A SURVEY OF THE MACHINE INTERFERENCE PROBLEM

Lani Haque and Michael J Armstrong

Sprott School of Business, Carleton University

Ottawa, ON, K1S 5B6, Canada

Abstract

This paper surveys the research published on the machine interference problem since the 1985 review by Stecke & Aronson. After introducing the basic model, we discuss the literature along several dimensions. We then note how research has evolved since the 1985 review, including a trend towards the modelling of stochastic (rather than deterministic) systems and the corresponding use of more advanced queuing methods for analysis. We conclude with some suggestions for areas holding particular promise for future studies.

***Keywords*:** maintenance, reliability, queuing, machine interference, machine repairman.

Acknowledgements

This work was supported in part by a Discovery Grant from Canada's Natural Sciences and Engineering Research Council (NSERC). The authors would also like to thank Josephine Dahan for her help in collecting the references surveyed in this paper.

Published as:

Haque L, Armstrong MJ, 2007, "A survey of the machine interference problem", *European Journal of Operational Research* 179 #2, 469-482, doi:10.1016/j.ejor.2006.02.036.

http://www.sciencedirect.com/science/article/pii/S037722170600244X

A SURVEY OF THE MACHINE INTERFERENCE PROBLEM

Abstract

This paper surveys the research published on the machine interference problem, also called the machine repairman problem, in which machines interfere with each other's service. Our emphasis is on work that has appeared since the 1985 review by Stecke & Aronson. After describing the basic model and the scope of our study, we discuss the literature along several dimensions. We describe some of the more interesting papers, and offer some suggestions for topics holding particular promise for future studies. We conclude with comments about how the research has evolved over the past two decades.

***Keywords*:** queueing, maintenance, reliability, machine interference problem, machine repairman problem.

1. Introduction

Consider the simple system consisting of *n* machines and *r* workers shown in Figure 1. Each machine operates for a period of time until it breaks or otherwise requires attention, at which point it is sent to the service facility. A worker there spends a period of time servicing the machine and then returns it to operation. If there are more machines than workers, $n>r$, it will occasionally happen that all the workers are already busy when another machine needs service, i.e. the machines *interfere* with each other's service. Thus this system is a simple example of what is referred to as the *machine interference problem* (MIP) or alternatively as the *machine repairman problem*.

The MIP model can be found in even introductory-level textbooks in business, engineering, and mathematics, because it can be used to model a wide variety of real systems. For example, in [49] a group of numerically controlled machine tools rely on a single human operator to load the parts being machined; interference occurs whenever the loading needs of multiple machines overlap. Another manufacturing example from [35] is an "autoconer" machine containing several spindles for winding yarn in a textile factory. Whenever the yarn breaks off from one of these spindles, it is re-attached by a robotic knotting device; but since there is only one robot serving several spindles, the need for knotting at one spindle can lead to delays in knotting at another spindle.

Other applications include telecommunications [88] and computer networks [32]. Here the MIP models a system consisting of a large number of terminals or desktop computers ("clients") that submit requests ("jobs") for processing service to a small number of more powerful computers ("servers"). Interference arises in these client-server systems whenever a queue of jobs forms at the servers, causing delays in the work being done at the client computers. Further non-manufacturing applications include coal shipment [134] and aircraft maintenance [28].

A survey of MIP research by Stecke & Aronson [7] was produced in 1985, but the two decades since then have seen the publication of a large number of additional papers spread across a variety of journals. Many of these are listed within the more general queuing bibliography prepared by Sztrik [9], and some are discussed in the relevant sections of [2] and [10]. Our objective in this article is to provide an updated survey that consolidates much of this diverse body of more recent MIP research.

Given the large number of published papers, for practical reasons we had to make some choices about which articles to include, particularly when they overlapped the MIP and another topic area. For example, we excluded maintenance models involving larger multi-echelon systems or more general queuing networks, such as those used to model systems of repairable inventory, as in e.g. [6]. Another borderline situation is that of "self-service" situations where $n=r$ so that there is no interference for service; here the steady state probabilities can be derived using the multinomial distribution. Most of the research we discuss includes this latter situation as a limiting case, but we did not include studies that focus mostly or solely on it since they involve no interference as such.

The core MIP is related to several other models. For example, most of the work on *k-out-of-n:F* systems includes either no repair at all, or else assumes repair capacity large enough to avoid any interference. A few studies however, such as [90], do assume a limited repair facility and its corresponding interference potential, and consequently these are related to studies of the MIP with spare machines. We have included a sample of recent papers in this latter category, but do not claim to be complete in our coverage of *k-out-of-n:F* research. In a similar manner, we have included only a few examples of work on completely deterministic systems, as in e.g. [12], as here the main issue is the *optimal scheduling* of a known and finite set of service requirements in advance.

The *economic machining problem* and the *tool wear problem* (see e.g. [1]) are somewhat similar to the MIP, except that the machine tools in the system gradually wear out rather than suddenly breaking down. Since there is generally no queue of idle machines awaiting maintenance, our survey herein does not cover this topic. Our survey also excludes research on *opportunistic maintenance* for systems in which it is advantageous to perform repairs on many machines at once, as in e.g. [8] where there is a large fixed cost for any repair but only a small variable cost per machine repaired. Whereas MIP policies try to *avoid* simultaneous shut downs, opportunistic maintenance policies *prefer* to have multiple machines shut down at the same time.

We should also note that the term "machine interference" is sometimes used (e.g. in mechanical engineering) to describe situations where machines may physically get in each other's way *during operation*, rather than while *awaiting service*. This includes for example the mechanical motion of robotic arms that are closely spaced on an assembly line. We do not consider these subjects or those that deal with "machine repair" in more general settings. For the later we refer the reader to the survey of maintenance research for multi-machine systems by Cho & Parlar [2].

We begin in the next section by describing the basic MIP and its analysis more fully. In the sections that follow we discuss the existing literature along a number of different dimensions. We then highlight several ways in which MIP research has evolved since the previous review, and conclude with a discussion of potential future work.

2. MIP modelling

Analysis of a MIP model typically begins by deriving the steady-state probability distribution p_i, $i \in \{0,1,...n\}$, that describes the long-run probabilities of i machines being in the failed or "down" state at any given point in time. For example, suppose that in the system shown in Figure 1 the machine operating lifetimes are exponentially distributed with a mean time of $1/\lambda$, and that the service durations are likewise exponentially distributed with a mean time of $1/\mu$. Further assume that the service facility has ample buffer space for machines to queue up while awaiting service, machines are served in First-Come, First Served (FCFS) order, and that a machine returns to operation "as good as new" after being served. Under these assumptions the system is easily modelled as a finite population M/M/r queue with n sources (sometimes denoted M/M/r/n/n) and analysed as a birth-death process (see e.g. [3]). The steady-state distribution can be used in turn to derive a variety of performance measures for the system, such as the probability that at least k machines are in the operational or "up" state, $P[Up \geq k] = p_0 + ... + p_{n-k}$, or the mean number of machines down, $E[Down] = \Sigma\, i\, p_i$.

This *descriptive* analysis takes the model parameters $\{n, r, \lambda, \mu\}$ as given and then describes the system's performance according to certain metrics. Many researchers take the analysis a step further with a *prescriptive* study in which they seek to optimise a selected performance measure by choosing the best value for one or more model parameters, perhaps subject to some constraints. This approach addresses system management questions such as:

- How many machines should be in the system in total? How many of them should we hold as spares? At what speed should they be run?
- How much service (repair) capacity is needed in the system? How much of this capacity should be held in reserve ("on call")? How quickly should it serve?
- Which machines should be assigned to which servers for service? In what order should a given queue of machines be serviced?
- For each of these questions, is it sufficient to follow a *static* policy in which all parameters remain fixed, or is it worthwhile to use a more complex *dynamic* policy that adapts to the current state of the system?

As in the example system of Figure 1, all MIP models have a finite number of machines sharing a service facility with limited capacity, which is where the interference arises. One way to view the system is as a closed queuing network consisting of two stations. One station represents the service facility with r servers in parallel, while the other has "servers" that represent the machines that are working or standing-by as spares. The n machines flow back and forth between the two stations as they shut down and are subsequently returned to operation.

3. Categorization by queuing model

One useful way to organize the research done on the MIP is to look at the core of each model in queuing terms, i.e. the basic assumptions concerning the machine operating time distribution, the service time distribution, and the number of servers present. The most commonly used distribution for either the machine lifetimes or the service durations has been the exponential (symbol M as in Markov), as this greatly simplifies mathematical analysis. Some researchers have broadened the applicability of their work by generalizing this to the Erlang (E), hypo-exponential, hyper-exponential, or phase (PH) distribution. Other studies permit any general (G) distribution, while a few have focussed on the deterministic (D) case.

Table 1 shows a breakdown of papers by the distributions and the number of servers used. Note that the structures shown are the "underlying" ones defined before taking into account other features of a particular study; for example, if a model uses an exponential distribution for machine failure time distribution but also includes balking or retrials, we have shown it in the "M" category even though the resulting inter-arrival pattern seen by the server is no longer exponential as such.

The work in [145] is distinctive in being a case study of an actual manufacturing process, in which historical data is used to create empirical distributions for a computer simulation of the process. Compound distributions to model bulk arrivals of machines for service are used in [110] (M^x/E/1) and in [71] (G^x/G/r). In [47] each machine is itself a system of components. Each component has its own lifetime and repair time, so the overall service times depend on how many components have failed by the time the service is performed. Also of note is that whereas most MIP studies are done using continuous time models, the analysis in both [97] and [104] is done in discrete time.

We think it could be beneficial to have more work relating to deterministic service times, e.g. M/D/1 or G/D/1 models. The use of phase distributions might be one way to approximate this deterministic aspect. These would be useful for representing systems where the service times are relatively consistent, perhaps because the service is automated or standardized. Aircraft maintenance, for example, involves many standard checklists for components that need regular inspection for safety reasons.

4. Categorization by model features

In the subsections that follow we consider the literature along several dimensions: server availability, service disciplines, machines types, arrival patterns, state-dependent times, and performance metrics. In each case we list the papers that have used a particular feature, and we briefly describe some of the more distinctive examples.

4.1 Server availability

One way in which a system's service capacity can vary is through the employment of *reserve* servers who temporarily assist the regular *r* servers whenever the queue of waiting machines becomes too large. This is one example of a MIP using dynamic control, wherein a system parameter is varied according to the evolving state of that system. In [75] there are *r* regular servers plus *s* reserve servers; the reserves are gradually added to the facility as the queue length grows, and they cease working (return to reserve status) as the queue later subsides. [76] likewise allows *r* regular servers but only 1 reserve. Other studies use fewer servers: [26] allows for 2 regular servers plus 1 reserve, while [144] allows 1 regular and 2 reserves. [107] and [27] allow only 1 regular server plus 1 reserve, whereas [68], [73], and [87] have 0 regular and 1 reserve. The simulation in [145] includes a pool of reserves that float between several machine groups (parallel MIPs) as required.

The reverse situation, where the regular server may be temporarily absent, has also been studied. Under the *vacation* model of [61] a server can become unavailable by shutting-down for a "vacation" of random length at the moment it becomes empty (i.e. when there are no machines in need of tending). Whenever a service worker returns from vacation and finds the facility still empty, one of several policies may then be followed:

- Multiple vacation policy: The worker immediately takes another vacation.
- Single vacation policy: The worker does not take another vacation until after tending at least one machine and then finding the facility empty.
- Hybrid vacation policy: The worker waits for a random time and if there are still no machines needing service, he takes another vacation.

In [91] and [125] the facility offers *gated service*: when the server returns from vacation, she only serves those machines that are already waiting. Any new arrivals must wait for the server to return from her next vacation. In [41] the server returns to work either at the end of her random vacation or when the queue exceeds some control limit, whichever is *earlier*; i.e. she returns early if enough work has accumulated. In [128] the server returns at the *later* of these two times, i.e. she extends her vacation if there is not much work waiting. To represent set-up time that may be required before a server can begin work, [95] allows server vacations to start at the moment a machine arrives.

Another way for a server to be unavailable is for it to be *unreliable,* i.e. the server itself could break-down. Studies with this feature typically assume that the failed server resumes functioning (or "self repairs") after a random amount of time. The models in [40], [41], [130], and [137] have a single unreliable server; while [81], [133], and [136] each include several unreliable servers. The studies in [130] and [133] compare the case where an unreliable server can fail at any time (even if idle) to an alternative case where the facility can fail only when it is in use. In [96] the unreliable server is composed of several components: all components must be functioning for the server to function, and each component can fail and self repair independently.

The models used in [18], [20], [21], [22], [23], and [24] take the idea of unreliable servers a step further by modelling a "two level" MIP. They consider a computer system composed of multiple terminals supported by a single central processing unit (CPU). Each terminal can generate one job at a time for the CPU to process, and these may have to wait until the CPU finishes jobs from other terminals. Thus the normal operation of the computer network behaves as a "software" MIP, with electronic jobs lining-up for service from the CPU. But in these models the terminals and the CPU are themselves machines that may fail and need repair by a human worker. Thus there is simultaneously a second "hardware" MIP in which the computer terminals and CPU are the machines that line up for service from the human worker. The work in [20] is similar, except that only the terminals can fail, not the CPU.

A few studies have allowed servers to be *heterogeneous*, i.e. to each have a different service rate. [57] has *r* servers with different rates, while [26] has 2 regular servers plus 1 reserve server, all with different rates.

The model is [49] has a more complicated server consisting of several stages in series. Machines that need service must pass through all of these stages; each stage can handle only 1 machine at a time, but there can be machines in different stages simultaneously (e.g. a two-stage server could have one machine at stage 1 and another machine at stage 2).

To date the amount of research done with variable numbers of servers (as in reserves and vacations) has been relatively small, at least in comparison to the work that has been done with spare machines (covered later in this survey), so there may be potential for more work on this topic.

4.2 Service disciplines

When machines arrive for service, the service facility needs to make two decisions: which worker will service each machine (the *loading* decision), and in what order will the machines be served (the *sequencing* decision). These decisions are made in accordance with the *service discipline* in use at the facility. The service discipline most commonly used in MIP research is FCFS, and in a *homogeneous* system where the servers and machines are all identical FCFS is as good as any other discipline. On the other hand, if the machines and/or servers differ in some way (a *heterogeneous* system), an alternative discipline could provide better performance. Some of these alternatives are Service In Random Order (SIRO), Shortest Processing Time (SPT) and Smallest Failure Rate (SFR). If priority is given to certain machines or classes of machines, then this may be on either a non-pre-emptive or a pre-emptive basis. With *non-pre-emptive priority*, a higher priority machine that enters the service facility will go to the head of the queue but will not interrupt the service of the machine currently being served. In contrast, with *pre-emptive priority* the unfinished service of a lower priority machine would be interrupted so that the higher priority machine could begin service immediately.

The work in [22], [42], [43], [55], [56], [78], [94], [102], and [124] compares the performance of different service disciplines in systems with a single server. Comparisons involving multiple servers are found in [32], [54], [100], [101], and [117]. The sequence of repairs is also relevant if machines have several failure modes, each with their own service rate or cost; this issue is considered in [30], [31], [67], [77], [99], [103], [138], and [141]. [70]

examines the optimality within the MIP context of the $c\mu$ rule, which gives priority to machines that are more costly to have idle and/or are quicker to repair.

More complicated priority schemes use *dynamic* rules that choose the sequence of machines for service based upon the current state of the system. [56], [126], and [127] give priority in part to machines that have waited the longest in queue for service. The systems in [15], [44], [89], [90], and [143] are *k-out-of-n:F*, i.e. the failure of more than *k* of the *n* machines will cause the entire system to fail. These studies look for repair sequences that offer the greatest reduction in this risk of system failure. [69] considers which of several production machines to service next, given the current inventory levels of the products that they produce.

Models with a *patrolling server* (similar to a queuing system with *polling*) examine a different kind of service discipline traditionally found within the textile industry. One version of this model has the machines evenly spaced around a circle; a worker walks around the circle from machine to machine and tends failed machines in the order that he comes to them. An alternative layout has the machines in a straight line and a worker who walks back and forth along this line (*bi-directional* patrolling). Examples with patrolling include [21], [33], and [97].

Other patrolling studies are found in [35], [36], and [48], where there is a possibility that a repair may not succeed: if one repair attempt is unsuccessful, the worker can either retry immediately or resume their patrol. Repairs may also be unsuccessful in [47] which compares a variety of different patrolling schemes, including some that involve server vacations. Patrolling workers in [30] must deal with two types of machine failures, one of which has non-pre-emptive priority for repair over the other.

Somewhat removed from the main focus of this survey is the scheduling of a server for a set of service requirements of known (deterministic) duration, as in the tending of machine tools for parts production; herein we discuss just a few examples that involve 2 machines and 1 server. [84] and [86] evaluate a *look ahead scheduling* heuristic that chooses the next machine for service so as to minimize interference at the moment when the chosen machine will need its next service; i.e. the heuristic looks ahead one operating lifetime into the future. The studies in [12] and [85] search farther ahead to find an optimal sequence for the entire set of jobs awaiting production. This optimisation problem is known to be NP-hard, so algorithms are proposed for finding good (but not necessarily optimal) sequences to minimize the makespan.

4.3 Machine types

MIP research often assumes that all machines are identical, but many studies have included machines that differ in operating lifetimes, service durations, and/or costs. Work involving these *heterogeneous* machines includes: [19], [21], [22], [23], [24], [32], [37], [39], [43], [47], [48], [50], [54], [55], [56], [70], [78], [94], [100], [101], [102], [115], [116], [117], [119], [120], [122], [123], and [124].

Machines can also differ based upon whether they are currently in full operation or are sitting ready as *spares*. In models with spares, at most k of the available machines are in productive use at any time, while the remaining s are spares; this gives a total of $n = k+s$ machines in the system. When an operating machine fails and is sent for repair, one of the spare machines takes its place. Whenever a machine finishes being repaired, it normally goes into the pool of spare machines but it can go directly into the operating group if there are less than k machines in operation. Machines in the spare pool usually have a lower failure rate λ_s than the failure rate λ of the operating machines, but this depends on the model:

- A *cold* spare does not fail at all, so $\lambda_s = 0$;
- A *hot* spare fails at the same rate as an operating machine, $\lambda_s = \lambda$;

- A *warm* spare fails at some intermediate rate, i.e. $0 < \lambda_s < \lambda$.

Models that include only cold spares are found in: [26], [27], [46], [59], [64], [65], [66], [71], [77], [93], [108], [110], [129], [132], and [138]. The following works consider only warm spares: [16], [60], [61], [68], [72], [73], [75], [76], [111], [112], [113], [114], [133], [134], [140], and [141].

The systems studied in [74] and [131] have two classes of spares, cold and warm. When an operating machine fails, a warm spare (if available) moves into operation, and a cold spare (if available) takes the warm spare's place. When a failed machine finishes its repair, it normally becomes a cold spare but it can become a warm spare or an operating machine if needed. The work in [118] has both warm & hot spares, while that in [135] considers a system with two types of warm spares with different failure rates.

Sometimes it is important to always have at least *k* machines available for operations. In e.g. [72] and [75], if all available spares are being used and there are still less than *k* operating machines, the system is said to be *short*, and this can result in a failure of the overall system. This is an example where the research on the MIP is overlapped by that on *k-out-of-n:G* (or *k-out-of-n:F*) systems. Some recent examples in this category include [15], [41], [44], [87], [89], [90], [123], [128], [105], and [143].

4.4 Arrival patterns

In most MIP models a machine requiring attention is immediately sent to the service facility and remains there until it has been restored; however, some models allow machines to *balk* and/or *renege*. Balking occurs when a failed machine arrives at the service facility but chooses not to enter because all the servers are already busy. Similarly, reneging occurs when a machine that is currently waiting in line for service decides to leave the facility without being served. In either case, the machine in question immediately returns to normal operation.

Papers with both balking and reneging include [13], [27], [77], [81], [107], [108], and [109]. The models in [25], [26], and [75] incorporate just balking, while [73] contains only reneging.

Related to balking is the idea of *retrials*. Whenever a machine needs attention it makes an initial visit to the service facility; if all technicians are busy at the moment of the visit, the machine does not enter the queue but instead waits a random amount of time before making a repeat visit, called a *retrial*. While waiting between retrials the machine is said to be *blocked* or *in orbit*. MIP retrial models are typically used in the study of computer networks, and are considered in [17], [19], [29], [51], [52], [95] and [98].

Machines needing service may also delay entering the repair queue under the *double age* preventive repair policy briefly introduced in [28]. Under this policy, a still-functioning machine that reaches its normal age for overhaul will try to undergo preventive repair. If however all the service workers are busy at that time, it instead stays in operation and only enters the repair facility when a worker becomes available, or when the machine reaches its second (higher) age limit, or when the machine breaks down and requires corrective repair.

Service for machines in MIP is normally performed while the machine is shut down, but in some cases a service can be provided without taking the machine out of operation (these are sometimes distinguished as *internal* and *external* service activities). One step in this direction was taken in [16], which includes preventive maintenance as a control parameter: performing more preventive maintenance while the machine is in operation decreases the machine failure rate, but also increases the maintenance cost, so a suitable trade-off needs to be found.

This idea of "internal" service (which we might alternatively call "concurrent" service, since it occurs concurrently with the machine's operation) may in fact be a good topic for future study. For example, it is sometimes possible to perform certain tending activities for machine tools while the machines are still functioning, e.g. preparing a pallet of parts for loading. Both of these situations would imply that a server would be engaged in service while the machine is still operating, i.e. the "queue" is empty but the server is busy. How would one efficiently model this?

Another topic for future research is to consider the physical and economic implications of balking, reneging, or retrials in the MIP context. For example, in a conventional (open or infinite source) queuing system, a customer that balks is typically assumed to go elsewhere for service and is lost; but what happens in the closed system of a MIP? In maintenance contexts, is the failed machine immediately repaired by an external service agency, and if so, how do we account for the time and cost of this external repair? Likewise in the computer network context, when a user's job from a terminal balks or goes into a series of retrials, what is the impact on user satisfaction and thus system revenues?

4.5 State-dependent distributions

Several of the model features that we mentioned in previous sections implicitly involve having the machine lifetimes and/or the service durations depend on the state of the system in some way. In this section we mention several papers where the explicit focus of the work is on systems with such *state-dependent* distributions.

The models in [27], [135], [136], and [140] have exponential service times that can use either of 2 rates depending on the current length of the queue; e.g. service can switch to a higher speed if all servers are busy. [46], [63], [79], and [92] allow several discrete service speeds, while [55] allows the speed to be chosen from a continuous range. In [129] the service rate follows a general distribution with 2 possible speeds; the speed chosen can depend on both the queue length and also on the amount of work required to perform the next service task.

The research of [87], [89], [90], and [105] involves the topic of *load sharing*. Here the machine failure rate follows an exponential distribution, but the rate parameter increases as the queue length increases: when fewer machines remain in operation, they must work harder in order to handle the fixed demand for their combined output.

The models in [116] and [117] allow the machine lifetimes (general distribution) and the service times (exponential distribution) to all depend on the queue length. [72] and [75] have generally distributed arrival rates and service rates; each of these switches between 2 speeds depending on the queue length. In [74] the distributions are only exponential, but the rate parameters are fully controllable.

The model of [80] has a service rate that depends on the queue length; as well, both the machine failure rate and the service rate are allowed to change over time, so that e.g. the machine failure rate could increase as the system gets older.

In [118], [119], [120], [121], [122], [123], and [124] the exponential machine arrival rates and the exponential service rates each depend on the state of their "random environments". These environments are modelled as Markov chains, in which the state reflects such exogenous factors as weather or consumer demand.

We believe there is potential for more research on dynamic operating policies in which parameter values can be varied depending on e.g. how busy the repair facility is at any given point in time. The objective here is to account for system managers who actively intervene to improve system performance when needed (by e.g. deploying reserve servers), rather than simply letting the system run "hands-off" in a "random" fashion. The use of state-dependent rates might be one way to work in this direction, and it might also be possible to borrow ideas from other queuing topics, as in e.g. [11]. One interesting issue for such studies is the evaluation of the net benefits of dynamic policies; e.g. is the increase in operating performance large enough to justify the increased complexity of managing such a system?

4.6 Performance measures and optimisation

A wide variety of system performance measures have been considered in MIP research. These may be broadly categorised as measures of either *operational* performance (e.g. average number of machines waiting in the queue), or of *economic* performance (e.g. average total cost per year). For economic measures a number of different costs may be included, such as the cost of service, the cost of shortage (e.g. lost revenue while a machine is unavailable), and the cost of operating (e.g. fuel usage while a machine is running). Each of these costs may be either deterministic or random, and lump-sum amounts (e.g. dollars per repair) or time-based rates (e.g. dollars per day spent repairing). These performance measures are most commonly calculated as long run average rates per unit time (e.g. dollars per day) but an expected discounted total such as net present value is sometimes used instead. One purpose of these measures is to find optimal or near-optimal values for one or more decision variables. Decision variables commonly used include the number of servers r, the service rate μ per server, and the number of spare machines s.

The most common optimisation target in MIP research has been to minimize some kind of average cost rate, or equivalently to maximize an average profit rate. Papers in this category include: [28], [30], [40], [45], [46], [59], [66], [67], [69], [70], [77], [81], [87], [111], [112],

[129], [130], [131], [132], [133], [135], [136], [137], [138], [140], [141], and [146]. The work in [16] differs by using discounted total cost as its economic performance measure.

The maximization of the average machine availability is the objective in [102], whereas [54] seek to maximize the expected discounted total availability. Papers that considered *k-out-of-n:G* (or *k-out-of-n:F*) systems, such as [15], [41], [44], [73], [87], [89], [90], [105], [123], [128], and [143], tend to focus on the reliability of the overall system, e.g. the probability of having at least k machines functioning, though they implement this in different ways.

Instead of evaluating the steady state of the system, the research in [58], [82], and [106] considers its *transient* behaviour over time. The work in [34] and [83] also takes this view in analysing the system's busy period, while [80] examines transient behaviour in the context of aging machines that become less reliable over time. The deterministic models in [12], [84], and [85] deal with minimizing the makespan for the set of jobs currently facing the system.

In the system studied in [16] the machines' failure rate λ can be reduced by paying for more extensive preventive maintenance (e.g. oiling and minor tune-ups), and thus the failure rate λ indirectly becomes a decision variable to be optimised. The work in [28] explores the use of *age repair* policies and so seeks an optimal age t at which to shut down machines for preventive repair. In [39] the total average processing capacity (the sum of processing capacities of all operating machines) of a system of heterogeneous machines is maximised through the choice of repair policy.

The study of [14] is distinctive in looking at the overall relationships between several performance measures for the MIP in general rather than analysing any particular system in detail. The author considers a fairly general system of heterogeneous machines and homogeneous repair technicians, allows failure times and repair times to be generally distributed, and assumes that the service discipline is work conserving. The results of this approach are independent of the more specific modelling assumptions that other papers make, and so are more broadly applicable.

A factor common to most MIP research to date is the tendency to focus on the *mean* value of system performance measures, such as the average number of operational machines. We think it would be of interest to look beyond these averages to also examine some measure of the *variation* in system performance (e.g. standard deviation, 95th percentile, etc.). The objective here is to consider the "risk" implied by an operating decision. For example, a system manager might prefer a service policy that provides a smaller average number of operational machines if it is able to provide those machines more consistently. Unfortunately, derivations of the variance tend to be less tractable than those of the mean in most queuing models, so this topic may prove difficult.

5. Combining the MIP with other models

Although the MIP has proven to be a rich subject for research in and of itself, in actual applications it often forms just part of a larger system. Consequently some studies have considered combinations of the MIP with models from other areas of research. In these cases the question is not so much how to operate the MIP subsystem for its own sake, but rather how to run it so as to support the performance of the larger system.

For example, in [69] a MIP is combined with inventory considerations to model a production process that produces a number of different goods to satisfy external demands. The management objective is to minimise the average sum of inventory holding costs and shortage penalty costs. The two decisions that need to jointly be made are the inventory holding policy for the goods produced by each machine (given the possibility of machine downtime), and the assignment of servers to failed machines (given the remaining inventory levels of each product).

Another interesting idea is the combination of two interconnected MIP models. As mentioned earlier, the work in [18], [20], [21], [22], [23], and [24] deals with a client-server computing system in which there is both a software MIP where a CPU processes requests from terminals, and also a hardware MIP in which a human worker repairs terminals and the CPU.

Also related to computing systems is [39]. The basis of this model is a single MIP system in which the machines are heterogeneous CPUs that are repaired by a service worker when required. The complication is that these CPUs face a shared queue of processing jobs coming from an external Poisson source, and thus the goal is to assign the worker so as to maximize the performance of this open multi-server queue. The work in [142] is somewhat similar, but there each CPU faces its own external demand, so that each one is part of an open single-server queue.

The model in [50] has a single server covering two classes of demands: one class arrives from an external Poisson source, while the other one comes from n local machines. Thus it combines aspects of a finite source MIP queue with those of a conventional (infinite source) queue.

The computer simulation in [145] models a manufacturing plant in which the machinery is laid out in several MIP-style groups in parallel. This case study includes many other complicating features that reflect actual factory operations.

We believe that the idea of combining the MIP along with other issues in operations management has a great deal of promise and is worth exploring in future research. To draw a parallel, one might consider how the work on inventory theory has developed over the past decade due to its consideration of the "bigger picture" questions posed by supply chain management.

6. The evolution of MIP research

Since the review by Steck & Aronson [7] there have been some noticeable changes in MIP research. In the papers covered in that survey, both stochastic systems and deterministic systems were major categories of research; the works published since then almost all consider stochastic systems of some sort, though they vary in their stochastic nature (e.g. Markov versus General distributions). Purely deterministic models still appear but they are less common and are now often thought of as belonging more naturally with the scheduling literature rather than with that of the MIP.

Related to this increase in stochastic modelling is the continued growth of queuing theory and computer simulation to become the dominant methodologies for MIP research. Specific queuing approaches include the use of birth-death processes, integro-differential equations, embedded Markov chains, and matrix-geometric analysis (see e.g. [4,5]). For problems that do not lend themselves to exact analysis, diffusion approximations have become fairly common, while fluid approximations are more rarely used. The increased use of simulation for numerical work is partially attributable to advances in computer technology that make it relatively straightforward to perform. Simulation can be used as the main analysis method, but it is more commonly used along with some other approach, e.g. to verify the results of a diffusion approximation.

One new modelling feature that has appeared over the past two decades is the use of spare machines, and these were included in about one quarter of the papers we surveyed. Balking, reneging, and retrials by machines also appear in some studies now, whereas neither was mentioned previously. For servers, the main new feature is their potential unavailability due to vacations or breakdowns, and the option of adding or withdrawing them from reserve status. One model feature that has become less common in recent publications is the use of patrolling servers.

All of the above features can be thought of as enhancements within the MIP model itself. As already noted, some researchers have begun to explore beyond the traditional boundaries of the MIP by incorporating it as part of larger models of operational decisions. We believe that motivation for future MIP research could be generated by looking outside of the traditional set of "engineering-oriented" topics: i.e. beyond manufacturing, maintenance, and computer systems. Are there some interesting service operations "out there" in the real world that could benefit from the existing MIP research base while propelling it forward with new challenges? The use of novel case studies might be one way to initiate new research in these directions.

Acknowledgements

We thank the anonymous referees for their suggestions on improving this survey, in particular for pointing out a number of additional references for inclusion. We would also like to thank Jose Dahan for her help in gathering many of the papers that we surveyed here. This research has been supported in part by a Discovery Grant from Canada's Natural Sciences & Engineering Research Council (NSERC).

References

General References

[1] E. Berke, K. Moinzadeh, Analysis of maintenance policies for machines with deteriorating performance, IIE Transactions 32 (5) (2000) 433-445.

[2] D. Cho, M. Parlar, A survey of maintenance models for multi-unit systems, European Journal of Operational Research 51 (1991) 1-23.

[3] D. Gross, C.M. Harris, Fundamentals of Queueing Theory, 3rd Ed., John Wiley & Sons Inc., 1998.

[4] M.F. Neuts, Matrix-Geometric Solutions In Stochastic Models: An Algorithmic Approach, Johns Hopkins University Press, Baltimore, USA, 1981.

[5] M.F. Neuts, Structured Stochastic Matrices Of M/G/1 Type And Their Applications, Marcel Dekker, New York, USA, 1989.

[6] Y. Perlman, A. Mehrez, M. Kaspi, Setting repair policy in a multi-echelon repairable-item inventory system with limited repair capacity, Journal of the Operational Research Society 52 (2001) 198-209.

[7] K.E. Stecke, J.E. Aronson, Review of operator/machine interference models, International Journal of Production Research 23 (1985) 129-151.

[8] D. Stengos, L.C. Thomas, The blast furnaces problem, European Journal of Operational Research 4 (1980) 330-336.

[9] J. Sztrik, Finite source queueing systems and their applications: a bibliography. Working paper, Institute of Mathematics & Informatics, University of Debrecen, Debrecen, Hungary, 2002. Downloadable from web at: http://it.math.klte.hu/user/jsztrik/research/fsqreview.pdf.

[10] H. Takagi, Queueing Analysis: A Foundation Of Performance Evaluation, Volume 2: Finite Systems, North-Holland, Amsterdam, The Netherlands, 1993.

[11] J. Teghem, Control of the service process in a queuing system, European Journal of Operational Research 23 (1986) 141-158.

Machine Interference Problem Papers

[12] A.H. Abdekhodaee, A. Wirth, Scheduling parallel machines with a single server: some solvable cases and heuristics, Computers & Operations Research 29 (2002) 295-315.

[13] M.O. Abou El-Ata, An interarrival hyperexponential machine interference with balking and reneging, Microelectronics Reliability 32 (1992) 1251-1256.

[14] S.R. Agnihothri, Interrelationships between performance measures for the machine-repairman problem, Naval Research Logistics 36 (1989) 265-271.

[15] S. Akhtar, Reliability of k-out-of-n:G systems with imperfect fault coverage, IEEE Transactions on Reliability 43 (1) (1994) 101-106.

[16] S.C. Albright, Optimal maintenance-repair policies for the machine repair problem, Naval Research Logistics Quarterly 27 (1980) 17-27.

[17] A.S. Alfa, K.P.S. Isotupa, An M/PH/k retrial queue with finite number of sources, Computers & Operations Research 31 (2004) 1455-1464.

[18] B. Almasi, Response time for finite heterogeneous nonreliable queueing systems, Computers & Mathematics with Applications 31 (11) (1996) 55-59.

[19] B. Almasi, G. Bolch, J. Sztrik, Heterogeneous finite-source retrial queues, Journal of Mathematical Sciences 121 (5) (2004) 2590-2596.

[20] B. Almasi, J. Sztrik, A queueing model for a non-homogeneous terminal system subject to breakdowns, Computers & Mathematics with Applications 25 (4) (1993) 105-111.

[21] B. Almasi, J. Sztrik, A queueing model for a nonreliable multiterminal system with polling scheduling, Journal of Mathematical Sciences 92 (4) (1998) 3974-3981.

[22] B. Almasi, J. Sztrik, The effects of service disciplines on the performance of a nonreliable terminal system, Journal of Mathematical Sciences 92 (4) (1998) 3982-3989.

[23] B. Almasi, J. Sztrik, Optimization problems on the performance of a nonreliable terminal system, Computers & Mathematics with Applications 38 (1999) 13-21.

[24] B. Almasi, J. Sztrik, Reliability investigations of heterogeneous terminal systems using MOSEL, Journal of Mathematical Sciences 123 (1) (2004) 3795-3801.

[25] R.O. Al-Seedy, The service Erlangian machine interference with balking, Microelectronics Reliability 32 (5) (1992) 705-710.

[26] R.O. Al-Seedy, The truncated queue: M/M/2/m/m + Y with balking, spares, machine interference and an additional server for longer queues, Microelectronics Reliability 35 (1995) 1423-1427.

[27] R.O. Al-Seedy, F.M. Al-Ibraheem, An interarrival hyperexponential machine interference with balking, reneging, state-dependent, spares, and an additional server for longer queues, International Journal of Mathematics & Mathematical Science 27 (12) (2001) 737-749.

[28] M.J. Armstrong, Age repair policies for the machine repair problem, European Journal of Operational Research 138 (2002) 127-141.

[29] J.R. Artalejo, Retrial queues with a finite number of sources, Journal of the Korean Mathematical Society 35 (3) (1998) 503-526.

[30] A.A. Bahnasawi, M.S. Mahmoud, S.Z. Eid, Sensitivity analysis of machine interference in manufacturing systems, Computers & Industrial Engineering 30 (1996) 753-764.

[31] A.A. Bahnasawi, M.S. Mahmoud, S.Z. Eid, Priority analysis of patrolled machines, Simulation Practice & Theory 5 (1997) 535-552.

[32] B.D. Bunday, H.M. Bokhari, The efficiency of two groups of heterogeneous stations cared for by r operatives with a priority group, Applied Mathematical Modelling 21 (1997) 42-47.

[33] B.D. Bunday, W.K. El-Badri, A model for a textile winding process, European Journal of Operational Research 15 (1984) 55-62.

[34] B.D. Bunday, W.K. El-Badri, The busy period for the M/M/1 machine interference model, Stochastic Analysis & Applications 3 (1) (1985) 1-13.

[35] B.D. Bunday, W.K. El-Badri, Machine interference - a comparison of two patrolling methods, International Journal of Production Research 23 (4) (1985) 795-805.

[36] B.D. Bunday, W.K. El-Badri, Machine Interference - A comparison of two repair strategies, International Journal of Production Research 23 (6) (1985) 1115-1120.

[37] B.D. Bunday, E. Khorram, A closed form solution for the G/M/r machine interference model, International Journal of Production Research 26 (1988) 1811-1818.

[38] D.G. Carmichael, Machine interference with general repair and running times, Zeitschrift Operations Research 31 (1987) B115-B133.

[39] R. Chakka, I. Mitrani, Heterogeneous multiprocessor systems with breakdowns: performance and optimal repair strategies, Theoretical Computer Science 125 (1994) 91-109.

[40] S.R. Chakravarthy, A. Agarwal, Analysis of a machine repair problem with an unreliable server and phase type repairs and services, Naval Research Logistics 50 (2003) 462-480.

[41] S.R. Chakravarthy, A. Krishnamoorthy, P.V. Ushakumari, A k-out-of-n reliability system with an unreliable server and phase type repairs and services: the (n,t) policy, Journal of Applied Mathematics & Stochastic Analysis 14 (4) (2001) 361-380.

[42] M.J. Chandra, A study of multiple finite source queuing models, Journal of the Operational Research Society 37 (1986) 275-283.

[43] M.J. Chandra, R.G. Sargent, A numerical method to obtain the equilibrium results for the multiple finite source priority queueing systems, Management Science 29 (1983) 1298-1308.

[44] K. Cheng, Y.L. Zhang, Analysis of a consecutive k-out-of-n:F repairable system with priority in repair, International Journal of Systems Science 32 (5) (2001) 591-598.

[45] W.K. Ching, Machine repairing models for production systems, International Journal of Production Economics 70 (2001) 257-266.

[46] T.B. Crabhill, Optimal control of a maintenance system with variable service rates, Operations Research 22 (1974) 736-745.

[47] T.K. Das, M.A. Wortman, Performance of N machine centers of k-out-of-m:G type maintained by a single repairman, Naval Research Logistics 39 (1992) 919-936.

[48] T.K. Das, M.A. Wortman, Analysis of asymmetric patrolling repairman systems, European Journal of Operational Research 64 (1993) 45-60.

[49] P. Desreuelle, H.J. Steudel, A queueing network model of a single operator manufacturing workcell with machine/operator interference, Management Science 42 (1996) 576-590.

[50] S. Drekic, W.K. Grassmann, An eigenvalue approach to analyzing a finite source priority queueing model, Annals of Operations Research 112 (2002) 139-152.

[51] G.I. Falin, A multiserver retrial queue with a finite number of sources of primary calls, Mathematical & Computer Modelling 30 (1999) 33-49.

[52] G.I. Falin, J.R. Artalejo, A finite source retrial queue, European Journal of Operational Research 108 (1998) 409-424.

[53] A.E. Ferdinand, An analysis of the machine interference model, IBM Systems Journal 10 (2) (1971) 129-142.

[54] E. Frostig, Optimal policies for machine repairmen problems, Journal of Applied Probability 30 (1993) 703-715.

[55] E. Frostig, Jointly optimal allocation of a repairman and optimal control of service rate for machine repairman problem, European Journal of Operational Research 116 (1999) 274-280.

[56] D.P. Gaver, J.A. Morrison, R. Silveira, Service adaptive multitype repairman problems, SIAM Journal of Applied Mathematics 53 (2) (1993) 459-470.

[57] L.C. Goheen, On the optimal operating policy for the machine repair problem when failure and repair times have Erlang distributions, Operations Research 25 (1977) 484-492.

[58] M.N. Gopalan, N. Anantharaman, Stochastic analysis of machine interference of a production system with identical machines, Stochastic Analysis & Applications 9 (1) (1991) 41-52.

[59] D. Gross, H.D. Kahn, J.D. Marsh, Queueing models for spares provisioning, Naval Research Logistics Quarterly 24 (4) (1977) 521-536.

[60] S.M. Gupta, Interrelationship between queueing models with balking and reneging and machine repair problem with warm spares, Microelectronics Reliability 34 (2) (1994) 201-209.

[61] S.M. Gupta, Machine interference problem with warm spares, server vacations and exhaustive service, Performance Evaluation 29 (1997) 195-211.

[62] U.C. Gupta, T.S.S. Srinivasa Rao, A recursive method to compute the steady state probabilities of the machine interference model: (M/G/1)/K, Computers & Operations Research 21 (6) (1994) 597-605.

[63] U.C. Gupta, T.S.S. Srinivasa Rao, Computing steady state probabilities in $\lambda(n)/G/1/K$ queue, Performance Evaluation 24 (1996) 265-275.

[64] U.C. Gupta, T.S.S. Srinivasa Rao, On the M/G/1 machine interference model with spares, European Journal of Operational Research 89 (1996) 164-171.

[65] Haryono, B.D. Sivazlian, Analysis of the machine repair problem: a diffusion process approach, Mathematics & Computers in Simulation 27 (1985) 339-364.

[66] J.E. Hilliard, An approach to cost analysis of maintenance float systems, AIIE Transactions 8 (1) (1976) 128-133.

[67] Y-C. Hsieh, Optimal assignment of priorities for the machine interference problems, Microelectronics Reliability 37 (1997) 635-640.

[68] Y-C. Hsieh, K-H. Wang, Reliability of a repairable system with spares and a removable repairman, Microelectronics Reliability 35 (2) (1995) 197-208.

[69] S.M.R. Iravani, I. Duenyas, T.L. Olsen, A production/inventory system subject to failure with limited repair capacity, Operations Research 48 (2000) 951-964.

[70] S.M.R. Iravani, B. Kolfal, When does the cμ rule apply to finite population queueing systems?, Operations Research Letters 33 (2005) 301-304.

[71] M. Jain, Diffusion equation for $GI^x/G/r$ machine interference problem with spare machines, Microelectronics Reliability 33 (1993) 1415-1418.

[72] M. Jain, An (m, M) machine repair problem with spares and state dependent rates: a diffusion process approach, Microelectronics Reliability 37 (1997) 929-933.

[73] M. Jain, Rakhee, S. Maheshwari, N-policy for a machine repair system with spares and reneging, Applied Mathematical Modelling 28 (2004) 513-531.

[74] M. Jain, G.C. Sharma, S. Moses, The M/M/C interdependent machining system with mixed spares and controllable rates of failure and repair, Journal of the Rajasthan Academy of Physical Science 3 (2) (2004) 91-100.

[75] M. Jain, G.C. Sharma, M. Singh, Diffusions process for multi-repairmen machining system with spares and balking, International Journal of Engineering Transactions B: Applications 15 (1) (2002), 49-56.

[76] M. Jain, G.C. Sharma, M. Singh, G/G/r machine repair problem with spares and additional repairman, International Journal of Engineering Transactions B: Applications, 15 (1) (2002) 57-62.

[77] M. Jain, G.C. Sharma, M. Singh, M/M/R machine interference model with balking, reneging, spares and two modes of failure, OPSEARCH 40 (1) (2003) 24-41.

[78] H. Kameda, A finite-source queue with different customers, Journal of the Association for Computing Machinery 29 (2) (1982) 478-491.

[79] K. Karmeshu, Queueing models with threshold phenomena: with special reference to machine interference problems, Sadhana 15 (4&5) (1990) 319-330.

[80] K. Karmeshu, N.K. Jaiswal, A non-linear stochastic model for the machine interference problem, International Journal of Systems Science 12 (1981) 293-303.

[81] J.C. Ke, K.H. Wang, Cost analysis of the m/m/r machine repair problem with balking, reneging, and server breakdowns, Journal of the Operational Research Society 50 (1999) 275-282.

[82] C. Knessl, On the transient behavior of the repairman problem, Advances in Applied Probability 23 (1991) 327-354.

[83] C. Knessl, On the distribution of the maximum number of broken machines for the repairman problem, SIAM Journal of Applied Mathematics 54 (2) (1994) 508-547.

[84] C.P. Koulamas, A bi-criterion algorithm for minimizing machine interference and reducing job waiting time, International Journal of Systems Science 23 (7) (1992) 1229-1235.

[85] C.P. Koulamas, Scheduling two parallel semiautomatic machines to minimize machine interference, Computers & Operations Research 23 (1996) 945-956.

[86] C.P. Koulamas, M.L. Smith, Look-ahead scheduling for minimizing machine interference, International Journal of Production Research 26 (1988) 1523-1533.

[87] A. Krishnamoorthy, P.V. Ushakumari, K-out-of-n:G system with repair: the D-policy, Computers & Operations Research 28 (2001) 973-981.

[88] N. Kryvinska, Intelligent network analysis by closed queuing models, Telecommunication Systems 27 (1) (2004) 85-98.

[89] Y. Lam, H.K.T. Ng, A general model for consecutive-k-out-of-n: F repairable system with exponential distribution and (k-1)-step Markov dependence, European Journal of Operational Research 129 (2001) 663-682.

[90] Y. Lam, Y.L. Zhang, Repairable consecutive-k-out-of-n: F system with Markov dependence, Naval Research Logistics 47 (2000) 18-39.

[91] C. Langaris, A. Katsaros, An M/G/1 queue with finite population and gated service discipline, Journal of the Operations Research Society of Japan 40 (1) (1997) 133-139.

[92] A.A. Lazar, Optimal flow control of a class of queueing networks in equilibrium, IEEE Transactions on Automatics Control 28 (11) (1983) 1001-1007.

[93] H-W. Lee, S-H. Yoon, S-S. Lee, Continuous approximations of the machine repair system, Applied Mathematical Modelling 19 (1995) 550-559.

[94] T. Lehtonen, On the optimal policies of an exponential machine repairman problem. Naval Research Logistics Quarterly 3 (1) (1984) 173-181.

[95] H. Li, T. Yang, A single server retrial queue with server vacations and a finite number of input sources, European Journal of Operational Research 85 (1995) 149-160.

[96] B. Liu, J. Cao, A machine service model with a service station consisting of r unreliable units, Microelectronics Reliability 35 (4) (1995) 683-690.

[97] M. Mittler, C. Kern, Discrete-time approximation of the machine interference problem with generally distributed failure, repair, and walking times, European Journal of Control 3 (1997) 254-267.

[98] H. Ohmura, Y. Takahashi, An analysis of repeated call model with a finite number of sources, Electronics & Communications in Japan part 1 68 (6) (1985) 112-121.

[99] J. Palesano, J. Chandra, A machine interference problem with multiple types of failures, International Journal of Production Research 24 (3) (1986) 567-582.

[100] A. Posafalvi, J. Sztrik, On the heterogeneous machine interference problem with priority and ordinary machines, European Journal of Operational Research 41 (1989) 54-63.

[101] A. Posafalvi, J. Sztrik, A numerical approach to the repairman problem with two different types of machine, Journal of the Operational Research Society 40 (9) (1989) 797-803.

[102] R. Righter, Optimal policies for scheduling repairs and allocating heterogeneous servers, Journal of Applied Probability 33 (1996) 536-547.

[103] M.D. Rossetti, G. Clark, Evaluating a queueing approximation for the machine interference problem with two types of stoppages via simulation optimization, Computers & Industrial Engineering 34 (1998) 655-668.

[104] K.C. Seal, Spreadsheet simulation of a queue with arrivals from a finite population - the machine repair problem, International Journal of Operations & Production Management 15 (6) (1995) 84-100.

[105] J. Shao, L.R Lamberson, Modeling a shared load k-out-of-n:G system, IEEE Transactions on Reliability 40 (2) (1991) 205-209.

[106] O.P. Sharma, J. Dass, Transient analysis of a M/M/r machine interference model, Stochastic Analysis & Applications 6 (2) (1988) 205-216.

[107] A.I. Shawky, The single-server machine interference model with balking, reneging and an additional server for longer queues, Microelectronics Reliability 37 (2) (1997) 335-357.

[108] A.I. Shawky, The machine interference model: M/M/C/K/N with balking reneging and spare, OPSEARCH 36 (1) (2000) 25-35.

[109] A.I. Shawky, An interarrival hyperexponential machine interference model: Hr/M/c/k/N with balking and reneging, Communications of the Korean Mathematical Society 16 (4) (2001) 659-666.

[110] S.K. Singh, G.C. Sharma, Markovian analysis of the $M^x/E_k/1$ machine repair problem with spares, Microelectronics Reliability 33 (1993) 493-495.

[111] B.D. Sivazlian, K.H. Wang, Economic analysis of the M/M/R machine repair problem with warm standbys, Microelectronics Reliability 29 (1) (1989) 25-35.

[112] B.D. Sivazlian, K.H. Wang, System characteristics and economic analysis of the G/G/R machine repair problem with warm standbys using diffusion approximation, Microelectronics Reliability 29 (5) (1989) 829-848.

[113] B.D. Sivazlian, K.H. Wang, Diffusion approximation to the G/G/R machine repair problem with warm standby spares, Naval Research Logistics 37 (1990) 753-772.

[114] T.S.S. Srinivasa Rao, U.C. Gupta, Performance modeling of the M/G/1 machine repair problem with cold-, warm-, and hot-standbys, Computers & Industrial Engineering 38 (2000) 251-267.

[115] J. Sztrik, On the finite source G/M/r queue, European Journal of Operational Research 20 (1985) 261-268.

[116] J. Sztrik, The G/M/r/FIFO machine interference model with state-dependent speeds, Journal of the Operational Research Society 39 (2) (1988) 201-207.

[117] J. Sztrik, The G/M/r SIRO machine interference model with state-dependent speeds, Serdica 16 (1990) 210-216.

[118] J. Sztrik, Asymptotic analysis of the reliability of a complex renewable standby system with fast repair, Theory of Probability & Applications 37 (1) (1992) 101-104.

[119] J. Sztrik, Markov-modulated finite source queueing models and their applications, Journal of Mathematical Sciences 111 (6) (2002) 3895-3900.

[120] J. Sztrik, B.D. Bunday, Asymptotic analysis of the heterogeneous machine interference problem with random environments, Applied Mathematical Modelling 17 (1993) 105-110.

[121] J. Sztrik, B.D. Bunday, Machine interference problem with a random environment, European Journal of Operational Research 65 (1993) 259-269.

[122] J. Sztrik, C.S. Kim, Markov-modulated finite-source queueing models in evaluation of computer and communication systems, Mathematical & Computer Modelling 38 (2003) 961-968.

[123] J. Sztrik, O. Möller, Stochastic simulation of Markov-modulated finite-source queueing systems, Journal of Mathematical Sciences 105 (6) (2001) 2615-2625.

[124] J. Sztrik, O. Möller, Simulation of machine interference in randomly changing environments, Yugoslav Journal of Operations Research 12 (2) (2002) 1-10.

[125] H. Takagi, Analysis of an M/G/1//N queue with multiple server vacations and its application to a polling model, Journal of the Operations Research Society of Japan 35 (1992) 300-315.

[126] P. Tosirisuk, M.J. Chandra, Multiple finite source queueing model with dynamic priority scheduling, Naval Research Logistics 37 (1990) 365-381.

[127] P. Tosirisuk, M.J. Chandra, An iterative algorithm for a multiple finite-source queueing model with dynamic priority scheduling, Journal of the Operational Research Society 46 (1995) 905-912.

[128] P.V. Ushakumari, A. Krishnamoorthy, k-out-of-n system with repair: the max(N,T) policy, Performance Evaluation 57 (2004) 221-234.

[129] F. Van Der Duyn Schouten, P. Wartenhorst, A two machine repair model with variable repair rate, Naval Research Logistics 40 (1993) 495-523.

[130] K-H. Wang, Profit analysis of the machine repair problem with a single service station subject to breakdowns, Journal of the Operational Research Society 41 (1990) 1153-1160.

[131] K-H. Wang, Cost analysis of the M/M/R machine repair-problem with mixed standby spares, Microelectronics Reliability 33 (1993) 1293-1301.

[132] K-H. Wang, Profit analysis of the machine repair problem with cold standbys and two modes of failure, Microelectronics Reliability 34 (10) (1994) 1635-1642.

[133] K-H. Wang, Profit analysis of the M/M/R machine repair problem with spares and server breakdowns, Journal of the Operational Research Society 45 (1994) 539-548.

[134] K-H. Wang, Comparative analysis of the $M/E_k/1$ machine repair problem with spares, Computers & Industrial Engineering 26 (1994) 765-774.

[135] K-H. Wang, An approach to cost analysis of the machine repair problem with two types of spares and service rates, Microelectronics Reliability 35 (1995) 1433-1436.

[136] K-H. Wang, L.Y Hsu, Cost analysis of the machine repair problem with r non-reliable service stations, Microelectronics Reliability 35 (6) (1995) 923-934.

[137] K-H. Wang, M-Y. Kuo, Profit analysis of the $M/E_k/1$ machine repair problem with a non-reliable service station, Computers & Industrial Engineering 32 (1997) 587-594.

[138] K-H. Wang, H-C. Lee, Cost analysis of the cold-standby M/M/R machine repair problem with multiple modes of failure, Microelectronics Reliability 38 (1998) 435-441.

[139] K-H. Wang, B.D. Sivazlian, Comparative analysis for the G/G/R machine repair problem, Computers & Industrial Engineering 18 (1990) 511-520.

[140] K-H. Wang, B.D. Sivazlian, Cost analysis of the M/M/R machine repair problem with spares operating under variable service rates, Microelectronics Reliability 32 (1992) 1171-1183.

[141] K-H. Wang, J-D. Wu, Cost analysis of the M/M/R machine repair problem with spares and two modes of failure, Journal of the Operational Research Society 46 (1995) 783-790.

[142] P. Wartenhorst, N parallel queueing systems with server breakdown and repair, European Journal of Operational Research 82 (1995) 302-322.

[143] R.C.M. Yam, M.J. Zuo, Y.L. Zhang, A method for evaluation of reliability indices for repairable circular consecutive k-out-of-n:F systems, Reliability Engineering & System Safety 79 (2003) 1-9.

[144] M. Yamashiro, Y. Yuasa, Repair system where the repairmen change depending on the failed machines, Microelectronics Reliability 36 (1996) 231-234.

[145] T. Yang, H-P. Fu, C. Yang, A simulation-based dynamic operator assignment strategy considering machine interference – a case study on integrated circuit chip moulding operations, Production Planning & Control 13 (6) (2002) 541-551.

[146] M.A. Younis, A.F. Hamed, A stochastic approach to machine interference problems in robot-served manufacturing systems, International Journal of Systems Science 28 (6) (1997) 563-570.

Figure 1. A MIP system of *n* machines and *r* servers.

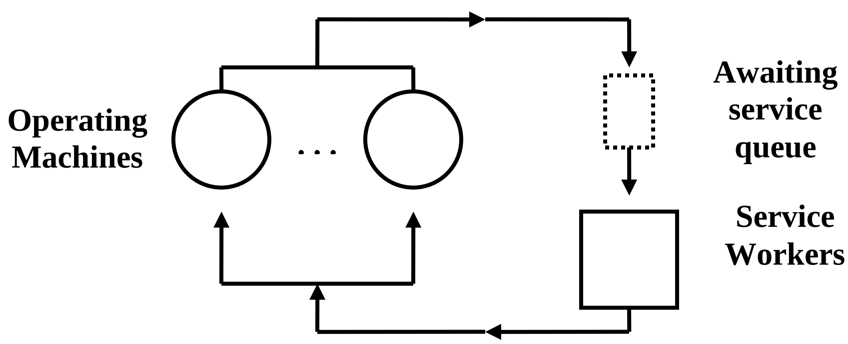

Table 1. Basic MIP queuing structures.

M/M/1:

15, 18, 19, 20, 21, 22, 23, 24, 30, 31, 34, 39, 44, 46, 50, 53, 55, 56, 61, 67, 68, 70, 73, 78, 79, 82, 87, 89, 90, 92, 94, 100, 101, 102, 107, 119, 121, 122, 128, 130, 143, 144, 146

M/M/r:

16, 26, 51, 54, 59, 60, 66, 69, 74, 77, 80, 81, 93, 103, 105, 106, 108, 111, 118, 120, 131, 132, 133, 135, 136, 138, 140, 141, 142

PH, E, M, Hyper, Hypo, with 1 server:

25, 27, 38, 40, 41, 42, 99, 123, 124, 126, 127, 134, 137

PH, E, M, Hyper, Hypo, with r servers:

13, 17, 45, 57, 109

M/G/1:

43, 48, 49, 52, 58, 62, 63, 64, 83, 88, 91, 95, 96, 98, 104, 114, 125

M/G/r:

29

G/M/r:

28, 32, 37, 115, 116, 117

G/G/1:

97, 129

G/G/r:

14, 65, 72, 75, 76, 112, 113, 139

M/D/1:

33, 35, 36

D/D/1:

12, 84, 85, 86

Other:

47, 71, 110, 145