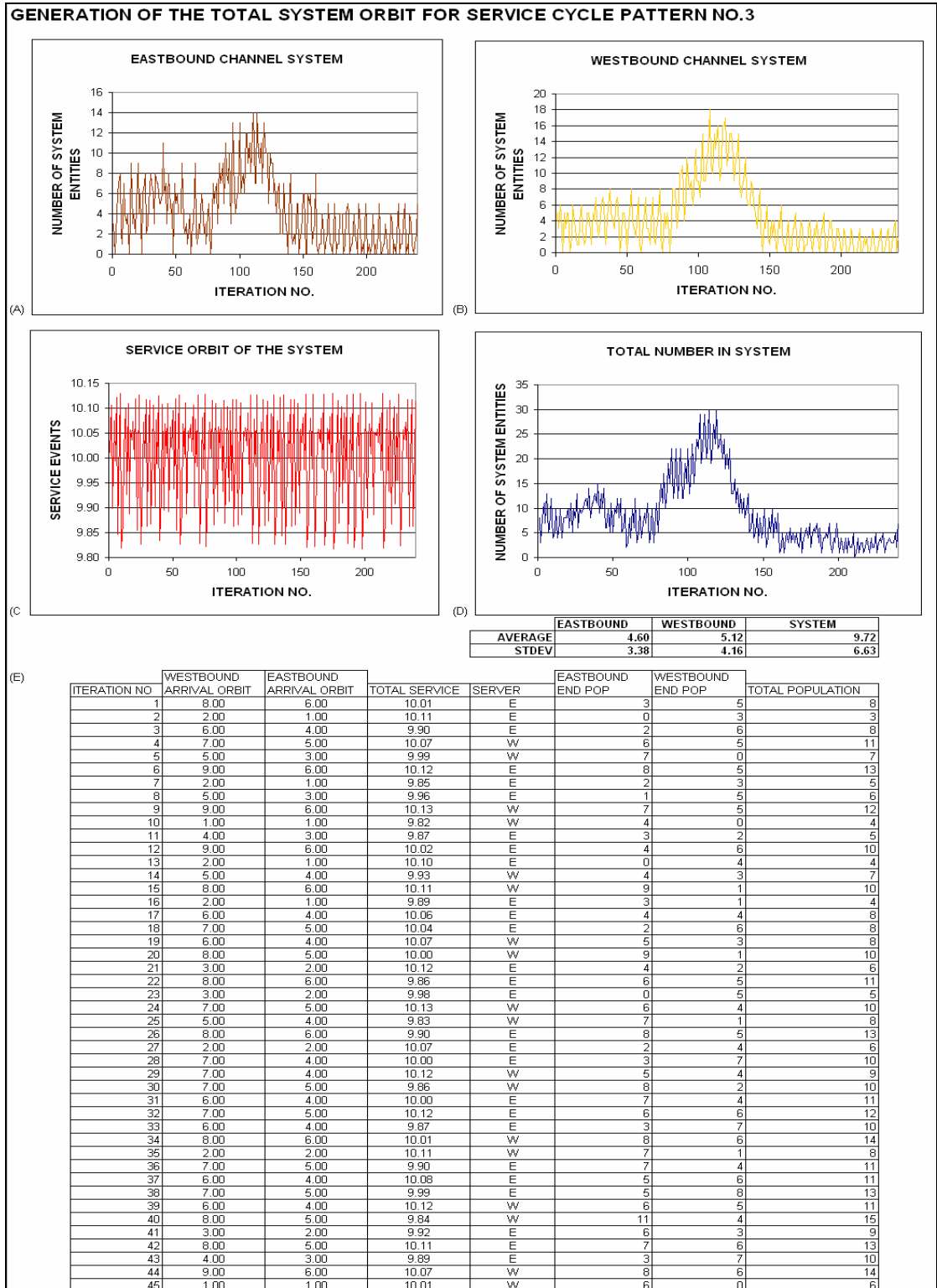


Fig. 7.2.2.9 GENERATION OF THE TOTAL SYSTEM ORBIT FOR SERVICE CYCLE PATTERN NO.3

7.2.4 Using realtime feedback to improve system performance

As an alternative to the foregoing attempt to minimize the total number of entities in the system by the use of service cycle pattern adjustment the use of a simplistic feedback system may be considered. Assume that the feedback system could influence the automated signalling system by comparing the number of eastbound and westbound entities in the system at the beginning of each 2-minute interval and then assigning the single service channel to the direction which contains the greater number of entities.

The results of analysing the system with realtime feedback are shown in Fig. 7.2.2.10 (D). When compared to the results of the system without feedback shown in Fig. 7.2.2.6 it is obvious that feedback can beneficially affect the degree of congestion by dramatically lowering the average number of entities in the total system from the initial value of 26.40 given in Fig 7.2.2.6 to 8.09 given in Fig 7.2.2.10.

7.2.5 The effect of the size of system waiting area on system performance

A further measure which may be considered to improve congestion is to limit the total number of entities in the system for the eastbound and westbound traffic.

If one were to consider the system described in par. 7.2.2 and were to constrain the total eastbound population to 15 entities and impose the same limitation on the westbound population the results of analysing the system are shown in Fig. 7.2.2.11.

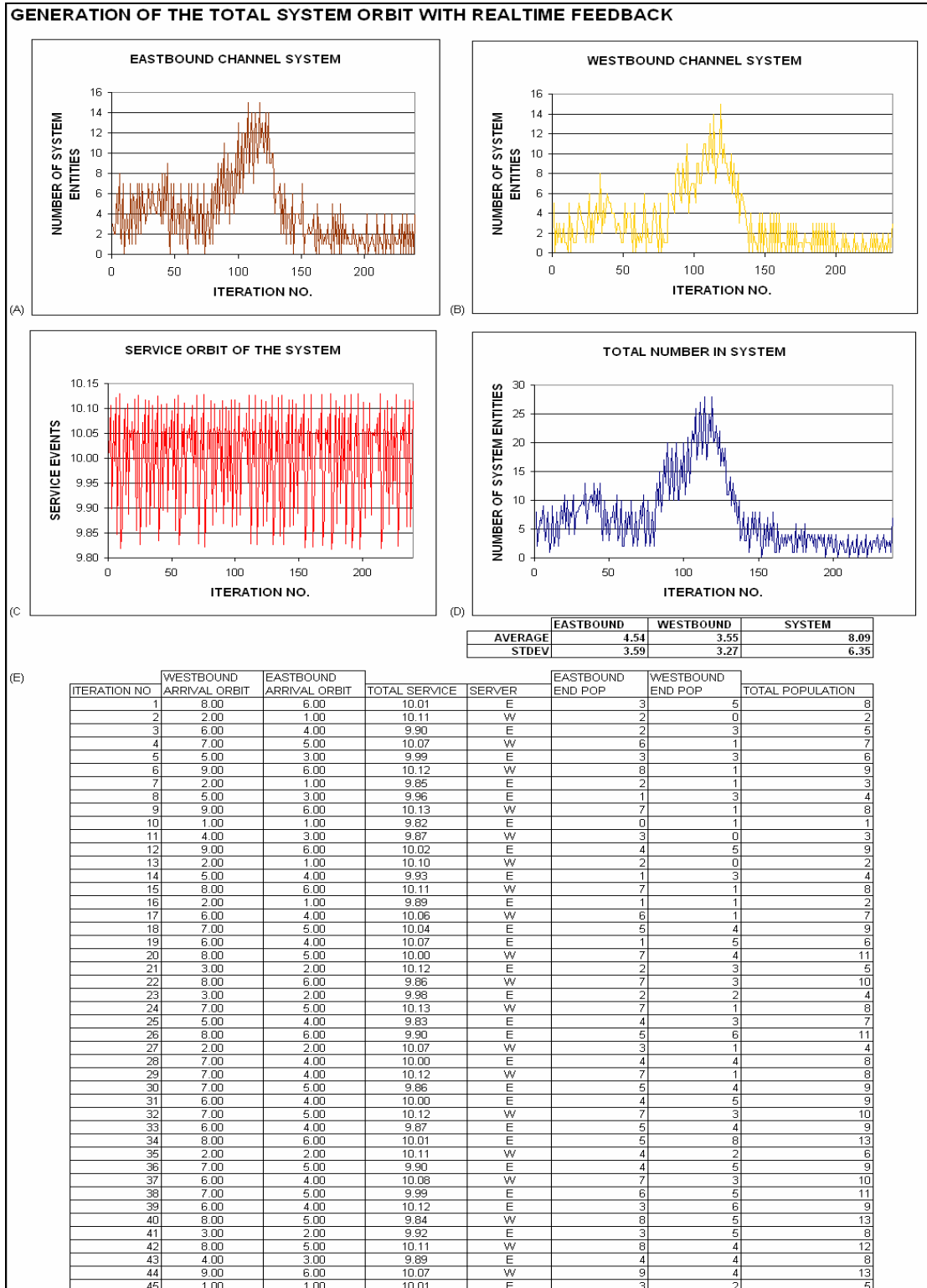


Fig. 7.2.2.10 GENERATION OF THE TOTAL SYSTEM ORBIT WITH REALTIME FEEDBACK

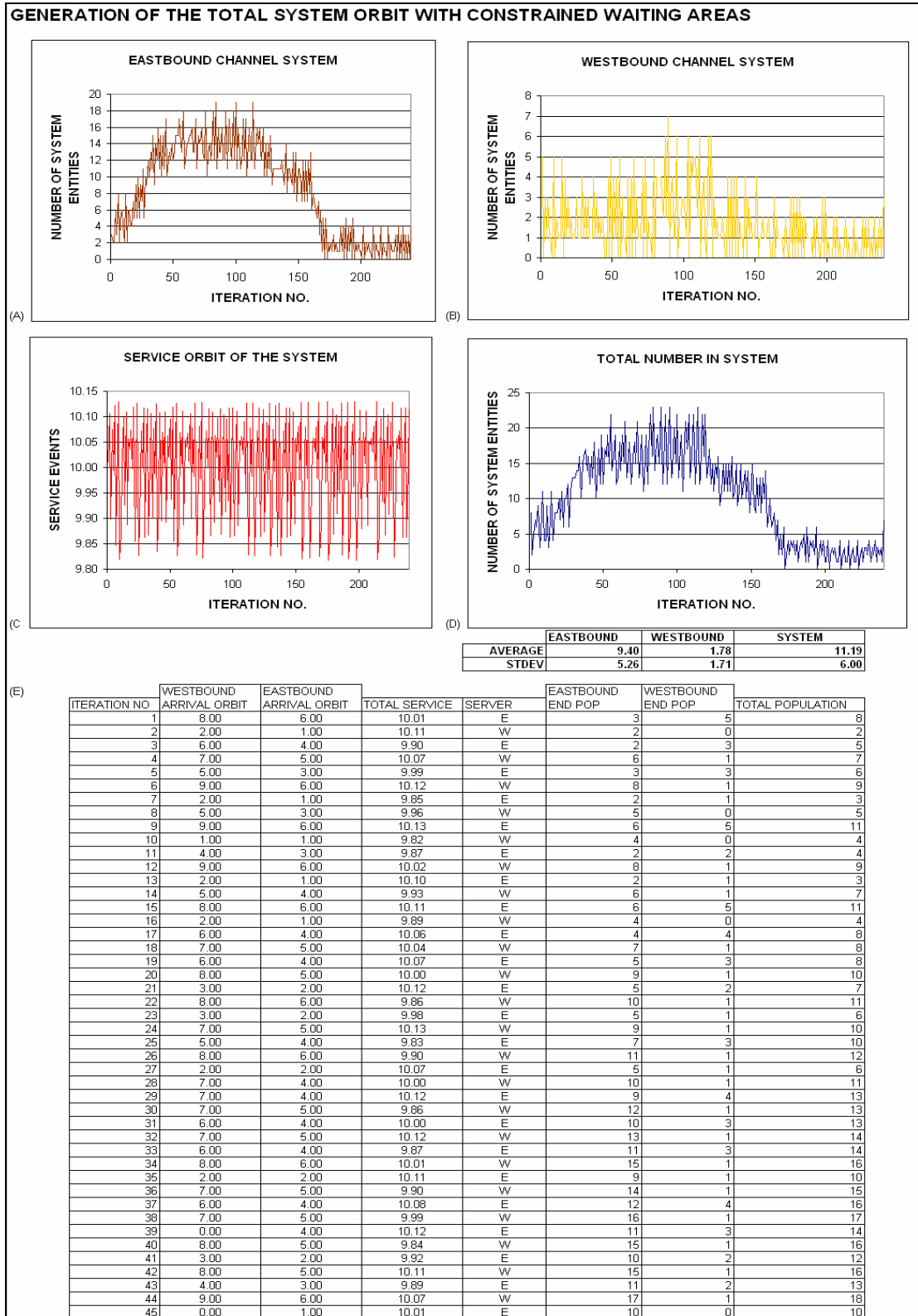


Fig. 7.2.2.11 GENERATION OF SYSTEM ORBIT WITH CONSTRAINED WAITING AREAS

7.2.6 Concluding comments on System No. 1

The foregoing analysis indicates that in attempting to depict the operation of System No. 1 one may investigate several alternative solutions and only be constrained by a lack of originality and imagination. Nevertheless one is often required to remain within the bounds of reality which implies that suggested improvement of the system under consideration must be practically feasible.

7.3 SYSTEM NO. 2

7.3.1 System scenario

System No. 2 is an example of a typical toll plaza (Pumalani Plaza South Africa) on a national highway serving vehicles in two directions between Pretoria and Polokwane. The specific system under investigation serves traffic in a North and Southbound direction. The system experiences congestion in the Northbound and Southbound direction on Fridays. At the end of the month the Northbound area of the toll plaza becomes heavily congested. The study will focus on this peak traffic situation.

The System of Congestion has four lanes for normal traffic and one lane dedicated to heavy vehicles. Lanes 1 and 2 serve more vehicles than lanes 3 and 4. The heavy vehicle lane serves a total of 14% of the total vehicle flow. The flow distribution pattern over the five lanes is shown in Fig. 7.3.2.3 (A). The traffic intensity increases over the period from 17h00 to 18h45 and then decreases over the period from 18h45 to 20h00.

7.3.2 The system model

Verhulst orbit generation of the northbound arrivals is shown in Fig. 7.3.2.1. The average arrival rate is determined for consecutive 15-minute intervals. The 15 minute intervals are divided into 1-minute intervals for purposes of orbit generation. The arrival orbit agrees with the actual observations gathered on site.

Service orbit generation of a single lane is shown in Fig. 7.3.2.2. The average service time was measured at 6 seconds per vehicle resulting in an average service rate of 10 vehicles per minute.

The total system was modelled using the arrival orbits and distributing the arrivals to different service lanes to match actual conditions on site. The

modelled result is shown in Fig. 7.3.2.3. It shows the arrival rate and the distribution of arrivals as percentages to the different service lanes in Fig. 7.3.2.3 (A). The queues that are generated ahead of each service lane and the total number of vehicles in the system are shown in Fig. 7.3.2.3 (B). The average number of vehicles in the system for the period from 17h00 to 19h00 is 53.

GENERATION OF THE ARRIVAL ORBIT OF NORTHBOUND TRAFFIC

(A)

ORBIT GENERATION FOR FIRST 42 INTERVALS			
ITERATION NO	x	F(x)	ARRIVAL ORBIT
1	0.400	0.936	25
2	0.936	0.234	14
3	0.234	0.698	21
4	0.698	0.822	23
5	0.822	0.571	20
6	0.571	0.955	25
7	0.955	0.167	13
8	0.167	0.543	19
9	0.543	0.968	26
10	0.968	0.122	13
11	0.122	0.417	17
12	0.417	0.948	25
13	0.948	0.191	14
14	0.191	0.604	20
15	0.604	0.933	25
16	0.933	0.243	19
17	0.243	0.718	27
18	0.718	0.790	28
19	0.790	0.647	26
20	0.647	0.891	30
21	0.891	0.380	21
22	0.380	0.918	30
23	0.918	0.292	20
24	0.292	0.806	28
25	0.806	0.609	25
26	0.609	0.928	31
27	0.928	0.259	19
28	0.259	0.749	28
29	0.749	0.734	27
30	0.734	0.761	28
31	0.761	0.709	32
32	0.709	0.805	34
33	0.805	0.611	30
34	0.611	0.927	36
35	0.927	0.265	24
36	0.265	0.759	33
37	0.759	0.714	32
38	0.714	0.797	34
39	0.797	0.631	31
40	0.631	0.908	36
41	0.908	0.325	25
42	0.325	0.856	35

(C)

ARRIVAL RATE PATTERN		
TIME INTERVAL		ARRIVAL RATE
17:00	17:15	20
17:15	17:30	25
17:30	17:45	30
17:45	18:00	35
18:00	18:15	45
18:15	18:30	49
18:30	18:45	49
18:45	19:00	45
19:00	19:15	40
19:15	19:30	30
19:30	19:45	25
19:45	20:00	20

(B)

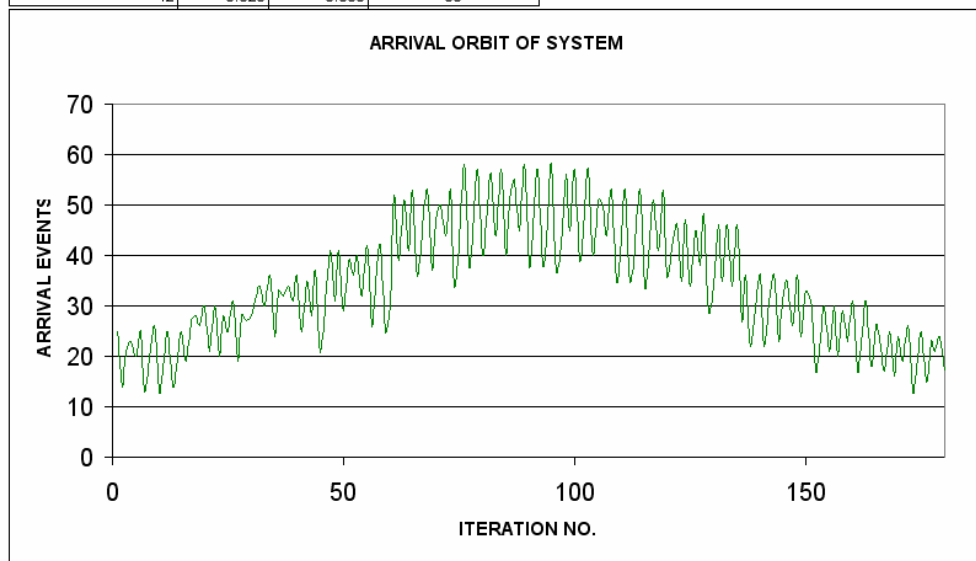


Figure 7.3.2.1 GENERATION OF ARRIVALS OF NORTHBOUND TRAFFIC

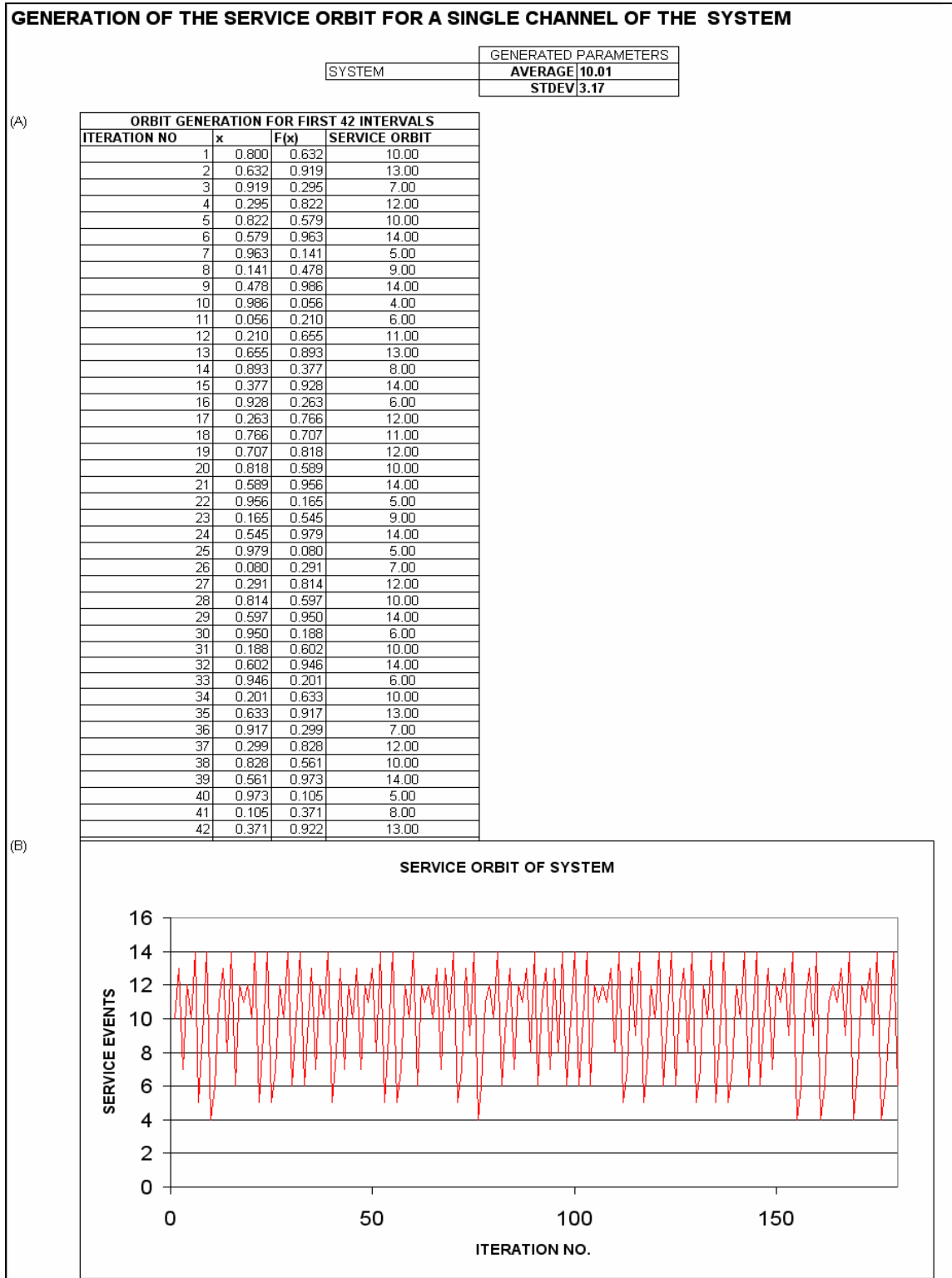


Figure 7.3.2.2 GENERATION OF A TYPICAL SERVICE ORBIT FOR A SINGLE SERVICE LANE

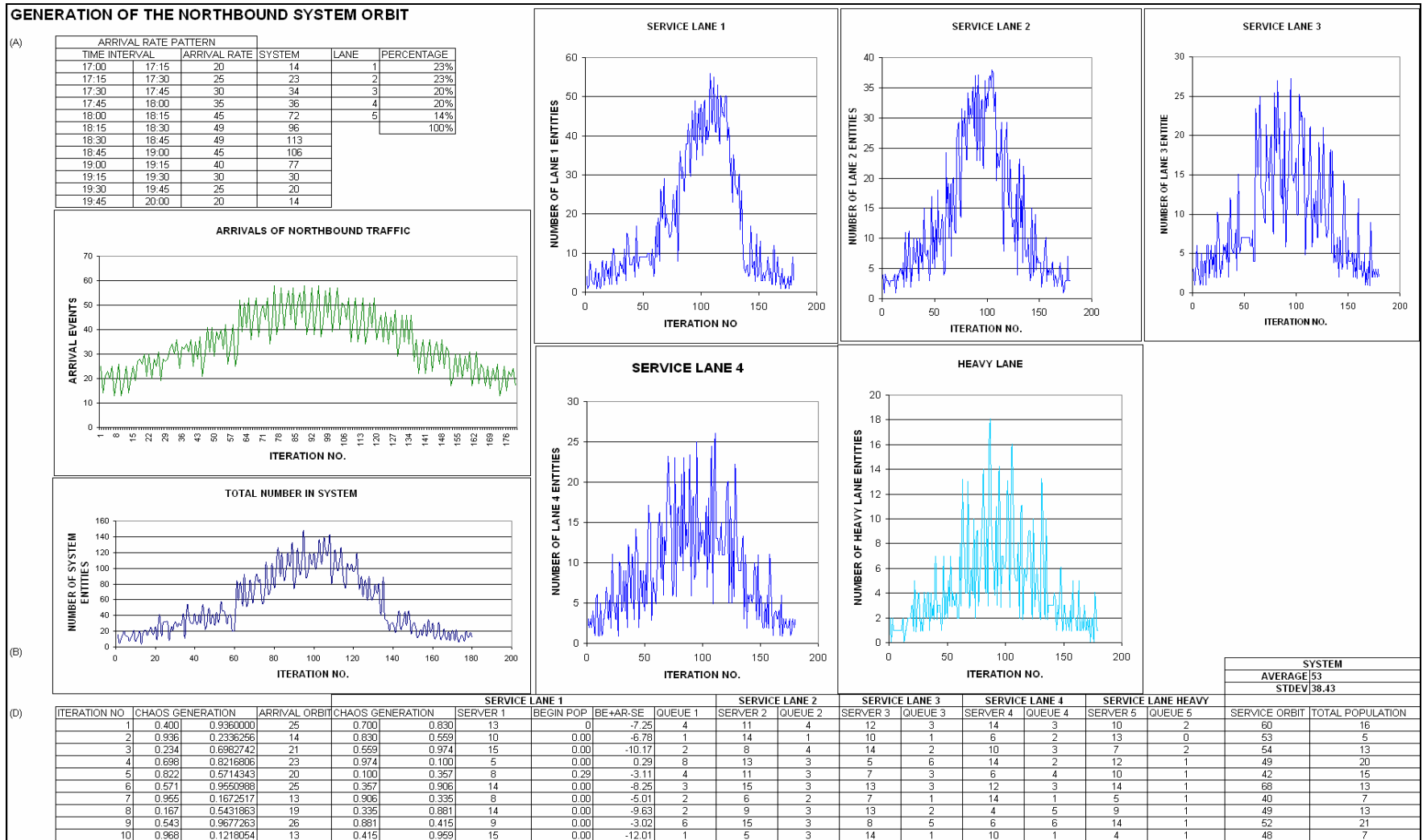


Figure 7.3.2.3 GENERATION OF THE TOTAL NORTHBOUND SYSTEM ORBIT

7.3.3 Diagnosis of the model results

The results of Fig. 7.3.2.3 closely emulate the system behaviour. Lanes 1 and 2 have the greatest number of entities in their queues. The heavy vehicle lane peaks at 18 trucks in the queue. The maximum number of entities reaches a peak close to 150. The focus is on the average number in the system. One could attempt to improve the current system to reduce the average number in the system.

7.3.4 Using realtime feedback to improve system performance

The physical system attributes suggest room for improvements such as urging drivers to choose the lane that is least congested. Such a solution is however not without physical hazard due to the jockeying (switching of lanes) that will take place.

A new technology that is available is the “e-tag” system that will improve the service. Certain of these electronic payment systems are already employed by some plazas. The designers of the system have made provision for multi-tasking by some of the lanes. This implies that a lane could be used to serve either northbound or southbound traffic as system conditions may dictate.

The use of a standby lane has been modelled and the results are shown in Fig 7.3.2.4. The model shows the use of the standby lane when the total number of vehicles in the system is above 80 vehicles. Fig 7.3.2.4 (A) shows the distribution of traffic during the peak congestion period. The standby lane only operates for periods of 70 consecutive iterations. These changes reduce the average number in the total system from 53 to 34 vehicles.

Other innovative improvement measures may be considered, the only limitation for improving the system being physical system constraints and financial implications.

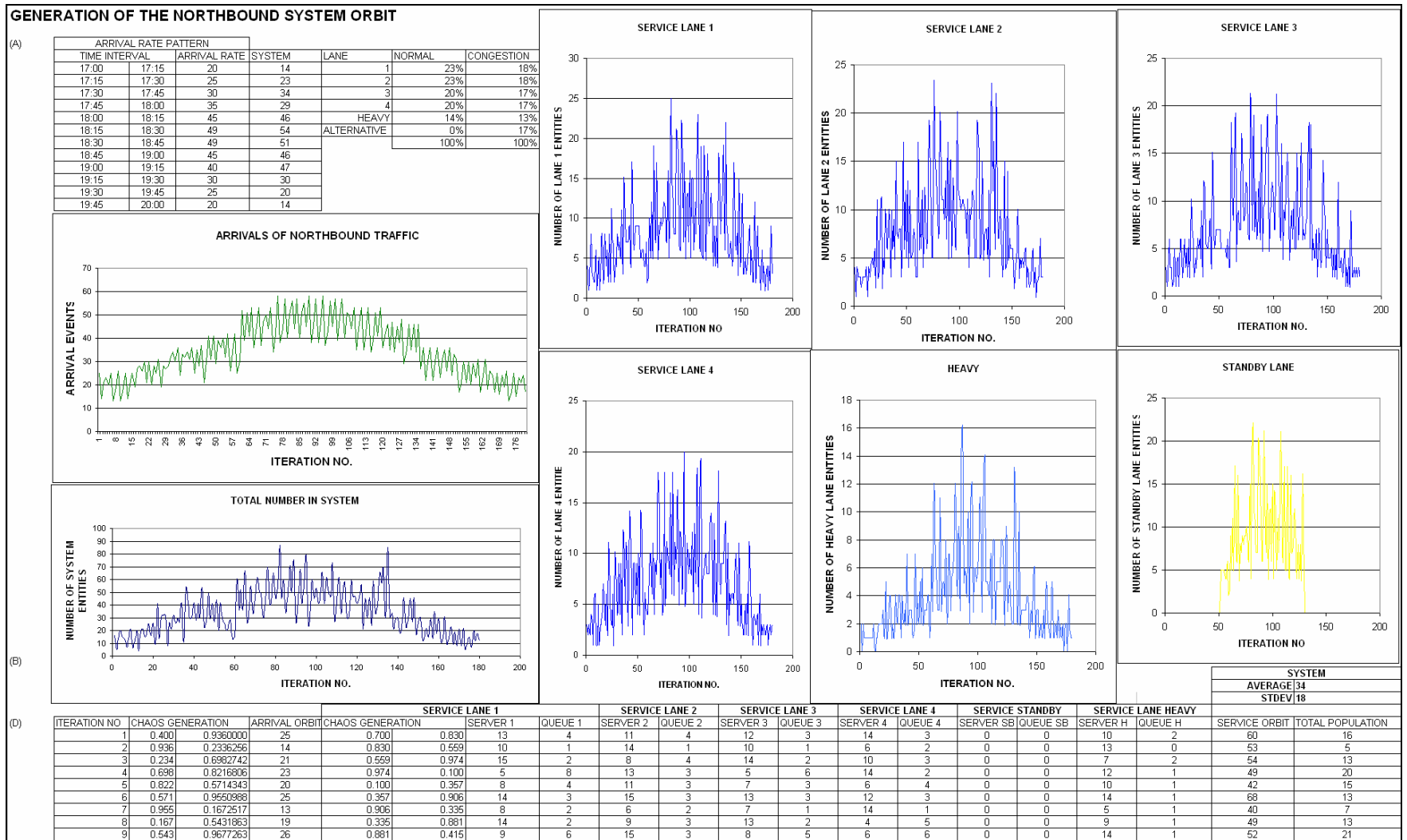


Figure 7.3.2.4 GENERATION OF SYSTEM ORBIT WITH REALTIME FEEDBACK

7.3.6 Concluding comments on System No. 2

System No. 2 has been successfully modelled by means of the chaos orbit generation method. This implies that the basic building blocks which emanate from Chapter 6 may be used to model a multi-channel system. The modelling of the system has also demonstrated versatility in modelling improvements of the system via realtime feedback and other system adjustment methods.

7.4 SYSTEM NO. 3

7.4.1 System scenario

The system under consideration is a large *FERRIS*¹ wheel which is used for entertaining tourists in a large European city. The system is specifically designed to afford viewing of the entire city skyline. On clear days one can see up to 20km from the apex of the wheel. The wheel is equipped with 30 equally spaced cabins which can each accommodate 25 adult passengers.

The wheel makes 2 revolutions per hour. Each cabin completes a single revolution in 30 minutes and upon completion thereof discharges the passengers at ground level. The wheel diameter is 150 metres resulting in a peripheral speed of 0,26 metres per second.

The system described above is an “approximate” facsimile of the same order of magnitude as the actual system in respect of physical size and operational parameters. This has intentionally been done to avoid infringement of design copyright. The facsimile system is shown schematically in Fig.7.4.1.1. During certain periods of the day in the peak tourism season the system is a prime example of a System of Congestion.

7.4.2 The system model

Whenever one attempts to model a System of Congestion it is wise to consider the simplest model which would generate credible dynamic operation. The use of “designer equations” is also facilitated in terms of extent and excessive complication of modelling. Consequently the system is modelled as a single channel queue which serves an arrivals process according to the average arrival rate pattern shown in Fig. 7.4.1.2. (A). The service rate is approximately 50 entities per two-minute interval. The

¹After George Washington Gale Ferris: American Engineer who designed a wheel for an Exposition in Chicago in 1893: An amusement device consisting of a large power-driven wheel having suspended seats which maintain a horizontal position while the wheel rotates in a vertical plane

generation of arrival and service orbits is based on actual observations on site.

The portrayal of the system event dynamics is shown in Fig. 7.4.1.2 (C) which agrees with “actual” observations on site.

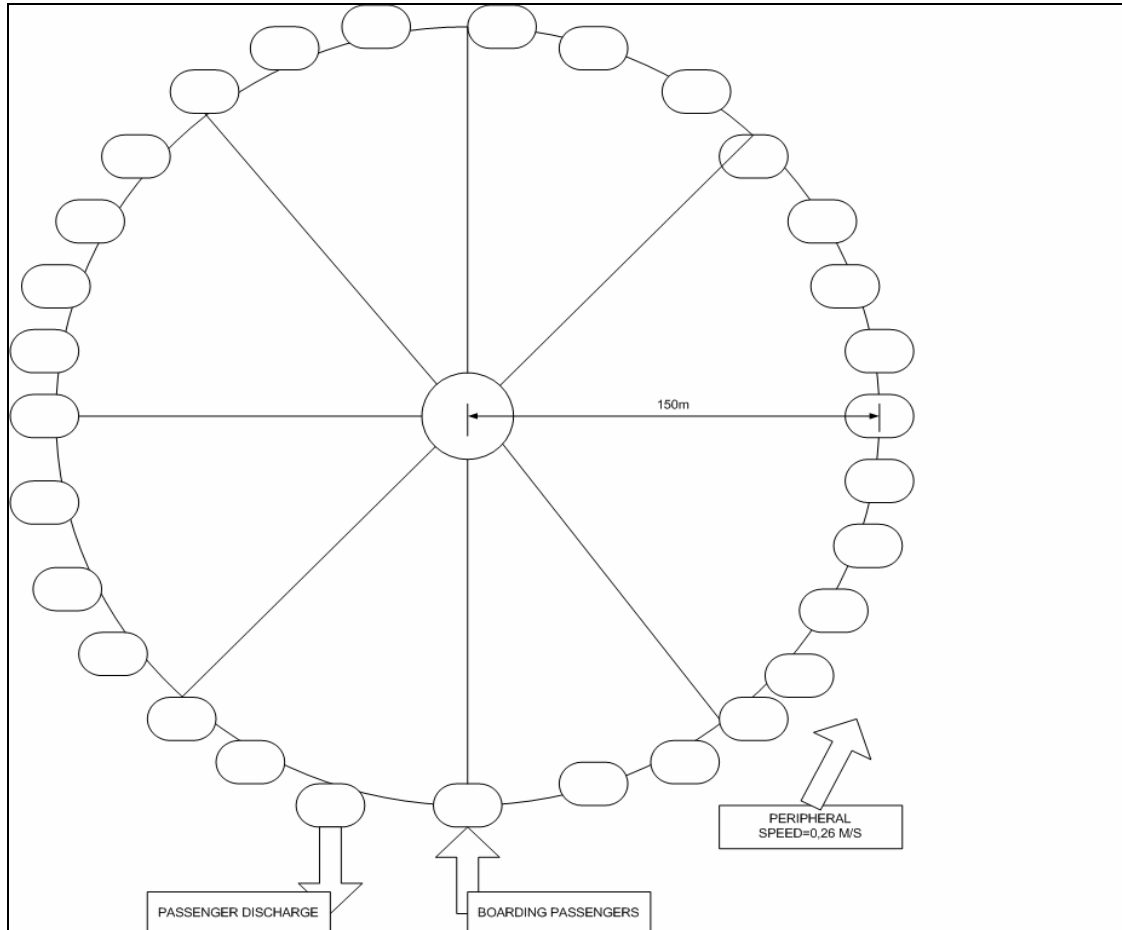


Figure 7.4.1.1 SCHEMATIC REPRESENTATION OF FERRIS WHEEL SYSTEM

7.4.3 Diagnosis of the model results

The system population values of Fig. 7.4.1.2 (C) which closely agree with observations on site indicate serious congestion, for example when the average system arrival rate shown in Fig. 7.4.1.2 (A) is 50 entities per 2 minute interval between 11h00 and 13h30 (fixed by conditions upstream of the Ferris wheel waiting area) the total system population is often 1500

entities which implies that approximately 750 entities are waiting for service on a FIRST COME FIRST SERVED (FCFS) basis for 30 minutes. At ground zero on site serious congestion occurs and is to be seen to be believed.

As is normally the case with systems suffering from congestion one should consider some-or-other action to improve system performance.

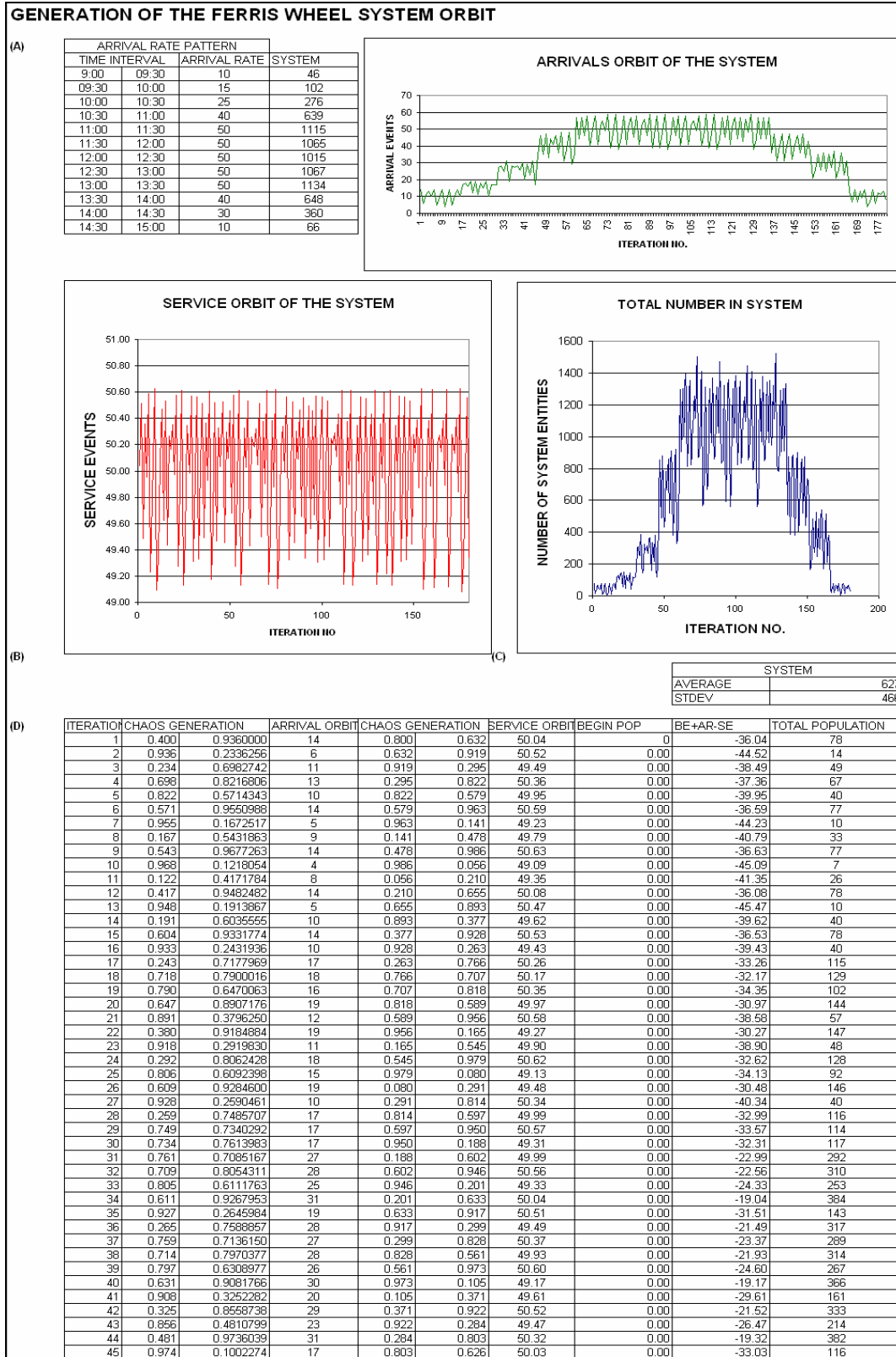


Figure 7.4.1.2 GENERATION OF THE FERRIS WHEEL SYSTEM ORBIT

7.4.4 Using realtime feedback to improve system performance

To decrease the number of entities in the system one may consider increasing the service rate by increasing the peripheral speed of the Ferris wheel and/or by limiting the average arrival rate under peak demand conditions. To demonstrate the effect of realtime feedback the operation of the system could be geared to increase the service rate when the peak population exceeds a predetermined value.

The effect of such a realtime feedback arrangement could be tested by using the following feedback rule: “as soon as the system population exceeds 1200 alter the average service rate to 60 entities per 2 minute interval and reset the average service rate to 50 entities per 2 minute interval as soon as the system population becomes less than 600”.

The results of analysing the system with realtime feedback are shown in Fig. 7.4.1.3 which indicates that the degree of congestion is considerably diminished. One may pose the question whether one could not gain greater congestion improvement by a further increase of the service rate. The maximum feasible service rate is however 60 entities per 2 minute interval for physical (ergonomic) reasons of loading and unloading at ground level.

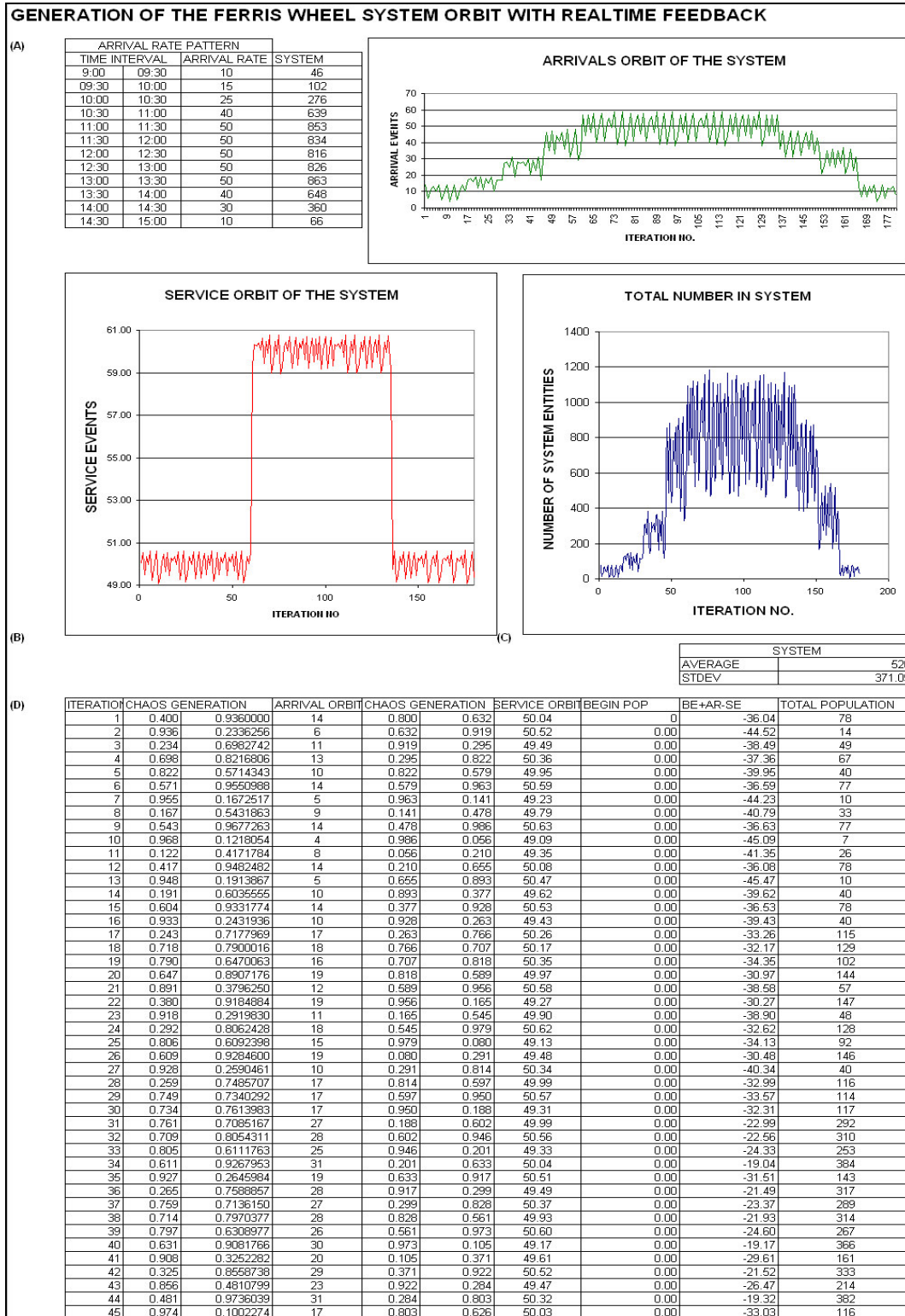


Fig. 7.4.1.3 GENERATION OF THE FERRIS WHEEL SYSTEM ORBIT WITH REAL TIME FEEDBACK

7.4.5 Concluding comments on System No. 3

The analysis of the system confirms that attempting to improve the operation of System No. 3 does not afford much leeway in effort since one is constrained to optimize within bounds such as:

- structural reasons in respect of peripheral speed,
- ergonomic reasons relating to loading and unloading of entities while the Ferris wheel is in motion,
- passenger value-for-money by making use of the wheel for visual entertainment for a period of time, and last but not least
- specified economic performance of the system.

7.5 SYSTEM NO. 4

7.5.1 System scenario

System No. 4 is an example of a typical municipal call centre that handles enquiries and problem reporting by a given urban population. The call centre has replaced some previously existing service centres. The specific system under consideration handles general enquiries and is the reporting centre for interruptions of service. Occasionally it occurs that the centre becomes congested. The resulting peak conditions of congestion will be the focus of the study.

The call centre operates in the following way. Entities phone the centre, a computer answers the call and the entity has a range of options to choose from. The entity has the following options:

- Report a failure or general enquiries.
- Choose a region of failure.
- Listen to a scenario of reported failures.

The average delay is 1.6 minutes and 16% of entities abandon the call during this process. During the following process an operator serves entities on a FCFS basis. They form a queue if the operators are all busy. The specific congestion period that will be modelled is when an infrastructure failure occurs in one municipal region from 16h00 to 19h00.

7.5.2 The system model

Verhulst orbit generation for the arrivals is show in Fig. 7.5.2.1. The average arrival rate is determined for consecutive 15-minute intervals. The arrivals orbit is similar to the conditions that prevail on site.

The service orbit for a single operator is shown in Fig. 7.5.2.2. The actual call duration is described by a general distribution with a mean service rate of 20 calls per hour and a standard deviation of 1.6. The total system was modelled

using the same multi-channel process as for System No. 2 of par. 7.3. The results are shown in Fig. 7.5.2.3 for two service lanes. It shows the arrivals, the abandonment rate and the number of operators in use. The average number of calls in the system is 6 for the period from 16h00 to 19h00.

GENERATION OF THE ARRIVAL ORBIT AT THE CALL CENTRE

(A)

ORBIT GENERATION FOR FIRST 42 INTERVALS			
ITERATION NO	x	F(x)	ARRIVAL ORBIT
1	0.400	0.936	8
2	0.936	0.234	1
3	0.234	0.698	6
4	0.698	0.822	7
5	0.822	0.571	5
6	0.571	0.955	8
7	0.955	0.167	1
8	0.167	0.543	4
9	0.543	0.968	8
10	0.968	0.122	0
11	0.122	0.417	3
12	0.417	0.948	8
13	0.948	0.191	1
14	0.191	0.604	5
15	0.604	0.933	8
16	0.933	0.243	1
17	0.243	0.718	6
18	0.718	0.790	7
19	0.790	0.647	5
20	0.647	0.891	8
21	0.891	0.380	3
22	0.380	0.918	8
23	0.918	0.292	2
24	0.292	0.806	7
25	0.806	0.609	5
26	0.609	0.928	8
27	0.928	0.259	2
28	0.259	0.749	6
29	0.749	0.734	6
30	0.734	0.761	6
31	0.761	0.709	6
32	0.709	0.805	7
33	0.805	0.611	5
34	0.611	0.927	8
35	0.927	0.265	2
36	0.265	0.759	6
37	0.759	0.714	6
38	0.714	0.797	7
39	0.797	0.631	5
40	0.631	0.908	8
41	0.908	0.325	2
42	0.325	0.856	7

(C)

ARRIVAL RATE PATTERN		
TIME INTERVAL		ARRIVAL RATE
16:00	16:15	5
16:15	16:30	5
16:30	16:45	5
16:45	17:00	8
17:00	17:15	12
17:15	17:30	25
17:30	17:45	25
17:45	18:00	15
18:00	18:15	10
18:15	18:30	5
18:30	18:45	5
18:45	19:00	5

(B)

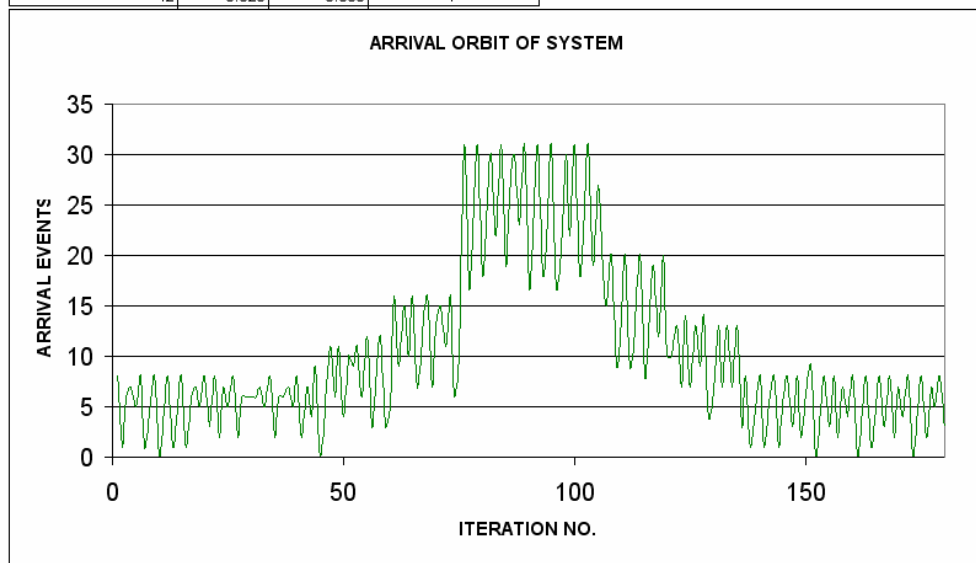


Figure 7.5.2.1 GENERATION OF ARRIVALS AT THE CALL CENTRE

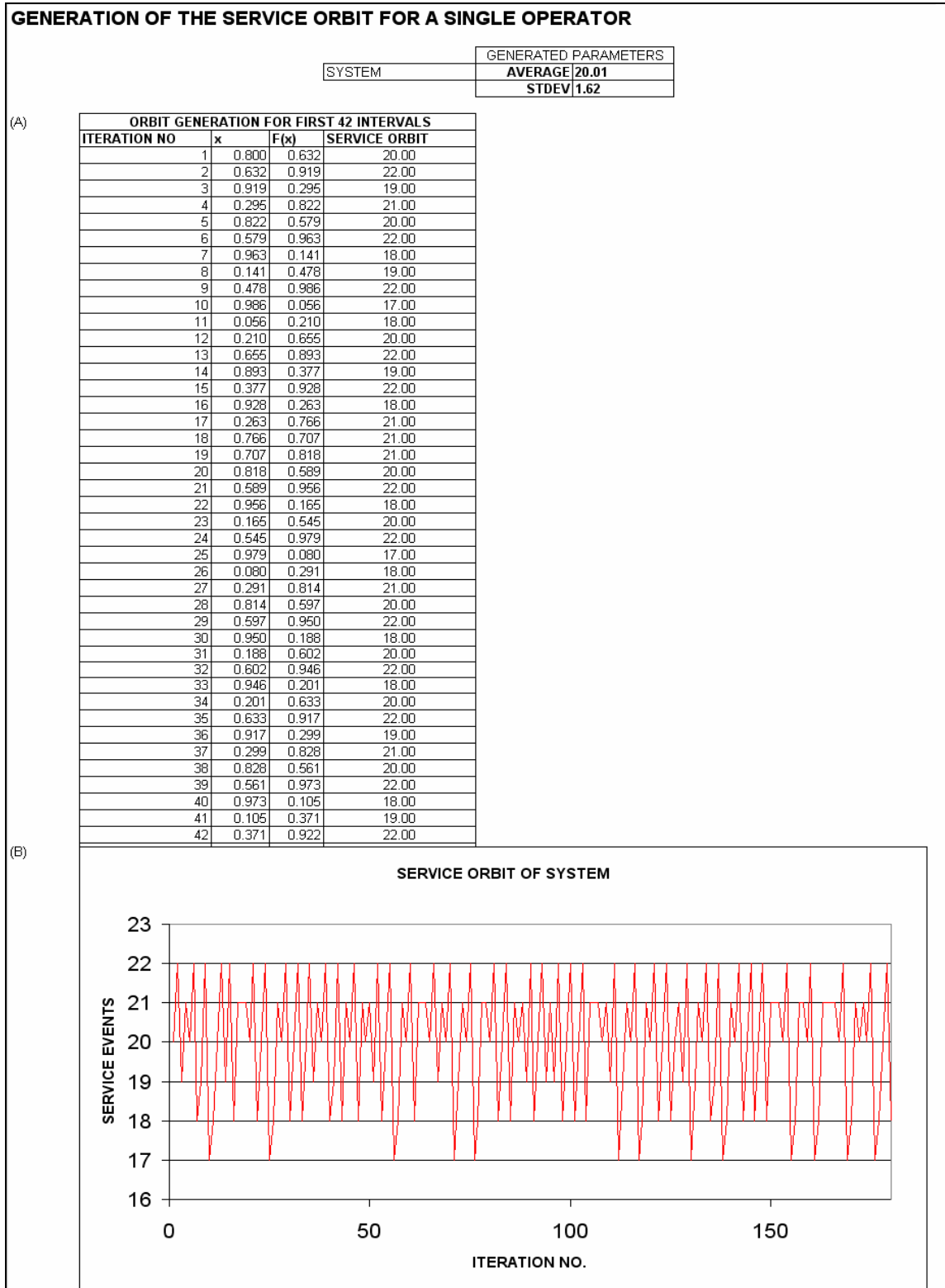


Figure 7.5.2.2 GENERATION OF A TYPICAL SERVICE ORBIT FOR A SINGLE OPERATOR

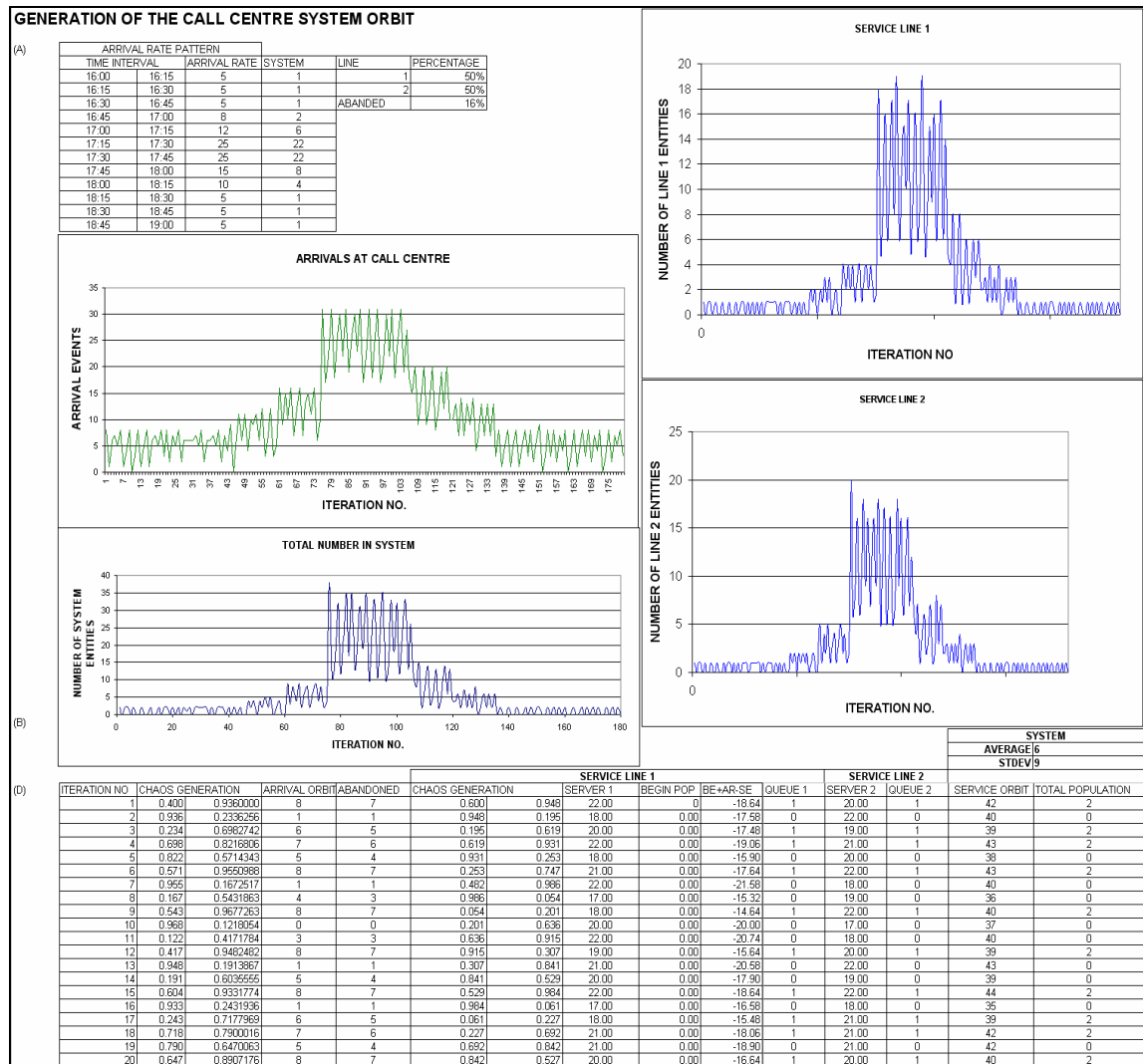


Figure 7.5.2.3 GENERATION OF THE CALL CENTRE SYSTEM ORBIT

7.5.3 Diagnosis of the model results

The system population value of Fig 7.2.2.3 is closely related to the “actual” situation. This system arrival pattern differs from the previous models in that the sudden peak in arrival rate is more dramatic. If the system is operating under normal operating conditions the average number of entities in the queue is one. During the sudden increase the queue length increases to 36.

The next step is to improve the system by focussing on the average number in the system and the standard deviation. The latter value must also be decreased.

7.5.4 Using realtime feedback to improve system performance

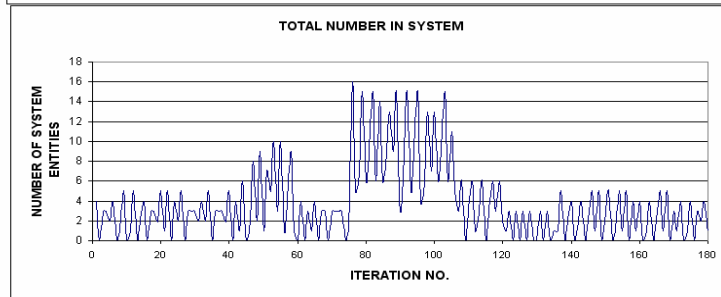
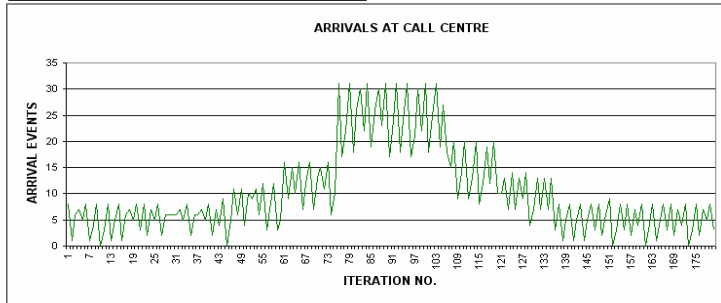
When one analyses the system one could attempt to decrease the peak load on the system by varying the number of service channels. The other concern is the idle time that system operators may have during normal uncongested operating conditions. The difficulty with the system is that one cannot predict when peak conditions will occur. It will also help when a major infrastructure failure of service occurs, that the computer communication menu be adapted to cause an increase of the call abandonment rate by offering the scenario of a history of reported failures first. The system management will have to use multi-tasking to limit the operating cost.

These improvements have been modelled and the results are shown in Fig. 7.5.2.4. This improvement has decreased the average number in the system and the standard deviation to 4 and 4 respectively.

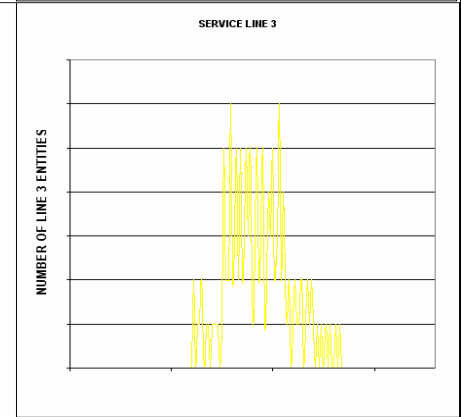
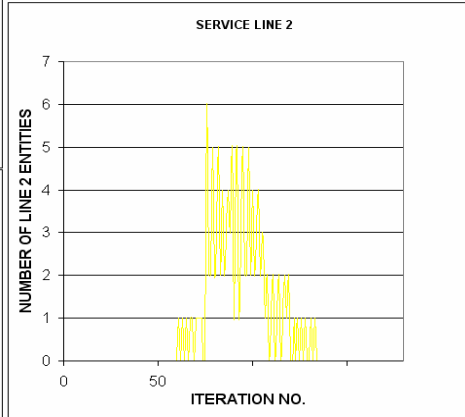
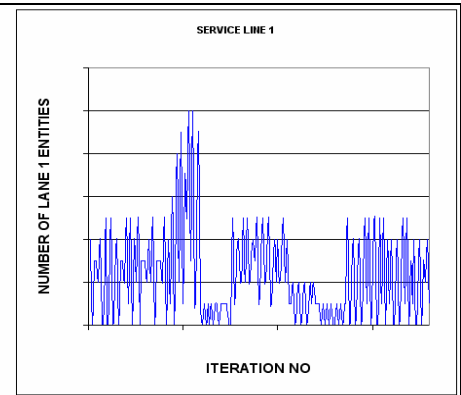
GENERATION OF THE CALL CENTRE SYSTEM ORBIT

(A)

ARRIVAL RATE PATTERN		SYSTEM	LINE	PERCENTAGE	CONGESTED	
TIME INTERVAL	ARRIVAL RATE					
16:00	16:15	5	2	1	100%	33%
16:15	16:30	5	3	2	0%	33%
16:30	16:45	5	3	3	0%	34%
16:45	17:00	8	5	ABANDONED	16%	20%
17:00	17:15	12	2			
17:15	17:30	25	9			
17:30	17:45	25	10			
17:45	18:00	15	4			
18:00	18:15	10	1			
18:15	18:30	5	2			
18:30	18:45	5	2			
18:45	19:00	5	2			



(B)



(D)

ITERATION NO	CHAOS GENERATION	ARRIVAL ORBIT	ABANDONED	NO. OF SERVER	CHAOS GENERATION	SERVICE LINE 1			SERVICE LINE 2		SERVICE LINE 3		SYSTEM		
						SERVER 1	BEGIN POP	BE+AR-SE	QUEUE 1	SERVER 2	QUEUE 2	SERVER 3	QUEUE 3	SERVICE ORBIT	TOTAL POPULATION
1	0.400	0.9360000	8	7	1	0.600	0.948	22.00	0	-15.28	4	0.00	0	18	4
2	0.936	0.2336256	1	1	1	0.948	0.195	18.00	0.00	-17.16	0	0.00	0	20	0
3	0.234	0.6982742	6	5	1	0.195	0.619	20.00	0.00	-14.96	3	0.00	0	22	3
4	0.698	0.8216806	7	6	1	0.619	0.931	22.00	0.00	-16.12	3	0.00	0	18	3
5	0.822	0.5714343	5	4	1	0.931	0.253	18.00	0.00	-13.80	2	0.00	0	21	2
6	0.571	0.9550988	8	7	1	0.253	0.747	21.00	0.00	-14.26	4	0.00	0	22	4
7	0.955	0.1672517	1	1	1	0.482	0.986	22.00	0.00	-21.16	0	0.00	0	17	0
8	0.167	0.5431863	4	3	1	0.986	0.054	17.00	0.00	-13.64	1	0.00	0	18	1
9	0.543	0.9677263	8	7	1	0.054	0.201	18.00	0.00	-11.28	5	0.00	0	20	5
10	0.968	0.1218054	0	0	1	0.201	0.636	20.00	0.00	-20.00	0	0.00	0	22	0

Figure 7.5.2.4 GENERATION OF CALL CENTRE SYSTEM ORBIT WITH REALTIME FEEDBACK

7.5.5 Concluding comments on System No. 4

System No. 4 was modelled successfully. The system orbit depicts useful results. The uniqueness of the system is clearly shown in the results. Realtime feedback will definitely improve the system congestion.

7.6 EVALUATION OF THE MODELLING METHODS AND ACHIEVEMENT OF DYNAMIC OPERATION RESULTS OF COMPLEX SYSTEMS OF CONGESTION

This chapter of the thesis demonstrates the application of chaos orbit generation methods to Systems of Congestion which are more complex in nature than those studied in Chapter 6. The orbit generation methods are deployed in each case in a system wide fashion that matches/ serves the accuracy of modelling of each of the individual systems studied.

In each instance the actual dynamic performance of the system was acceptably replicated by the chaos orbit generation method. These encouraging results consequently paved the way for improvement of the congested conditions by fashioning feedback bouquets. The use of chaos generation methods is therefore supported to such an extent that the **initial conjecture** that these methods could **possibly be used effectively** has achieved the status of **an assertion**.

CHAPTER 8

CONCLUSION

If we offend, it is with our good will, that you should think, we come not to offend, but with good will. To show our simple skill, that is the true beginning of our end. Shakespeare [54].

The research has attained a measure of achieving the formal objective of the work as set out and implied in the first chapter of this thesis, i.e. to address and investigate the phenomenon of congestion as and where it occurs. One may categorically state that such an investigation could hardly have been attempted without using the methods of analysis established by previous human generations, albeit that such method only emerged from an embryonic state after the Middle Ages.

The field of study which would attract the attention of theorists was Queueing Theory which would develop over the past century by leaps and bounds. The analytical investigation of stochastic processes has continued unabated up to the present time. The adjectival use of the term “stochastic” means that a probability function is generating an ordered sequence of events which in the context of Systems of Congestion means that the ordering is usually related to time.

The goal of the thesis as described in Chapter 1 has been to develop analytical skills related to real-world systems. The worth of models is then measured by their utility in dealing with real practical systems rather than by their mathematical elegance.

Chapter 2 deals with the necessary introductory matter of parameter estimation of the random phenomena of stochastic models. The methods used are demonstrated in respect of two queueing models which are subjected to inference investigation. The essence of the chapter lies therein that the value/accuracy of a model must be established in a prescribed fashion.

Chapters 3, 4 and 5 of the thesis demonstrate how system characteristics are determined for three selected Systems of Congestion and how model

usefulness may be assessed. The systems are modelled via the classical birth-death postulates which are adapted to fit the particular system configuration.

The models offer closed-form steady state solutions and transient analysis of system behaviour in response to sudden changes in input or service capabilities. Notwithstanding their elegance the resulting equations are overly complex for use by an average practitioner, or too cumbersome or awkward for useful manipulation. This statement may be supported by referring to the case of a simple M/M/1 queue which is the most “easily” solved of all queueing models. To initially find the transient solution, and once available, to use the solution, is sufficiently daunting to discourage all but curious and skilled practitioners from further exploration.

The contention may consequently be expressed that even if closed-form models may be created, and that they may be used with considerable computational burden, it is prudent to resort to simulation studies in an attempt to conveniently analyze the time varying behaviour of Systems of Congestion.

Chapter 6 attempts to emulate the generation of an ordered sequence of arrival and service events for simple queues via Chaos-based generation of arrival and service orbits, which orbits are then used to model the dynamics of the system population via a sequence of model iterations. The chosen modelling method, which inter alia makes use of “designer equations”, achieves acceptable performance without the fuss and bother of intricate and tedious manipulation.

Chapter 7 extends the applicability of the modelling approach to more practical and complex real-life Systems of Congestion and indicates how system congestion may be alleviated by modifying the system under consideration via a symbiotic partnership between model and modeller. Manipulation of the selected models is emphasized by spreadsheet iteration results which graphically exhibit the system dynamics as a matter of course.

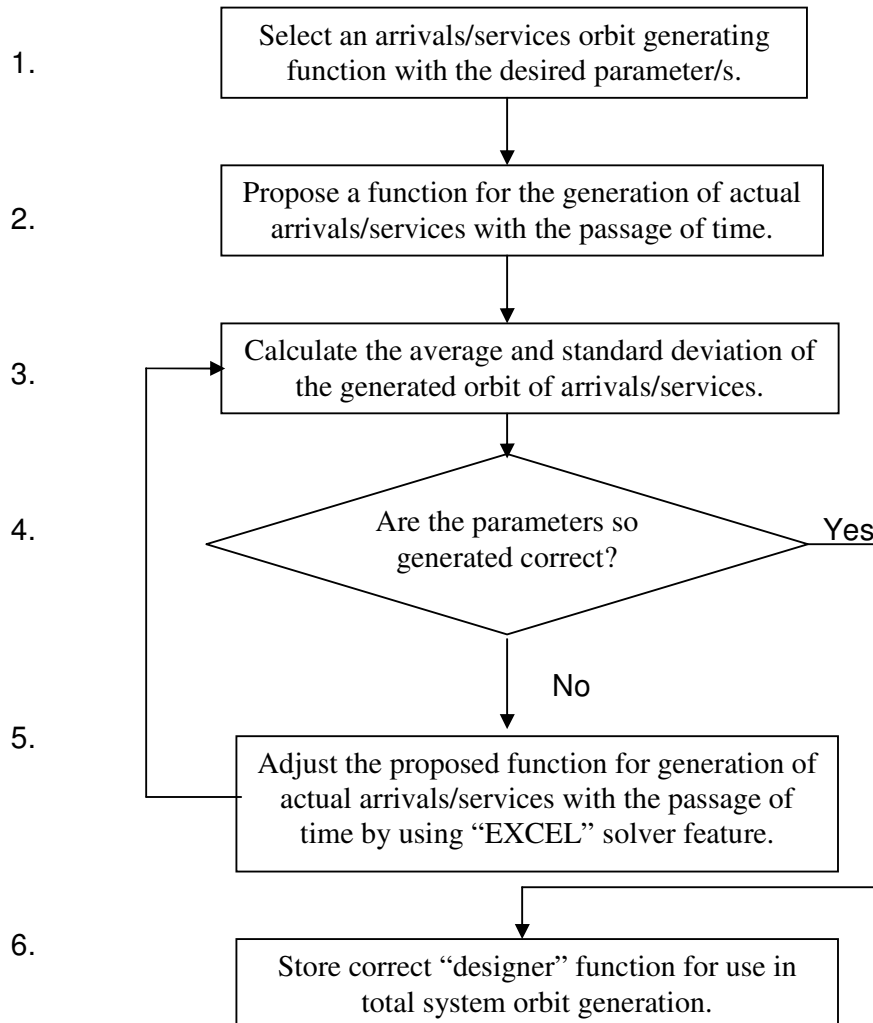
Construction of an iterative dynamical system model requires that a practitioner should:

- comprehend the structure of the System of Congestion in its entirety,
- place the necessary emphasis on those facets of the system structure which must be modelled accurately,
- identify those facets of the system structure which cannot significantly affect the system dynamics,
- assemble the available data on system operation and manipulate such information so that it proves to be useful during model iteration,
- assess how and where interaction with the modelled system may via amendment be most beneficial,
- select suitable methods of system orbit generation,
- ensure that the required designer equations are employed to shape the system orbit, and
- finally construct a desired system operation objective (minimize waiting time, minimize total system entities, maximize system “efficiency”, maximize customer pleasure etc.) by means of manipulation/adjustment of the system structure.

In conclusion one may submit the supposition that searching for robust simple models which deliver credible and useful solutions to complex design and operation problems of Systems of Congestion does have considerable merit.

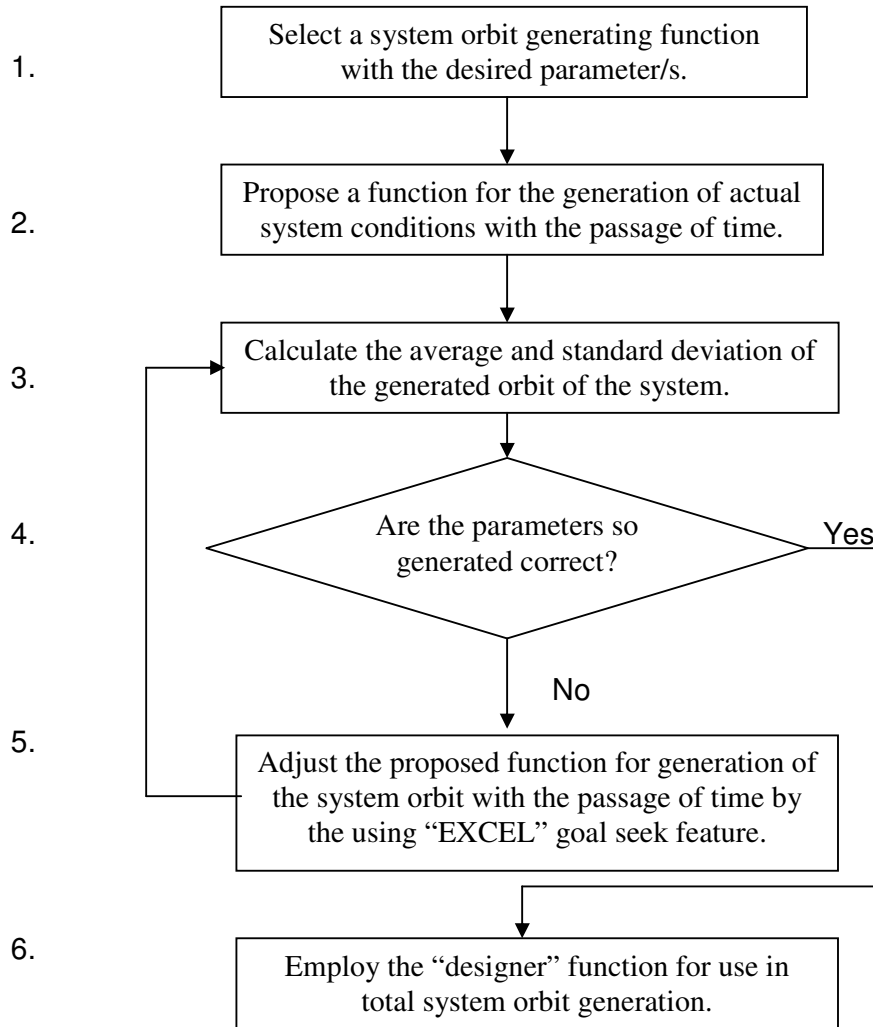
APPENDIX A

Flow diagram for the design of an arrivals/service orbit generating function.



APPENDIX B

Flow diagram for the design of a system orbit generating function.



REFERENCES

- [1] **Aristotle**, Volume II, "On Poetics", 1460a, University of Chicago, 1952.
- [2] **Gross, D. and Harris, G.** 1974. "Fundamentals of queueing theory", Wiley, New York
- [3] **Taha, H.A.** Seventh Edition 2003. "Operations Research. An introduction", Prentice Hall, New Jersey
- [4] **Grosh, D. L.** 1989. "A Primer of reliability theory", p.41, John Wiley and Sons, New York
- [5] **Giffin, C.** 1978. "Queueing: basic theory and applications", Grid Inc, Columbus, Ohio
- [6] **Pollaczek, F.** "Concerning an analytic method for the treatment of queueing Problems." *Proceedings of the Symposium on Congestion Theory*, University of North Carolina, Chapel Hill, (1964) 1-42.
- [7] **Bailey, N. T. J.** "A continuous time treatment of a simple queue using generating functions." *Journal of the Royal Statistical Society, Ser. B.*, 16, (1956) 288-291.
- [8] **Lederman, W and Reuter, G. E. G** (1954) "Spectral theory for the differential equations of simple birth and death processes" *Phil. Trans. Royal Society, London, Sev. A, Vol. 246.* pp321-369.
- [9] **Champernowne, D. G.** (1956) "An elementary method of solution of the queueing problem with a single server and constant parameter" *Journal of the Royal Statistic Society, Sev. B, Vol 18,* pp125-pp128.
- [10] **Kendall, D. G.** "Some problems in the theory of queues." *Journal of the Royal Statistical Society, Ser. B.*, 13, (1951) 151-185.
- [11] **Cox, D. R.** "The analysis of non-markovian stochastic processes by the inclusion of supplementary variables." *Proceedings of Cambridge Philosophical Society*, 51, (1955) 443-441
- [12] **Bhat, U. N.** "Sixty years of queueing theory," *Management Science.* 15, (1969) B280-B292
- [13] **Keilson, J. and Kooharian, A.** "On time dependant queueing processes." *Annals Of Mathematical Statistics*, 31, (1960) pp104-112

- [14] **Takacs, L.** *Introduction to the theory of queues*. New York: Oxford University Press, 1962.
- [15] **Borthakur, A** (1975), "On busy period of a bulk queueing system with a general rate for bulk service, *Operational Research*, Vol. 12, pp40-46.
- [16] **Madan, K.C.** (1991), "A single server bulk input queue with two kinds of service" *ACTACIENCIA INDICA*, Vol. XVII, pp25-32.
- [17] **Madan, K.C.** (1994a), "Balking and reneging in a vacation queueing model" *Re Vista Colombiana Estatica*, Vol. 20, pp355-369.
- [18] **Madan, K.C.** (1994b), "An M/G1 queueing system with additional optional service and no working capacity, *Microelectronics & Relay*, Vol. 34, pp521-528.
- [19] **Sapna, K. P.** (1996) "An M/G/1 type queueing system with perfect servers and no waiting capacity", *Microelectronics & Relays*, Vol. 36, pp697-700.
- [20] **Fuhrmann, S. W.** (1984), "A note on the M/G/1 queue with server vacations," *Operational Research*, Vol. 32, pp1368-1373.
- [21] **Doshi, B. T.** (1986), "Queueing systems with vacation a survey," *Queueing Systems*, Vol. 1, pp29-66.
- [22] **Levi, Y. and Yechiali, V.** (1976), "An M/M/s queue with servers vacation," *INFOR*, Vol. 14, pp153-163.
- [23] **Fuhrmann, S. W. and Cooper, R. B.** (1985), "Stochastic decompositions in the M/G/1 queue with generalised vacations," *Operational Research*, Vol. 33, No. 5, pp1117-1129.
- [24] **Shanthikumar, J. G.** (1988), "On stochastic decomposition in the M/G/1 queue with generalised vacation, *Operational Research*, Vol. 36, pp566-569.
- [25] **Keilson, J. and Servi, L. D.** (1990), "The distributional form of Little's Law and the Fuhrmann and Cooper decomposition, *Operational Research Letters*, Vol. 9, pp239-248.
- [26] **Keelson, J. and Ramaswamy, R** (1988), "The backlog and depletion time process for M/G/1 vacation models with exhaustive service discipline, *Jour. Application Problems*, Vol. 25, pp404-412.

- [27] **Latouche, G. and Ramaswami, V** (1999), "Introduction to matrix analytic methods in stochastic modeling", SAIM, Philadelphia, Pennsylvania, and ASA, Alexandria, Virginia.
- [28] **Levy, H. and Kleinrock, L.** (1986), "A queue starter and a queue with vacation delay analysis by decomposition," Operational Research, Vol. 34, pp426-436.
- [29] **Leung, K. K.** (1992), "On the additional delay in an M/G/1 queue with generalised vacations and exhaustive service," Operational Research, Vol. 40, pp5272-5283.
- [30] **White, H. and Chritie, L. S.** (1958). "Queueing with pre-emptive priorities or with breakdown", Operational. Research. Vol. 6, pp. 79-95.
- [31] **Jaiswal, N. K.** (1968), "Priority queues," Academic Press, New York.
- [32] **Heathcote, C. R.** (1961) "Pre-emptive priority queues," Biometrika, Vol. 48, pp57-63
- [33] **Thiruvengadam, K** (1963), "Queueing with breakdowns," Operational Research, Vol. 11, pp62-71
- [34] **Devaney, R. L.** (1992) "A first course in chaotic dynamical systems", Westview Press, Colorado.
- [35] **Gleick, James**, 1988, "Chaos", Vintage, London.
- [36] **Malthus, Thomas Robert**, 1798. An essay on the principle of population.
- [37] **Verhulst, Pierre-Francois:** *Récherches mathématiques sur la loi d'accroissement de la population*; Nouv. Mém. de L'Acad. Roy. des Sciences et Belles-Lettres de Bruxelles, XVIII.8, 1-38.
- [38] **Stewart, Ian.** *Does God play dice?*; Penguin; 1990
- [39] **Schroeder, Manfred.** *Fractals, chaos, power laws*; W.H. Freeman; New York; 1991.
- [40] **Feigenbaum, Mitchell J.** *Quantitative universality for a class of nonlinear transformations*; J. Stat. Phys. 19 (1978), 25-52.
- [41] **Bhat, U. N.** (2003) Parameter estimation in $M|G|1$ and $GI|M|1$ queues using queue length data, stochastic point processes (S.K. Srinivasan and A. Vijayakumar, Eds.), Narosa Publishing House, New Delhi, pp 96-107.

- [42] **Rao, C. R.** (1974) *Linear statistical inference and its applications*, Wiley, New York
- [43] **Sinha, S. K.** (1986) *Reliability and life testing*, Wiley Eastern Ltd., New Delhi
- [44] **Yadavalli, V.S.S.** Chandrasekhar, P. and Deepa, S. P. (2004)- Confidence limits for expected waiting time of two queueing models (to appear in ORION, South Africa)
- [45] **Yadavalli, V. S. S.** Chandrasekhar, P. and Natarajan, R. (2004)- Confidence limits for expected waiting time of M/M/C/ ∞ and M/M/C/N queueing models (communicated for publication)
- [46] **Takagi, 1991**, *Stochastic analysis of computer and communication systems*, New York: North-Holland New York, N.Y., U.S.A. , Elsevier Science Pub. Co, 1991.
- [47] **Widder, D.V.** (1941), "The Laplace transform," Princeton University Press, Princeton, V.S.A.
- [48] **Little, J. D. C.** "A Proof for the queueing formula $L=\lambda W$." Operational Research, 9, (1961) pp383-387
- [49] **Madan, K. C.** (1992), "A queueing system with two types of failure and priority for vital repairs," Journal Mathematics Analytical Applications, Vol. 25, pp88-97
- [50] **Jain, M. and Sharma** (1984), "An iterative approach for solving machine interference problem with interrupted service including priority", Journal of Mathematics Analytics Applications, Vol. 19, pp76-84.
- [51] **Reddy, C. R.** (1993), "Optimisation of K out of N systems subject to common cause failure with repair provision," Microelectronics & Relays, Vol. 33, pp175-183.
- [52] **Sridharan and Jayashree, P. J.** (1996), "Some characteristics on a finite queue with normal, partial and total failures," Microelectronics & Relays, Vol. 32, No.2, pp265-267.
- [53] **Saaty, T. L.** Elements of queueing theory. New York: McGraw-Hill, 1961.
- [54] **Shakespeare, W** "A *midsummer-night's dream*", Act 5, Scene 1, Prologue

The following publications represent a variety of treatments of chaos and fractals:

- [1] **Shub, M.** Global Stability of Dynamical Systems, Springer-Verlag, New York, 1986
- [2] **Tufillaro, N., Abbott, T., and Reilly, J.** An Experimental Approach to Non-linear Dynamical Systems, Addison-Wesley, Redwood City, CA, 1992
- [3] **Parker, T. and Chua, L.** Practical Numerical Algorithms for Chaotic Systems, Springer-Verlag, New York, 1989.
- [4] **Baker, G. L. and Gollub, J. P.** Chaotic Dynamics: An Introduction, Cambridge University Press, Cambridge, 1990
- [5] **Beardon, A.** Iteration of rational Functions, Springer-Verlag, New York, 1991
- [6] **Abraham, R. and Shaw, C.** Dynamics: The Geometry of Behavior, Addison-Wesley, Redwood City, CA, 1992.
- [7] **Devaney, R. L.** Chaos, Fractals, and Dynamics: Computer Experiments in Mathematics, Addison-Wesley, Menlo Park, 1990.
- [8] **Sandefur, J.** Discrete Dynamical Systems: Theory and applications, Clarendon Press, Oxford, 1990.
- [9] **Hall, R.** Queueing Methods for Service and Manufacturing, Englewood Cliffs, N.J.: Prentice Hall, 1991.
- [10] **Tanner, M.** Practical Queueing Analysis, New York: McGraw-Hill, 1995.
- [11] **Green, J.** "Managing a Telephone System Demands Skill" The Office (November 1987): pp144-145.