GEORGIA INSTITUTE OF TECHNOLOGY | OFFICE OF CONTRACT ADMINISTRATION

PROJECT ADMINISTRATION DATA SHEET

[X] ORIGINAL [ ] REVISION NO. ______

Project No. E-24-628 GTRI/~~GIT~~ DATE 10 / 2 / 84

Project Director: R.F. Serfozo School/~~Lab~~ ISyE

Sponsor: Air Force Office of Scientific Research
Building 410 Bolling AFB, D.C. 20552

Type Agreement: Grant No. AFOSR-84-0367

Award Period: From 9/30/84 To ~~9/30/85~~ 9-29-86 (Performance) 10/30/85 (Reports)

Sponsor Amount: This Change | Total to Date

Estimated: $ ______ $ ______

Funded: $ 69,387 $ 69,387

Cost Sharing Amount: $ – Cost Sharing No: –

Title: Extreme Values of Queues, Point Processes and Stochastic Networks

ADMINISTRATIVE DATA

OCA Contact Ralph Grede X4820

1) Sponsor Technical Contact:

Captain Brian W. Woodruff
AFOSR/NM
Building 410
Bolling AFB, D.C. 20332

2) Sponsor Admin/Contractual Matters:

AFOSR/PKZ
Building 410
Bolling AFB D.C. 20332

Defense Priority Rating: None

Military Security Classification: None
(or) Company/Industrial Proprietary: n/a

RESTRICTIONS

See Attached Grant Supplemental Information Sheet for Additional Requirements.

Travel: Foreign travel must have prior approval – Contact OCA in each case. Domestic travel requires sponsor approval where total will exceed greater of $500 or 125% of approved proposal budget category.

Equipment: Title vests with GIT - However none proposed.

COMMENTS:

The indirect rate is fixed for the period of this project at 55.3%.

COPIES TO:

Project Director
Research Administrative Network
Research Property Management
Accounting

Procurement/EES Supply Services
Research Security Services
Reports Coordinator (OCA)
Research Communications (2)

GTRI
Library
Project File
Other A. Jones

FORM OCA 4:383

GEORGIA INSTITUTE OF TECHNOLOGY

OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION/CLOSEOUT SHEET

Date December 10, 1986

Project No. E-24-628

SchooXXXX ISyE

Includes Subproject No.(s) N/A

Project Director(s) R. F. Serfozo

GTRC / XXX

Sponsor AFOSR

Title Extreme Values of Queues, Point Processess and Stochastic Networks

Effective Completion Date: 9/29/86 (Performance) 11/29/86 (Reports)

Grant/Contract Closeout Actions Remaining:

- [ ] None
- [x] Final Invoice or Final Fiscal Report
- [ ] Closing Documents
- [ ] Final Report of Inventions
- [ ] Govt. Property Inventory & Related Certificate
- [ ] Classified Material Certificate
- [ ] Other ______

Continues Project No. ______

Continued by Project No. E-24-659

COPIES TO:

Project Director
Research Administrative Network
Research Property Management
Accounting
Procurement/GTRI Supply Services
Research Security Services
~~Reports Coordinator (OCA)~~
Legal Services

Library
GTRC
Research Communications (2)
Project File
Other I. Lashley
A. Jones
R. Embry

FORM OCA 69.285

Industrial and Systems Engineering
Report Series No. J-85-10
July, 1985

STOCHASTIC BOUNDS ON DISTRIBUTIONS OF OPTIMAL VALUE FUNCTIONS WITH APPLICATIONS TO PERT, NETWORK FLOWS AND RELIABILITY

Gideon Weiss
Georgia Institute of Technology
and
Tel Aviv University

This research was partly carried out during a visit to RWTH Aachen, supported by the DAAD.

Partial support for this research was provided by Grant AFOSR 84-0367.

STOCHASTIC BOUNDS ON DISTRIBUTIONS OF OPTIMAL VALUE FUNCTIONS

WITH APPLICATIONS TO PERT, NETWORK FLOWS AND RELIABILITY

Gideon Weiss

Georgia Institute of Technology and Tel Aviv University

Abstract

Meilijson and Nadas [1979] have obtained stochastic bounds in the convex majorisation sense to the critical path length of a project network with random activity durations. In this paper we present those results in a more general framework and, using similar techniques, obtain bounds for shortest route, maximal flow and reliability system lifetime.

Subject classification: #488 Bounds for stochastic networks

#672 Convex majorisation of project critical path length.

#725 Stochastic majorisation of reliability system lifetime.

Consider a set $I = \{1, \ldots, n\}$ of n nodes, the base set. Let $I_1, \ldots, I_k$ be subsets whose union is I, and no two of which are ordered by inclusion; $\{I_j\}$ $1 < j < k$ is a clutter over I. The blocking clutter to $\{I_j\}$ is a clutter $J_1, \ldots, J_\ell$ such that $I_r \quad J_s \neq \phi$ for all r, s, and J_j are minimal sets with this property, cf. Edmonds and Fulkerson [1970]. In a directed acyclic graph or in a two terminal network, the paths and cuts are an example of a pair of blocking clutters. We call I a system, and $\{I_j\}$ $1 \leq j \leq k$, $\{J_j\}$ $1 \leq j \leq \ell$ the paths and cuts of the system. Let a weight X_i be associated with each node i of system I. In many combinatorial optimization problems the system has an optimal value function, a function of $X_1, \ldots, X_n$, which is defined by the clutter of paths or of cuts. To illustrate:

- Critical path of a PERT network (Elmaghraby [1977]): the nodes represent activities, the weights activity durations, the network the precedence constraints. The critical path length is the shortest time needed to complete the project, given by

$$M = \max_{1 \leq j \leq k} \sum_{i \in I_j} X_i \tag{1}$$

over the clutter of paths.

- Maximal flow (Ford and Fulkerson [1962], Lawler [1976]): the nodes represent pipelines, the weights maximal flow capacities. The maximal flow through a network from source to sink is:

$$L = \min_{1 \leq j \leq \ell} \sum_{i \in J_j} X_i \tag{2}$$

over the clutter of cuts.

- Shortest route (Ford and Fulkerson [1962], Lawler [1976]): the nodes represent sections of routes, the weights their lengths, the network their connections; the shortest route from source to sink is given by L, over the clutter of paths.

- Reliability system lifetime (Barlow and Proschan [1975]): the nodes represent components, the weights their lifetimes. The system lifetime can be expressed in terms of the paths $\{I_j\}$ or the cuts $\{J_j\}$ as:

$$T = \max_{1 \leq j \leq k} \min_{i \in I_j} X_i = \min_{1 \leq j \leq \ell} \max_{i \in J_j} X_i \tag{3}$$

The formulation of M, L, T via clutters applies equally well to structures more general than networks, e.g. precedence relations among project activities can be defined by any partial order, and a reliability system can be defined by any Boolean coherent structure function. The equality (3) holds for any pair of blocking clutters, cf. Edmonds and Fulkerson [1970].

The stochastic behaviour of the optimal value functions M, L and T is introduced as follows. Let the weights $X_1, \ldots, X_n$ be random variables, with marginal distribution functions $F_1, \ldots, F_n$ and a joint distribution P. Then M, L, T are random variables. It is extremely difficult to obtain the distributions of M, L, T - this is so even in the case where $X_1, \ldots, X_n$ are independent, since different I_j's will in general have nodes in common and not be independent. Nor is it any easier to determine single values such as $E(V)$, $P(V > y)$, $E(V - y)^{\pm}$ where V is any of M, L or T (here $z^+ = \max(z, o)$, $z^- = (-z)^+$). Let $\mathfrak{p}$ denote the family of all the joint distributions of $X_1, \ldots, X_n$ with the given marginal distributions $F_1, \ldots, F_n$. The subject of this paper is the investigation of:

$$\begin{aligned} \Psi(x) &= \sup_{\mathfrak{p}} E(M-x)^+ \\ \eta(x) &= \sup_{\mathfrak{p}} E(L-x)^- \\ \alpha(x) &= \sup_{\mathfrak{p}} P(T>x) \\ \beta(x) &= \sup_{\mathfrak{p}} P(T<x) \end{aligned} \tag{4}$$

We show how each of the functions Ψ, η, α, β can be calculated as the solution to an appropriate mathematical programming problem which is in general substantially easier than the calculation of $E(M-x)^+$, $E(L - x)^-$ or $P(T > x)$ for a particular $P \in \mathfrak{p}$. The suprema in (4) are attained for every x, that is, for every x there exists a joint distribution P for which $\Psi(x) = E(M - x)^+$, and similar distributions attain the supremum for

η, α, β. The joint distributions which attain these suprema can be chosen to have a special form. Define the inverse of a distribution function as

$$F^{-1}(u) = \inf \{x \mid F(x) > u\}$$

and let U be a uniform random variable on (0, 1). Then $X_1, \ldots, X_n = F_1^{-1}(\phi_1(U)), \ldots, F_n^{-1}(\phi_n(U))$, where U is common to all the nodes, and the functions $\phi_1, \ldots, \phi_n$ are piecewise linear, with a finite, not exceeding max (k, ℓ), number of discontinuities. $\phi_1, \ldots, \phi_n$ are obtained explicitly from the solution of the mathematical programming problems, together with additional structural information about the system.

The functions Ψ, η, α, β can be used to define random variables $\bar{M}$, $\underline{L}$, $\bar{T}$ and $\underline{T}$ as follows:

$$\begin{aligned} E(\bar{M} - x)^+ &= \Psi(x) \\ E(\underline{L} - x)^- &= \mu(x) \\ P(\bar{T} > x) &= \alpha(x) \\ P(\underline{T} < x) &= \beta(x) \end{aligned} \tag{5}$$

By their definition (4, 5), $\bar{M}$ is convexly greater, $\underline{L}$ is concavely smaller, $\bar{T}$ ($\underline{T}$) is stochastically greater (smaller) than M, L, T respectively, for any distribution $P \in \mathfrak{p}$.

The definitions of X stochastically greater than Y ($X >_{ST} Y$) and of X convexly (concavely) greater than Y ($X >_c Y$, ($X <_k Y$)) are (cf Stoyan [1983]):

$X >_{ST} Y \iff \forall x \quad P(X > x) > P(Y > x) \iff Eh(x) > Eh(Y) \quad \forall h$ nondecreasing

$X >_c Y \iff \forall x \quad E(X - x)^+ > E(Y - x)^+ \iff Eh(X) > Eh(Y) \quad \forall h$ convex nondecreasing

$X >_k Y \iff \forall x \quad E(X - x)^- < E(Y - x)^- \iff Eh(X) > Eh(Y) \quad \forall h$ concave nondecreasing

$\iff \forall x \quad E(x - X)^+ < E(x - Y)^+ \iff Eh(X) < Eh(Y) \quad \forall h$ convex nonincreasing

$\iff -X <_c -Y$

We say that the random variables $\bar{M}$, $\underline{L}$, $\bar{T}$, ($\underline{T}$) are convex upper, concave lower and stochastic upper (lower) bounds for M, L, T. Clearly by (4, 5) they are sharp bounds, in the sense that if for example $Z \geq_c M$ for every $P \in \mathfrak{p}$, then $Z \geq_c \bar{M}$. By the properties of $\geq_{ST}$, $\geq_c$, $\geq_k$, $Eh(\bar{M})$, $Eh(\underline{L})$, $Eh(\bar{T})$, $Eh(\underline{T})$ provide bounds for Eh(M), Eh(L), Eh(T) for every $P \in \mathfrak{p}$, whenever h has the appropriate monotonicity and convexity properties; these bounds are not necessarily sharp, unless $\bar{M}$, $\underline{L}$, $\bar{T}$, $\underline{T}$ are obtained within $\mathfrak{p}$.

In general, $\bar{M}$, $\underline{L}$, $\bar{T}$ and $\underline{T}$ are not obtained within $\mathfrak{p}$. If however the system is series parallel, then there exist joint distributions in $\mathfrak{p}$ for which $\bar{M} = M$, or $\underline{L} = L$ or $T = \bar{T}$, $T = \underline{T}$; bounds for series parallel systems are discussed in section 1, together with a discussion of modular decomposition. In sections 2, 3, 4 we discuss each of the optimal value functions, M, L and T separately. We conclude in section 5 with some general remarks on the type of bounds presented in this paper, and with a comparison with other types of bounds which appear in the literature.

The present work is based on a paper of Meilijson and Nadas [1979], who derived the properties of $\Psi(x)$. Some of the results on $\alpha(x)$, $\beta(x)$ have been previously obtained by Zemel [1982]. A brief summary of the present paper appeared in Weiss [1984]. Some related results and extensions appeared in klein Haneveld [1982], and Meilijson [1984].

1. Bounds for Series Parallel Systems.

The pure series system with nodes 1,..., n has a single path $I_1 = I = \{1, \ldots, n\}$ and n singleton cuts, $J_1 = \{1\}, \ldots, J_n = \{n\}$. The pure parallel system has paths $I_1 = \{1\}, \ldots, I_n = \{n\}$, and a single cut $J_1 = \{1, \ldots, n\}$. For the pure series system, $M = \sum_{i=1}^{n} X_i$, $L = T = \min_{1 \leq i \leq n} X_i$. For the pure parallel system $M = T = \max_{1 \leq i \leq n} X_i$, $L = \sum_{i=1}^{n} X_i$.

The following three special joint distributions of $X_1, \ldots, X_n$ are essential in this paper; they provide the bounds for the pure series and the pure parallel systems $\big($U is a uniform random variable on (0, 1), $\bar{F}(x) = 1 - F(x) = P(X > x)\big)$:

- The "perfect tracking" distribution P^*: $X_1, \ldots, X_n = F_1^{-1}(U), \ldots, F_n^{-1}(U)$.

- The "max antithetic" distribution P^{**} defined inductively for $n = 2$ by $X_1, X_2 = F_1^{-1}(U), F_2^{-1}(1-U)$ and, given P^{**} for $X_1, \ldots, X_{n-1}$ and $Y_{n-1} = \max\limits_{1 \le i \le n-1} X_i$, X_n and Y_{n-1} are distributed by $Y_{n-1}, X_n = F_{Y_{n-1}}^{-1}(U), F_n^{-1}(1-U)$.

- The "min antithetic" distribution P^{***} defined similarly to P^{**} with $Z_{n-1} = \min\limits_{1 \le i \le n-1} X_i$ replacing Y_{n-1}.

It is easy to check that P^* (P^{**}) stochastically minimises (maximises) $\max X_i$, and $P^*(P^{***})$ stochastically maximises (minimises) $\min\limits_{1 \le i \le n} X_i$ over $\mathfrak{P}$ by achieving equality in:

$$\max_{1 \le i \le n} F_i(x) \le P\left(\min_{1 \le i \le n} X_i \le x\right) \le \min\left(1, \sum_{i=1}^{n} F_i(x)\right)$$

$$\max_{1 \le i \le n} \bar{F}_i(x) \le P\left(\max_{1 \le i \le n} X_i > x\right) \le \min\left(1, \sum_{i=1}^{n} \bar{F}_i(x)\right).$$

Also, P^* convexly maximizes and at the same time concavely minimizes $\sum\limits_{i=1}^{n} X_i$ as is seen by the following argument: For every x and v and every $P \in \mathfrak{P}$:

$$\left(\sum_{i=1}^{n} X_i - x\right)^+ \le \left(\sum_{i=1}^{n} v_i - x\right)^+ + \sum_{i=1}^{n} (X_i - v_i)^+$$

$$\left(x - \sum_{i=1}^{n} X_i\right)^+ \le \left(x - \sum_{i=1}^{n} v_i\right)^+ + \sum_{i=1}^{n} (v_i - X_i)^+$$

On the other hand, note that $\sum_{i=1}^{n} F_i^{-1}(u)$ is left continuous non-decreasing in u, so for given x we can choose u_o such that $\sum_{i=1}^{n} F_i^{-1}(u_o) < x < \sum_{i=1}^{n} F_i^{-1}(u_o+)$, and we can then choose v_i for $i = 1,\ldots, n$ for which $F_i^{-1}(u_o) < v_i < F_i^{-1}(u_o+)$, so that $\sum_{i=1}^{n} v_i = x$. Using those v_i, for $X_i = F_i^{-1}(u)$, $i = 1,\ldots, n$ as in P*, the above inequalities hold as equalities. These properties of P*, P**, P*** ensure that the various bounds are

obtained within $\mathfrak{p}$ in the pure series and in the pure parallel case.

Theorem 1.1: For the pure series system, $\bar{M}$ and $\bar{T}$ are obtained by P^*, $\underline{L}$ and $\underline{T}$ are obtained by P***, and $\underline{L} <_{ST} L$ for all $P \in \mathfrak{p}$. For the pure parallel system, $\bar{M}$ and $\bar{T}$ are obtained by P**, $\underline{L}$ and $\underline{T}$ by P*, and $\bar{M} >_{ST} M$ for all $P \in \mathfrak{p}$.

A useful concept in the theory of networks or clutters is that of decomposition into modules (or autonomous sets), as discussed by Barlow and Proschen [1975] and by Mohring and Radermacher [1984]. Consider a set I^*, $I^* \subseteq I$, and let $I_1^*,\ldots, I_m^*$ be all the different subsets of I^* of the form $I^* \cap I_j$, $1 < j < k$. Then, I^* is a module of the system I if:

(i) $I_1^*,\ldots, I_m^*$ form a clutter.

(ii) Whenever $I_i \cap I^* = I_j^*$ it follows that for every r, $1 < r < m$ there exists an s, $1 < s < k$, such that $(I_i - I_j^*) \cup I_r^* = I_s$.

The module I^* is called nontrivial if it has more than 1 and less than n nodes. The quotient system I/I^* is formed by replacing all the nodes of I^* in I by a single new node o, with a similar replacement in each path of the clutter $\{I_j\}$. It is maybe more intuitive to think of a system, module, and quotient system in the reverse order: Start with the quotient system and the module (those can be any two systems), choose a node in the quotient system (node o can be any node) and replace this node by the base set of the module; then augment the clutter of the quotient system, by replacing each path which contains o with m new paths in which o is replaced by $I_1^*, \ldots, I_m^*$. For a nontrivial module I^*, call I a modular composition of I^*, I/I^* and call I^*, I/I^* a modular decomposition of I.

Mohring and Radermacher [1984] discuss the preservation of M, L, T under modular composition. Let V represent any of the optimal value functions M, L or T. For weights $x_1, \ldots, x_n$, let V, V^* be the optimal values for the system I and the module I^*. Then V can also be calculated in steps: Obtain V^*, assign the value V^* as the weight of node o in I/I^*, calculate the optimal value for I/I^*. For $X_1, \ldots, X_n$ random with joint distribution $P \in \mathfrak{p}$, V and V^* are random variables. The distribution of V can be calculated in steps: Obtain the joint distribution of V^*, assigned to node o, joint with the weights of the other nodes of I/I^*, and obtain the distribution of the optimal value of I/I^* for that joint distribution.

In the following sections we prove that modular composition also preserves the bounds $\bar{M}$, $\underline{L}$, $\bar{T}$, $\underline{T}$. We show for each of the optimal value functions that:

<u>Theorem 1.2</u>: If module I^* is replaced by the single node o, with weight X_o that has as its marginal distribution the distribution of the bound for I^*, then the bounds for I/I^* and for I are identical.

A general series parallel system is defined (inductively in the number of nodes n) as a system which is either pure series or pure parallel or has a nontrivial module I^* and quotient system I/I^* both of which are series parallel. Combining theorems 1.1 and 1.2 we have:

<u>Theorem 1.3</u>: For a series parallel system the bounds $\bar{M}$, $\underline{L}$, $\bar{T}$, $\underline{T}$ are obtained by joint distributions within $\mathfrak{p}$.

<u>Proof:</u> Combining theorems 1.2 and 1.1 provides a direct construction of the joint distributions for which M, L and T are extremal.

2. Convex Upper Bounds for Critical Path Length

In this section we discuss the optimal value function

$$M = \max_{1 \leq j \leq k} \sum_{i \in I_j} X_i$$

where $I = \{1, \ldots, n\}$, $I_1, \ldots, I_k$ is a clutter over I, and $X_1, \ldots, X_n$ have marginal distributions $F_1, \ldots, F_n$ and a joint distribution $P \in \mathfrak{p}$ (the dependence of M on P is suppressed to simplify notation). We start by quoting the results of Meilijson and Nadas [1979].

Let $\Psi(x)$ be defined by:

$$\Psi(x) = \inf_v \{(\max_{1 \leq j \leq k} \sum_{i \in I_j} v_i - x)^+ + \sum_{i \in I} E(X_i - v_i)^+\} \tag{6}$$

and let $x_o = \inf \{x \mid \Psi'(x) > -1\}$. It turns out that the calculation of (6) for $x \geq x_o$ is equivalent to the solution of the following mathematical program with a separable convex objective function and linear constraints:

$$\Psi(x) = \min_v \sum_{i \in I} E(X_i - v_i)^+ \tag{7}$$

$$\text{s.t.} \sum_{i \in I_j} v_i \leq x \qquad j = 1, \ldots, k.$$

Denote by $\lambda_1, \ldots, \lambda_k$ the Lagrange multipliers (dual variables) of the constraints.

Theorem 2.1:

(i) $\Psi(x) = \sup_{\mathfrak{p}} E(M - x)^+$

(ii) There exists a random variable $\bar{M}$ such that for all x, $\Psi(x) = E(\bar{M} - x)^+$, and $\bar{M} \geq_c M$ for all $P \in \mathfrak{p}$.

(iii) For every x there exists a $P \in \mathfrak{p}$ for which $E(M - x)^+ = \Psi(x)$.

(iv) A particular $P \in \mathfrak{p}$ satisfying (iii) is of the form $X_1, \ldots, X_n = F_1^{-1}(\phi_1 (U)), \ldots, F_n^{-1}(\phi_n(U))$, where $U \sim U(0, 1)$, and ϕ_i have at most k discontinuities and are linear inbetween.

(v) For every x and P as in (iv), the Lagrange multipliers of (7) satisfy:

$$\lambda_j \le P(M \ge x, \sum_{i \in I_j} X_i = M)$$

With equality if all F_i's are non atomic (absolutely continuous).

(vi) The constant $\max\limits_{1 \le j \le k} \sum\limits_{i \in I_j} E(X_i)$ is convexly smaller than M for all $P \in \mathfrak{P}$; in particular it is $\le E(M)$.

Outline of the proof: (i) For every x and every vector v, for every joint distribution $P \in \mathfrak{P}$ and every realisation $X_1, \ldots, X_n$ drawn from P:

$$(M - x)^+ \le (\max_j \sum_{i \in I_j} v_i - x)^+ + \sum_{i=1}^{n} (X_i - v_i)^+ \qquad (8)$$

which shows that the right hand side of (6) is $\ge E(M - x)^+$ for every $P \in \mathfrak{P}$. Equality to the supremum over $\mathfrak{P}$ follows from (iii).

(ii) Examination of (6) shows that $\Psi(x)$ is convex nonincreasing in x with slopes tending to -1 and 0 as x tends to $-\infty$ and ∞. Hence $\Psi(x)$ defines a random variable $\bar{M}$ according to (5), and $\bar{M} \ge_c M$ for all $P \in \mathfrak{P}$. For $x > x_o$, v which minimises (6) satisfies $\max\limits_{1 \le j \le k} \sum\limits_{i \in I_j} v_i = x$, and so it minimises (7), and (6) and (7) are equivalent. The solution of (7) at x_o, say v^o, minimises (6) for all $x \le x_o$.

(iii, iv) Consider the Lagrangean of (7):

$$\Psi(v, \lambda, x) = \sum_{i=1}^{n} E(X_i - v_i)^+ + \sum_{j=1}^{k} \lambda_j \left(\sum_{i \in I_j} v_i - x \right) \qquad (9)$$

with $\lambda_j \ge 0$. The Kuhn Tucker saddle point conditions for it are:

$$\begin{aligned} &\sum_{i \in I_j} v_i \le x \text{ and } \sum_{i \in I_j} v_i < x \text{ implies } \lambda_j = 0 \\ &P(X_i > v_i) \le \sum_{j \mid i \in I_j} \lambda_j \le P(X_i \ge v_i) \\ &P(\bar{M} > x) \le \sum_{j=1}^{k} \lambda_j \le P(\bar{M} \ge x). \end{aligned} \qquad (10)$$

For a given $x > x_o$, let $v_1, \ldots, v_n, \lambda_1, \ldots, \lambda_k$ be an optimal solution and a set of multipliers of (7,9). Let $\lambda_{k+1} = 1 - \sum_{j=1}^{k} \lambda_j$, $I_{k+1} = \phi$, $\alpha_i = \sum_{j|i\epsilon I_j} \lambda_j$, $i = 1, \ldots, n$. The joint distribution $P \in \mathfrak{p}$ stated in (iv) is defined by the functions ϕ_i, $i = 1, \ldots, n$ which for $m = 1, \ldots, k+1$ and $\sum_{j=1}^{m-1} \lambda_j < u \leq \sum_{j=1}^{m} \lambda_j$ have the value:

$$\phi_i(u) = \begin{cases} (1-\alpha_i) + \alpha_i \left(u - \sum_{j=1}^{m-1} \lambda_j\right)/\lambda_m & i \in I_m \\ (1-\alpha_i)\left(u - \sum_{j=1}^{m-1} \lambda_j\right)/\lambda_m & i \notin I_m \end{cases} \tag{11}$$

(v) For $\lambda_m \neq 0$, with probability λ_m, $\sum_{i=1}^{m-1} \lambda_j < U < \sum_{i=1}^{m} \lambda_j$ in which case $1 - \alpha_i < \phi_i(U) < 1$, and by (10) $X_i = F_i^{-1}(\phi_i(U)) \geq v_i$, for all $i \in I_m$, while at the same time, $X_i \leq v_i$ for all $i \notin I_m$. By (8), we see that in this case $M - x = \sum_{i\epsilon I_m} (X_i - v_i)$ and $\sum_{i\epsilon I_m} v_i = x$, so $M = \sum_{i\epsilon I_m} X_i \geq x$. The required inequality follows, and equality for nonatomic distributions follows similarly. Finally, (vi) holds by Jensen's inequality.

<u>Corollary 2.2.</u> <u>Modular decomposition:</u> Theorem 1.2 holds for the function M.

<u>Proof:</u> Let $I^* \subseteq I$ with clutter $I_1^*, \ldots, I_m^*$ be a module of I, and let I/I^* be the quotient system, with set of nodes $I^- = (I - I^*) \cup \{o\}$ and clutter of paths $I_1^-, \ldots, I_\ell^-$. Let Ψ, Ψ_o, Ψ_1 and $\bar{M}$, $\bar{M}_o$, $\bar{M}_1$ denote the bounds for the systems I, I^*, I/I^* respectively. We look at the program (7) and the two additional programs:

$$\Psi_o(y) = \min_u \sum_{i\epsilon I^*} E(X_i - u_i)^+$$

$$\text{s.t.} \quad \sum_{i\epsilon I_j^*} u_i \leq y \quad j = 1, \ldots, m \tag{12}$$

and:

$$\Psi_1(x) = \min_w \sum_{i \varepsilon I-I^*} E(X_i - w_i)^+ + \Psi_o(w_o)$$

$$\text{s.t.} \sum_{i \varepsilon \bar{I}_j} w_i \leq x \qquad j = 1,\ldots, \ell \tag{13}$$

Since $\Psi_o(w_o) = E(\bar{M}_o - w_o)^+ = E(X_o - w_o)^+$, $\Psi_1(x)$ is the bound for the module I/I^*. We need to show that $\Psi(x) = \Psi_1(x)$ for all x.

(i) $\Psi(x) \geq \Psi_1(x)$: Let v be an optimal solution of (7).

Define:

$$w_o = \max_{1 \leq j \leq m} \sum_{i \varepsilon I_j^*} v_i$$

$$w_i = v_i \quad i \varepsilon I - I^*.$$

Because I^* is a module, and v is feasible for (7), w is feasible for (13).

The value of the objective function (13) for w is $\sum_{i \varepsilon I-I^*} E(X_i - v_i)^+ +$ $\Psi_o(w_o)$. But $\{v_i\}$ $i \varepsilon I^*$ is feasible for (12) with $y = w_o$, and so $\Psi_o(w_o)$ $\leq \sum_{i \varepsilon I^*} E(X_i - v_i)^+$, so the value of the objective of (13) for w is $\leq \Psi(x)$, and therefore $\Psi_1(x) \leq \Psi(x)$.

(ii) $\Psi_1(x) \geq \Psi(x)$: Let w be an optimal solution of (13). Let u be an optimal solution of (12), with $y = w_o$. Let $v_i = u_i$, $i \varepsilon I^*$, and $v_i = w_i$, $i \varepsilon I - I^*$. Because I^* is a module, v is feasible for (7). The objective value of (7) for v is

$$\sum_{i \varepsilon I} E(X_i - v_i)^+ = \sum_{i \varepsilon I-I^*} E(X_i - w_i)^+ + \sum_{i \varepsilon I^*} E(X_i - u_i)^+$$

$$= \sum_{i \varepsilon I-I^*} (X_i - w_i)^+ + \Psi_o(w_o) = \Psi_1(x),$$

thus $\Psi(x) \leq \Psi_1(x)$.

<u>Monotonicity:</u>

<u>Corollary 2.3:</u> If X_i are replaced by Z_i so that $Z_i \geq_c X_i$, $i = 1,\ldots, n$

then the bounds $\bar{M}_1$, $\bar{M}$ obtained for $Z_1, \ldots, Z_n$ and $X_1, \ldots, X_n$ satisfy $\bar{M}_1 >_c \bar{M}$.

Proof: Let $\Psi_1(x)$, $v_1^{(1)}, \ldots, v_n^{(1)}$ be the solution of (7) with Z_i replacing X_i. By $Z_i >_c X_i$, $\hat{\Psi}(x) = \sum_{i=1}^{n} E(X_i - v_i^{(1)})^+ \leq \Psi_1(x)$. Minimising (7) with X_i, we get $\Psi(x) \leq \hat{\Psi}(x) \leq \Psi_1(x)$, so for all x, $E(\bar{M} - x)^+ \leq E(\bar{M}_1 - x)^+$.

Computational Aspects: Nadas [1979] discusses the computational aspects of solving the mathematical program (7), which with its linear constraints and separable convex objective function is relaively easy. If $E(X_i - v_i)^+$ is approximated from above by $\xi_i(v_i)$ piecewise linear and convex, the program can be solved as a linear program, and provide an upper bound for $\Psi(x)$. The approximation is equivalent to replacing each F_i by an approximating discrete distribution, and it can be chosen so that $0 \leq \xi_i(v) - E(X_i - v)^+ \leq \delta$ for any given $\delta > o$, uniformly for all v.

In the project planning application, the nodes represent activities and the clutter $I_1, \ldots, I_k$ is defined by the partial ordering of activities, and consists of all the paths from the start to the finish of the job. In that case the program (7) has the following deterministic interpretation: Find activity durations $v_1, \ldots, v_n$ so as to complete the whole project by time x at minimal cost, where doing activity i in duration v_i costs $E(X_i - v_i)^+$. This is the project cost curve problem, solved by Fulkerson [1961]. The solution is effected, parametrically for all x, by formulating the dual problem which is a minimal cost flow problem, and solving it parametrically for all flow values; this can be done by the very efficient out of kilter method, cf. Lawler [1976]. The minimal cost flow problem that arises from the dual to (7) is: For any total flow value A, find flows α_i through the nodes i, i = 1,..., n which yield total flow A, at minimal cost, that is:

$$\min_{\alpha} \sum_{i=1}^{n} h_i\,(\alpha_i)$$
$$\text{s.t.} \quad \min_{1 \le j \le \ell} \sum_{i \in J_j} \alpha_i = A \qquad i = 1, \ldots, n \tag{14}$$

Where $J_1, \ldots, J_\ell$ is the blocking clutter of cuts, and where:

$$h_i(\alpha) = \int_0^\infty \max\,(\bar{F}_i(t) - \alpha,\ 0)\ dt = \int_0^{1-\alpha} F_i^{-1}\,(u)\ du. \tag{15}$$

The total flow value A, and the flows through the nodes α_i, which are obtained from the solution of (14), are related to the λ's in (9), (10) through $A = \sum_{j=1}^{k} \lambda_j$, $\alpha_i = \sum_{j|i \epsilon I_j} \lambda_j$. The corresponding values of x and the v_i's in (7) can be obtained from (10).

Redesign of a PERT network: It is quite usual when designing a project with a PERT network to have a target date x for the completion of the project, and a nondecreasing convex penalty function C(y) for values M = y > x. For such a penalty function,

$$\bar{E}(C) = C'(x)\ \Psi(x) + \int_x^\infty C''(y)\ \Psi(y)dy \tag{16}$$

where C', C" are the 1st and 2nd derivatives of C, is an upper bound on the expected penalty.

For the target date x, the expected tardiness $E(M - x)^+$ is bounded sharply by $\Psi(x)$, and the solution of (7) provides a construction for the worst case distribution with respect to that tardiness. It also provides a host of additional information on that worst case distribution which can be used to redesign the project. Let $v = v(x) = v_1(x), \ldots, v_n(x)$ be the values of the solution of (7), $\lambda_1(x), \ldots, \lambda_k\,(x)$ the Lagrange multipliers, and $\alpha_i(x) = \sum_{j|i \epsilon I_j} \lambda_j\,(x)$, $i = 1, \ldots, n$.

The values $v_i(x)$ provide target durations for the activities with respect to the general target date x. If we let $\Psi_i(x) = E(X_i - v_i(x))^+$, then $\Psi_i(x)$ is the expected contribution of node i (activity i) to the total tardiness. Similarly, for a module I^* we get by solving (13) for I/I^* and due date x, a value $w_o(x)$ which is the target duration of the module I^* with respect to the general target date x, and we can get $\Psi_{I^*}(x) = E(\bar{M}_o - w_o(x))^+$ (obtained by solving (12), with $y = w_o(x)$), as the expected contribution of module I^* to the total tardiness. If $\Psi_i(x)$ or $\Psi_{I^*}(x)$ is inserted in (16) instead of Ψ, we obtain $\bar{E}_i(C)$ and $\bar{E}_{I^*}(C)$ which are the worst case bounds on the expected contribution of i or I^* to the penalty. Thus the $v_i(x)$ and $w_o(x)$ provide a way of assigning tardiness and penalties to each activity or module (on the basis of a worst case analysis).

The values $\lambda_j(x)$ provide, for the worst case distribution, the probability that tardiness beyond x occurs, and that the longest path is I_j (at least if all X_i's are continuous random variables), as stated in theorem 2.1. It is also easy to see from the proof of theorem 2.1 that $\alpha_i(x)$ is the probability that tardiness beyond x occurs and that node i is on the longest path.

Similar quantities can be calculated for a module I^*. Solution of (12) with $y = w_o(x)$ provides λ's and α's within I^*. Solution of (13) for I/I^*, provides by the value $\alpha_o(x)$ the probability that tardiness beyond x occurs and the longest path passes through I^*.

3. Concave Lower Bounds for Maximal Flow and Shortest Route

In this section we discuss the optimal value function

$$L = \min_{1 \leq j \leq \ell} \sum_{i \in J_j} X_i$$

where $I = \{1, \ldots, n\}$, $J_1, \ldots, J_\ell$ is a clutter over I, and $X_1, \ldots, X_n$ have marginal distributions $F_1, \ldots, F_n$ and a joint distribution $P \in \mathfrak{p}$. When $J_1, \ldots, J_\ell$ are the clutter of paths in a network, L is the shortest route; when $J_1, \ldots, J_\ell$ are the clutter of cuts in a network, L is the maximal flow. The results about L exactly mirror the results about M in section 2. This is due to the duality between the various pairs of concepts occuring here: path-cuts, series-parallel, min-max, convex-concave, $P(X < x) - P(X < x)$, and $E(X - x)^+ - E(X - x)^-$.

The function $\eta(x)$ in (4) is given by:

$$\eta(x) = \inf_v \{(x - \min_{1 \leq j \leq \ell} \sum_{i \in J_j} v_i)^+ + \sum_{i=1}^{n} E(v_i - X_i)^+\} \tag{17}$$

and for $x \leq x_o = \sup\{x | \eta'(x) < 1\}$, (17) is equivalent to

$$\eta(x) = \min_v \sum_{i \in I} E(v_i - X_i)^+ \tag{18}$$

$$\text{s.t.} \sum_{i \in J_j} v_i \geq x \qquad j = 1, \ldots, \ell$$

with Lagrange multipliers $\lambda_1, \ldots, \lambda_\ell$.

Theorem 3.1:

(i) $\eta(x) = \sup_{\mathfrak{p}} E(x - L)^+$

(ii) There exists a random variable $\underline{L}$ such that for all x, $\eta(x) = E(x - \underline{L})^+$, and $\underline{L} \leq_k L$ for all $P \in \mathfrak{p}$.

(iii) For every x there exists a $P \in \mathfrak{p}$ for which $E(x - L)^+ = \eta(x)$.

(iv) A particular $P \in \mathfrak{p}$ satisfying (iii) is of the form: $X_1, \ldots, X_n = F_1^{-1}(\phi_1(U)), \ldots, F_n^{-1}(\phi_n(U))$, where $U \sim U(0, 1)$, and ϕ_i have at most ℓ discontinuities and are linear in between.

(v) For every $x \leq x_o$ and P as in (iv), the Lagrange multipliers of (18) satisfy:

$$\lambda_j \leq P(L \leq x, \sum_{i \in J_j} X_i = L)$$

with equality if all F_i's are non atomic (absolutely continuous).

(vi) The constant $\min_{1\leq j\leq \ell} \sum_{i\in J_j} E(X_i)$ is concavely larger than L for all

$P \in \mathfrak{p}$; in particular it is $\geq E(L)$.

Proof: This is a corollary of theorem 2.1, if the problem is reformulated in terms of $-X_i$, with $-L = \max_{1\leq j\leq \ell} \sum_{i\in I_j}(-X_i)$.

The modular decomposition theorem 1.2 and monotonicity (with respect to $\geq_k$) hold for L, in analogy with M.

Computational Aspects: The program (18) has a separable convex nondecreasing objective function and linear constraints, and can be approximated by a linear program, like (7).

For the shortest route application, the solution of (18) can be obtained by using $Y_i = - X_i$, and solving (7).

For the maximal flow application, when $J_1, \ldots, J_\ell$ are cuts, problem (18) has the following deterministic interpretation: Find flows (or capacities) v_i for nodes $i = 1, \ldots, n$, so as to obtain a flow (maximal flow) of x, at minimal cost, where the cost of flow v_i in node i is given by $E(v_i - X_i)^+$, which is convex nondecreasing in v_i. This deterministic problem is very similar to the dual problem for the project planning application, given by (14). It can be solved parametrically for all flow values x, using the out of kilter method, cf Lawler [1976].

In applications to shortest route problems one may have a design value x and a convex decreasing reward function C(y) for values of L = y. In applications to maximal flow problems one may have a target flow x and a convex decreasing penalty function C(y) for value of $L = y \leq x$. $\eta(x)$ and $\underline{L}$ provide upper bounds for the expected shortfall below x, $E(x - L)^+$, and of $E(C(L))$. The solution of (18) provides similar information for redesign as in the critical path applications.

4. Stochastic Upper and Lower Bounds for Reliability System Lifetime

In this section we discuss the optimal value functions

$$T' = \max_{1<j<k} \ \min_{i \in I_j} X_i \tag{19}$$

and

$$T'' = \min_{1<j<\ell} \ \max_{i \in J_j} X_i \tag{20}$$

where $I = \{1,\ldots,n\}$, $I_1,\ldots, I_k$ and $J_1,\ldots, J_\ell$ are two clutters over I, and $X_1,\ldots, X_n$ have marginal distributions $F_1,\ldots, F_n$ with joint distribution function $P \in \mathfrak{P}$. If $I_1,\ldots, I_k$ are the paths and $J_1,\ldots, J_\ell$ are the cuts of a reliability system (defined through a network or through a general Boolean coherent structure function as in Barlow and Proschan [1975]), and also for any other pair of blocking clutters as shown by Edmonds and Fulkerson [1970], $T' = T''$. For a reliability system, if nodes $1,\ldots, n$ represent components, and $X_1,\ldots, X_n$ are the component lifetimes then $T = T' = T''$ is the system lifetime.

We will show that the supremum functions $\alpha(x)$ and $\beta(x)$ of (4) are given by solution of the following linear programming problems:

$$\begin{aligned} \alpha(x) = \max_{\lambda} \ & \sum_{j=1}^{k} \lambda_j \\ \text{s.t.} \ & \sum_{j \mid i \in I_j} \lambda_j < \bar{F}_i(x) \qquad i = 1,\ldots,n \\ & \sum_{j=1}^{k} \lambda_j < 1 \\ & \lambda_j > 0 \end{aligned} \tag{21}$$

and

$$\begin{aligned} \beta(x) = \max_{\mu} \ & \sum_{j=1}^{\ell} \mu_j \\ \text{s.t.} \ & \sum_{j \mid i \in J_j} \mu_j < F_i(x) \qquad i = 1,\ldots, n \\ & \sum_{j=1}^{\ell} \mu_j < 1 \\ & \mu_j > 0 \end{aligned} \tag{22}$$

where $\bar{F}_i(x) = 1 - F_i(x) = P(X_i > x)$. The analogy with the programs (7) and (18) is seen in the dual programs to (21), (22):

$$\alpha(x) = \min_v \sum_{i=1}^{n} \bar{F}_i(x)\, v_i + w$$

$$\text{s.t.} \quad \sum_{i \varepsilon I_j} v_i + w \geq 1 \qquad j = 1,\ldots, k \tag{23}$$

$$w, v_i \geq 0$$

$$\beta(x) = \min_u \sum_{i=1}^{n} F_i(x) u_i + w$$

$$\text{s.t.} \quad \sum_{i \varepsilon J_j} u_i + w \geq 1 \qquad j = 1,\ldots, \ell \tag{24}$$

$$w, u_i \geq 0$$

The following theorem is implied in parts by Zemel [1982].

<u>Theorem 4.1</u>:

(i) $\alpha(x) = \sup_{\mathfrak{p}} P(T' > x)$, $\beta(x) = \sup_{\mathfrak{p}} P(T'' \leq x)$.

(ii) There exist random variables $\bar{T}$ and $\underline{T}$ such that for all x, $\alpha(x) = P(\bar{T} > x)$, $\beta(x) = P(\underline{T} \leq x)$, i.e. $\bar{T} \geq_{ST} T'$ and $\underline{T} \leq_{ST} T''$ for all $P \varepsilon \mathfrak{p}$.

(iii) For every x there exist P', P'' $\varepsilon\ \mathfrak{p}$ for which $P(T' > x) = \alpha(x)$, $P(T'' \leq x) = \beta(x)$.

(iv) In particular P', P" $\varepsilon\ \mathfrak{p}$ satisfying (iii) exist which are of the form $X_1,\ldots, X_n = F_1^{-1}(\phi_1(U)),\ldots, F_n^{-1}(\phi_n(U))$, where $U \sim U(0, 1)$ and ϕ_i have at most max(n, k) (max(n, ℓ)) discontinuities and are linear inbetween.

(v) For every x and P', P'' as in (iv), the solutions (21), (22) satisfy

$$\lambda_j \leq P(T' > x, \min_{i \varepsilon I_j} x_i = T')$$

$$\mu_j \leq P(T'' \leq x, \max_{i \varepsilon J_j} x_i = T'')$$

with equality if all F_i's are non atomic (absolutely continuous).

Proof: (ii) From the form of the objective functions of (21), (22) $\alpha(x)$ is nonincreasing and $\beta(x)$ is nondecreasing. From (23), (24) $0 \le \alpha(x) \le 1$ and $0 \le \beta(x) \le 1$. From (23) (24) one obtains $\beta(-\infty) = \alpha(\infty) = 0$, and from (21), (22) $\alpha(-\infty) = \beta(\infty) = 1$. $\alpha(x)$ and $\beta(x)$ depend continuously on $F_i(x)$, which are continuous from the right, hence $\alpha(x)$ and $\beta(x)$ are continuous from the right. Thus $1 - \alpha(x)$ and $\beta(x)$ are distribution functions, defining $\bar{T}$ and $\underline{T}$.

(iii)(iv), We shall describe the construction of members of P as stated in (iv). For given x, let λ (μ) be the optimal basic solution of (24) ((22)).

Let $\lambda_{k+1} = 1 - \sum_{j=1}^{k} \lambda_j$, $\mu_{\ell+1} = 1 - \sum_{j=1}^{\ell} \mu_j$, $I_{k+1} = J_{\ell+1} = \phi$,

and let:

$$\alpha_i = \sum_{j|i\varepsilon I_j} \lambda_j \le \bar{F}_i\ (x)$$

$$\beta_i = \sum_{j|i\varepsilon J_j} \mu_j \le F_i\ (x),$$

and define for each $i = 1,\dots, n$ the following functions for $0 < u \le 1$:

If $\sum_{j=1}^{m-1} \lambda_j < u \le \sum_{j=1}^{m} \lambda_j$, and $\lambda_m \neq 0$, $(1 \le m \le k + 1)$:

$$\phi'(u) = \begin{cases} (1 - \alpha_i) + \alpha_i\ (u - \sum_{j=1}^{m-1} \lambda_j)/\lambda_m & i \in I_m \\ (1 - \alpha_i)\ \ (u - \sum_{j=1}^{m-1} \lambda_j)/\lambda_m & i \notin I_m \end{cases}$$

and if $\sum_{j=1}^{m-1} \mu_j < u \le \sum_{j=1}^{m} \mu_j$, and $\mu_m \neq 0$, $(1 \le m \le \ell + 1)$:

$$\phi_i''(u) = \begin{cases} \beta_i\ (u - \sum_{j=1}^{m-1} \mu_j)/\mu_m & i \in J_m \\ \beta_i + (1 - \beta_i)\ (u - \sum_{j=1}^{m-1} \mu_j)/\mu_m & i \notin J_m \end{cases}$$

It is clear from the construction that $F_i^{-1}(\phi_i'(U))$ and $F_i^{-1}(\phi_i''(U))$ are both distributed as X_i if $U \sim U(0,1)$ and we have $P', P'' \in \mathfrak{p}$. With probability λ_m (μ_m), $\sum_{j=1}^{m-1} \lambda_j < U \le \sum_{j=1}^{m} \lambda_j$ $(\sum_{j=1}^{m-1} \mu_j < U \le \sum_{j=1}^{m} \mu_j)$, and in that case, $\phi_i'(U) > 1 - \alpha_i \ge F_i(x)$ $(\phi_i''(U) \le \beta_i \le F_i(x))$ so $X_i > x$ $(X_i \le x)$ for all $i \in I_m$ $(i \in J_m)$. But then $\min_{i \in I_m} X_i > x$ $(\max_{i \in I_m} X_i \le x)$ and so $T' > x$ $(T'' \le x)$. Thus $\alpha(x) \le P(T' > x)$ and $\beta(x) \le P(T'' \le x)$ for P' and P'' respectively, with equality actually occuring because of (i).

(i) We prove (i) for $\alpha(x)$, the proof for $\beta(x)$ is analogous. Define for each $j = 1,\ldots, k$, $A_j = \bigcap_{i \in I_j} \{X_i > x\}$, A_j^c the complement event of A_j, and define for each $\phi \neq K \subseteq \{1,\ldots, k\}$ $B_K = \bigcap_{j \in K} A_j \bigcap_{j \notin K} A_j^c$. Let $\lambda_K = P(B_K)$ for an arbitrary $P \in \mathfrak{p}$. Then $P(T > x) = P(\bigcup_j A_j) = P(\bigcup_K B_K) = \sum_K \lambda_K$. Let i be some node in I. Then if $i \in I_j$ and $j \in K$, $B_K \subseteq \{X_i > x\}$, hence

$$\Sigma\lambda_K \le \bar{F}_i(x) \qquad i = 1,\ldots, n \tag{25}$$

where the summation is over all K's such that $i \in I_j$, $j \in K$ for some j. If we set up the program

$$\max \sum_K \lambda_K \tag{26}$$

$$\text{s.t. (25) and } \sum_K \lambda_K \le 1,\ \lambda_K \le 0 \tag{26}$$

then the solution is $\ge P(T' > x)$ for all $P \in \mathfrak{p}$. The linear program (26) has a variable λ_K for each subset $K \subseteq \{1,\ldots, k\}$. In particular this includes $\lambda_1,\ldots, \lambda_k$ which correspond to the singletons.

Consider any K which is not a singleton, say $K \supseteq \{i, j\}$. Then the variable λ_K appears in every constraint in which λ_i or λ_j appear. It is then clear (by examining the dual problem) that λ_K need not be basic in an optimal solution, and so it is a redundant variable. Thus (21) has the same solution as (26) and so $\alpha(x) \geq P(T' > x)$. This completes the proof.

Modular decomposition and Monotonicity: hold for $\alpha(x)$, $\beta(x)$ as for Ψ, η.

Computational aspects: The main difficulty in solving the LP's (21), (22), is the number of variables, one for each path, which may grow exponentially with the number of nodes. Consider solution by the simplex method. Then one can avoid the necessity of handling such a large number of variables, by using column generation, as follows: Let λ be a basic feasible solution to (21) and let v be its simplex multipliers; assume $\Sigma \lambda_j < 1$ and so $w = 0$. This solution is optimal if v is dual feasible, that is

$$\sum_{I_j} v_i \geq 1 \qquad j = 1, \ldots, k$$

to check that, one can solve the shortest route problem with the weights v_i. If the solution is < 1, it gives a new path I_j for which λ_j should enter the basis.

The solution for all values of x can again be done using parametric LP.

Zemel [1982] has shown that if the shortest route problem, $\min\limits_{1 \leq j \leq k} \sum\limits_{I_j} v_i$, can be solved in polynomial time then so can the LP (21), by employing an ellipsoid type algorithm. In particular, for a two terminal network the shortest route problem is indeed polynomial. This is in sharp contrast to the fact that the calculation of the reliability of a two terminal network with independent components is NP hard, as shown by Rosenthal [1975], Ball [1980] and Valiant [1979]. This great difference in difficulty of calculation between the independent case and the bounds of the present paper is the most remarkable feature of these bounds.

Redesign of a reliability system: The solution of (21, 22, 23, 24) provides similar information for redesigning the reliability system as we had for M and L. Typically one may have a target system lifetime x and be interested in $P(T > x)$. Else one may have a monotone decreasing penalty function $C_1(y)$ for $T = y < x$, or a monotone increasing reward function $C_2(y)$ for $T = y > x$, with (C'_1, C'_2 are the derivatives of C_1, C_2):

$$\begin{aligned} \bar{E}(C_1) &= C_1(x)\beta(x) + \int_{-\infty}^{x} -C_1'(y)\beta(y)\,dy \\ \bar{E}(C_2) &= C_2(x)\alpha(x) + \int_{x}^{\infty} C_2'(y)\alpha(y)\,dy \end{aligned} \tag{27}$$

providing upper bounds for penalty or reward.

If $\beta(x)$, $\alpha(x)$ are replaced by $u_i(x)\ F_i(x)$ and by $v_i(x)\ F_i(x)$ in (27) where $u(x)$, $v(x)$ are the solutions to (24), (23), one gets an assignment of penalties or rewards to the various nodes for the extreme cases. Similar assignments are obtained for modules. The probabilistic interpretations of λ_j, μ_j, α_i, β_i can also be used in redesign, as in section 2.

5. Discussion

In this paper we have developed various bounds of various types for the optimal value functions M, L, T. It is fortunate that each of those bounds is exactly of the type most useful in application. The bounds for T are stochastic, so that given a target date x we have upper and lower bounds on the system reliability $P(T > x)$ at the date x. The bound for M is an upper bound in the convex majorisation sense, which in the critical path length application gives an upper bound to the expected tardiness beyond a target day x, $E(M - x)^+$. The bound on L is a lower bound in the concave majorisation sense, which in the maximal flow application gives an upper bound on the expected shortfall of the flow below a target level x, $E(x - L)^+$.

For aesthetic reasons, or for some unforeseen applications one would nevertheless desire the bounds not obtained here, i.e. stochastic bounds on M, L, convex lower and concave upper bounds on M and L respectively, and convex and concave bounds for T. In the following discussion we explain why these are unlikely to possess the same nice properties as the bounds already obtained.

In section 1 we introduced three special joint distributions, the perfect tracking distribution, P^*, and the max and min antithetic distributions, P^{**}, P^{***}, which achieve the various bounds for the pure series and the pure parallel system. We state the following simple lemma:

Lemma 5.1: Let $\mathfrak{M}$ be a family of distributions. If $M^* \in \mathfrak{M}$ satisfies $M^* >_c M$ for all $M \in \mathfrak{M}$, then there is no member $\bar{M} \in \mathfrak{M}$ for which $\bar{M} >_{ST} M$ for all $M \in \mathfrak{M}$ unless $\bar{M} = M^*$.

Proof: Assume existence of $\bar{M}$. Then $\bar{M} >_{ST} M^* >_c \bar{M}$, so $\bar{M} >_c M^* >_c \bar{M}$, so

$E_{\bar{M}}(X - x)^+ = E_{M^*}(X - x)^+$ for all x, and $\bar{M} = M^*$.

Since P^* does not in general stochastically majorize or stochastically minorize ΣX_i (the only exclusions one can think of are when n-1 of the variables are deterministic and $\mathfrak{p}$ has only one member), stochastic bounds on M (L) in the purely series (purely parallel) case will not be obtained within $\mathfrak{p}$.

We do not know how to construct convex lower bounds or concave upper bounds for $\sum_{i=1}^{n} X_i$. Such bounds if they exist within   will minimize the variance of $\sum_{i=1}^{n} X_i$ and will provide the optimum distributions for variance reduction in Monte Carlo simulation, as discussed by Hammerseley and Mauldon [1950], Handscomb [1958] and Whitt [1976].

Next we note that M is nondecreasing convex, L is nondecreasing concave and T', T'' are nondecreasing but neither convex nor concave. The generalization of the bounds from the pure series or pure parallel case to general series parallel systems and the property of modular decomposition follow from the fact that $>_c$, $>_k$, $>_{ST}$ are preserved respectively by nondecreasing convex, nondecreasing concave and nondecreasing functions. We cannot however expect the same to hold for concave bounds on M, convex bounds on L or convex and concave bounds on T, T'. The stochastic bounds on T if obtained within $\mathfrak{P}$ are of course also sharp convex and concave bounds. If they are obtained outside $\mathfrak{P}$ we do not know how to construct sharp convex or concave bounds.

The bounds discussed in the present paper provide one way of circumventing the impossible problem of calculating the exact distributions of M, L or T. We conclude by mentioning other approaches that appear in the literature. These are based on special models for which the distributions of M, L and T are tractable, namely:

- Markovian systems: If the weights have independent exponential distributions then M, L, T have phase type distributions in the sense of Neuts [1981], and their distributions can be evaluated, though at considerable computational effort, e.g. Kulkarni and Adlakha [1984].

- Series Parallel systems with independent weights: The distributions of M, L, T are obtained by convolutions products and complementations.

- "Perfect tracking" (Nadas [1979]): Each weight X_i is of the form $X_i = a_i + b_i Z$, $i = 1,\ldots, n$, where Z is a single random variable, common to all the X_i's. If $p = P(Z \leq z)$, then the p percentiles of M, L, T are obtained from the values $x_i = a_i + b_i z$.

By approximating $F_1, \ldots, F_n$ with phase type distributions one can presumably get bounds on M, L or T using the first approach.

Series parallel systems form the basis for calculations of bounds on values of $P(V > y)$ where V is M, L or T - see Robillard & Trahan [1977], Shogan [1977], Devroye [1979] for PERT and shortest route applications, and Barlow and Proschan [1975], Esary & Proschan [1970], Shogan [1978], Natvig [1980] for reliability applications.

The approach of the present paper utilises the third approach of perfect tracking.

Acknowledgements: I am most grateful to I. Meilijson, R.H. Mohring, F.J. Radermacher and R. Serfozo for useful discussions and suggestions. This research was partly carried out during a visit to RWTH Aachen, supported by the DAAD. Partial support for this research was provided by Grant AFOSR 84-0367.

References

1. Ball, M.O. 1980. The Complexity of Network Reliability Computations. Networks 10, 153-165.

2. Barlow, R.E. and F. Proschan. 1975. Statistical Theory of Reliability and Life Testing: Probability Models. Holt Rhinehart Winston, New York.

3. Devroye, L.P. 1979. Inequalities for Completion Times of Stochastic Pert Networks. Math. Oper. Res. 4, 441-447.

4. Edmonds, J. and D.R. Fulkerson. 1970. Bottleneck Extrema. J. Combinatorial Th. 8, 299-306.

5. Elmaghraby, S.E. 1977. Activity Networks, Project Planning and Control by Network Models. Wiley, New York.

6. Esary, J.D. and F. Proschan. 1970. A Reliability Bound for Systems of Maintained, Interdependent Components. J. Amer. Statist. Assoc. 65, 329-338.

7. Ford, L.R. and D.R. Fulkerson. 1962. Flows in Networks. Princeton Univ. Press, Princeton.

8. Fulkerson, D.R. 1961. A Network Flow Computation for Project Cost Curves. Man. Sci. 7, 167-178.

9. Hammersley, J.M. and J.G. Mauldon. 1956. General Principles of Antithetic Variables. Proc. Camb. Philos. Soc. 52, 476-481.

10. Handscomb, D.C. 1958. Proof of the Antithetic Variate Theorem for $n > 2$. Proc. Camb. Philos. Soc. 54, 300-301.

11. Klein Haneveld, W.K. 1982. Distributions with Known Marginals and Duality of Mathematical Programming with Application to PERT. Technical Report 90 (OR-8203), Rijksuniversiteit, Groningen.

12. Kulkarni, V.A. and V.G. Adlakha. 1984. Markov and Markov Regenerative Pert Networks. Technical Report U. North Carolina, Chapel Hill, UNC/ORSA/TR-84-4.

13. Lawler, L.E. 1976. Combinatorial Optimization: Networks and Matroids. Holt Rhinehart Winston, New York.

14. Meilijson, I. and A. Nádas. 1979. Convex Majorization with an Application to the Length of Critical Paths. J. Appl. Prob. 16, 671-677.

15. Meilijson, I. 1984. Sharp Bounds on the Largest of Some Linear Combinations of Random Variables with Given Marginal Distributions. Technical Report.

16. Möhring, R.H. and F.J. Radermacher. 1984. Substitution Decomposition for Discrete Structures and Connections with Combinatorial Optimization. Ann. Discrete Math. 19, 257-356.

17. Nádas, A. 1979. Probabilistic PERT. IBM J. Res. Develop. 23, 339-347.

18. Natvig, B. 1980. Improved Bounds for the Availability and Unavailability in a Fixed Time Interval for Systems of Maintained, Interdependent Components. Adv. Appl. Prob. 12, 200-221.

19. Neuts, M.F. 1981. Matrix Geometric Solutions in Stochastic Models. Johns Hopkins, Baltimore.

20. Robillard, P. and M. Trahan. 1977. The Completion Time of PERT Networks. Opns. Res. 25, 15-29.

21. Rosenthal, A. 1975. A Computer Scientist Look at Reliability Computations. In Reliability and Fault Tree Analysis R.E. Barlow, J.B. Fussel, D.D. Singpurwalla, Eds. SIAM Philadelphia, 133-152.

22. Shogan, A.W. 1977. Bounding Distributions for a Stochastic PERT Network. Networks 7, 359-381.

23. Shogan, A.W. 1978. A Decomposition Algorithm for Network Reliability Analysis. Networks 8, 231-251.

24. Stoyan, D. 1983. Comparison Methods for Queues and Other Stochastic Models. Wiley, New York.

25. Valiant, L.G. 1979. The Complexity of Enumeration and Reliability Problems. SIAM J. Comput. 8, 410-421.

26. Weiss, G. 1984. Stochastic Bounds on Distributions of Optimal Value Functions with Applications to Pert, Network Flows and Reliability. Ann. Oper. Res. 1, 59-65.

27. Whitt, W. 1976. Bivariate Distributions with Given Marginal. Ann. Statist. 4, 1280-1289.

28. Zemel, E. 1982. Polynomial Algorithms for Network Reliability. Networks 12.

OPTIMAL IDLE AND INSPECTION PERIODS FOR M/G/1 QUEUES

by

Sung Schick Kim
Korea University

and

Richard F. Serfozo
Georgia Institute of Technology

Abstract

We consider an M/G/1 queue that operates under a (T,N)-policy: whenever the system becomes empty, the server is idle for a time T and then it inspects the queue continuously without serving customers until there are N customers waiting - thereupon the server is activated for service and serves customers continuously until the system becomes empty. This idle-inspection-service cycle is repeated indefinitely. There are costs for inspecting the queue, activating and running the server, and holding customers in the system. We present a computational procedure for determining the design parameters (T,N) that minimize the average cost.

This research was supported in part by the grant AFOSR 84-0367.

Introduction

Intermittent rather than continuous service is characteristic of service systems in which servers must be absent periodically for other duties or for rejuvenation. Intermittent service is also used in systems where short queues are tolerable, or where short busy periods for servers are uneconomical. In designing such systems, a natural question is: How long should the server be absent without observing the queue and at what queue length should the server start serving customers?

In this paper, we address this question for an M/G/1 queue that operates under a (T,N)-policy as follows. Customers arrive by a Poisson process $\{A(t);\ t \geq 0\}$ with rate λ, and the service times have a mean $\mu > \lambda$ and a finite variance. For simplicity, we assume the system begins at time zero with the server deactivated and no customers in the queue. The server remains idle in the time interval $[0,T]$ and, at time T, the queue is inspected which reveals $A(T)$ customers waiting. If $A(T)$ is less than a number N, then the queue is inspected continuously until the time S_N of the N-th arrival. Thereupon the server is activated for service and serves customers continuously until the system becomes empty, at which time the server is deactivated. On the other hand, if at time T it is found that $A(T) \geq N$, then the server is immediately activated for service and serves customers until the system becomes empty, as in the previous case. This idle-inspection-service cycle is repeated indefinitely.

Associated with this (T,N)-policy are costs for inspecting the queue, for activating and running the server, and for holding customers in the system. The aim is to find the design parameters (T,N) that minimize the average cost of operating the system.

We begin our analysis in Section 1 by deriving an expression for the average cost of a (T,N)-policy. Then in Section 2 we present a method for computing an optimal (T,N)-policy. We also give some insights into how the optimal policy changes as the parameters of the model change.

The special (T,N)-policy with T=0 (no idle time) is the well-known N-policy studied by Yadin and Naor (1963), Heyman (1968), and Sobel (1969); related works are Bell (1971), Balachandran (1973), Levy and Yechiali (1975), Tijms (1976), Talman (1979), Shanthikumar (1981), Kimura (1982), and Lu and Serfozo (1984). Also, the special (T,N)-policy with N=1 (no inspection period) is essentially the T-policy studied in Heyman (1977) (Heyman and Sobel (1982) discuss the N- and T-policy as well). In Heyman's model, when the server completes an idle period and finds no customers waiting, then the server takes another idle period; in our model the server is committed to serve after each idle period. We show how our analysis can be easily modified to conform to the former assumption.

1. The Average Cost of a (T,N)-Policy

In this section, we derive an expression for the average cost of operating the M/G/1 queue under a fixed (T,N)-policy. We begin by introducing more notation.

Associated with the idle-inspection-service cycle described above, we let ι denote the length of time that the queue is inspected after time T, while no services are being performed. Namely, $\iota = \max\{0, S_N - T\}$. At time $T + \iota$ the server begins a busy period. The number of customers waiting at the start of this busy period is $\nu = \max\{A(T),N\}$. We let B_ν denote the length of the busy period starting with ν customers. Then the total duration of the idle-inspection-service cycle is $S = T + \iota + B_\nu$.

We assume that the costs of operating the system are as follows:

K = cost per cycle for activating and deactivating the server

v = cost per unit time of inspecting (viewing) the queue

r = cost per unit time of running the server

h = cost per unit time of holding one customer in the system.

Then the total cost for a cycle is

$$Z = K + v\iota + rB_\nu + h\int_0^Z X(t)dt,$$

where $\{X(t);\ t \geq 0\}$ is the number of customers in the system over time. The integral is the total customer waiting time.

Our main concern is the average cost per unit time over the infinite horizon, which we denote by $C(T,N)$. Since the traffic intensity $\rho = \lambda/\mu$ is below one, the queueing process is regenerative, and so it is well known that $C(T,N) = EZ/ES$.

An expression for this cost is given in the following result. For this, we let τ denote the length of a busy period for a standard M/G/1 queue started with one customer, and let W denote the total waiting time of the customers present in this busy period. It is known (e.g. see p. 447 in Heyman and Sobel (1982)), that

$$E\tau = 1/(\mu - \lambda) \tag{1.1}$$

$$EW = \lambda\sigma^2/(2(1 - \rho)^2) + 1/(\mu - \lambda), \tag{1.2}$$

where σ^2 is the variance of the service time. We also let

$$\phi_1(T,N) = \sum_{n=0}^{N-1} (N - n)(\lambda T)^n e^{-\lambda T}/n!$$

$$\phi_2(T,N) = \sum_{n=0}^{N-1} (N^2 - n^2)(\lambda T)^n e^{-\lambda T}/n.$$

Theorem. Under the preceding assumptions,

(1.3) $ES = [T+\lambda^{-1}\phi_1(T,N)]/(1-\rho)$

(1.4) $EZ = K + [h(\lambda T)^2 + h\phi_2(T,N) + \phi_1(T,N)(2v(1-\rho) - h)]/[2\lambda(1-\rho)]$

(1.5) $$C(T,N) = (h/2)\{[a\lambda K + (\lambda T)^2 + (av-1)\phi_1(T,N) + \phi_2(T,N)]/[\lambda T + \phi_1(T,N)]\} + r\rho + h\lambda(1-\rho)EW,$$

where $a = 2(1-\rho)/h$.

Proof. By the definition of the cycle time S, we have

(1.6) $$ES = T + E\iota + EB_\nu.$$

Clearly

(1.7) $$E\iota = E[E(\max\{0, S_N - T\}|A(T))] = \sum_{n=0}^{N-1} E(S_{N-n})P(A(T)=n) = \phi_1(T,N)/\lambda.$$

Next, we can write $B_\nu=\Sigma_{k=1}^{\nu}\tau_k$, where $\tau_1,\tau_2,\ldots$ are independent copies of τ that are independent of ν. Then by Wald's identity and (1.1), we have

(1.8) $$EB_\nu = E\nu E\tau = E\nu/(\mu - \lambda),$$

where

(1.9) $$E\nu = E[A(T) + \max\{0, N - A(T)\}] = \lambda T + \phi_1(T,N).$$

Combining (1.6) - (1.9) yields expression (1.3).

Now consider the expected cycle cost

(1.10) $$EZ = K + vE\iota + rEB_\nu + hE\int_0^Z X(t)dt.$$

We already have expressions for $E\iota$ and EB_ν. It remains to find an expression for the expectation of the waiting time

(1.11) $$\int_0^Z X(t)dt = \int_0^T A(t)dt + \int_T^{T+\iota} A(t)dt + \int_{T+\iota}^Z X(t)dt.$$

By Fubini's theorem, we have

$$(1.12) \qquad E\int_0^T A(t)dt = \int_0^T EA(t)dt = \int_0^T \lambda t dt = \lambda T^2/2.$$

Next, observe that

$$(1.13) \qquad E\int_T^{T+\iota} A(t)dt = E[I(A(T) < N) \sum_{n=A(T)}^{N-1} nY_n]$$

$$= \lambda^{-1}E[I(A(T) < N) \sum_{n=A(T)}^{N-1} n],$$

where $Y_1, Y_2, \ldots$ are independent exponential variables with mean λ^{-1} that are independent of $A(T)$, and I is the indicator function. Then applying the identity

$$\sum_{N=a}^{N-1} n = [(N-1)N - a(a+1)]/2 = [(N^2-a^2) - (N-a)]/2$$

to (1.13), and recalling the definitions of ϕ_1 and ϕ_2, we obtain

$$(1.14) \qquad E\int_T^{T+\iota} A(t)dt = [\phi_2(T,N) - \phi_1(T,N)]/(2\lambda).$$

Finally, we can write

$$\int_{T+\iota}^{Z} X(t)dt = \sum_{n=1}^{\nu} \int_{a_n}^{a_n+\tau_n} X(t)dt = \sum_{n=1}^{\nu} [W_n + (\nu-n)\tau_n],$$

where $a_n = T + \iota + \tau_1 + \ldots + \tau_{n-1}$ and (τ_1,W_1), $(\tau_2,W_2), \ldots$ are independent copies of (τ,W) that are independent of ν. Then

$$(1.15) \qquad E\int_{T+\iota}^{Z} X(t)dt = E\sum_{k=1}^{\nu} W_k + E\{E[\sum_{k=1}^{\nu} (\nu-k)\tau_k | \nu]\}$$

$$= E\nu EW + E\tau E[\sum_{k=1}^{\nu} (\nu-k)]$$

$$= E\nu EW + E\tau E(\nu^2-\nu)/2,$$

where

$$(1.16) \qquad E\nu^2 = E[A(T)^2 + \max\{0, N^2-A(T)^2\}]$$

$$= \lambda T + (\lambda T)^2 + \phi_2(T,N).$$

Substituting (1.11) - (1.16) into (1.10) yields expression (1.4). Then expression (1.5) follows from (1.3), (1.4) and $C(T,N) = EZ/ES$.

2. Computation of Optimal (T,N)-Policies

In this section, we address the problem of finding a (T,N)-policy that minimizes the average cost $C(T,N)$.

As a first step, consider the subproblem of minimizing $C(T,N)$ over N for T fixed. This is of interest in itself when one is designing a system in which the idle time T is preset and cannot be varied. The solution to this subproblem is as follows.

Theorem 2.1. For each T, the cost $C(T,N)$ has a unique minimum over N, which is attained at the value

$$N(T) = \min\{N \geq 1 : D(T,N) > 0\}, \tag{2.1}$$

where

$$\begin{aligned} D(T,N) = {} & \lambda T(av-1) - a\lambda K - (\lambda T)^2 + \\ & (2N+1)[\lambda T + \phi_1(T,N)] - \phi_2(T,N). \end{aligned} \tag{2.2}$$

Proof. It is easily seen that

$$\begin{aligned} \phi_1(T,N+1) &= \phi_1(T,N) + F(N) \\ \phi_2(T,N+1) &= \phi_2(T,N) + (2N+1)F(N), \end{aligned} \tag{2.3}$$

where F is the Poisson distribution with mean λT. Using these expressions and (1.3) - (1.5), one can show that

$$C(T,N+1) - C(T,N) = D(T,N)(h/2)F(N)/[(\lambda T + \phi_1(T,N))(\lambda T + \phi_1(T,N+1))].$$

The terms following $D(T,N)$ are positive, and so $C(T,N)$ will have a unique minimum over N at the value (2.1) if $D(T,N)$ is strictly increasing in N. But this is true since one can show that

$$D(T,N+1) - D(T,N) = 2[\lambda T + \phi_1(T,N)] > 0. \tag{2.4}$$

Computation of Optimal N(T) Policies. The optimum N(T) in (2.1) can be obtained by computing $D(T,N)$ recursively by the following formulas based on (2.3) and (2.4):

$$\phi_1(T,N) = \phi_1(T,N-1) + F(N-1) \tag{2.5}$$

$$D(T,N) = D(T,N-1) + 2(\lambda T + \phi_1(T,N-1)), \quad N \geq 2$$

where $\phi_1(T,1) = e^{-\lambda T}$.

Our computations show that $N(T)$, as a function of T, is nonincreasing and then nondecreasing. This was as anticipated: For T near zero, $N(T)$ is moderate since it is the major control parameter; as T grows, $N(T)$ can be reduced, but it eventually tends to ∞.

Remark. Recall that the (T,N)-policy with $T=0$ is the N-policy. In this case, $\phi_1(0,N) = N$, $\phi_2(0,N) = N^2$ and $D(0,N) = N^2 + N - a\lambda K$; and so the optimal $N(0)$ is the smallest integer greater than $(1/4+a\lambda K)^{1/2} - 1/2$. This is consistent with Heyman (1968).

Now consider the problem of finding an optimal (T,N)-policy. This problem can be expressed, with Theorem 2.1 in mind, as

$$\min_{T,N} C(T,N) = \min_{T} \min_{N} C(T,N) = \min_{T} C(T,N(T)). \tag{2.6}$$

If the function $C(T) = C(T,N(T))$ were to have a unique minimum, say at T^*, then it would follow from (2.6) that $(T^*,N(T^*))$ is the unique optimal (T,N)-policy. Because the function $C(T)$ is rather intractable, we were not able to prove that it has a unique minimum. However, extensive computations showed that $C(T)$ does indeed have a unique minimum; we enumerated hundreds of functions and each one had a unique minimum.

Computation of Optimal (T,N)-Policies. From the preceding comments, it follows that an optimal (T,N)-policy can be computed as follows. Compute $N(T)$ and $C(T)$ by the recursion (2.5) for a grid of T-values $T_1,T_2,\ldots$

as fine as desired. Do this for successive T_1, T_2,... until the time T^* $= \min\{T_k: C(T_k) > 0\}$. The resulting $(T^*,N(T^*))$ is the optimal (T,N)-policy. (Alternatively, one may find the T^* that minimizes $C(T_k)$ by a Fibonacci or Golden Section Search Procedure, where $N(T)$ and $C(T)$ are computed at each stage by (2.5). However, the saving of computation time by this procedure is negligible.)

This procedure is very easy to implement. Examples of optimal (T,N)-policies computed by it are shown in Table 1. For these computations, we set $\lambda = 1$ (which is equivalent to λ being the time unit), and we set $a = 2(1-\rho)/h = 1$ (which is equivalent to a^{-1} being the monetary unit).

The average cost associated with an optimal (T,N)-policy is

$$C(T^*,N(T^*)) = hC^*/2 + r\rho + h\lambda(1-\rho)EW,$$

where C^* denotes the expression in braces in (1.5), which is the only term relevant to the optimization (the other terms do not depend on (T,N)). Some of the values of C^* associated with Table 1, for v=30, are as follows:

K=100	200	300	400	500
C*=20	28	35	40	45

These C^* values are rounded to the nearest integer. The corresponding values of C^* for v below 30 are not more than one unit below these values for v=30. The C^* is obviously increasing in v and K. Note that the optimal policies do not depend on the cost r of running the server or on the variance σ^2 of the service time.

Remark. If there is no cost for inspecting the queue (v=0), then it is optimal to continually inspect the queue and have no idle time (T^*=0).

This intuitively obvious result follows since one can show that $\frac{\partial C(T,N)}{\partial T} =$ 0 when T=0.

It is of interest to know whether the optimal policy (T*, N(T*)) is nonincreasing or nondecreasing in a particular input parameter. For example, Table 1 shows that T* $\uparrow$ in v, but N(T*) $\downarrow$ in v. Here is a formal result in this regard.

Theorem 2.2.

(i) T* is strictly $\uparrow$ in each of the parameters K, v and μ.

(ii) N(T*) $\uparrow$ in K and $\downarrow$ in v.

(iii) N(T*) $\uparrow$ in μ for $\mu \leq \mu_0 = \inf\{\mu\colon T^* > v/\lambda^2 K\}$, and N(T*) $\downarrow$ in μ for $\mu > \mu_0$.

Proof. These properties are based on the following result. Consider an optimization problem, like ours, of the form

$$\min_{x \in S} f(x,v),$$

where S is a subset of the line or plane and $v \geq 0$ is a parameter of interest. Suppose f(x,v) has a minimum over $x \in S$ at the point $x^*(v)$; when there are several minima we assume there is a smallest one and call it $x^*(v)$. That is, we assume the following minimum exists

$$x^*(v) = \min\{x\colon f(x,v) = \min_{x'} f(x',v)\}.$$

It is known (see for instance [8]) that $x^*(v)$ $\uparrow$ or $\downarrow$ in v according to whether $\frac{\partial f}{\partial v}(x,v)$ $\downarrow$ or $\uparrow$ in x; moreover $x^*(v)$ is strictly monotone when $\frac{\partial f}{\partial v}$ (x,v) is.

First consider T* and N(T*) as functions of μ. One can show that

$$\frac{\partial^2 C(T,N)}{\partial T \partial \mu} = - v\lambda^3 \mu^{-2}[F(N-1) + \phi_1(T,N)(1 - F(N-1))] / [\lambda T + \phi_1(T,N)]^2 < 0.$$

Thus, it follows by the preceding comments that T* is strictly ↑ in μ (as asserted in (i)). Similarly,

$$\frac{\partial C(T,N+1)}{\partial \mu} - \frac{\partial C(T,N)}{\partial \mu} = \lambda\mu^{-2}F(N)(KT\lambda^2 - v) \\ /[\lambda T + \phi_1(T,N+1)][\lambda T + \phi_1(T,N)].$$

This expression is negative or positive according to whether T is $<$ or $>$ $v/K\lambda^2$, and T* is strictly increasing in μ. This proves assertion (iii). Assertion (ii) and the rest of (i) follow by similar arguments.

Remarks. Our model assumes that after each idle period, even when there are no customers waiting, the server is committed to an inspection-service period. A variation is that when a server completes an idle period and finds no customers waiting, then it takes another idle period. The results above also apply to this setting: just replace $\phi_1(T,N)$ and $\phi_2(T,N)$ by $\phi_1(T,N) - Ne^{-\lambda T}$ and $\phi_2(T,N) - N^2e^{-\lambda T}$, respectively. Note that the (0,N)-policy in this setting is not the N-policy, whereas in our model it is.

References

[1] Balachandran, K.R. (1973). Control policies for a single server system. Management Science 19, 1013-1018.

[2] Bell, C.E. (1971). Characterization and computation of optimal policies for operating an M/G/1 queueing system with removable server. Operations Res. 19, 208-218.

[3] Heyman, D.P. (1968). Optimal operating policies for M/G/1 queueing systems. Operations Res. 16, 362-382.

[4] Heyman, D.P. and M.J. Sobel (1982). Stochastic Models in Operations Research Vol. I. McGraw Hill.

[5] Heyman, D.P. (1977). The T-policy for the M/G/1 queue. Management Science 23, 775-778.

[6] Kimura, T. (1982). Further results on the N-policy for the M/G/1 queue. Technical report. Department of Information Sciences, Tokyo Institute of Technology.

[7] Levy, Y. and U. Yechiali (1975). Utilization of idle time in an M/G/1 queueing system. Management Science 22, 202-211.

[8] Lu, F.V. and R.F. Serfozo (1984). M/M/1 queueing decision processes with monotone hysteretic optimal policies. Operations Res. 32, 1116-1132.

[9] Shanthikumar, J.G. (1981). Optimal control of an M/G/1 priority queue via N-control. American Journal of Mathematical and Management Sciences 1, 191-212.

[10] Sobel, M.J. (1969). Optimal average cost policies for a queue with start-up and shut-down costs. Operations Res. 17, 145-162.

[11] Talman, A. (1979). A simple proof of the optimality of the best N-policy in the M/G/1 queueing control problem with removable server. Statist. Neerlandica 33, 143-150.

[12] Tijms, H.C. (1976). Optimal control of the workload in an M/G/1 queueing system with removable server. Math. Operationsforsch. Statist. 6, 933-943.

[13] Yadin, M and P. Naor (1963). Queueing systems with a removable service station. Operations Research Quarterly 14, 393-405.

Table 1

OPTIMAL (T,N)-POLICIES

The Table entries are T* N(T*).

K \ V	0		5		10		15		20		25		30	
100	9	0	9.5	7	9.9	4	10.0	2	10.0	1	10.0	1	10.0	1
120	10	0	10.4	8	10.9	5	11.0	3	11.0	1	11.0	1	11.0	1
140	11	0	11.2	9	11.7	6	11.8	4	11.8	1	11.8	1	11.8	1
160	12	0	12.0	10	12.5	7	12.6	5	12.6	2	12.6	1	12.6	1
180	12	0	12.7	10	13.3	8	13.4	5	13.4	3	13.4	1	13.4	1
200	13	0	13.8	11	14.0	9	14.1	6	14.1	4	14.1	1	14.1	1
220	14	0	14.0	12	14.7	9	14.8	7	14.8	4	14.8	2	14.8	1
240	14	0	14.6	12	15.3	10	15.5	7	15.5	5	15.5	2	15.5	1
260	15	0	15.2	13	15.9	11	16.1	8	16.1	6	16.1	3	16.1	1
280	16	0	15.8	14	16.5	11	16.7	9	16.7	6	16.7	4	16.7	1
300	16	0	16.3	14	17.1	12	17.3	9	17.3	7	17.3	4	17.3	2
320	17	0	16.8	15	17.6	12	17.8	10	17.9	7	17.9	5	17.9	2
340	17	0	17.3	15	18.2	13	18.4	10	18.4	8	18.4	5	18.4	3
360	18	0	17.8	16	18.7	13	18.9	11	19.0	8	19.0	6	19.0	3
380	18	0	18.3	16	19.2	14	19.4	11	19.5	9	19.5	6	19.5	4
400	19	0	18.8	17	19.7	14	19.9	12	20.0	9	20.0	7	20.0	4
420	19	0	19.2	17	20.2	15	20.4	12	20.5	10	20.5	7	20.5	5
440	20	0	19.7	18	20.6	15	20.9	13	21.0	10	21.0	8	21.0	5
460	20	0	20.1	18	21.1	16	21.4	13	21.4	11	21.4	8	21.4	6
480	21	0	20.6	19	21.5	16	21.8	14	21.9	11	21.9	9	21.9	6
500	21	0	21.0	19	22.0	17	22.3	14	22.3	12	22.4	9	22.4	7

E-24-608

Research Progress and Forecast Report for

"Extreme Values of Queues, Point Processes and Stochastic Networks"

Grant No. AFOSR 84-0367

by

Professor Richard F. Serfozo
Industrial and Systems Engineering
Georgia Institute of Technology

June 15, 1985

1. Summary of Activities

Our research has been evolving as follows.

(a) Extreme Values of Queues. The work on this topic has been progressing steadily along the lines of our initial proposal. There are several subtle technical problems we have to resolve in order to present our results in as general and natural setting as possible. We expect to start documentation of this work next fall.

(b) Extreme Values of Point Processes. Little time was spent on this topic; we will get to this later.

(c) Extreme Values of Stochastic Networks. We are developing bounds for extreme values of specially structured networks. This should be completed by next fall. After that we are planning to change the emphasis on this topic to the following one.

(d) Optimal Control of Networks of Queues. Although this topic was not in our original proposal, we made a breakthrough in this area that we intend to pursue. A major problem in the control of queueing and inventory systems is to estalish the existence of nicely structured monotone optial policies. This problem has been solved for several standard one-dimensional processes, but the techniques used do not extend to multi-dimensional processes such as queueing networks. However, we have developed a new approach that works for multi-dimensional processes. We plan to develop this further during the next year.

(e) Point Processes Related to Extreme Values. The study of extreme values as in (a) and (b) is closely related to the convergence of certain point processes to Poisson processes. We have obtained some exciting results in this topic that are documented in the attached papers. Further discussion is given below.

2. Poisson and Compound Poisson Approximations

A standard approach for analyzing the asymptotic behavior of extreme values of a discrete-time stochastic process is via point processes. The key idea is to represent the cumulative maxima of the process as a functional of an approximate point process on the plane. One establishes the convergence of the point process to a monohomogeneous Poisson process, and then invokes the continuous mapping theorem to obtain the convergence of the maxima. This approach has been used for extreme values of independent variables and of stationary variables.

Our research on this topic was prompted by the following questions. What can one say about extremes of variables that are not independent or stationary? Is it possible that for certain dependent variables the associated point process converges to a compound Poisson or infinitely divisible point process rather than a Poisson process? What are natural conditions for this? What types of limiting distributions would the extremes have and how are they related to the three classical extreme-value distributions? Do the results hold for multi-dimensional processes? What are the rates of convergence of the extremes to their limits?

We were able to shed some light on these questions by developing several compound Poisson apprdoximations that are of fundamental interest. The attached papers contain our results on this theme. Here is more background on them.

Compound Poisson Approximations for Sums of Random Variables

There are a number of well-known Poisson approximations for sums of uniformly small random variables. When the random variables are generally small but not uniform, then one might expect the sum to be approximately compound Poisson or infinitely divisible. Freedman (1975) gave examples of sums that converge to compound Poisson variables but have wild oscillations. This suggests that there may not be omnibus compound Poisson approximations analogous to Poisson or normal approximations.

This is not the case. My paper with the title above gives a compound Poisson approximation for a large class of sums of dependent variables. The results also yield theorems on the convergence of point processes to compound Poisson processes. In operations research or computer systems, most flows of goods and services or data packets are compound Poisson (batch flows) rather than Poisson. Our results should be of use in justifying the assumption of compound Poisson flows in the same way that classical results justified assumptions of Poisson flows.

Partitions of Point Processes: Multivariate Poisson Approximations

A classical result is that a sum of a large number of thin point processes is approximately Poisson. What about the reverse? When a point process is partitioned into a large number of subprocesses, are these subprocesses approximately Poisson? When might they be compound Poisson? These are the issues I address in the second paper.

Annual Technical Report

Project Title: Extreme Values of Queues, Point Processes and Stochastic Networks

Grant No.: AFOSR-84-0367

Project Director: Professor Richard F. Serfozo
Industrial and Systems Engineering
Georgia Institute of Technology

November 1985

Extreme Values of Queues, Point Processes
and Stochastic Networks

During the last year we worked on four research topics. Our progress on these topics is described in the following discussion.

1. Modeling of Stochastic Flows in Networks: Compound Poission Approximations

Our most significant accomplisment last year was the development of compound Poisson approximations for random variables and Point processes. Such approximations are instrumental in the modeling of stochastic flows in networks. Being fundamental in nature, our results apply to other settings as well. The following papers described our work; further discussion is given below.

Serfozo, R. F. (1985). Compound Poisson Approximations for Sums of Random Variables. To appear in Ann. Probability.

Serfozo, R. F. (1985). Partitions of Point Processes: Poisson Approximations. To appear in Stochastic Processes and Their Applications.

Compound Poisson Approximations for Sums of Point Processes. A basic theme in probability is the characterization of the behavior of sums of random variables and point processes. Many physical quantities can be viewed as a sum of a large number of small quantities (e.g. an SAT score is the sum of scores from individual questions, or a company's revenue in a month is the sum of the revenues from its many sales). Moreover, any random sequence S_n can be viewed as the sum of its increments:

$S_n = \sum_{k=1}^{n} (S_k - S_{k-1}) - S_0$. The classical central limit theorem for a sum S_n of independent identically distributed random variables asserts that the distribution of S_n, for large n, is approximately normal, and the quality of this approximation is described by the Berry-Esseen inequality. It is also known that S_n under slightly different conditions, may be approximately Poisson, compound Poisson or infinitely divisible, and there are known error estimates for the Poisson approximation.

The Poisson approximation is frequently used in the operational analysis of telecommunications networks. For instance, the number of telephone calls that arrive to a switching station in an hour from a large number of subscriber lines, as shown below, is typically modeled as a Poisson random variable. More generally, the flow of calls over time from each subscriber is viewed as a "thin" point process and the sum of these point processes that enters the station is modeled as a Poisson process.

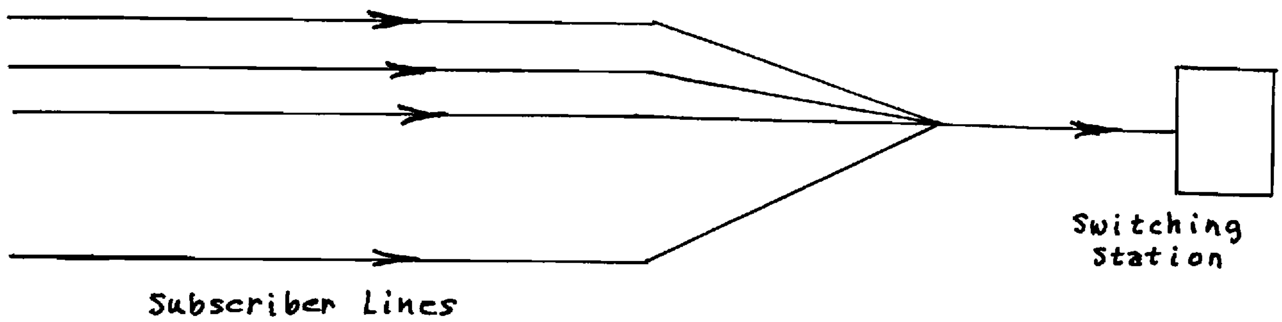

Figure 1. Flows in a Telecommunications Network

This type of merging or summation of point processes occurs in other networks such as (i) flows of data packets in computer networks, (ii) flows of material and parts in automated production plants, and (iii) flows of goods in distribution networks. (These are the principle application areas for our results.) Although Poisson processes are used for modeling flows in these networks, they are inappropriate when the flows have certain natural groupings of points (e.g. a series of data-packets consitutte a message, or a group of parts constitute a delivery). In such instances, which are evidently more common than not, a compound Poisson model may be more appropriate than a Poisson model. This raises the questions: under what conditions can a sum of random variables or point processes be approximated by compound Poisson random variables or point processes? This question is what motivated our research.

During the last twenty years, several ad hoc theorems had been proved on the convergence of sums of independent random variables to compound Poisson variables, but little was known about the error in their attendant approximations. D. Freedman (1973) gave some examples that seemed to imply that one could not develop compound Poissosn approximations that would be as natural or universal as normal or Poisson approximations.

In spite of this dire evidence, we have been fortunate enough to develop such approximations. We have found rather general conditions under which sums of dependent random variables or sums of dependent point processes are asymptotically compound Poisson. More important, we have established bounds on the errors involved in these approximations. Our results are applicable, for instance, for constructing compound Poisson models of merging of flows in networks as described above. These models could be used in conjunction with queueing models to analyze the delay or throughput of the flows. Another major application of our results is described next.

Partitions of Point Processes. The preceding discussion was on the merging of stochastic flows in networks. Another related operation is the partitioning of a single flow into many subflows as shown below.

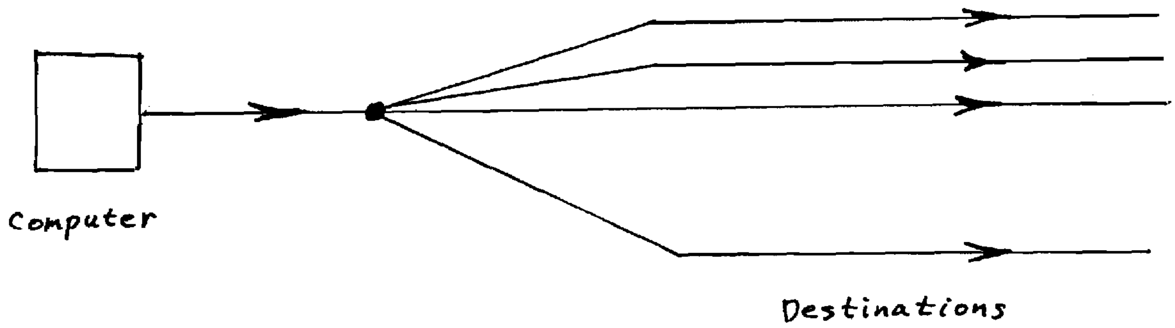

Figure 2. Subflows in a Computer Network

Here a stochastic flow of computer data packets on a network line is entering a computer that directs the packets to several other computers depending on the packets' respective instructions. In other words, the initial flow is randomly partitioned into several subflows. When the number of subflows is large so that each subflow is relatively thin, then one would suspect that the subflows may be modeled as multi-variate Poisson or compound Poisson point processes. Using the results described above, we have been able to shed light on this phenomena. We have found several types of random partitions whose resulting subflows are approximately Poisson or compound Poisson, and we have obtained bounds on the errors in these approximations.

Partitions of point processes, like sums, are fundamental to a variety of contexts other than networks. For instance, consider a point process over time in which each point has one of several attributes (e.g. insurance claims over time may be categorized as small, medium or large in size), then the numbers of points with these attributes form a partition of the parent process. Our results are useful for analyzing the dependency among such subflows as well as the characteristics of each subflow.

2. Extremal Problems in Stochastic Networks

We have obtained a family of bounds for the distributions of certain generic random variables associated with networks. These random variables represent critical path lengths in PERT networks, maximum flows in networks, and lifetimes of systems. This work is documented in:

Weiss, G. (1985). Stochastic Bounds on Distributions of Optimal Value Functions with Applications to PERT, Network Flows and Reliability. Technical report, Georgia Tech.

Description of the Study. We consider a network with nodes $\{1,\ldots,n\}$ and random variables $X_1,\ldots,X_n$ associated with the nodes. Let $I_1,\ldots,I_k$ denote the sets of paths and let $J_1,\ldots,J_\ell$ denote the sets of cuts of the network. We focus on the following random variables.

(a) Critical Path of a PERT network: The nodes represent activities, the X_i are activity durations, and the network structure depicts the precedence constraints. The critical path length is the shortest time needed to complete the project, namely

$$M = \max_{1 \le j \le k} \sum_{i \in I_j} X_i .$$

(b) Maximal Flow in a Network: The nodes represent pipelines and the X_i are maximal flow capacities. The maximal flow through the network from source to sink is

$$L = \min_{1 \le j \le \ell} \sum_{i \in J} X_i .$$

(c) Reliability of a System: The nodes represent components and the X_i are their lifetimes. The system lifetime is

$$T = \max_{1 \le j \le k} \min_{i \in I_j} X_i = \min_{1 \le j \le \ell} \max_{i \in J_j} X_i .$$

It is generally impossible to obtain tractable expressions for the distributions of M, L, T in terms of the joint distribution P of the $X_1, \ldots, X_n$. Consequently, it is natural to seek partial information or bounds on M, L, T. In this regard, we consider worse-case bounds of M, L, T. Specifically, we address the question: What joint distributions P on the $X_1, \ldots, X_1$ solve the following optimizaton problems

$$\max_P E(M - x)^+, \qquad \max_P E(L - x)^-$$

$$\max_{P} P(T > x), \qquad \sup_{P} P(T < x),$$

where the optimization is over all joint distributions P with the fixed marginal distributions $F_1,\ldots,F_n$? We answer this question by presenting mathematical programming algorithms for optimizing P in these problems. This gives us worse-case networks in which the distributions of M, L, T can be computed. These distributions are then bounds for M, L, T in any network.

3. Optimization of Queueing Systems

Two major problem areas in the optimization of queueing systems are as follows:

Optimal Design of Queueing Systems. In designing a service system involving queueing, one typically is able to choose some of the basic parameters of the system (such as numbers of servers or arrival and service rates) from a range of possible values. The problem is to select the parameters so as to minimize the total cost of the system, including the operational cost of the system over its lifetime. This is a static optimization in that the parameters are chosen at the inception of the system and are thereafter fixed for the system's lifetime.

Optimal Dynamic Control of Queueing Systems. In some queueing systems, the basic parameters can be changed dynamically as the queues evolve. For instance in telecommunications systems, the service rate or numbers of servers change as the queue lengths change. The problem is to determine a policy for dynamically regulating the system parameters, based on the queue length, so as to minimize the total cost of operating the system.

We have begun work on several problems in these areas; our progress on these is discussed below. This work compliments our analysis of extreme values of queues discussed in Section 4 in that here we are seeking economical ways to control or dampen extremes of queus.

Optimal Idle and Inspection Periods for M/G/1 Queues

In a standard M/G/1 queueing system, a Poisson stream of customers arrives to a single server who serves them on a first-come-first-serve basis and the service times are independent and identically distributed. In this system, the server is alwasy available for service. In actual systems, however, a server may have to be absent periodically for other duties or for rejuvenation (e.g. a computer may do file maintenance in addition to its standard processing of jobs). In such systems, the customers are served intermittently rather than continuously. Intermittant service is also characteristic of service systems in which short queues are tolerable or when short busy periods for servers is uneconomical. In designing such a system, a natural question is: How long should the server be absent without serving customers and how large should the queue be before the server starts serving customers?

We have solved this problem for an M/G/1 queue that operates under a (T,N)-policy described as follows. Whenever the system becomes empty, the server is idle for a time T and then it inspects the queue continuously without serving customers until there are N customers waiting - thereupon the server is activated and serves customers continuously until the system becomes empty. This idle-inspection-service cycle is repeated indefinitely. There are costs for inspecting the queue, for activating and running the server, and for holding customers in the system. We have developed a nonlinear programming model for determining the parameters (T,N) that minimize the average cost. This is documented in the following paper.

Kim, S. S. and R. F. Serfozo (1985). Optimal Idle and Inspection Periods for M/G/1 Queues. Technical Report, Georgia Tech.

Optimal Control of Networks of Queues. Service systems in manufacturing and telecommunications usually involve random flows of customers among a network of queueing systems.

We have begun a study of the dynamic control of a network in which the service rates at the nodes and the routing of the customers through the network are subject to control each time a customer moves in the network. Whenever a customer moves, the queues in the entire network are observed and, based on the observation, the service rates and routing probabilities are selected until the next customer movement. This is repeated indefinitely. Our aim is to establish the existence of certain natural monotone optimal policies in which the service rates are increasing functions of the queue lengths and the routing probabilities have monotonicity properties such that large queues are avoided. The knowledge of the existence of such policies leads to efficient computational procedures for optimal policies. Furthermore, monotone policies are more natural for implementation in actual systems.

Our approach to this problem area is as follows. We characterize the queueing network process as a multi-dimensional Markov process whose transitions are determined by a family of "transition operators". (As a simple example, a birth and death process has two operators: an upward unit jump and a downward unit jump.) We first establish certain optimal monotonicity properties for these operators, and then translate these into monotonicity properties for the parameters under control. The analysis involves transforming the Markov process into a simpler process and introducing and exploiting the notion of convexity and submodularity of functions with respect to the transition operators. We plan to start documenting our results on this next year.

4. Extreme Values of Queues and Point Processes

Although much of our effort this year has been spent on this topic, we have not reached the documentation stage yet. The research is proceeding along the lines of our proposal, which need not be reviewed here.

There are several technical issues that we are striving to understand more fully: (i) Our major results show that queueing processes have an asymptotic extreme-value distribution that is different from the three classical ones. To shed light on why this is so, we are attempting to prove our results by another approach, possibly via limits of diffusion processes. (ii) We are seeking a more complete characterization of the types of service times in queues for which our results apply. (iii) We are attempting to obtain necessary as well as sufficient conditions for our limit properties of queues and point processes.

We will give a more extensive review of this work in our next progress report.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: NA

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY: NA

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: NA

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S):

6a. NAME OF PERFORMING ORGANIZATION: Georgia Institute of Technology

6b. OFFICE SYMBOL *(If applicable)*:

7a. NAME OF MONITORING ORGANIZATION: AFOSR/NM

6c. ADDRESS *(City, State and ZIP Code)*:
School of Industrial and Systems Engineering
Atlanta, GA 30332

7b. ADDRESS *(City, State and ZIP Code)*:
Bldg 410
Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR

8b. OFFICE SYMBOL *(If applicable)*:

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR 84-0367

8c. ADDRESS *(City, State and ZIP Code)*:
Bldg 410
Bolling AFB, DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
6.1102F	2304		

11. TITLE *(Include Security Classification)*: Final report on Extreme Values of Queues Point Processes and Networks

12. PERSONAL AUTHOR(S):

13a. TYPE OF REPORT: Final

13b. TIME COVERED: FROM 10/1/84 TO 9/30/86

14. DATE OF REPORT *(Yr., Mo., Day)*: November 30, 1986

15. PAGE COUNT:

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*:

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*:

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL: Richard F. Serfozo

22b. TELEPHONE NUMBER *(Include Area Code)*: 404 - 894-2305

22c. OFFICE SYMBOL:

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

FINAL REPORT

Project Title: Extreme Values of Queues, Point Processes and Stochastic Networks

Grant No.: AFOSR 84-0367

Project Director: Professor Richard F. Serfozo
Industrial and Systems Engineering
Georgia Institute of Technology

November 1986

Extreme Values of Queues, Point Processes
and Stochastic Networks

This two-year research grant consisted of several themes. The following is a summary of our results.

1. Compound Poisson Approximations

Our work on this topic just appeared in Serfozo, R.F. (1986). Compound Poisson Approximations for Sums of Random Variables *Ann. Probability*, October issue.

During the last 20 years, several theorems have been proved on the convergence of sums of independent random variables to compound Poisson variables. Little was known about how close the sums are to being compound Poisson. Examples were published that seemed to indicate that one could not develop compound Poisson approximations that would be as natural as normal or Poisson approximations.

However, we have been able to develop such approximations. We have proved that a sum of dependent random elements is approximately compound Poisson when the variables are rarely nonzero and, given they are nonzero, their conditional distributions are nearly identical. We have given several upper bounds on the total-variation distance between the distribution of such a sum and a compound Poisson distribution. These bounds are analogous to Berry-Esseen bounds for normal approximations. Our results apply to general random elements such as unions of random sets and sums of random measures or point processes. Our results appear to be useful for characterizing high-level exceedances of dependent variables. We hope to pursue this in the near future.

2. Partitions of Point Processes

Our work on this appeared in Serfozo, R.F. (1985). Partitions of Point Processes: Multivariate Poisson Approximations. Stoch. Process Appl. 20, 281-294.

We proved that when a point process is partitioned into certain sparse subprocesses, then the subprocesses are asymptotically multivariate Poisson or compound Poisson. Using results described above, we derived bounds for the total-variation distance between the subprocesses and their limits. We did this for several types of partitioning rules including independent, Markovian and batch assignment of points. Partitions of point processes are omnipresent in flows of parts in manufacturing networks and distribution systems, and flows of data packets in computer networks.

3. Extreme Values of Birth and Death Processes and Queues

Our work on this has been documented in:

Serfozo, R.F. (1987). Extreme Values of Birth and Death Processes and Queues. Stoch. Processes Appl. (to appear).

Serfozo, R.F. (1986). Extreme Values of Queue Lengths in M/G/1 and GI/M/1 Systems. Technical Report.

In these papers, we solve the long-standing problem of characterizing the asymptotic behavior of the maximum values of birth and death processes and queues over large time intervals. When these processes are positive recurrent, the distributions of their maxima do not converge to a non-degenerate distribution, in the usual way under linear normalizations. We show, however, that by varying the process parameters in a certain way as the time interval grows, then these maxima do indeed have three possible limit distributions. Two of them are classical extreme value distributions and the other one is a new distribution.

Our results on birth and death processes include conditional limit theorems for maxima of transient processes conditioned that they are finite. For the M/G/1 and GI/M/1 queues, the analysis was more complicated since a certain basic distribution was known only indirectly in terms of ratios of integrals of complex valued functions.

4. Stationary and Reversible Processes

Our work on this appeared in Serfozo, R.F. (1986). Heredity of Stationary and Reversible Processes. Adv. Appl. Probability.

The notion of reversibility plays an important role in characterizing the equilibrium behavior of a network of queues. There are a number of processes associated with a queueing network that are important by themselves. Examples are the number of customers in a certain sector of the network and the point processes of customer flows between pairs of nodes. One frequently confronts the problem: Are these related processes stationary, reversible or ergodic when the network process has these properties? In other words, if a process is stationary, reversible or ergodic, then what functionals of the process have these properties. We answer this by identifying a large class of such functionals. In doing so, we generalize a fundamental result for the heredity of stationarity and we provide an efficient characterization of reversibility that can be used for general random elements such as point processes and random sets.

5. Extremal Problems in Stochast Networks

Our work on this will soon appear in Weiss, G. (1987). Stochastic Bounds on Distributions of Optimal Value Functions With Applications to PERT, Network Flows and Reliability. Operations Research, to appear.

It is generally impossible to obtain tractable expressions for the probability distributions of (1) Critical path lengths in a PERT network, (2) Maximal flows in a network, or (3) Lifetimes of complicated systems. Consequently, it is natural to seek partial information on worse-case bounds of these variables. This can be formulated as a mathematical programming problem. We present an algorithm for solving this problem. The solution can then be used to obtain bounds for general networks.