# SOME EXTENSIONS TO RELIABILITY MODELING AND OPTIMIZATION OF NETWORKED SYSTEMS

**PENG RUI**

(*B.Sc., University of Science and Technology of China*)

**A THESIS SUBMITTED**
**FOR THE DEGREE OF DOCTOR OF PHILOSOPHY**
**DEPARTMENT OF INDUSTRIAL & SYSTEMS ENGINEERING**
**NATIONAL UNIVERISITY OF SINGAPORE**
**2011**

# ACKNOWLEDGEMENTS

First I would like to thank my main supervisor Prof. Xie Min for his patient guidance and enthusiastic assistance during my whole Ph. D candidature. His suggestions and encouragement helped me overcome my fear when I felt uncertain and braced me up when I stumbled. He taught me many things that will benefit my entire life. I am also deeply indebted to my co-supervisor Dr. Ng Szu Hui for her patient help and warmhearted advices. Without their great help, this dissertation is impossible.

I am grateful to Department of Industrial and Systems Engineering for its nice facilities. I would like to thank Prof. Poh and Dr. Kim for attending my oral qualifying examination and giving constructive comments on my research and thesis writing. I wish to thank Prof. Goh for his suggestions on giving tutorials. I also owe a lot to Ms. Ow Lai Chun, Ms. Tan Ai Hua and the ISE computing lab technician Mr. Cheo for their great technical support.

I would like to thank Dr. Hu Qingpei for his suggestions and collaboration. I would also like to express my sincere gratitude to Long Quan and Li Yanfu for their advices and encouragements. I would like to thank my friends Li Xiang, Wu Jun, Xie Yujuan, Xiong Chengjie, Yao Zhishuang, Yin Jun, Jiang Jun, Ren Xiangyao and Ye Zhisheng for their friendship.

I would like to thank Dr. Levitin from the Electrical Corporation of Israel. I learnt a lot from our discussion and cooperation.

At last I present my full regards to my parents and my sister for their love and support. They have brought me a lot of joy and strength.

# TABLE OF CONTENTS

# SUMMARY

The purpose of this thesis is to model the reliability of some networked systems and study the related optimization problems. The reliability of a system is usually dependent on the structure of the system and the resources spent on the maintenance and protection of the system. Appropriate configuration of system structure and allocation of different kinds of resources are effective measures to increase system reliability and reduce the cost.

In many critical applications, fault tolerance has been an essential architectural attribute for achieving high reliability. However, faults in some elements of the system can remain undetected and uncovered, which can lead to the failure of the total system or its subsystem. As a result, the system reliability could decrease with the increase of redundancy over some particular limit if the system is subjected to imperfect fault coverage. Therefore the optimal system structure problem arises. The optimal structure of multi-state series-parallel systems with consideration of different kinds of imperfect fault coverage is studied. The linear multi-state consecutively connected system (LMCCS) is important in signal transmission and other network systems. The reliability of LMCCS has been studied in the past restricted to the case when each system element is associated with a constant reliability. In practice, a system usually contains elements with increasing failure rates and the availabilities of system elements are dependent on the maintenance actions taken. Different from existing works, the optimal component allocation and maintenance strategy in a linear multistate consecutively connected system is studied.

Besides system with internal failures, this dissertation also studies the defense of system subjected to external attacks. For systems under external intentional attacks, protecting system elements and deploying false targets are two measures for system reliability enhancement. The protection is a technical or organizational measure which is aimed to reduce the vulnerability of protected system elements. The objective of a false target is to distract the attacker so that genuine elements are harder to locate. Existing papers have studied the efficiency of perfect false targets which are restricted. To move towards reality, system defense with imperfect false targets is studied. One work studies the defense of simple series and parallel systems with imperfect false targets. It is assumed that the detection probability of each false target is a constant. Another work studies the defense of a single object with imperfect false targets by assuming that the detection probability is a function of the attacker's intelligence effort and the defender's disinformation effort. For systems subjected to both internal failures and external impacts, maintenance and protection are two measures intended to enhance system availability. A tradeoff exists between investments into system maintenance and its protection. This dissertation proposes a framework to study the optimal maintenance and protection strategy for series-parallel systems. The methodology used can be extrapolated to study the protection and maintenance of other networked systems.

# LIST OF TABLES

# LIST OF FIGURES

# NOMENCLATURE

| | |
|---|---|
| IFC | imperfect fault coverage |
| ELC | element level coverage |
| FLC | fault level coverage |
| PDC | performance dependent coverage |
| MSS | multi-state system |
| UGF | universal generating function (or u-function) |
| GA | genetic algorithm |
| BIT | built-in test |
| WSG | work sharing group (group of elements affected by uncovered failures) |
| LMCCS | linear multi-state consecutively connected systems |
| GE | genuine element |
| FT | false target |
| $\Phi_m$ | set of elements belonging to subsystem $m$ |
| $\Phi_{mi}$ | set of elements belonging to the $i$-th WSG of subsystem $m$ |
| $c_m(\lvert\Phi_{mi}\rvert, j)$ | the fault coverage probability in the case of $j$-th failure in WSG $i$ in subsystem $m$ |

| | |
|---|---|
| $r_{mi}(k)$ | the probability that WSG $i$ in subsystem $m$ does not fail after $k$ failures have consecutively occurred |
| $l_{mi}(g)$ | the fault coverage probability of WSG $i$ in subsystem $m$ in the case that the entire group performance is $g$ |
| $\lambda_i(t)$ | expected number of failures of the element $i$ during time interval $(0,t]$ |
| $\Lambda$ | total number of considered replacement frequency alternatives |
| 1(x) | unity function: 1(TRUE)=1, 1(FALSE)=0 |
| $\lfloor y \rfloor$ | the greatest integer not greater than $y$ |

# CHAPTER 1 INTRODUCTION

This dissertation focuses on the reliability modeling and optimization of some networked systems. The reliability of a system is usually dependent on the structure of the system and the resources spent on the maintenance and protection of the system. Different kinds of networked systems are investigated in this dissertation, which involve series-parallel systems with imperfect fault coverage, linear multi-state consecutively connected systems comprising of elements with increasing failure rates, simple series and parallel systems exposed to external intentional attacks, and series-parallel systems subjected to both internal failures and external attacks.

The organization of this chapter is as follows. The introductory part first provides the background in section 1.1, and then states the motivation of research in section 1.2. Section 1.3 presents some important techniques used, which include universal generating function and genetic algorithm. The research scope and the organization of this dissertation are given in section 1.4.

## 1.1. Background

Reliability is the probability that a system will perform satisfactorily for at least a given period of time when used under stated condition. It is an important measure of how well a system meets its design objective. As many of today's systems are large and complicated, the reliability analysis of such systems has drawn much attention, see Cook and Ramirez-Marquez (2009), Yeh and Lin (2009) and Huang and Xu (2010).

A system is a collection of independent and interrelated components connected as a unity to perform some specified functions. System reliability is usually evaluated by reliability block diagram, which is a graphic representation of the logic connections of system components within a system. Some common networked systems are single component systems, series systems, parallel systems, series-parallel systems, parallel-series systems, and *k*-out-of-*n* partially redundant systems. Series and parallel are the two basic elements of logic connections, from which more complicated configurations can be formed.

A system is said to be a series system if the failure of any element results in the failure of the entire system. In other words, a series system functions only when all the elements function. The reliability of a series system is the product of the reliabilities of all the components within the system. For this reason the system reliability is no more than the reliability of any component. And the system reliability decreases drastically with the increase of the number of components.

A system is said to be a parallel system if the system manages to work if at least one element is operational. The unreliability, one minus reliability, of a parallel system is the product of the unreliabilities of all the components. In contrast to a series system, the reliability of a parallel system increases with the number of components within the system. Thus parallel configuration is usually implemented in safety-critical systems such as aircraft and spaceships. However, parallel configuration is often restricted by other factors, such as cost and weight constraints.

There are situations in which series and parallel configurations are mixed in a system design to achieve functional and reliability requirements. The combinations form series-parallel and parallel-series configurations. A series-parallel system is comprised of $n$ subsystems in series with $m_i$ ($i$=1,…,$n$) components in parallel in subsystem $i$. The configuration is sometimes called the low-level redundancy design. A parallel-series system is comprised of $m$ subsystems in parallel with $n_i$ ($i$=1,…,$m$) components in series in subsystem $i$. The configuration is sometimes called the high-level redundancy design.

A $k$-out-of-$n$ system is a partially redundant system, which succeeds if and only if at least $k$ ($1 \le k \le n$) out of $n$ components function. A series system can be regarded as an $n$-out-of-$n$ system whereas a parallel system can be regarded as a 1-out-of-$n$ system. This kind of $k$-out-of-$n$ systems is also noted as $k$-out-of-$n$: G systems, where G stands for "good". To the contrary, a $k$-out-of-$n$: F system, where F stands for "failure", fails if and only if at least $k$ components out of $n$ components fail. The reliability of $k$-out-of-$n$ systems has been studied in many papers, such as Ding et al. (2010), Tian et al. (2009), and Chakravarthy and Gómez-Corral (2009).

As a kind of generalized *k*-out-of-*n* systems, the reliability of the consecutive-*k*-out-of-*n*: F system has aroused a lot of attention, see Pekoz and Ross (1995) and Cluzeau et al. (2008). The usual definition of a consecutive-*k*-out-of-*n*: F system is a line of *n* components where the system fails if and only if any *k* consecutive components fail. One way to interpret such a system is to add a component 0 (source) and a component *n*+1 (sink) to the system and that each component, if working, is directly connected to the subsequent *k* components (or all remaining components if the number is less than *k*), and that the source and sink always work. The system works if and only if a flow can be sent from the source to the sink. A consecutive-*k*-out-of-*n*: F system can be either a linear system or a circular system, depending on whether the components are arranged in a line or on a circle.

## 1.2. Motivation

### 1.2.1. Imperfect fault coverage

In many critical applications, fault tolerance has been an essential architectural attribute for achieving high reliability (Lee and Na, 2009; Perhinschi et al. 2006; Tian et al. 2008). However, faults in some elements of the system can remain undetected and uncovered, which can lead to the failure of the total system or its subsystem (Amari et al., 2004; Xing, 2007; Myers 2008). The optimal work sharing structure of a multi-state series-parallel system has been studied in Levitin (2008) with the incorporation of imperfect fault

coverage. The coverage model considered in Levitin (2008) applies only to element level coverage (ELC), that is, a particular fault coverage probability is associated with each element. In practice, there are different kinds of fault coverage models corresponding to different fault coverage techniques used. In order to adapt to different situations, we have studied the optimal work sharing structure problem with consideration of different kinds of fault coverage mechanisms.

### 1.2.2. Linear multi-state consecutively connected systems

The linear multi-state consecutively connected system (LMCCS) is important in signal transmission and other network systems. The system consists of $N$+1 linearly ordered positions (nodes). Each node can provide a connection between its position and the next few positions. The system fails if the first node (source) is not connected with the final node (sink). The reliability of LMCCS has been studied in the past restricted to the case when each system element is associated with a constant reliability (Malinowski and Preuss 1996; Levitin 2003). In practice, a system usually contains elements with increasing failure rates and the availabilities of system elements are dependent on the maintenance actions taken (Lisnianski et al., 2008; Ding et al., 2009; Rao and Naikan, 2009). Different from existing works, we have studied the combined optimal maintenance and allocation strategy of the elements in LMCCS which minimizes the system maintenance cost restricted by a pre-specified system availability requirement.

### 1.2.3. Defending systems against intentional attacks

Protecting against intentional impacts is fundamentally different from protecting against unintentional impacts, such as naturally occurring events or technological accidents. Adaptive strategy allows the attacker to target the most sensitive part of a system. Thus it is important for the defender to take into account the attacker's strategy when it decides how to allocate its resources among several defensive measures (Azaiez and Bier, 2007; Dighe et al., 2009; Powell, 2007a; Powell, 2007b). For systems against intentional attacks, protecting system elements and deploying false targets are two important measures for system reliability enhancement.

The efficiency of false targets in defense strategy has been studied in Levitin and Hausken (2009a), which assumes the attacker cannot distinguish the genuine object from the false targets. In practice the false targets are after all different from the genuine object, and they are possible to be detected by the attacker. Different from Levitin and Hausken (2009a), we assume that there is a probability that a false target can be detected by the attacker. The detection probability of a false target is assumed to be either a constant or a function of the attacker's intelligence effort and the defender's disinformation effort. Frameworks of solving the optimal defense strategy are proposed.

### 1.2.4. Optimal replacement and protection strategy

Many systems contain elements with increasing failure rates and the availabilities of the system elements are dependent on the maintenance actions taken (Lisnianski et al., 2008;

Ding et al., 2009; Rao and Naikan, 2009). For systems containing elements with increasing failure rates, preventive replacement of the elements is an efficient measure to increase the system availability (Levitin and Lisnianski, 1999). Besides internal failures, an element may also fail due to external impacts, say, natural disasters (Zhuang and Bier, 2007). In order to increase the survivability of a system element under external impacts, defensive investments can be made to protect the system element. A tradeoff exists between investments into the maintenance and the protection of system elements. For multistate systems, the system availability is a measure of the system's ability to meet the demand (required performance level). In order to provide the required availability with minimum cost, the optimal maintenance and protection strategy is studied.

## 1.3. Some important techniques

### 1.3.1. Universal generating function

The universal generating function (also called u-function or UGF) representing the pmf of a discrete random variable $X$ is defined as a polynomial

$$u_X(z) = \sum_{h=0}^{H} \varepsilon_h z^{x_h}, \tag{1.1}$$

where the variable $X$ has $H$+1 possible values and $\varepsilon_h = \Pr\{X = x_h\}$.

To obtain the UGF representing the pmf of a function of two independent random variables $\varphi(X,Y)$ the following composition operator is used:

$$U_{\varphi(X,Y)}(z) = u_X(z) \underset{\varphi}{\otimes} u_Y(z) = \left(\sum_{h=0}^{H} \varepsilon_h z^{x_h}\right) \underset{\varphi}{\otimes} \left(\sum_{d=0}^{D} \varepsilon_d z^{y_d}\right) = \sum_{h=0}^{H} \sum_{d=0}^{D} \varepsilon_h \varepsilon_d z^{\varphi(x_h, y_d)} \qquad (1.2)$$

The polynomial U(z) represents all of the possible mutually exclusive combinations of realizations of the variables by relating the probabilities of each combination to the value of function $\varphi(X,Y)$ for this combination.

The UGF is a convenient tool for evaluating the reliability and performance of multi-state systems (MSS). In the case of MSS, UGF

$$u_j(z) = \sum_{h=0}^{k_j} p_{jh} z^{g_{jh}}, \qquad (1.3)$$

represent the pmf of random performances of system elements $(g_j, p_j)$. If, for any pair of elements connected in series or in parallel, their cumulative performance is defined as a function of individual performances of the elements, then the pmf of the entire system performance can be obtained using the following recursive procedure (Levitin, 2005).

1) Find any pair of system elements (*i* and *j*) connected in parallel or in series in the MSS.

2) Obtain the UGF of the pair using the corresponding composition operator $\underset{\varphi}{\otimes}$ over two UGF of the elements:

$$U_{\{i,j\}}(z) = u_i(z) \underset{\varphi}{\otimes} u_j(z) = \sum_{h=0}^{k_i} \sum_{d=0}^{k_j} p_{ih} p_{jd} z^{\varphi(g_{ih}, g_{jd})}, \quad (1.4)$$

where the function $\varphi$ is determined by the nature of interaction between the elements' performances.

3) Replace the pair with a single element which has the UGF obtained in step 2.

4) If the MSS contains more than one element, return to step 1.

### 1.3.2. Genetic algorithm

In many optimization problems, the solution space is too large that an exhaustive examination of all possible solutions is not realistic, considering reasonable time limitations. As in most combinatorial optimization problems, the quality of a given solution is the only information available during the search for the optimal solution. Therefore, a heuristic search algorithm is needed, which uses only estimates of solution quality, and which does not require derivative information to determine the next direction of the search.

The genetic algorithm (GA) has proven to be an effective optimization tool for a large number of complicated problems in reliability engineering (Coit and Smith, 1996, Levitin et al., 1998; Huang et al., 2009). Basic notions of GA are originally inspired by biological genetics. GA operates with "chromosomal" representation of solutions, where crossover, mutation, and selection procedures are applied. Unlike various constructive optimization

algorithms that use sophisticated methods to obtain a good singular solution, the GA deals with a set of solutions (population), and tends to manipulate each solution in the simplest manner. “Chromosomal” representation requires the solution to be coded as a finite length string.

Detailed information on GA and its basic operators can be found in Goldberg (1989), Gen and Cheng (1997), and Lisnianski and Levitin (2003). The basic structure of the version of GA referred to as GENITOR is as follows (Whitley, 1989).

First, an initial population of $N_s$ randomly constructed solutions (strings) is generated. Within this population, new solutions are obtained during the genetic cycle by using crossover, and mutation operators. The crossover produces a new solution (offspring) from a randomly selected pair of parent solutions, facilitating the inheritance of some basic properties from the parents by the offspring. Mutation results in slight changes to the offspring’s structure, and maintains a diversity of solutions. This procedure avoids premature convergence to a local optimum, and facilitates jumps in the solution space.

Each new solution is decoded, and its objective function (fitness) values are estimated. These values, which are a measure of quality, are used to compare different solutions.

The comparison is accomplished by a selection procedure that determines which solution is better: the newly obtained solution, or the worst solution in the population. The better solution joins the population, while the other is discarded. If the population contains equivalent solutions following selection, redundancies are eliminated, and the population size decreases as a result.

After new solutions are produced $N_{rep}$ times, new randomly constructed solutions are generated to replenish the shrunken population, and a new genetic cycle begins.

The GA is terminated after $N_c$ genetic cycles. The final population contains the best solution achieved. It also contains different near-optimal solutions which may be of interest in the decision-making process. To apply the genetic algorithm to a specific problem, a solution representation and decoding procedure must be defined.

## 1.4. Research objective and scope

The purpose of this thesis is to model the reliability of networked systems with different structures and study the related optimization problems. The structure of this thesis is illustrated by Figure 1.1.

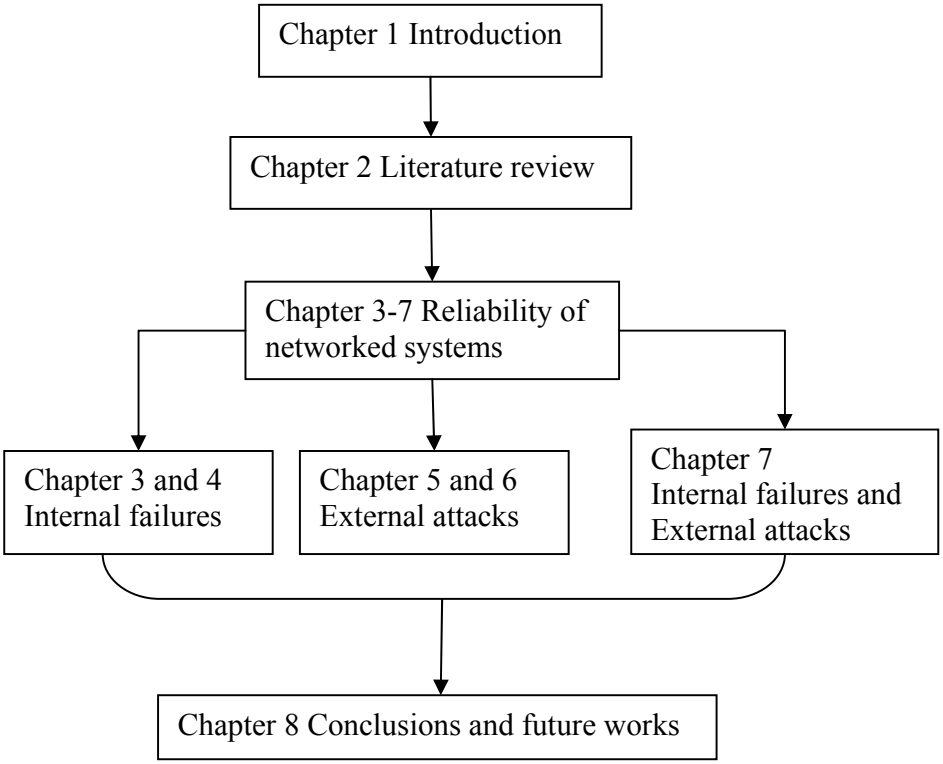


Figure 1.1 The structure of this thesis

Chapter 2 provides a brief literature review on the reliability of the selected systems and some other relevant issues.

Chapter 3 and 4 focus on networked systems subjected to only internal failures. Chapter 3 studies the optimal structure of multi-state series-parallel systems with consideration of different kinds of imperfect fault coverage. The components in the same subsystem can be allocated into different redundant work sharing groups in order to achieve reliability and performance requirement. An uncovered failure makes a whole

work sharing group fail and the fault coverage factor depends on the specific coverage technique used. A framework is proposed to solve the optimal allocation of components into different work sharing groups in order to maximize the system reliability. Chapter 4 studies the optimal elements allocation and maintenance strategy in linear multistate consecutively connected systems. The objective is to minimize the total maintenance cost through optimal elements allocation onto different nodes when the system is subjected to pre-specified availability requirements. A framework is proposed to solve the combined elements allocation and maintenance strategy.

Chapter 5 and 6 focus on system defense against external attacks. Chapter 5 studies the defense of simple series and parallel systems with imperfect false targets. It is assumed that the detection probability of a false target is constant. The contest between defender and attacker is modeled as a two period game, where the defender moves first and the attacker attacks thereafter. The defender aims to minimize the expected system damage while the attacker aims to maximize the expected system damage. A framework is presented to solve the optimal attack and defense strategies. Different from Chapter 5, Chapter 6 studies the defense of a single object with imperfect false targets by assuming that the detection probability of a false target is a function of the attacker's intelligence effort and the defender's disinformation effort. A framework is presented to solve the optimal resource allocation into intelligence/disinformation actions and different kinds of defense/attack actions.

Both internal failures and external attacks are considered in Chapter 7, which studies the optimal elements maintenance and protection strategy in series-parallel systems. It is

assumed that the system consists of elements with increasing failure rates. Replacement of system elements can reduce their failures rates, and thus increase system availability. Besides internal failures, the system elements can be destroyed by external attacks, say, natural disasters. In order to achieve system availability requirement with minimum cost, the optimal trade-off between system maintenance and protection is studied.

Chapter 8 makes conclusions and suggests some potential future works.

# CHAPTER 2 LITERATURE REVIEW

According to different configurations, networked systems can be classified as single component systems, series systems, parallel systems, series-parallel systems, parallel-series systems, etc. Besides system structure, there are some other factors that have impacts on system reliability, such as imperfect fault coverage and external attacks. A lot of research has been done to study the reliability of different systems with different features.

This chapter reviews some important works related to reliability studies of networked systems. The remainder of this chapter is organized as follows: Section 2.1 reviews the literatures on imperfect fault coverage. Section 2.2 focuses on the literatures related to linear consecutively connected systems. Section 2.3 reviews literatures on system defense strategies against external intentional attacks.

## 2.1. Different kinds of imperfect fault coverage techniques

Redundancy is widely used to enhance system reliability, especially for systems with stringent reliability requirements, such as nuclear power controllers and flight control systems (Lee and Na, 2009; Perhinschi et al. 2006; Tian et al. 2008). Usually the fault tolerance is implemented by providing sufficient redundancy and using automatic fault and error handling mechanisms (detection, location, and isolation of faults/failures). However, as the fault and error handling mechanisms themselves can fail, some failures can remain undetected or uncovered, which can lead to the total failure of the entire system or its sub-systems (Bouricius et al., 1969; Arnold, 1973; Xing, 2007). Examples of this effect of uncovered faults can be found in computing systems, electrical power distribution networks, pipe lines carrying dangerous materials etc. (Amari et al., 2004; Chang et al., 2005).

The probability of successfully covering a fault (avoiding fault propagation) given that the fault has occurred is known as the coverage factor (Bouricius et al., 1969). The models that consider the effects of imperfect fault coverage are known as imperfect fault coverage models or simply fault coverage models or coverage models (Amari, 1997). Depending on the type of fault tolerant techniques used, there are mainly three kinds of fault coverage models: 1. Element Level Coverage (ELC). A particular coverage factor value is associated with each element. This value is independent of the statuses of other elements. 2. Fault Level Coverage (FLC). The coverage factor value depends on the number of good elements that belong to a specific group (i.e., the statuses of other elements). 3. Performance Dependent Coverage (PDC). The coverage factor value depends on the

cumulative performance of the available group elements at the moment when the failure occurs.

The ELC model is appropriate when the selection among the redundant elements is made on the basis of a self-diagnostic capability of the individual elements. Such systems typically contain a built-in test (BIT) capability. Amari et al. (1999) studied the reliability of different systems with imperfect fault coverage. The systems considered include parallel, parallel-series, series-parallel, and *k*-out-of-*n* systems. Levitin (2007a) suggested a modification of the generalized reliability block diagram (RBD) method for evaluating reliability and performance indices of multi-state systems with uncovered failures. The fault coverage functions considered in these papers are performed at element level.

The FLC model is appropriate for modeling systems in which the selection among redundant elements varies between initial and subsequent failures. In the HARP terminology (Bavuso et al., 1994), ELC models are known as single-fault models, whereas FLC models are known as multi-fault models. Multi-fault models have the ability to model a wide range of fault tolerant mechanisms. An example is a majority voting system among the currently known working elements, see Myers and Rauzy (2008). A system with three or more redundant elements can be designed to assure extremely high levels of coverage so long as a mid value select voting strategy can be applied. However, selection from among the last two remaining elements, whose outputs do not agree by an amount in excess of some predetermined fault detection threshold, cannot be done with the same high level of coverage. In this case, the redundancy management process is unable to determine which element is the failed one. For this one-on-one case, redundancy

management function is typically accomplished by using built-in test, as done for ELC systems. Since the coverage for the initial faults is very close to unity and only the one-on-one fault has a coverage level typical of an ELC system, and, as a result, FLC systems can be designed to achieve much lower levels of failure probability. For this reason, most digital aircraft flight control systems (typically designed to have a probability of failure on the order of $10^{-7} - 10^{-9}$ per flight hour) are designed as FLC systems. Levitin and Amari (2008b) proposed a universal generating function based methodology to calculate the reliability of complex multi-state systems with fault level coverage.

The performance dependent coverage considered in Levitin and Amari (2008a) takes place when the fault detection and recovery functions are performed by system elements in parallel with their main functions. The proposed model is suitable for systems that cannot change the states during task execution, such as alarm systems and data processing systems performing short tasks. The systems usually remain in idle mode, thus fault detection and coverage can be performed only during task execution. When the task arrives, the system can be in one of various states, depending on availability of its elements. Therefore, the coverage probability depends only on the performance available at the moment of task arrival and does not depend on the history of failures.

Due to imperfect fault coverage, the system reliability can decrease with increase in redundancy over some particular limit (Amari et al., 2003; Levitin and Amari, 2008b). As a result the system structure optimization problems arise. Myers (2008) discussed the optimal redundancy level of *k*-out-of-*n* systems with the consideration of both element level coverage and fault level coverage. Levitin (2008) presented a model of series-

parallel multi-state systems with two types of task parallelization: parallel task execution with work sharing, and redundant task execution. A framework is proposed to solve the optimal balance of the two kinds of parallelization which maximizes the system reliability based on the assumption that the ELC applies in each work sharing group. Myers and Rauzy (2008) proposed a binary decision diagram based algorithm to analyze the reliability of redundant systems with the consideration of imperfect fault coverage.

## 2.2. Linear multi-state consecutively connected systems

A linear multi-state consecutively connected systems (LMCCS) consists of $N$+1 consecutively ordered positions (nodes) $C_n$, $n$=1,…,$N$+1. The first node $C_1$ is the source and the last node $C_{N+1}$ is the sink. The system fails if the first node (source) is not connected with the final node (sink). The LMCCS was first introduced by Hwang and Yao (1989) as a generalization of linear consecutive-$k$-out-of-$n$: F systems and linear consecutively connected systems with two-state elements (Shanthikumar 1987; Eryilmaz and Tutuncu, 2009). The basic assumptions are that the transmission range of each component is a random variable and the states of all the components are statistically independent. A recursive approach is proposed for obtaining the reliability of a LMCCS. The evaluation of LMCCS reliability was also studied in Zuo (1993) and Kossow and Preuss (1995). Zuo (1993) proposed an algorithm to evaluate the reliability of a LMCCS with two-state components with the consideration of the relevancy of the components to

the whole system reliability. A component is regarded as irrelevant to the system reliability if all the previous components that can reach the component can reach farther than the component. A universal generating function based approach was proposed in Levitin (2001) for the reliability evaluation of a linear multi-state consecutively connected signal transmission system with consideration of the possible delay of re-transmitters. When the re-transmitter delay is considered, the reliability of a LMCCS is defined as the probability that signal can be transmitted from the source to the sink within a pre-specified time.

Due to the structure of LMCCS, the reliability of a LMCCS is not only related to the respective reliability/performance of each element but also largely dependent on the allocation of the elements onto different nodes. The problem of optimal element allocation in LMCCS was first formulated by Malinowski and Preuss (1996). In this problem, elements with different characteristics should be allocated in different positions in such a way that maximizes the system reliability. It only studied the case when one and only one element can be allocated onto each node. The near-optimal components arrangement is solved by recursively changing the positions of two components to maximize the system reliability. As proved in Levitin (2003), even for $M=N$, greater reliability can be achieved if some of the $M$ elements are gathered in the same position providing redundancy (in hot standby mode) than if all the $M$ elements are evenly distributed between all the positions. The LMCCS considered in Levitin (2003) allows the system elements to be allocated onto the first $N$ positions arbitrarily so that some positions may have multiple elements whereas the other positions may have no elements. A universal generating function is adopted for

system reliability evaluation and a genetic algorithm is employed to solve the optimal element allocation strategy.

## 2.3. System defense strategies against intentional attacks

There are three measures of passively defending objects against intentional attacks: 1) providing redundancy (and separating redundant elements, which makes it impossible to destruct multiple elements by a single impact); 2) protecting the system elements (where protection presumes actions aimed at reducing the destruction probability of an element in the case of any external impact); 3) deploying false targets (which dissipates the attacker resources among greater number of targets and reduces its per-target effort). Measure 1 makes the system parallel (though each redundant object may have complex structure, it can be considered as a single target that can be destroyed/incapacitated by an impact from the defender's and attacker's points of view). The protection is a technical or organizational measure which is aimed to reduce the vulnerability of protected system element. The vulnerability of each element is its destruction probability when it is attacked. Besides direct protections, deploying false targets is another effective measure to defense systems against intentional attacks. The objective of a false target, sometimes referred to as a decoy, is to give the appearance that the element is something else than it actually is. A false target conceals or distracts something else, i.e. the genuine object, which the attacker actually searches for.

### 2.3.1. Redundancy and protection

The pioneering works Bier and Abhichandani (2002) and Bier et al. (2005) studied the optimal protection resource allocation onto different system components in simple series and parallel systems. Whereas Bier and Abhichandani (2002) assumes that the attacker will maximize the success probability of an attack, Bier et al. (2005) assumes that the attacker will maximize the expected damage on the system. It has proposed a revised objective function which incorporates the inherent values of system components. Zhuang and Bier (2007) studied the equilibrium strategies for both attacker and defender in a fully endogenous model of resource allocation for countering terrorism and natural disasters. Although these models have demonstrated a general approach and suggested some useful recommendations, these models failed to consider some important aspects, such as the possibility of the destruction of several elements by a single attack and the damage caused by partial system incapacitation.

Levitin (2007b) considered the defense of a series-parallel system against intentional attacks with protection cases. The system consists of some subsystems connected in series, where each subsystem contains some parallel elements. It is assumed that the elements within the same subsystem can be separated and protected in different protection cases so that a single attack can at most destroy the elements in a single protection case. The defense and attack contest is modeled as a two period min-max game. The defender builds the infrastructure in the first period assuming that the attacker will use the most harmful attack strategy and the attacker attacks the system in the second period in order to incur

maximum system damage. A framework is proposed to solve the optimal allocation of different elements into different levels of protection cases, which aims to minimize the total expected system damage. In this paper, the optimal protection strategy is studied assuming that the system structure is fixed. Sometimes the defender needs to determine both the structure of the system and the protection strategy in order to maximize the system reliability. Levitin and Hausken (2008) studied the optimal resource allocation between deploying separated redundant elements and protecting these elements against external intentional attacks. In this case the defender needs to determine both the number of elements to construct and the number of elements to protect. Hausken and Levitin (2008) studied the efficiency of even separation of parallel elements. A framework is proposed to solve the optimal resource allocation between separation and protection of the system elements. It has also considered the possibility of the change of contest intensity after the separation of elements. Hausken (2008) studied the protection and attack strategies of series-parallel and parallel-series systems. The defense and attack of the systems are modeled as a simultaneous game. A framework is proposed to solve the optimal distribution of the defender's protection resource and the attacker's attack resource. Ramirez-Marquez et al. (2009) studied the optimal protection of general source-sink networks via evolutionary techniques. It is assumed that the attacker has evenly distributed some attacking resource among all the links. The optimal allocation of defense resource onto the links which maximizes the survivability of the network is studied.

### 2.3.2. Deploying false targets

Blanks (1994) provides historical examples for the use of decoys in WWII and the 1990-1991 Operation Desert Storm, and writes that the U.S. Army (at one point prior to 1994) invested $7.5M into fielding multispectral tactical decoys. Although “initially, many company commanders were reluctant to include the decoys in their tactical planning,” Blanks (1994) “concludes that decoys do enhance combat effectiveness when decoy employment is incorporated into the tactical scheme of maneuver.” NATO commander Wesley Clark publicly admits that during the 1998-1999 Kosovo war the Serbs "did skillfully deploy lots of decoys”. Clark points out that very few damaged or destroyed vehicles have been found in Kosovo. The Serbs evidently fooled NATO airmen into attacking false tanks made from wooden frames covered with tarpaulins or plastic sheeting.

The aim of deploying false targets is to mislead the attacker so that the genuine target will be attacked with less probability or less attacking effort. The efficiency of false targets in defense strategy has been studied in Levitin and Hausken (2009a), which assumes that there is a single genuine target to be protected and false targets can be deployed to distract the attacker. When both the defender’s and the attacker’s resources are limited, the defender may consider whether it is more cost effective to spend more resources on protecting the genuine target to reduce its vulnerability or to spend more resource on deploying false targets to reduce the probability of attack against the genuine target. For variable defender’s and attacker’s resources, the defender and the attacker have

their own utility functions. The Nash equilibrium defense and attack strategies are solved. Levitin and Hausken (2009b) studied the optimal resource allocation between constructing redundant genuine elements, protecting these elements and deploying false targets. Hausken and Levitin (2009) studied the optimal resource allocation in protecting system elements and deploying false targets in series systems. It is assumed in these papers that the attacker cannot distinguish the genuine object from false targets, that is, it has no preference in attacking the genuine object and a false target.

# CHAPTER 3 SYSTEM RELIABILITY WITH IMPERFECT FAULT COVERAGE

Due to imperfect fault coverage (IFC), the system reliability can decrease with increase in redundancy over some particular limit (Myers 2008). As a result the system structure optimization problems arise. Some of these problems have been formulated and solved for parallel systems, *k*-out-of-*n* systems (Amari, 1997; Amari et al., 2004). Levitin (2008) presented a model of series-parallel multi-state systems with two types of task parallelization: parallel task execution with work sharing, and redundant task execution. A framework to solve the optimal balance of the two kinds of parallelization which maximizes the system reliability is proposed based on the assumption that the ELC applies in each work sharing group. Considering the different types of fault handling mechanisms in practice, the ELC model alone cannot adapt to all the cases.

Depending on the type of fault tolerant techniques used, there are mainly three kinds of fault coverage models: 1. Element Level Coverage (ELC). A particular coverage factor value is associated with each element. This value is independent of the status of other elements. 2. Fault Level Coverage (FLC). The coverage factor value depends on the

number of good elements that belong to a specific group (i.e., the status of other elements). 3. Performance Dependent Coverage (PDC). The coverage factor value depends on the cumulative performance of the available group elements at the moment when the failure occurs. The ELC model is appropriate when the selection among the redundant elements is made on the basis of a self-diagnostic capability of the individual elements. Such systems typically contain a built-in test (BIT) capability. The FLC model is appropriate for modeling systems in which the selection among redundant elements varies between initial and subsequent failures. In the HARP terminology (Bavuso et al., 1994), ELC models are known as single-fault models, whereas FLC models are known as multi-fault models. Multi-fault models have the ability to model a wide range of fault tolerant mechanisms. An example is a majority voting system among the currently known working elements, see Myers and Rauzy (2008). The performance dependent coverage considered in Levitin and Amari (2008a) takes place when the fault detection and recovery functions are performed by system elements in parallel with their main functions. The proposed model is suitable for systems that cannot change the states during task execution, such as alarm systems and data processing systems performing short tasks. When the task arrives, the system can be in one of various states, depending on availability of its elements. Therefore, the coverage probability depends only on the performance available at the moment of task arrival and does not depend on the history of failures.

In this chapter, the problem of finding the optimal balance between the two kinds of parallelization has been extended to the cases of FLC and PDC. Section 3.1 presents the model. Section 3.2 describes a universal generating function (UGF)-based algorithm for evaluating the reliability of series-parallel MSS with FLC and PDC. Section 3.3 discusses

the optimization procedures with the genetic algorithm technique. Several numerical examples are shown in section 3.4 to illustrate the possible applications of the results.

## 3.1. Model description and problem formulation

### 3.1.1. General model and assumptions

Consider a system consisting of $M$ subsystems connected in series. Each subsystem $m$ contains $E_m$ different elements connected in parallel. Any system element $j$ can have $k_j$+1 different states corresponding to the performance rates, represented by the set $\boldsymbol{g}_j = \{g_{j0}, g_{j1}, ..., g_{jk_j}\}$, where $g_{jh}$ is the performance rate of element $j$ in the state $h$, $h \in \{0, 1, \ldots, k_j\}$. The performance rate $G_j$ of element $j$ at any time instant is a random variable that takes its values from $g_j$: $G_j \in g_j$. The probability associated with the different states of the system for a given element $j$ can be represented by the set

$$\boldsymbol{p}_j = \{p_{j0}, p_{j1}, ..., p_{jk_j}\} \tag{3.1}$$

where

$$p_{jh} = \Pr\{G_j = g_{jh}\} \tag{3.2}$$

The state 0 corresponds to the total element failure, and other $k_j$ states correspond to the working states with full or partial performance.

The pmf of the performance of any system element $j$ can be represented by the pair of vectors $\boldsymbol{g}_j$, $\boldsymbol{p}_j$. Since the element is always in one and only in one of the $k_j$+1 states, we have

$$\sum_{h=0}^{k_j} p_{jh} = 1 \tag{3.3}$$

The basic assumptions of our model are listed as follows:

1) The states of different system elements are mutually independent.

2) The elements belonging to the same subsystem can be separated into independent work sharing groups (WSG). The number of WSG in a subsystem $m$ can vary from 1 where all the elements belong to the same group, to $E_m$ where each element constitutes a separate group.

3) The available elements belonging to a WSG share their work in an optimal way that maximizes the performance of the entire group. In the case of detected failures of some elements, the redundancy management system is able to redistribute the task among the available elements. An undetected failure of any element belonging to a WSG cannot be covered within this WSG, and causes the failure of the entire group.

4) Different WSG belonging to the same subsystem perform the same task in parallel providing the task execution redundancy.

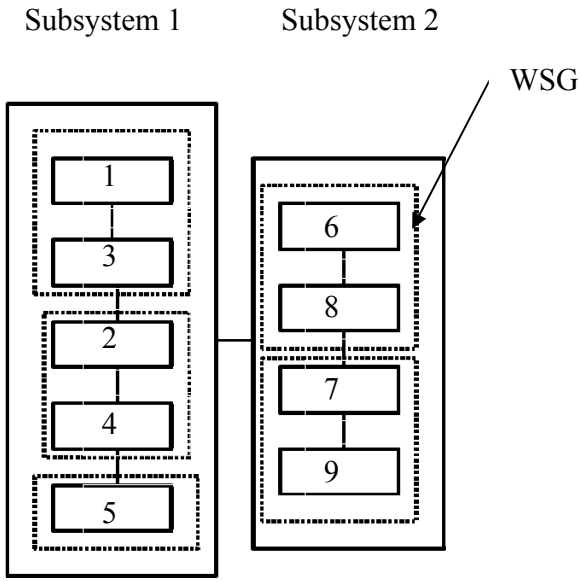


Figure 3.1 An illustrative series-parallel system with two types of parallelization

Figure 3.1 is shown for illustration. At each moment, the system elements have certain performance rates corresponding to their states. The performance rate of the entire system is unambiguously determined by its structure, and by the performance rates of its elements. Assume that the entire system has $K$+1 different states, and that $v_i$ is the entire system performance rate in state $i$. The MSS performance rate is a random variable $V$ that takes values from the set $\{v_0,\ldots,v_K\}$. The system structure function $V=\phi(G_1,\ldots,G_n)$, which maps the spaces of the elements' performance rates into the space of the system's performance rates, is determined by the system structure. In our model, the system structure function is affected by the distribution of elements among WSG in each subsystem. A real example is the data transmission system with multiple channels

connected in parallel in each subsystem. Each subsystem can be divided into some WSGs to transmit the data in parallel. If an element in a WSG fails and the failure is uncovered, the data assigned to the element is lost and the whole WSG fails to transmit the correct data.

### 3.1.2. The formulation of elements distribution

The elements' distribution among WSG in each component $m$ can be considered as a problem of partitioning a set $\Phi_m$ of $E_m$ items into a collection of $E_m$ mutually disjoint subsets $\Phi_{mi}$, i.e. such that

$$\bigcup_{i=1}^{E_m} \Phi_{mi} = \Phi_m \tag{3.4}$$

$$\Phi_{mi} \bigcap \Phi_{mj} = \varnothing \quad , i \neq j \tag{3.5}$$

Each set $\Phi_{mi}$ can contain from 0 to $E_m$ elements. The partition of the set $\Phi_m$ can be represented by the vector $\boldsymbol{\alpha}_m=\{\alpha_{mj}, 1\leq j\leq E_m\}$, where $\alpha_{mj}$ is the index of the subset to which element $j$ belongs.

Concatenation of vectors $\boldsymbol{\alpha}=\{\boldsymbol{\alpha}_1,\ldots, \boldsymbol{\alpha}_M\}$ determines the distribution of elements among the WSG for the entire system. For any given $\boldsymbol{\alpha}$, and given pmf of the system elements, one can obtain the pmf of the entire system performance $V$ in the form

$$Q_i,\ v_i, 0 \leq i \leq K, \text{where } Q_i = \Pr\{V = v_i\}. \tag{3.6}$$

### 3.1.3. The formulation of system reliability

The acceptability of a system state can usually be defined by the acceptability function $f(V,\theta^*)$, representing the desired relation between the system performance $V$, and some limit value named system demand ($f(V,\theta^*)=1$ if the system performance is acceptable, and $f(V,\theta^*)=0$ otherwise). The MSS reliability is defined as its expected acceptability, the probability that the MSS satisfies the demand (Levitin, 2005). Having the pmf of system performance (3.6), one can obtain its reliability as

$$R(\theta^*) = \sum_{i=1}^{K} Q_i f(v_i, \theta^*) \tag{3.7}$$

For example, in applications where the system performance is defined as a task execution time, and $\theta^* = T^*$ is the maximum allowed task execution time, (3.7) takes the form

$$R(T^*) = \sum_{i=1}^{K} Q_i 1(v_i < T^*) \tag{3.8}$$

whereas in applications where the system performance is defined as its productivity/capacity, and $\theta^* = C^*$ is the minimum allowed capacity, (3.7) takes the form

$$R(C^*) = \sum_{i=1}^{K} Q_i 1(v_i > C^*) \tag{3.9}$$

### 3.1.4. The formulation of the entire problem

The problem of solving the optimal elements allocation strategy in a multi-state series-parallel system with imperfect fault coverage is formulated as follows.

Find vector $\boldsymbol{\alpha}^*(\theta^*)=\{\boldsymbol{\alpha}_1,\ldots, \boldsymbol{\alpha}_M\}$, which maximizes the multi-state system reliability $R(\theta^*)$ for a given demand $\theta^*$,

$$\boldsymbol{\alpha}^*(\theta^*) = \arg\max\{R(\boldsymbol{\alpha},\theta^*)\}. \tag{3.10}$$

## 3.2. Evaluating reliability of series-parallel MSS with uncovered failures

### 3.2.1. Incorporating uncovered failures in WSG into the UGF technique

The UGF representing the pmf of a discrete random variable $X$ is defined as a polynomial (Ushakov, 1987)

$$u_X(z) = \sum_{h=0}^{H} \varepsilon_h z^{x_h}, \tag{3.11}$$

where the variable $X$ has $H$+1 possible values and $\varepsilon_h = \Pr\{X = x_h\}$.

To obtain the UGF representing the pmf of a function of two independent random variables $\varphi(X,Y)$, the following composition operator is used:

$$\begin{aligned} U_{\varphi(X,Y)}(z) &= u_X(z) \underset{\varphi}{\otimes} u_Y(z) \\ &= \left(\sum_{h=0}^{H} \varepsilon_h z^{x_h}\right) \underset{\varphi}{\otimes} \left(\sum_{d=0}^{D} \varepsilon_d z^{y_d}\right) \\ &= \sum_{h=0}^{H} \sum_{d=0}^{D} \varepsilon_h \varepsilon_d z^{\varphi(x_h, y_d)} \end{aligned} \tag{3.12}$$

The polynomial $U(z)$ represents all of the possible mutually exclusive combinations of realizations of the variables by relating the probabilities of each combination to the value of function $\varphi(X, Y)$ for this combination.

### A. The UGF of a WSG in the case of FLC

In the case when multi-fault coverage takes place in each WSG, one needs to incorporate the coverage probabilities depending on the number of failed elements into the performance distribution of any WSG. Thus one has to know not only entire group performance but also the total number of failed elements in each state of this group (combination of states of its elements). To obtain both these indices, the performance distribution for system elements is described by a modified UGF as

$$\tilde{u}_j(z) = \sum_{h=0}^{k_j} p_{jh} z^{s_{jh}, g_{jh}}, \tag{3.13}$$

where $s_{jh}$ represents the realization of the random number of failed elements in state $h$. The UGF of an individual element takes the form

$$\tilde{u}_j(z) = p_{j0} z^{1, g_{j0}} + \sum_{h=1}^{k_j} p_{jh} z^{0, g_{jh}}, \tag{3.14}$$

where $g_{j0}$ corresponds to the case of failure of the element (1 failure), $g_{jh}$ ($1 \leq j \leq k_j$) corresponds to the $h$-th working state of element $j$ (0 failure).

Applying the operator

$$\tilde{U}_{\{i,j\}}(z) = \tilde{u}_i(z) \underset{\omega}{\otimes} \tilde{u}_j(z) = \sum_{h=0}^{k_i} \sum_{d=0}^{k_j} p_{ih} p_{jd} z^{s_{ih}+s_{jd}, \omega(g_{ih}, g_{jd})} \tag{3.15}$$

recursively one can obtain the UGF of the entire WSG $i$ in subsystem $m$ in the form.

$$\tilde{U}_{mi}(z) = \sum_{h=0}^{n_{mi}} P_{mih} z^{s_{mih}, g_{mih}} \tag{3.16}$$

that represents the distribution of the number of failed elements and the corresponding performance of the WSG. Here $w$ is the performance composition function for elements connected in parallel with work sharing, $P_{mih}$ is the probability that WSG $i$ in subsystem $m$ contains exactly $s_{mih}$ failed elements and functions at the performance level $g_{mih}$ given all the failures are covered ($g_{mi0}$ correspond to the failures of all the elements in the group).

We assume that the coverage probability of a failure is determined by the total number of elements in the WSG and the number of failed elements in this group (which affects the load on the monitoring system). Let $c_m(|\Phi_{mi}|, j)$ be the fault coverage probability in the

case of $j$-th failure in WSG $i$ in subsystem $m$ (when $j$-1 elements are already unavailable), and $r_{mi}(k)$ be the probability that the group does not fail after $k$ failures have consecutively occurred. It can be seen, that

$$r_{mi}(k) = \prod_{j=0}^{k} c_m(|\Phi_{mi}|, j) \tag{3.17}$$

By definition $r_{mi}(0) = c_m(|\Phi_{mi}|, 0) = 1$ and $c_m(|\Phi_{mi}|, |\Phi_{mi}|) = 0$.

The unconditional probability that the WSG $i$ in subsystem $m$ can work with performance $g_{mih}$ ($h$=1,…,$n_{mi}$) after $s_{mih}$ elements have failed is

$$P_{mih} r_{mi}(s_{mih}) = P_{mih} \prod_{j=0}^{s_{mih}} c_m(|\Phi_{mi}|, j) \tag{3.18}$$

Thus the uncovered failures can be incorporated into the UGF by applying the following operator $\varepsilon$:

$$\begin{aligned} U_{mi}(z) &= \varepsilon(\tilde{U}_{mi}(z)) = \varepsilon\left(\sum_{h=0}^{n_{mi}} P_{mih} z^{s_{mih}, g_{mih}}\right) \\ &= \sum_{h=0}^{n_{mi}} P_{mih} r(s_{mih}) z^{g_{mih}} + [1 - \sum_{h=0}^{n_{mi}} P_{mih} r(s_{mih})] z^{g_{mi0}} \end{aligned} \tag{3.19}$$

This UGF represents the unconditional distribution of performance of entire WSG $i$ in subsystem $m$.

**B. The UGF of a WSG in the case of PDC**

In the case that the coverage probability of a WSG depends on the entire performance of the group, one can use $l_{mi}(g)$ to denote the fault coverage probability of WSG $i$ in subsystem $m$ in the case the entire performance of the group is $g$. By definition we have $l_{mi}(0)=0$. The uncovered failures can be incorporated into the UGF by applying the following operator:

$$\begin{aligned} U_{mi}(z) &= \psi(\tilde{U}_{mi}(z)) = \psi(\sum_{h=0}^{n_{mi}} P_{mih} z^{s_{mih}, g_{mih}}) \\ &= \sum_{h=0}^{n_{mi}} P_{mih} l_{mi}(g_{mih}) z^{g_{mih}} + [1 - \sum_{h=0}^{n_{mi}} P_{mih} l_{mi}(g_{mih})] z^{g_{mi0}} \end{aligned} \tag{3.20}$$

This UGF represents the unconditional distribution of performance of entire WSG $i$ in subsystem $m$.

**C. The UGF of a subsystem**

Applying (3.12) with $\varphi \equiv \varpi$

$$U_{\{mi,mj\}}(z) = U_{mi}(z) \underset{\varpi}{\otimes} U_{mj}(z) \tag{3.21}$$

recursively one can obtain the UGF of subsystem $m$ in the form.

$$U_m(z) = \sum_{h=0}^{n_m} P_{mh} z^{g_{mh}} \tag{3.22}$$

Here $\varpi$ is the performance composition function for elements connected in parallel without work sharing, $P_{mh}$ is the probability that the performance of subsystem $m$ equals to $g_{mh.}$

**D. The UGF of the entire system**

Applying (3.12) with $\varphi \equiv \pi$

$$U_{\{m,l\}}(z) = U_m(z) \underset{\pi}{\otimes} U_l(z) \tag{3.23}$$

recursively one can obtain the UGF of the entire system in the form.

$$U_s(z) = \sum_{h=0}^{n_s} P_h z^{g_h} \tag{3.24}$$

Here $\pi$ is the performance composition function for elements connected in series, $P_h$ is the probability that the performance of the entire system equals to $g_{h.}$

From the UGF $U_s(z)$ representing the pmf of the entire MSS performance (3.6), the system reliability can be obtained using (3.7).

### 3.2.2. Performance composition functions

The choice of functions $\varphi$ depends on the type of connection between the elements, and on the type of the system. In our model, we have to distinguish three different functions: for

redundant parallel connection without work sharing ($\varphi \equiv \varpi$), for parallel connection with work sharing ($\varphi \equiv \omega$), and for series connection ($\varphi \equiv \pi$).

Consider, for example, a task processing with performance defined as task completion time. Assume that each element $j$ can complete the task by random time $G_j$, the case of total failure of the element corresponds to $G_j=\infty$. If two elements $i$ and $j$ perform the same task in parallel, providing task execution redundancy, then the task completion time is equal to the time when the fastest element completes the task. The performance of the pair of elements in this case is determined by the function

$$\varpi(G_i, G_j) = \min(G_i, G_j). \tag{3.25}$$

As shown in Levitin (2005), if two parallel elements can share the work by dividing the task in proportion to their processing speed, the task completion time for the pair of elements is determined by the function

$$\omega(G_i, G_j) = \begin{cases} G_i G_j /(G_i + G_j) \text{ if } G_i < \infty, G_j < \infty \\ G_i \text{ if } G_j = \infty \\ G_j \text{ if } G_i = \infty \end{cases}. \tag{3.26}$$

If two elements consecutively execute different subtasks, represented by series connection of the elements, the entire task completion time for the pair of elements is equal to the sum of their individual execution times. The performance of the pair of elements in this case is determined as

$$\pi(G_i, G_j) = G_i + G_j. \tag{3.27}$$

Another example is a data transmission system with performance defined as transmission capacity (bandwidth). Assume that each element $j$ has a random data transmission capacity $G_j$, the case of total failure of the element corresponds to $G_j$=0. If two elements $i$ and $j$ transmit the same data, providing data transmission redundancy, the transmission capacity of the pair of elements is determined by the element with greater performance. The performance of the two elements is determined by the function

$$\varpi(G_i, G_j) = \max(G_i, G_j). \tag{3.28}$$

If the parallel elements share their work, then the entire capacity that they provide is equal to the sum of their individual capacities. The performance of the two elements is determined by the function

$$\omega(G_i, G_j) = G_i + G_j. \tag{3.29}$$

If data flow is transmitted by two consecutive elements, the bandwidth of the slowest element becomes the bottleneck of the system. Therefore, the performance of the two elements is determined by the minimum of their individual performances,

$$\pi(G_i, G_j) = \min(G_i, G_j). \tag{3.30}$$

## 3.3. Optimization technique

Equation (3.10) formulates a complicated combinatorial optimization problem. An exhaustive examination of all possible solutions is not realistic, considering reasonable time limitations. The genetic algorithm (GA) has proven to be an effective optimization tool for a large number of complicated problems in reliability engineering, and it is used for our optimization.

### 3.3.1. Solution representation

In the considered problem, element separation is determined by vector $\boldsymbol{\alpha}$ that contains $n = \sum_{m=1}^{M} E_m$ items corresponding to elements composing the entire system. In our GA, solutions are represented by integer strings $S=\{s_1,s_2,\ldots s_n\}$. For each $i = \sum_{k=1}^{m-1} E_k + j$, item $s_i$ of the string corresponds to item $\alpha_{mj}$ of the vector $\boldsymbol{\alpha}$, and determines the number of WSG to which the $j$-th element of the $m$-th subsystem belongs. Therefore, all the items $s_i$ of the string $S$, corresponding to component $m$ ($\sum_{k=1}^{m-1} E_m + 1 \le i \le \sum_{k=1}^{m} E_m$), should vary in the range (1, $E_m$). Because the random solution generation procedure can produce strings with elements randomized within the same range, to provide solution feasibility one must use a transformation procedure that makes each string element belonging to the proper range. This procedure determines the value of $\alpha_{mj}$ as $1+\mathrm{mod}_{Em}(s_i)$. The range of values produced by the random generation procedure should be (1, $\max_{m=1}^{M} E_m$).

### 3.3.2. Solution decoding procedure

The following procedure determines the fitness value for an arbitrary solution defined by integer string $S=\{s_1,s_2,\ldots s_n\}$.

1) For each subsystem $m=1,\ldots,M$:

1.1. Determine the number of WSG for each element of the $m$-th component:

$$\alpha_{mj} = 1 + \mathrm{mod}_{E_m}(s_{c+j}), 1 \le j \le E_m, \tag{3.31}$$

where $c = \sum_{k=1}^{m-1} E_k$.

1.2. For each WSG $i$ ($1 \le i \le E_m$), create set $\Phi_{mi}$ using the recursive procedure

$\Phi_{mi} = \emptyset$, for $i=1,\ldots, E_m$:

$$\text{if } \alpha_{mj}=i,\ \Phi_{mi} = \Phi_{mi} \bigcup \{c + j\}. \tag{3.32}$$

2) Determine the system reliability by the algorithm presented in section 3.2. Assign the obtained system reliability to the solution fitness.

### 3.3.3. Crossover and mutation procedures

The cross operator for given parent strings $\boldsymbol{P1}$, $\boldsymbol{P2}$ and the offspring string $\boldsymbol{O}$ is defined as follows: the $i$-th element ($1 \le i \le n$) of the string $\boldsymbol{O}$ is equal to the $i$-th element of either $\boldsymbol{P1}$ or $\boldsymbol{P2}$ both with probability 0.5.

The mutation procedure swaps elements initially located in two randomly chosen positions.

## 3.4. Illustrative examples

Consider a data transmission system consisting of two consecutive multi-channel communication lines. Each channel can have failure state with zero transmission capacity and two working states with full and reduced transmission capacity. The distributions of the performances (transmission capacities) of channels are presented in Table 3.1.

Any subset of channels belonging to the same line can compose a WSG in which the data packages are divided into sub-packages transmitted by different channels. Undetected failures within any WSG remain uncovered. Depending on system monitoring architecture the probability of failure detection can be represented by different functions of number of elements in the WSG, the number of failed elements, and the entire group performance.

Table 3.1 Performance distributions of data transmission channels

| Sub-system | Element | Performance levels | | | | | |
|---|---|---|---|---|---|---|---|
| | | Probability $p_{j0}$ | capacity $g_{j0}$ | probability $p_{j1}$ | capacity $g_{j1}$ | probability $p_{j2}$ | capacity $g_{j2}$ |
| 1 | 1 | 0.15 | 0 | 0.7 | 10 | 0.15 | 20 |
| | 2 | 0.15 | 0 | 0.65 | 12 | 0.20 | 20 |
| | 3 | 0.20 | 0 | 0.60 | 15 | 0.20 | 25 |
| | 4 | 0.15 | 0 | 0.60 | 18 | 0.25 | 25 |
| | 5 | 0.15 | 0 | 0.70 | 14 | 0.15 | 20 |
| | 6 | 0.10 | 0 | 0.80 | 11 | 0.10 | 24 |
| | 7 | 0.20 | 0 | 0.50 | 20 | 0.30 | 30 |
| 2 | 8 | 0.20 | 0 | 0.60 | 12 | 0.20 | 25 |
| | 9 | 0.20 | 0 | 0.60 | 14 | 0.20 | 24 |
| | 10 | 0.20 | 0 | 0.70 | 15 | 0.10 | 25 |
| | 11 | 0.15 | 0 | 0.65 | 20 | 0.20 | 30 |
| | 12 | 0.15 | 0 | 0.70 | 12 | 0.15 | 20 |
| | 13 | 0.10 | 0 | 0.80 | 18 | 0.10 | 30 |
| | 14 | 0.25 | 0 | 0.65 | 10 | 0.10 | 20 |

**A. FLC example 1**

In some occasions the load on monitoring system is proportional to the number of failed elements because it performs failure detection and monitoring actions and these actions are much more time consuming than monitoring the available elements. In this case it is reasonable to assume that $c_m(|\Phi_{mi}|,j)$ depends only on $j$ when $1 \le j \le |\Phi_{mi}|-1$. As an illustration we assume that the coverage (detection) probability of $j$-th failure in WSG $i$ in any subsystem $m$ is a decreasing function of $j$ as given in Table 3.2.

Table 3.2 Coverage probability after $j$-th failure in WSG with $|\Phi_{mi}|$ elements in FLC example 1

| $\|\Phi_{mi}\|$ / $j$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 2 | - | 0 | 0.63 | 0.63 | 0.63 | 0.63 | 0.63 |
| 3 | - | - | 0 | 0.36 | 0.36 | 0.36 | 0.36 |
| 4 | - | - | - | 0 | 0.22 | 0.22 | 0.22 |
| 5 | - | - | - | - | 0 | 0.15 | 0.15 |
| 6 | - | - | - | - | - | 0 | 0.08 |
| 7 | - | - | - | - | - | - | 0 |

Different WSG of the same line transmit the same data in parallel. The system transmission capacity should be greater than $C^*$. The system corresponds to the flow transmission model with composition functions (3.28)–(3.30), and reliability defined according to (3.9). The problem is to find the optimal system configuration (distribution of the channels among the WSGs) that can provide certain system transmission capacity $C^*$ with maximal reliability.

Table 3.3 Parameters of solutions in FLC example 1

| | | No sharing | $C^*$=20 | $C^*$=30 | $C^*$=40 | No redundancy |
|---|---|---|---|---|---|---|
| Max capacity | | 30 | 70 | 89 | 90 | 164 |
| R(0) | | ≈1.0 | 0.9996 | 0.9961 | 0.9958 | 0.6865 |
| R(20) | | 0.3640 | 0.9932 | 0.9915 | 0.9909 | 0.6865 |
| R(30) | | 0.0 | 0.9113 | 0.9834 | 0.9601 | 0.6865 |
| R(40) | | 0.0 | 0.5658 | 0.7999 | 0.9119 | 0.6863 |
| Structure | Subsystem 1 | (1),(2),(3), (4),(5),(6),(7) | (1,2,6), (3,4,5),(7) | (1,2,3,6), (4,5,7) | (1,6,7), (2,3,4,5) | (1,2,3,4, 5,6,7) |
| | Subsystem 2 | (8),(9),(10), (11),(12), (13),(14) | (8,10,14), (11,12, 13),(9) | (8,9,12,14), (10,11,13) | (8,11,14), (9,10,12, 13) | (8,9,10,11, 12,13,14) |

Table 3.3 contains the optimal system configurations for $C^*$=20 Kb/sec, $C^*$=30 Kb/sec, and $C^*$=40 Kb/sec obtained using the GA and characteristics of the corresponding transmission systems. This table also contains the characteristics of the system without work sharing, when all of the channels transmit the same data, and the characteristics of the system without redundancy, when all the channels within each line belong to a single WSG. Table 3.3 presents the maximal possible system capacity, the probability that system does not fail

totally $R(0)$, reliabilities for different values of $C^*$, and the system structure. The system without work sharing has the greatest reliability $R(0)$; however, it is not able to provide capacity greater than 30Kb/sec. On the contrary, the system without redundancy has the greatest possible performance of 164Kb/sec, but very low reliability. The structures optimal for different demands have intermediate values of maximal possible capacity, and $R(0)$, while providing the greatest reliabilities $R(C^*)$.

The system reliabilities as functions of the minimum allowed transmission capacity for all the five cases are presented in Figure 3.2.

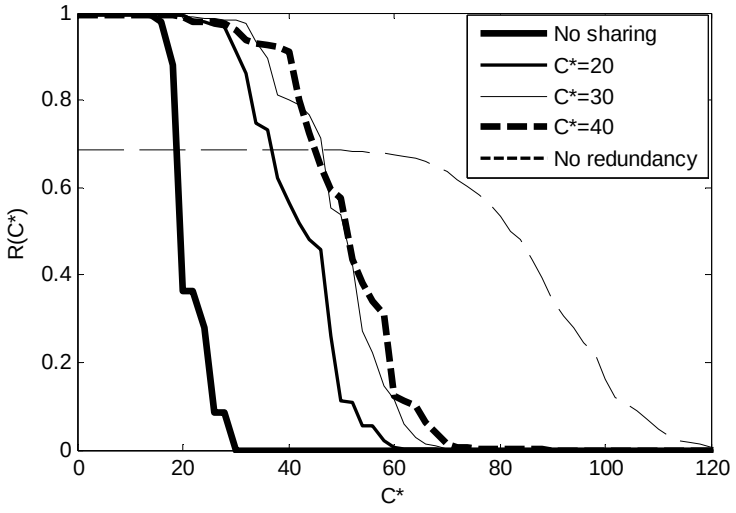


Figure 3.2 Function $R(C^*)$ for obtained configurations of the data transmission system in FLC example 1

**B. FLC example 2**

In some occasions the load on monitoring system is proportional to the number of available elements because it switches the failed elements off and does not monitor them. In this case it is reasonable to assume that $c_m(|\Phi_{mi}|,j)$ depends on $|\Phi_{mi}|$ - $j$ when $1 \le j \le |\Phi_{mi}|-1$.

Table 3.4 Coverage probability after $j$-th failure in WSG with $|\Phi_{mi}|$ elements in FLC example 2

| $\|\Phi_{mi}\|$ / $j$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0.99 | 0.63 | 0.36 | 0.22 | 0.15 | 0.08 |
| 2 | - | 0 | 0.99 | 0.63 | 0.36 | 0.22 | 0.15 |
| 3 | - | - | 0 | 0.99 | 0.63 | 0.36 | 0.22 |
| 4 | - | - | - | 0 | 0.99 | 0.63 | 0.36 |
| 5 | - | - | - | - | 0 | 0.99 | 0.63 |
| 6 | - | - | - | - | - | 0 | 0.99 |
| 7 | - | - | - | - | - | - | 0 |

As an illustration we assume that the coverage probability of the $j$-th failure in WSG $i$ in any subsystem $m$ decreases with $|\Phi_{mi}|$ and increases with $j$ as given in Table 3.4. Table 3.5 contains the optimal system configurations for $C^*$=20Kb/sec, $C^*$=30Kb/sec, and $C^*$=40Kb/sec obtained using the GA and characteristics of the corresponding transmission systems.

Table 3.5 Parameters of solutions in FLC example 2

| | | No sharing | $C^*$=20 | $C^*$=30 | $C^*$=40 | No redundancy |
|---|---|---|---|---|---|---|
| Max capacity | | 30 | 60 | 65 | 74 | 164 |
| $R(0)$ | | ≈1.0 | 0.9998 | 0.9998 | 0.9894 | 0.0952 |
| $R(20)$ | | 0.3640 | 0.9772 | 0.9733 | 0.9375 | 0.0952 |
| $R(30)$ | | 0.0 | 0.8461 | 0.9405 | 0.8656 | 0.0952 |
| $R(40)$ | | 0.0 | 0.2393 | 0.2650 | 0.7526 | 0.0952 |
| Structure | Subsystem 1 | (1),(2),(3), (4),(5),(6), (7) | (1,2,5), (3,7), (4,6) | (1,2,3), (4,5), (6,7) | (1,4,5), (2,6,7), (3) | (1,2,3,4, 5,6,7) |
| | Subsystem 2 | (8),(9),(10), (11),(12),(13), (14) | (8,9), (10,12,14), (11,13) | (8,9,14) (11,12), (10,13) | (8,9,13) (10,11,12), (14) | (8,9,10,11, 12,13,14) |

The system reliabilities as functions of the minimum allowed transmission capacity for all the five cases are presented in Figure 3.3.

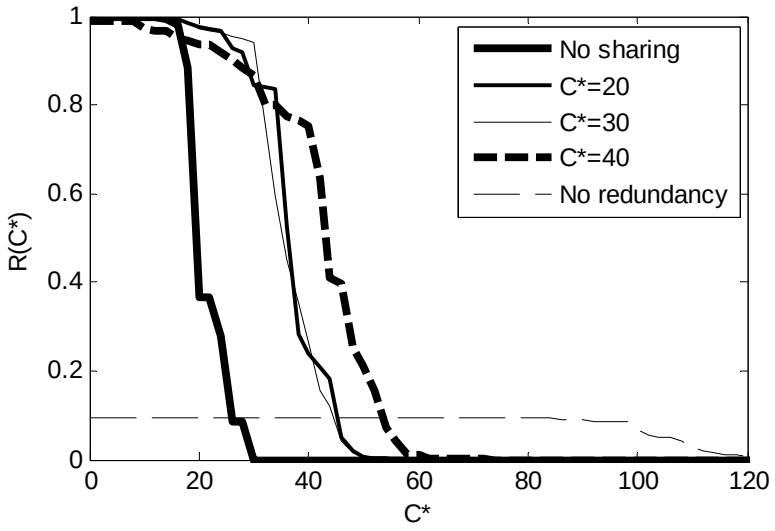


Figure 3.3 Function $R(C^*)$ for obtained configurations of the data transmission system in FLC example 2

Comparing the solutions presented in Table 3.3 and Table 3.5 one can see that in the case when the coverage probability in a WSG decreases with the increase of total number of elements in this group the optimal configurations consist of smaller WSGs.

### C. FLC example 3

In some occasions the monitoring system uses voting among the remaining components to detect failures. In this case it is reasonable to assume that $c_m(|\Phi_{mi}|,j)$ is determined by the number of remaining components. When there are at least 3 remaining components available, the coverage factor $c_m(|\Phi_{mi}|,j)$ can be regarded as 1. When there are only 2 remaining components, the monitoring system can no longer use voting to detect failures. In this case, failures can only be detected through built-in test (BIT) technology for each component with limited success probability. As an illustration we assume that the coverage probability of the *j*-th failure in WSG *i* in any subsystem *m* is as given in Table 3.6.

Table 3.7 contains the optimal system configurations for $C^*$=20Kb/sec, $C^*$=30Kb/sec, and $C^*$=40Kb/sec obtained using the GA and characteristics of the corresponding transmission systems.

Table 3.6 Coverage probability after *j*-th failure in WSG with $|\Phi_{mi}|$ elements in FLC example 3

| $\|\Phi_{mi}\|$ / $j$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0.9 | 1 | 1 | 1 | 1 | 1 |
| 2 | - | 0 | 0.9 | 1 | 1 | 1 | 1 |
| 3 | - | - | 0 | 0.9 | 1 | 1 | 1 |
| 4 | - | - | - | 0 | 0.9 | 1 | 1 |
| 5 | - | - | - | - | 0 | 0.9 | 1 |
| 6 | - | - | - | - | - | 0 | 0.9 |
| 7 | - | - | - | - | - | - | 0 |

Table 3.7 shows that in this case single WSG consisting of all elements is preferred in each subsystem, as all failures in these WSGs are always covered until two available elements remain.

Table 3.7 Parameters of solutions in FLC example 3

| | | No sharing | $C^*$=20 | $C^*$=30 | $C^*$=40 | No redundancy |
|---|---|---|---|---|---|---|
| Max capacity | | 30 | 164 | 164 | 164 | 164 |
| $R(0)$ | | ≈1.0 | ≈1.0 | ≈1.0 | ≈1.0 | ≈1.0 |
| $R(20)$ | | 0.3640 | 0.9998 | 0.9998 | 0.9998 | 0.9998 |
| $R(30)$ | | 0.0 | 0.9982 | 0.9982 | 0.9982 | 0.9982 |
| $R(40)$ | | 0.0 | 0.9921 | 0.9921 | 0.9921 | 0.9921 |
| Structure | Subsystem 1 | (1),(2),(3),(4),(5),(6),(7) | (1,2,3,4,5,6,7) | (1,2,3,4,5,6,7) | (1,2,3,4,5,6,7) | (1,2,3,4,5,6,7) |
| | Subsystem 2 | (8),(9),(10),(11),(12),(13),(14) | (8,9,10,11,12,13,14) | (8,9,10,11,12,13,14) | (8,9,10,11,12,13,14) | (8,9,10,11,12,13,14) |

The system reliabilities as functions of the minimum allowed transmission capacity for all the five cases are presented in Figure 3.4.

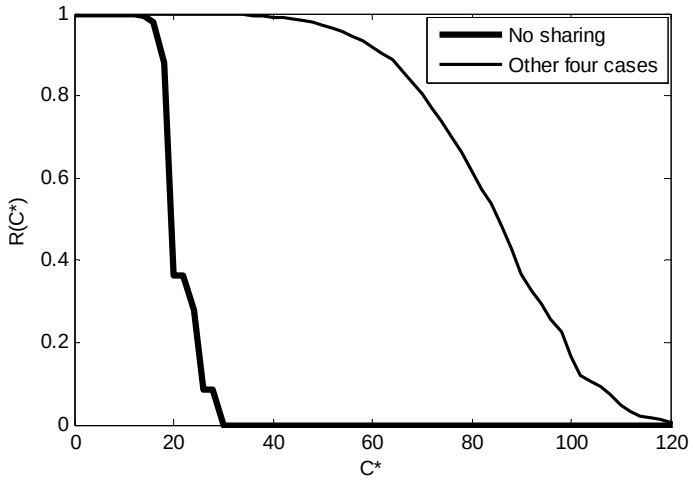


Figure 3.4 Function $R(C^*)$ for obtained configurations of the data transmission system in FLC example 3

**D. PDC example**

In the case when the fault detection and recovery functions in a WSG are performed by the system elements in the WSG, it is reasonable to assume that the fault coverage probability of a WSG depends on the entire group performance. As an example we assume that the fault coverage probability takes the form $l_{mi}(g) = \min\{1, 0.01g\}$ . Table 3.8 contains the optimal system configurations for $C^*$=20Kb/sec, $C^*$=30Kb/sec, and $C^*$=40Kb/sec obtained using the GA and characteristics of the corresponding transmission systems.

Table 3.8 Parameters of solutions in PDC example

| | | No sharing | $C^*$=20 | $C^*$=30 | $C^*$=40 | No redundancy |
|---|---|---|---|---|---|---|
| Max capacity | | 30 | 50 | 65 | 164 | 164 |
| $R$(0) | | 1.0 | 0.9937 | 0.9621 | 0.8336 | 0.8336 |
| $R$(20) | | 0.3640 | 0.9651 | 0.9486 | 0.8335 | 0.8335 |
| $R$(30) | | 0.0 | 0.6217 | 0.9353 | 0.8331 | 0.8331 |
| $R$(40) | | 0.0 | 0.0557 | 0.2633 | 0.8311 | 0.8311 |
| Structure | Subsystem 1 | (1),(2),(3), (4),(5),(6), (7) | (1,2),(3,4) (5,6),(7) | (1,2,3), (4,5), (6,7) | (1,2,3,4, 5,6,7) | (1,2,3,4, 5,6,7) |
| | Subsystem 2 | (8),(9),(10), (11),(12),(13), (14) | (8,14),(9), (10,11), (12,13) | (8,9,14), (10,13), (11,12) | (8,9,10,11, 12,13,14) | (8,9,10,11, 12,13,14) |

It can be seen from Table 3.8 that although the fault coverage probability function equals to 1 when $g$ is greater than 100, small WSGs are preferred in the cases $C^*$=20 and $C^*$=30. This is because redundancy prevents system failure even in the case, when the detected (covered) failures reduce the performance of certain WSG below the demand.

The system reliabilities as functions of the minimum allowed transmission capacity for all the five cases are presented in Figure 3.5.

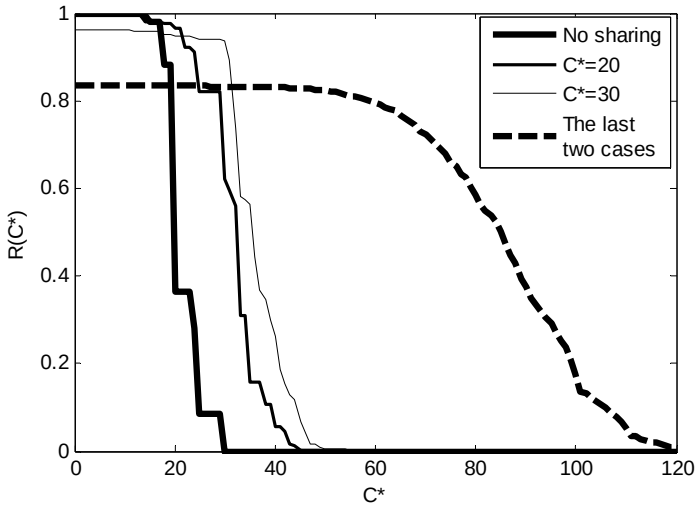


Figure 3.5 Function $R(C^*)$ for obtained configurations of the data transmission system in PDC example

## 3.5. Conclusions

This chapter extends the problem of finding optimal balance between redundancy and task sharing in multi-state systems with uncovered failures to the cases of multi-fault coverage and performance dependent coverage. It is assumed that the uncovered failures in the elements belonging to the group of elements sharing the same task can cause failure of the

entire group. Due to different fault covering mechanisms, the probability of such failure can be determined by different factors, such as the number of working elements in the group when the failure occurs, the number of failed elements in the group when the failure occurs, and the entire group performance. The procedures of finding the optimal system structure (distribution of different parallel elements among work sharing groups) have been described. The illustrative examples show the results obtained by the optimization algorithm for a data transmission system with performance defined as transmission capacity. Various assumptions of coverage factors are discussed to illustrate the application of the procedures in the cases of different fault covering mechanisms. It was shown that the greatest system reliability (defined as a probability of meeting a certain demand) can be achieved by proper balance between two types of task parallelization.

# CHAPTER 4 RELIABILITY OF LINEAR MULTI-STATE CONSECUTIVELY CONNECTED SYSTEMS

The linear multi-state consecutively connected system (LMCCS) consists of $N$+1 consecutively ordered positions (nodes) $C_n$, $n$=1,…,$N$+1. The first node $C_1$ is the source and the last node $C_{N+1}$ is the sink. At each position, elements from a set $E=\{e_1,\ldots,e_M\}$ can be allocated to provide a connection between the position in which it is allocated and the next few positions. The system fails if the first node (source) is not connected with the ($N$+1)th node (sink). Each system element $e_i$ in working state can connect the node it is located at with $g_i$ next nodes. Each element is also characterized by its lifetime distribution with an increasing failure rate.

An example of the LMCCS is a set of radio relay stations with a transmitter allocated at $C_1$ and a receiver allocated in $C_{N+1}$. Each station $C_n$ ($2\leq n\leq N$) can have retransmitters generating signals that reach the next few stations. The farthest station that can be reached by a station depends on the amplifier power of the retransmitters allocated on the station and on the random signal propagation conditions. The aim of the system is to provide propagation of a signal from the transmitter to the receiver.

The LMCCS was first introduced by Hwang and Yao (1989) as a generalization of linear consecutive-*k*-out-of-*n*: F systems and linear consecutively connected systems with two-state elements (Shanthikumar 1987; Eryilmaz and Tutuncu, 2009). The evaluation of LMCCS reliability was studied in Hwang and Yao (1989), Zuo (1993) and Kossow and Preuss (1995). Due to the structure of LMCCS, the reliability of a LMCCS is not only related to the respective reliability/performance of each element but also largely dependent on the allocation of the elements onto different nodes. The problem of optimal element allocation in LMCCS was first formulated by Malinowski and Preuss (1996). In this problem, elements with different characteristics should be allocated in different positions in such a way that maximizes the system reliability. It only studied the case when one and only one element can be allocated onto each node. As proved in Levitin (2003), even for $M=N$, greater reliability can be achieved if some of the $M$ elements are gathered in the same position providing redundancy (in hot standby mode) than if all the $M$ elements are evenly distributed between all the positions. In these works, the reliability of each element is assumed to be constant.

In practice, system elements usually fail with increasing failure intensity due to wear, rotting, deterioration, or aging effects (Lisnianski et al., 2008; Ding et al., 2009; Rao and Naikan, 2009; Wu et al., 2010). For systems containing elements with increasing failure rates, preventive replacement of the elements is an efficient measure to increase the system reliability (Nakagawa and Mizutani, 2009; Ambani et al., 2010; Liu et al., 2010). Replacing elements that have a high risk of failure, while reducing the chance of failure, can incur significant expenses, especially in systems with high replacement rates. Minimal repair, the less expensive option, enables the system element to resume its work after

failure, but does not affect its hazard rate (Beichelt and Fischer, 1980; Zhang and Jardine, 1998; Sheu and Chang, 2010). Since the element replacement reduces its failure rate, the more frequently an element is replaced the higher the availability of the element is. Therefore there is a trade-off between the availability of the system and the total system maintenance cost. Since the reliability of a LMCCS can be comprehensively increased by adjusting the positions of system elements, this property can also be utilized to reduce the maintenance cost needed for the system to meet availability requirement.

In this chapter, the combined maintenance and allocation problem is studied. Different from Malinowski and Preuss (1996) and Levitin (2003), the objective of element allocation is to minimize the maintenance cost subject to a pre-specified system availability requirement. Section 4.1 formulates the problem. Section 4.2 describes the universal generating function technique used for evaluating the LMCCS availability. A genetic algorithm is adopted for optimization in section 4.3. Illustrative examples are presented in section 4.4.

## 4.1. Problem formulation

### 4.1.1. General model and assumptions

The LMCCS consists of $N$+1 consecutively ordered positions (nodes) $C_n$, $n$=1,…,$N$+1. The lifetime for the system is denoted as $T_c$. At each position, elements from a set

$E=\{e_1,\ldots,e_M\}$ can be allocated to provide a connection (also called path or arc) between the position in which it is allocated and the next few positions. For each element $e_i$ located at node $j$ the connection of this node with nodes $j+1,\ldots, j+g_i$ is provided in the working state. The expected number of failures of element $e_i$ during time interval $(0,t]$ is denoted as $\lambda_i(t)$, which is an increasing function of $t$. It is assumed that the following two kinds of maintenance actions can be taken (Feldman and Chen, 1996; Sheu and Chang, 2009):

1) Preventive replacement. The $i$-th element is replaced when it reaches an age $T_i$. The cost $c_{pi}$ of each replacement is constant. The average time for each replacement of element $i$ is $t_{pi}$.

2) Minimal repair. This action is used when the element fails between two consecutive replacements. Minimal repair resumes the failed element to work without affecting its hazard function. The average cost for a minimal repair of element $i$ is $c_{mi}$. The average time for a minimal repair of element $i$ is $t_{mi}$.

Another important assumption is that repair and replacement times are significantly shorter than the time periods between failures.

### 4.1.2. The formulation of system maintenance cost

The expected total maintenance cost for an element during the system life cycle is the total expected preventive replacement cost and minimal repair cost for the element. The expected total maintenance cost is the sum of these costs for all $M$ elements. That is

$$C_{tot} = \sum_{i=1}^{M} [n_i c_{pi} + l_i], \tag{4.1}$$

where $n_i = \frac{T_c}{T_i} - 1$ is the number of preventive replacements during the system life cycle for element $i$, and $l_i$ is the expected minimal repair cost for element $i$.

The average number of failures during the period between replacements $\lambda_i(T_i)$ can be obtained by using the replacement interval $T_i$ for each element. Furthermore, the total expected number of failures of the element $i$ during the system life cycle can be obtained as

$$(n_i + 1)\lambda_i(T_i) = \frac{\lambda_i(T_i)T_c}{T_i}. \tag{4.2}$$

From (4.2), we can obtain the availability of element $i$ as

$$A_i = \frac{T_c - t_{pi}(T_c / T_i - 1) - t_{mi}\lambda_i(T_i)T_c/T_i}{T_c} \tag{4.3}$$

and the expected minimal repair cost for element $i$ as

$$l_i = \frac{c_{mi}\lambda_i(T_i)T_c}{T_i} \tag{4.4}$$

From (4.1) and (4.4), we can obtain the expected total maintenance cost during the system life cycle as

$$C_{tot} = \sum_{i=1}^{M} [c_{pi}(T_c / T_i - 1) + \frac{c_{mi}\lambda_i(T_i)T_c}{T_i}] \tag{4.5}$$

### 4.1.3. The formulation of elements allocation

The elements allocation problem can be considered as a problem of partitioning a set $E$ of $M$ elements into $N$ mutually disjoint subsets $E_n$ ($1 \le n \le N$) such that

$$\bigcup_{i=1}^{N} E_n = E, \tag{4.6}$$

$$E_i \bigcap \mathrm{E_j} = \Phi \ , \mathrm{i} \neq \mathrm{j} \tag{4.7}$$

where each set $E_i$ corresponds to LMCCS node $C_i$ and can contain from 0 to $M$ elements. The partition of the set $E$ can be represented by the vector $\boldsymbol{H}=\{h(i),1\le i\le M\}$, where $h(i)$ denotes the number of the subset to which element $i$ belongs. The cardinality of each subset $E_i$ can be easily obtained as

$$\left|E_i\right| = \sum_{j=1}^{M} 1(h(j) = i). \tag{4.8}$$

### 4.1.4. The combined optimization problem

The combined element allocation and maintenance optimization problem is to find the optimal positions and replacement intervals for the system elements which minimize the total system maintenance cost subject to a pre-specified system availability requirement. The general formulation of the problem can be presented as follows:

Find vectors $\boldsymbol{H}=\{h(1),\ h(2),\ldots,\ h(M)\}$ and $\boldsymbol{T}=\{T_1,T_2,\ldots,T_M\}$ that minimize the total maintenance cost.

$$\boldsymbol{H},\ \boldsymbol{T} = \arg\min\{C_{tot} = \sum_{i=1}^{M}[c_{pi}(T_c / T_i - 1) + \frac{c_{mi}\lambda_i(T_i)T_c}{T_i}]\} \quad \text{subject to}: A(\boldsymbol{H},\ \boldsymbol{T}) \geq A^* \tag{4.9}$$

where $A^*$ is some preliminary specified system availability requirement.

## 4.2. LMCCS availability estimation based on a universal generating function

The universal generating function (UGF) was introduced in Ushakov (1986) and proved to be extremely effective in evaluating reliability of complex multi-state systems. Much research has been done on incorporating UGF into reliability analysis of various *k*-out-of-*n* systems, series-parallel systems, weighted voting systems, acyclic information networks, and manufacturing systems (Ding et al. 2010; Li et al. 2010; Yeh, 2009; Youssef and Elmaraghy, 2008). The UGF of a discrete random value $X$ is defined as a polynomial

$$u(z) = \sum_{k=1}^{K-1} p_k z^{x_k}, \tag{4.10}$$

where the variable $X$ has $K$ possible values and $p_k$ is the probability that $X$ takes the value $x_k$.

### 4.2.1. UGF for group of elements allocated at the same position

Consider element $e_i$ located as position $C_n$. When the element is available, it connects the $n$-th node with the ($n$+1)-th, the ($n$+2)-th ,…, and the $\theta(n+g_i)$-th node, where $\theta(n+g_i)=\min\{n+g_i,N+1\}$. When the element is unavailable, it is not able to connect the $n$-th node with any further remote positions. Thus the states of the element can be represented by the following UGF

$$u_{in}(z)=\sum_{h=1}^{H} p_{inh} z^{k_{inh}}, \tag{4.11}$$

where $H$=2, $p_{in1}=(1-A_i)$, $k_{in1}=n$, $p_{in2}=A_i$, $k_{in2}=\theta(n+g_i)$.

Let random value $T_n$ be the number of the most remote position which can be reached by elements allocated on node $C_n$. When there are multiple elements allocated on $C_n$, the most remote position to which $C_n$ can be connected is determined by the available element which has the greatest connecting range. To capture this feature, the following composition operator $\underset{\max}{\otimes}$ is used to obtain the combined UGF of a pair of elements

$$\underset{\max}{\otimes}(u_{in}(z),u_{jn}(z))=\underset{\max}{\otimes}\left(\sum_{h=1}^{H} p_{inh} z^{k_{inh}},\sum_{l=1}^{L} p_{jnl} z^{k_{jnl}}\right)=\sum_{h=1}^{H}\sum_{l=1}^{L} p_{inh} p_{jnl} z^{\max(k_{inh},k_{jnl})}. \tag{4.12}$$

One can see that the operator $\underset{\max}{\otimes}$ satisfies the following conditions:

$$\underset{\max}{\otimes}(u_{in}(z),u_{jn}(z))=\underset{\max}{\otimes}(u_{jn}(z),u_{in}(z)) \tag{4.13}$$

and

$$\underset{\max}{\otimes}(\underset{\max}{\otimes}(u_{in}(z),u_{jn}(z)),u_{kn}(z)) = \underset{\max}{\otimes}(u_{in}(z),\underset{\max}{\otimes}(u_{jn}(z),u_{kn}(z))) \tag{4.14}$$

Therefore, the UGF $u_n(z)$ for the group of elements allocated at $C_n$ can be obtained by sequentially applying the composition operator $\underset{\max}{\otimes}$.

In the case when node $n$ contains no elements, no arc exists from $C_n$ to any other node. In this case, the corresponding $u_n(z)$ takes the form

$$u_n(z) = z^n. \tag{4.15}$$

### 4.2.2. UGF for the entire LMCCS

Let random value $Y_n$ be the number of the farthest position that can be reached by the elements allocated at the first $n$ positions. The probabilistic distribution of $Y_n$ is denoted by u- function $U_n(z)$. According to the definitions of $T_n$ and $Y_n$, it can be seen that $Y_1=T_1$ and $U_1(z)=u_1(z)$.

For an arbitrary pair of adjacent positions $C_n$ and $C_{n+1}$, the paths provided by the elements belonging to the first $n$ nodes can be continued by the elements allocated at position $n+1$ only if $Y_n \geq n+1$ (the path reaches $C_{n+1}$). If this condition is satisfied, the most remote position that can be reached by elements allocated on the first $n+1$ positions can be determined as $Y_{n+1}=\max\{Y_n,T_{n+1}\}$.

In order to consider only the combinations of states of elements from the first $n$ positions which make the path from $C_1$ to $C_{n+1}$ exist, the following $\phi$ operator is used to eliminate the term with $Y_n=n$ from $U_n(z)$

$$\phi(U_n(z)) = \phi(\sum_{j=n}^{N+1} q_{nj} z^j) = \sum_{j=n+1}^{N+1} q_{nj} z^j. \tag{4.16}$$

Having the distributions of $Y_n$ and $T_{n+1}$, represented by $U_n(z)$ and $u_{n+1}(z)$ respectively, the UGF $U_{n+1}(z)$ representing distribution of $Y_{n+1}$ can be determined as

$$U_{n+1}(z) = \underset{\max}{\otimes}(\phi(U_n(z)), u_{n+1}(z)) \tag{4.17}$$

By sequentially applying (4.17), one can obtain $U_N(z)$ containing two terms corresponding to $Y_N=N$ and $Y_N=N+1$. $\phi(U_N(z))$ has only one term corresponding to the probability that the path from $C_1$ to $C_{N+1}$ exists. The coefficient of this term is equal to LMCCS availability $A$.

### 4.2.3. Computational complexity analysis

Since the farthest position that can be connected by each element $i$ allocated at node $C_n$ has at most $N+2-n$ states (from $n$ to $N+1$), combining UGF of any pair of elements allocated at node $C_n$ by (4.12) has a computational complexity $O(N^2)$. As the number of elements allocated at node $C_n$ equals to $|E_n|$, calculating $u_n(z)$ by sequentially combining the UGF of elements allocated at node $C_n$ has a computational complexity $O(|E_n| \cdot N^2)$.

Furthermore, applying (4.16) and (4.17) to calculate each $U_{n+1}(z)$ based on $U_n(z)$ and $u_n(z)$ for $n$=1,…, $N$-1 has a complexity $O(N^2)$. Thus the calculation of $U_N(z)$ has a complexity

$$(N-1)O(N^2)+\sum_{n=1}^{N} O(|E_n|\cdot N^2) = O(N^2(M+N)).$$

## 4.3. Optimization technique

Equation (4.9) formulates a complicated combinatorial optimization problem. An exhaustive examination of all possible solutions is not realistic, considering reasonable time limitations. The genetic algorithm (GA) has proven to be an effective optimization tool for a large number of complicated problems in reliability engineering, and it is used for our optimization. To apply the GA to a specific problem the solution representation and the decoding procedures must be defined.

### 4.3.1. Solution representation

Each solution is represented by string $\boldsymbol{S}=\{s_1,s_2,\ldots,s_M\}$, where $s_i$ corresponds to element $i$ for each $i$=1,2,…,$M$.

Each number $s_i$ determines both the number of the node onto which element $i$ is allocated ($h(i)$) and the replacement interval of element $i$ ($T_i$). To provide this property all the numbers $s_i$ are generated in the range

$$0 \le s_i < N \cdot \Lambda \tag{4.18}$$

where $\Lambda$ is the total number of considered replacement interval alternatives.

### 4.3.2. Solution decoding procedures

Step 1: Obtain the vectors (***H***, ***T***) representing the position and replacement interval of each system element with the following procedures.

For a given $\boldsymbol{S}=\{s_1,s_2,\ldots,s_M\}$, calculate

$$h(i) = [s_i / \Lambda] + 1 \tag{4.19}$$

$$v_i = 1 + \mathrm{mod}_\Lambda s_i \tag{4.20}$$

where $v_i$ is the number of replacement interval alternative for element $i$, $[x]$ is the maximal integer not greater than $x$, and $\mathrm{mod}_x y = y - [y/x]x$. The possible replacement interval alternatives are ordered in vector $\boldsymbol{Q}=\{q_1,q_2,\ldots,q_\Lambda\}$ so that $q_i<q_{i+1}$, where $q_i$ represents the replacement interval that corresponds to alternative $i$. After obtaining $v_i$ from decoding the solution string, the replacement interval for element $i$ can be obtained as

$$T_i = q_{v_i} \tag{4.21}$$

Step 2: For each given pair of vectors (***H***,***T***), first determine the availability of each element $A_i$ using (4.3) and then calculate the total expected system maintenance cost $C_{tot}$ using (4.5).

Step 3: From the vector $\boldsymbol{H}$, determine the $N$ mutually disjoint subsets $E_n$ representing the elements allocated on the first $N$ nodes of LMCCS.

Step 4: Obtain the entire system availability index $A$ using the procedures presented in section 4.2.

Step 5: In order to let the genetic algorithm search for the solution with minimal maintenance cost, when $A$ is not less than the required value $A^*$, the solution quality (fitness) is evaluated as follows:

$$F(\boldsymbol{H},\boldsymbol{T}) = \omega \cdot (A^* - A) \cdot 1(A^* - A) + \sum_{i=1}^{M} \left[ c_{pi}(T_c / T_i - 1) + \frac{c_{mi}\lambda_i(T_i)T_c}{T_i} \right] \tag{4.22}$$

where $\omega$ is a sufficiently large penalty.

For solutions that meet the requirements $A \geq A^*$, the fitness of the solution is equal to its total cost.

### 4.3.3. Crossover and mutation procedures

The cross operator for given parent strings $\boldsymbol{P1}$, $\boldsymbol{P2}$ and the offspring string $\boldsymbol{O}$ is defined as follows: the $i$-th element ($1 \leq i \leq M$) of the string $\boldsymbol{O}$ is equal to the $i$-th element of either $\boldsymbol{P1}$ or $\boldsymbol{P2}$ both with probability 0.5.

The mutation procedure swaps elements initially located in two randomly chosen positions.

## 4.4. Illustrative example

Consider an LMCCS consisting of 9 nodes. $M$=8 elements are to be allocated onto the first 8 nodes. The lifetime of the system $T_c$ is 120 months. We assume that $\Lambda$=8 different replacement frequency alternatives are considered and the alternatives are $\boldsymbol{h}$={29,24,19,14,9,4,2,1}. The replacement intervals corresponding to these alternatives are 4 months, 4.8 months, 6 months, 8 months, 12 months, 24 months, 40 months and 60 months respectively. The problem is to find the optimal positions and replacement intervals for all the 8 system elements so that the total system maintenance cost is minimized and the system availability requirement $A^*$ is satisfied. The characteristics of the elements are presented in Table 4.1.

According to (4.18) we have

$$0 \le s_i < 64$$

For any given solution string, $\boldsymbol{H}$ and $\boldsymbol{T}$ can be decoded using (4.19), (4.20) and (4.21). Thereafter (4.22) can be used to obtain the fitness function $\boldsymbol{F}(\boldsymbol{H},\boldsymbol{T})$. The optimal element allocation and maintenance strategy $(\boldsymbol{H},\boldsymbol{T})$ which minimizes $F(\boldsymbol{H},\boldsymbol{T})$ can be found by genetic algorithm.

Table 4.1 The characteristics of the elements

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $g_i$ | 3 | 4 | 3 | 2 | 1 | 1 | 4 | 3 |
| $t_{pi}$(month) | 0.003 | 0.003 | 0.003 | 0.004 | 0.004 | 0.003 | 0.004 | 0.002 |
| $c_{pi}$ | 100 | 110 | 100 | 80 | 50 | 45 | 105 | 95 |
| $t_{mi}$(month) | 0.036 | 0.042 | 0.040 | 0.035 | 0.025 | 0.025 | 0.045 | 0.034 |
| $c_{mi}$ | 2 | 3 | 2 | 2 | 1 | 2 | 2 | 2 |
| $\lambda_i(4)$ | 0.8 | 0.72 | 0.64 | 0.6 | 0.72 | 0.7 | 0.6 | 0.5 |
| $\lambda_i(4.8)$ | 1.04 | 0.92 | 0.85 | 0.8 | 0.96 | 0.9 | 0.85 | 0.8 |
| $\lambda_i(6)$ | 1.6 | 1.5 | 1.4 | 1.2 | 1.5 | 1.4 | 1.4 | 1.4 |
| $\lambda_i(8)$ | 2.8 | 2.7 | 2.7 | 2.1 | 2.7 | 2.4 | 2.7 | 2.8 |
| $\lambda_i(12)$ | 5.5 | 5.2 | 5.4 | 4.2 | 5.3 | 4.8 | 5.4 | 5.8 |
| $\lambda_i(24)$ | 15 | 15 | 15 | 14 | 15 | 14 | 14 | 16 |
| $\lambda_i(40)$ | 33 | 32 | 31 | 32 | 33 | 32 | 31 | 33 |
| $\lambda_i(60)$ | 58 | 57 | 54 | 57 | 58 | 57 | 56 | 58 |

### 4.4.1. The fitness function for a given solution string

As an illustration, the fitness function for the solution string $\boldsymbol{S}$={3,15,21,46,47,34,54,57} is obtained by the following procedures:

Step 1: $\boldsymbol{S}$ is decoded into $\boldsymbol{H}$=(1,2,3,6,6,5,7,8) and $\boldsymbol{T}$= (8,60,24,40,60,6,40,4.8) using (4.19), (4.20) and (4.21).

Step 2: The availability for each system element is calculated using (4.3) as $A_1$=0.9870, $A_2$=0.9601, $A_3$=0.9749, $A_4$=0.9719, $A_5$=0.9758, $A_6$=0.9937, $A_7$=0.9651, and $A_8$=0.9939. The total maintenance cost is calculated using (4.5) as $C_{tot}$=6631.

Step 3: From the vector $\boldsymbol{H}$, the $N$ mutually disjoint subsets $E_n$ representing the elements allocated on the first $N$ nodes of LMCCS are obtained as $E_1$={1}, $E_2$={2}, $E_3$={3}, $E_4$= $\Phi$, $E_5$={6}, $E_6$={4, 5}, $E_7$={7}, $E_8$={8}. The allocation of elements in LMCCS is shown in Figure 4.1.

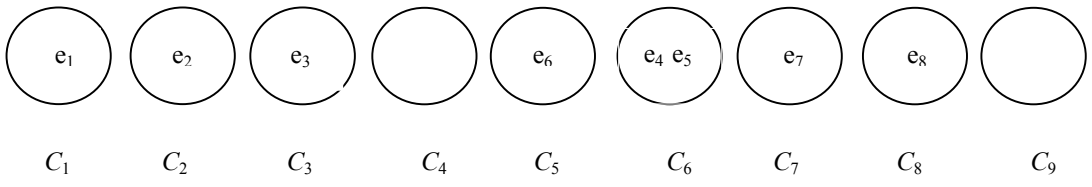


Figure 4.1 The structure of the LMCCS

Step 4: From (4.11) we can have the UGF for each element as

$$u_{11}(z) = (1-A_1)z^1 + A_1z^4,\ u_{22}(z) = (1-A_2)z^2 + A_2z^6, u_{33}(z) = (1-A_3)z^3 + A_3z^6,$$

$$u_{46}(z) = (1-A_4)z^6 + A_4z^8,\ u_{56}(z) = (1-A_5)z^6 + A_5z^7, u_{65}(z) = (1-A_6)z^5 + A_6z^6,$$

$$u_{77}(z) = (1-A_7)z^7 + A_7z^9,\ u_{88}(z) = (1-A_8)z^8 + A_8z^9$$

Furthermore, the UGF $u_n(z)$ for the group of elements allocated at $C_n$ can be obtained as

$$u_1(z) = u_{11}(z) = (1-A_1)z^1 + A_1z^4,\ u_2(z) = u_{22}(z) = (1-A_2)z^2 + A_2z^6,$$

$$u_3(z) = u_{33}(z) = (1-A_3)z^3 + A_3z^6,\ u_4(z) = z^4,\ u_5(z) = u_{65}(z) = (1-A_6)z^5 + A_6z^6,$$

$$\begin{aligned} u_6(z) &= \underset{\max}{\otimes}(u_{46}(z), u_{56}(z)) = \underset{\max}{\otimes}((1-A_4)z^6 + A_4z^8, (1-A_5)z^6 + A_5z^7) \\ &= (1-A_4)(1-A_5)z^6 + (1-A_4)A_5z^7 + A_4(1-A_5)z^8 + A_4A_5z^8 \\ &= (1-A_4)(1-A_5)z^6 + (1-A_4)A_5z^7 + A_4z^8 \end{aligned},$$

$$u_7(z) = u_{77}(z) = (1-A_7)z^7 + A_7z^9, u_8(z) = u_{88}(z) = (1-A_8)z^8 + A_8z^9$$

The UGF for the entire LMSSC can be obtained by applying (4.17) sequentially as

$$\begin{aligned} U_2(z) &= \underset{\max}{\otimes}(\phi(U_1(z)), u_2(z)) = \underset{\max}{\otimes}(\phi(u_1(z)), u_2(z)) \\ &= \underset{\max}{\otimes}(\phi((1-A_1)z^1 + A_1z^4), (1-A_2)z^2 + A_2z^6) \\ &= \underset{\max}{\otimes}(A_1z^4, (1-A_2)z^2 + A_2z^6) = A_1(1-A_2)z^4 + A_1A_2z^6 \end{aligned},$$

$$U_3(z) = \underset{\max}{\otimes}(\phi(U_2(z)), u_3(z)) = \underset{\max}{\otimes}(\phi(A_1(1-A_2)z^4 + A_1A_2z^6), (1-A_3)z^3 + A_3z^6)$$
$$= \underset{\max}{\otimes}(A_1(1-A_2)z^4 + A_1A_2z^6, (1-A_3)z^3 + A_3z^6) = A_1(1-A_2)(1-A_3)z^4 + [A_1(1-A_2)A_3 + A_1A_2]z^6$$

,

$$U_4(z) = \underset{\max}{\otimes}(\phi(U_3(z)), u_4(z)) = \underset{\max}{\otimes}(A_1(1-A_2)(1-A_3)z^4 + [A_1(1-A_2)A_3 + A_1A_2]z^6, z^4)$$
$$= A_1(1-A_2)(1-A_3)z^4 + [A_1(1-A_2)A_3 + A_1A_2]z^6 \quad ,$$

$$U_5(z) = \underset{\max}{\otimes}(\phi(U_4(z)), u_5(z)) = \underset{\max}{\otimes}([A_1(1-A_2)A_3 + A_1A_2]z^6, (1-A_6)z^5 + A_6z^6)$$
$$= [A_1(1-A_2)A_3 + A_1A_2]z^6 \quad ,$$

$$U_6(z) = \underset{\max}{\otimes}(\phi(U_5(z)), u_6(z)) = \underset{\max}{\otimes}([A_1(1-A_2)A_3 + A_1A_2]z^6,$$
$$(1-A_4)(1-A_5)z^6 + (1-A_4)A_5z^7 + A_4z^8) = [A_1(1-A_2)A_3 + A_1A_2](1-A_4)(1-A_5)z^6 ,$$
$$+ [A_1(1-A_2)A_3 + A_1A_2](1-A_4)A_5z^7 + [A_1(1-A_2)A_3 + A_1A_2]A_4z^8$$

$$U_7(z) = \underset{\max}{\otimes}(\phi(U_6(z)), u_7(z)) = \underset{\max}{\otimes}([A_1(1-A_2)A_3 + A_1A_2](1-A_4)A_5z^7 + [A_1(1-A_2)A_3 + A_1A_2]A_4z^8,$$
$$(1-A_7)z^7 + A_7z^9) = [A_1(1-A_2)A_3 + A_1A_2](1-A_4)A_5(1-A_7)z^7 + [A_1(1-A_2)A_3 + A_1A_2]A_4(1-A_7)z^8$$
$$+ \{[A_1(1-A_2)A_3 + A_1A_2]*[A_4A_7 + (1-A_4)A_5A_7]\}z^9$$
$$,$$

$$U_8(z) = \underset{\max}{\otimes}(\phi(U_7(z)), u_8(z)) = \underset{\max}{\otimes}([A_1(1-A_2)A_3 + A_1A_2]A_4(1-A_7)z^8$$
$$+ \{[A_1(1-A_2)A_3 + A_1A_2]*[A_4A_7 + (1-A_4)A_5A_7]\}z^9, (1-A_8)z^8 + A_8z^9) =$$
$$[A_1(1-A_2)A_3 + A_1A_2]A_4(1-A_7)(1-A_8)z^8 +$$
$$\{[A_1(1-A_2)A_3 + A_1A_2]*[A_4A_7 + (1-A_4)A_5A_7 + A_4(1-A_7)A_8]\}z^9$$

The system reliability which equals to the coefficient of $z^9$ in $U_8(z)$ can be calculated as

$$A = [A_1(1-A_2)A_3 + A_1A_2]*[A_4A_7 + (1-A_4)A_5A_7 + A_4(1-A_7)A_8]$$
$$= 0.9861*[0.9380 + 0.0264 + 0.0338) = 0.9843$$

Step 5: The fitness function is calculated as

$$F(\boldsymbol{H},\boldsymbol{T}) = \omega \cdot (A^* - A) \cdot 1(A^* - A) + C_{tot}$$
$$= \begin{cases} 6631, A^* < 0.9843 \\ 6631 + \omega \cdot (A^* - 0.9843), A^* \geq 0.9843 \end{cases}$$

### 4.4.2. The optimization problem

The problem is to find the optimal element allocation and maintenance strategy ($\boldsymbol{H}$,$\boldsymbol{T}$) which minimizes $F(\boldsymbol{H},\boldsymbol{T})$. In order to show the influence of element allocation, the optimization problem is solved for three different cases: 1) Fixed element allocation; 2) Even elements distribution among the nodes (no node contains more than one element); 3) Arbitrary allocation of the elements.

**Case 1: Fixed element allocation**

Table 4.2 contains the optimal solutions obtained for different values of $A^*$ with fixed $\boldsymbol{H}$=(1,2,3,4,5,6,7,8). Each solution was obtained as the optimal one among five different runs of the GA with different randomly generated initial populations. The coefficients of variation among the values of $F(\boldsymbol{H},\boldsymbol{T})$ obtained in the five runs are also presented in Table 4.2. The low values of this coefficient evidence the good consistency of the GA.

With the increase of the availability requirement the elements need to be replaced more frequently, thus the total maintenance cost increases. The minimal availability $A$=0.9026 is achieved when $\boldsymbol{T}$=(60,60,60,60,60,60,60,60) and the corresponding total maintenance cost is $C_{tot}$=2503. The maximum availability $A$=0.9794 is achieved when $\boldsymbol{T}$=(4,4,4,4,4,4,4,4) and the corresponding total maintenance cost is $C_{tot}$=20182.

Table 4.2 Examples of solutions obtained for fixed elements distribution

| Constraints | $\boldsymbol{H}$ | $\boldsymbol{T}$ | $A$ | $F(\boldsymbol{T},\boldsymbol{x})$ | variation |
|---|---|---|---|---|---|
| $A^*$=0.90 | (1,2,3,4,5,6,7,8) | (60,60,60,60,60,60,60,60) | 0.9026 | 2503 | 0 |
| $A^*$=0.95 | (1,2,3,4,5,6,7,8) | (24,60,60,40,60,6,12,24) | 0.9512 | 4340 | 0.46% |
| $A^*$=0.97 | (1,2,3,4,5,6,7,8) | (8,60,60,60,60,4.8,6,60) | 0.9709 | 6229 | 0.90% |

**Case 2: Even elements distribution among the nodes**

Table 4.3 contains the optimal solutions obtained for different values of $A^*$ when no more than one element is allowed to be allocated onto the same node. This is achieved by adding a large penalty to the fitness function when the corresponding $\boldsymbol{H}$ has at least two equal elements. Each solution was obtained as the optimal one among five different runs of the GA with different randomly generated initial populations. The coefficients of variation among the values of $F(\boldsymbol{H},\boldsymbol{T})$ obtained in the five runs are also presented in Table 4.3 to illustrate the good consistency of the GA.

With the increase of the availability requirement the elements need to be replaced more frequently, thus the total maintenance cost increases. Comparing with Table 4.2, less maintenance cost is needed to reach the same level of availability requirement. It can be seen that for $A^*$=0.90 and 0.95, only the least frequent replacements $\boldsymbol{T}$=(60,60,60,60,60,60,60,60) are needed to meet the availability requirement if the elements are allocated appropriately. In these cases the system reliability improvement is achieved solely by the optimal elements' allocation.

Table 4.3 Examples of solutions obtained for even elements distribution

| Constraints | $\boldsymbol{H}$ | $\boldsymbol{T}$ | $A$ | $F(\boldsymbol{T},\boldsymbol{x})$ | variation |
|---|---|---|---|---|---|
| $A^*$=0.90 | (2,8,4,6,7,1,3,5) | (60,60,60,60,60,60,60,60) | 0.9045 | 2503 | 0 |
| $A^*$=0.95 | (7,6,8,3,5,2,1,4) | (60,60,60,60,60,60,60,60) | 0.9547 | 2503 | 0 |
| $A^*$=0.97 | (1,6,4,5,3,7,8,2) | (24,60,60,60,60,60,60,60) | 0.9726 | 2721 | 0.48% |
| $A^*$=0.98 | (6,2,3,1,5,8,7,4) | (60,60,60,12,60,60,60,60) | 0.9835 | 2999 | 0 |
| $A^*$=0.99 | (5,4,6,1,8,7,2,3) | (60,60,60,6,60,60,60,60) | 0.9901 | 3763 | 0.88% |
| $A^*$=0.992 | (4,2,6,1,7,8,5,3) | (60,40,60,4.8,60,60,60,60) | 0.9921 | 4211 | 0.24% |

**Case 3: Arbitrary allocation of the elements**

Table 4.4 contains the optimal solutions obtained for different values of $A^*$ when multiple elements are allowed to be allocated onto the same node. Each solution was obtained as the optimal one among five different runs of the GA with different randomly generated initial populations. The coefficients of variation among the values of $F(\boldsymbol{H},\boldsymbol{T})$ obtained in the five runs are also presented in Table 4.4. The low values of this coefficient evidence the good consistency of the GA.

Comparing with Table 4.2 and Table 4.3, higher levels of availability requirements can be achieved with less maintenance cost. It can be seen that even for availability

requirement as high as $A^*$=0.998, only the least frequent replacements $\boldsymbol{T}$=(60,60,60,60,60,60,60,60) are needed if the elements are allocated appropriately. The extremely high availability requirement $A^*$=0.9999 can be achieved with just a maintenance cost as high as 8431.

Table 4.4 Examples of solutions obtained for arbitrary elements distribution

| Constraints | $\boldsymbol{H}$ | $\boldsymbol{T}$ | $A$ | $F(\boldsymbol{T},\boldsymbol{x})$ | variation |
|---|---|---|---|---|---|
| $A^*$=0.997 | (8,4,1,8,8,4,1,5) | (60,60,60,60,60,60,60,60) | 0.9971 | 2503 | 0 |
| $A^*$=0.998 | (1,4,5,4,8,8,1,6) | (60,60,60,60,60,60,60,60) | 0.9980 | 2503 | 0 |
| $A^*$=0.999 | (1,1,6,7,5,4,5,4) | (24,40,60,60,60,60,60,60) | 0.9991 | 2777 | 0.32% |
| $A^*$=0.9992 | (6,1,3,1,5,5,1,6) | (24,60,60,60,60,60,40,40) | 0.99925 | 2849 | 1.16% |
| $A^*$=0.9995 | (4,5,6,7,4,5,1,1) | (24,24,24,40,60,60,24,12) | 0.99950 | 4087 | 1.44% |
| $A^*$=0.9999 | (1,4,4,7,8,8,4,1) | (8,6,40,60,40,12,12,4.8) | 0.99990 | 8431 | 0.94% |

## 4.5. Conclusions

This chapter presents a framework to solve the joint element allocation and maintenance optimization problem for linear multi-state consecutively connected systems. An example of such system is a set of radio relay stations in which multi-state retransmitters with different characteristics are allocated. Since a linear consecutively connected system is not symmetrical, the availability of such a system is not only related to the respective availability of each element but also to the arrangement of the elements. It is shown that through optimally allocating the elements onto different nodes one can reduce the maintenance cost needed to meet a pre-specified availability requirement.

A universal generating function is used to evaluate the availability of the system and a genetic algorithm is adopted for the joint elements allocation and maintenance optimization. The application of the proposed framework is illustrated by numerical examples. The optimal elements allocation and maintenance strategy are found in the example for three different cases: 1) Fixed element allocation; 2) Even elements distribution among the nodes (no node contains more than one element); 3) Arbitrary allocation of the elements. For all the cases, the minimum maintenance cost increases with the increase of the availability requirement. It is revealed clearly in the results that the flexibility of element allocation enables the system to achieve much higher availability with less maintenance cost.

# CHAPTER 5 SYSTEM DEFENSE WITH IMPERFECT FALSE TARGETS

For a system under intentional attacks, the attacker can take advantage of its knowledge about the system to optimize its attacking strategy so as to incur maximum expected damage to the system (Bier et al., 2005; Patterson and Apostolakis, 2007; Xiao et al., 2008). Thus it is important for the defender to take into consideration the attacker's strategy when he decides how to allocate its resource among several defensive measures (Dighe et al., 2009; Powell, 2007a; Powell, 2007b).

As the two simplest systems, the protections of series systems and parallel systems against intentional attacks have been discussed in many papers, such as Bier and Abhichandani (2002), Bier et al. (2005), and Hausken (2008). The protection is a technical or organizational measure which is aimed to reduce the vulnerability of protected system elements. The vulnerability of each element is its destruction probability when it is attacked. It can be determined by an attacker-defender contest success function. The contest between the defender and the attacker is usually modeled as a two-period game (Azaiez and Bier, 2007; Levitin and Hausken, 2008). The defender moves first to

distribute its defending resource among different components to minimize the expected damage to the system assuming that the attacker will use the most harmful strategy to attack. When the attacker moves, it has full knowledge of the defender's resource allocation and it can optimally allocate its attacking resource so that the expected damage to the system is maximized. In these papers the optimal resource allocation problem is formulated as a minmax problem: the defender chooses its free choice variables to minimize the system vulnerability corresponding to the most harmful attacker's action.

Besides direct protections, deploying false targets (FTs) is another effective measure to defense systems under intentional attacks and it is an often employed strategy. The objective of a FT, sometimes referred to as a decoy, is to give the appearance that the element is something else than it actually is. A FT conceals or distracts something else, i.e. the genuine object, which the attacker actually searches for.

The aim of deploying FTs is to misinform the attacker so that the genuine element (GE) will be attacked with less probability or less attacking effort. Levitin and Hausken (2009a) has studied the efficiency of deploying FTs in defending a homogeneous parallel system. In Hausken and Levitin (2009), the defense strategy of deploying FTs in series systems is analyzed. Levitin and Hausken (2009b) studied the optimal resource allocation between constructing redundant genuine elements, protecting these elements and deploying false targets. All these papers assume that the FTs are perfect, that is, the attacker has no preference between attacking a genuine target and attacking a FT. In practice the false targets are after all different from the genuine target, and it is possible for the attacker to detect some of them.

In this chapter, we consider defense of simple series and parallel systems that includes both protecting the elements and deploying FTs to distract the attacker. These FTs are imperfect and the attacker can detect each of them with the same probability. From practical point of view, the detection probability of a false target can be estimated from past experiences or experiments. Once the attacker detects a certain number of FTs, it ignores them and chooses such number of remaining elements randomly to attack that maximizes the expected damage to the system. The defender decides how many FTs to deploy to minimize the expected damage caused by the attacks assuming that the attacker always uses the most harmful strategy to attack. The expected damage to a series system is proportional to the probability of system destruction. Depending on the type of the system, the expected damage to a parallel system can be defined in two ways: as proportional to the loss of demand probability (the probability that the demand is not met) or as the expected amount of the unsupplied demand.

Section 5.1 presents the general model. Section 5.2 analyzes the defense of a simple series system with imperfect FTs. Section 5.3 analyzes the defense of a homogeneous parallel system with imperfect FTs. Section 5.4 concludes.

## 5.1. The model

***Assumptions:***

1. The defender uses identical FTs with the same detection probability

2. The attacker can detect each FT independently from other FTs

3. The attacker knows the defender's effort distribution and number of GEs and FTs and decides how many elements to attack

4. The attacker distributes its resources evenly among the attacked elements

5. Each element is attacked separately. Single attack cannot destroy more than one element

6. In a parallel system the genuine elements have identical performance

7. The defender distributes its protection resources evenly among the genuine elements

A system consisting of $N$ identical genuine elements (GEs), which are connected either in series or in parallel. All system elements are exposed to intentional attacks. The defender and the attacker's resources, $r$ and $R$, are fixed. The unit costs for the attacker and the defender's efforts are $A$ and $a$ respectively. The defender distributes its resource among deploying $H$ FTs and protecting the GEs. Since an unprotected GE can be destroyed by an arbitrarily small but positive attack effort, we assume that the defender distributes its protection resource evenly among all the GEs. The cost for deploying one FT is $s$. The FTs are imperfect i.e. the attacker can detect each FT with probability $d$. If the attacker detects $k$ FTs (with probability $p_k = \binom{H}{k} d^k (1-d)^{H-k}$) it ignores the detected

FTs and attacks $Q_k$ randomly chosen elements out of $N$+$H$-$k$ remaining undetected elements, as shown in Figure 5.1.

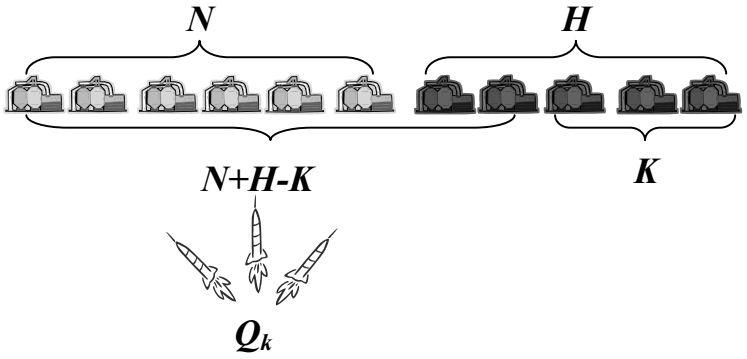


Figure 5.1 Graphical illustration of the model

The vulnerability (destruction probability) of the attacked object is determined by the attacker-defender contest success function modeled with the common ratio form (Tullock, 1980; Skaperdas, 1996; Hausken, 2005) as

$$v = \frac{T^m}{T^m + t^m} = \left[1 + (t/T)^m\right]^{-1} \tag{5.1}$$

where $T$ and $t$ are the efforts allocated to the element by the attacker and the defender respectively, and $m$ is a parameter that describes the intensity of the contest. Especially if an attacked element is without protection ($T$>0, $t$=0), the element will be destroyed with probability 1. When $m$=0, no matter what are the sizes of $T$ and $t$ the vulnerability of the element is 50%. When 0<$m$<1, there is a disproportional advantage of investing less than one's opponent. When $m$=1, the investments have proportional impact on the vulnerability.

When $m>1$, there is a disproportional advantage of investing more than one's opponent. When $m=\infty$, $v$ is a step function where "winner-takes-all".

For each $k$ the attacker solves the optimization problem and chooses the $Q_k$ which maximizes the expected damage to the system $D(Q_k,H)$. The entire expected damage is $D(H)=\sum_{k=0}^{H} p_k D(Q_k,H)$. The defender solves the minmax problem: finds $H$ that minimizes the maximal expected damage given that for any $H$ the attacker chooses vector $(Q_0,\ldots,Q_H)$ that maximizes the expected damage to the system $D(Q_k,H)$. Actually, this model applies also to the case when the attacker has no optimal strategy and chooses the value of $Q_k$ at random. The defender's most conservative strategy in this case is to anticipate the worst case scenario and assume that the attacker can guess the value of $Q_k$, which makes the attack most effective and harmful.

## 5.2. *N* genuine elements connected in series

The system consists of $N$ GEs connected in series. Destruction of any GE results in the destruction of the entire system. Since $H$ ($H\le r/s$) FTs are deployed, the defense effort exerted on each genuine target is $t=(r-Hs)/Na$. In the case when $k$ ($0\le k\le H$) FTs are detected by the attacker, it chooses $Q_k$ ($1\le Q_k\le N+H-k$) targets out of $N+H-k$ undetected elements to attack and the attack effort allocated onto each target is $T=R/(Q_kA)$. The vulnerability of each GE is

$$v = \frac{T^m}{T^m + t^m} = \frac{[R/(Q_k A)]^m}{[R/(Q_k A)]^m + [(r - Hs)/Na]^m} = \frac{R^m}{R^m + \varepsilon^m {Q_k}^m [(r - Hs)/N]^m} \tag{5.2}$$

where $\varepsilon = A/a$. For any $k$ and $Q_k$ the random number of attacked GEs can vary from max(0,$Q_k$–$H$+$k$) (all FTs are attacked) to min($N$,$Q_k$) (all genuine targets are attacked). The probability $\varphi(Q_k, i)$ that among $Q_k$ attacked elements $i$ elements are the genuine ones can be obtained using the hyper-geometric distribution:

$$\varphi(Q_k, i) = \frac{\binom{N}{i}\binom{H-k}{Q_k - i}}{\binom{N+H-k}{Q_k}} \tag{5.3}$$

The probability that at least one out of $i$ attacked GEs is destroyed is 1-(1-$v$)$^i$. The system destruction probability for any $k$ and $Q_k$ can be obtained as

$$D(Q_k, H) = \sum_{i=\max(0,Q_k-H+k)}^{\min(N,Q_k)} \varphi(Q_k, i) \cdot (1-(1-v)^i) = \sum_{i=\max(0,Q_k-H+k)}^{\min(N,Q_k)} \frac{\binom{N}{i}\binom{H-k}{Q_k - i}}{\binom{N+H-k}{Q_k}} \cdot \left\{1 - \left[\frac{1}{\frac{R^m N^m}{\varepsilon^m {Q_k}^m (r-Hs)^m} + 1}\right]^i\right\} \tag{5.4}$$

The attacker chooses the $Q_k$ which maximizes the expected damage to the system. Thus the most harmful $Q_k$ can be expressed by $Q_k^* = \arg\max_{1 \le Q_k \le N+H-k} D(Q_k, H)$.

Figure 5.2 presents the most effective attack strategies $Q_k^*$ (0≤$k$≤5) for $N$=5, $H$=5, $\varepsilon = 1$, $s$=0.1, $r$=1 and different $R$ and $m$. The attacker tends to attack more elements when

he has more resources. With the growth of the contest intensity it becomes more important for the attacker to achieve the effort superiority over the defender. Therefore the attacker concentrates greater per-target efforts by attacking fewer targets and $Q_k$ tends to decrease with the increase of $m$.

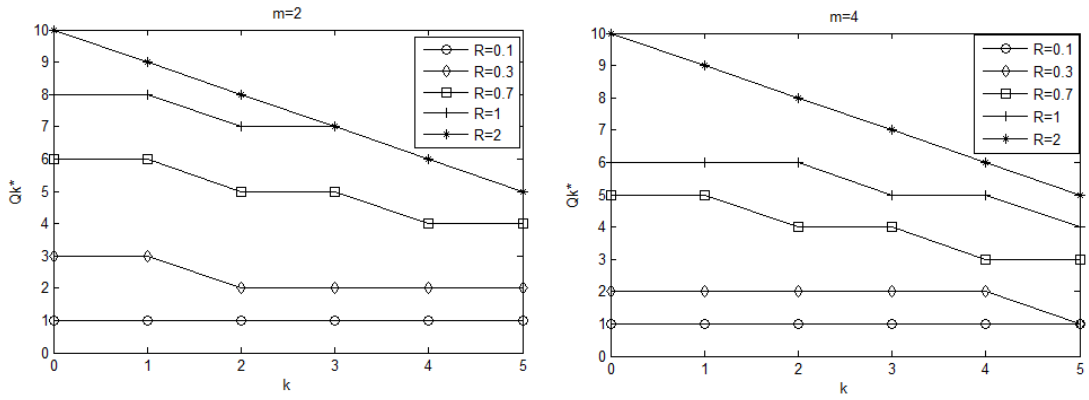


Figure 5.2 Optimal number of attacked targets for series systems

The total expected damage to the entire system is

$$D(H)=\sum_{k=0}^{H} p_k D(Q_k^*,H)=\sum_{k=0}^{H}\binom{H}{k} d^k (1-d)^{H-k} D(Q_k^*,H) \tag{5.5}$$

The defender chooses the $H$ which minimizes $D(H)$, thus we have the optimal number of deployed FTs $H^*=\arg\min_{0\le H\le \lfloor r/s \rfloor} D(H)$ .

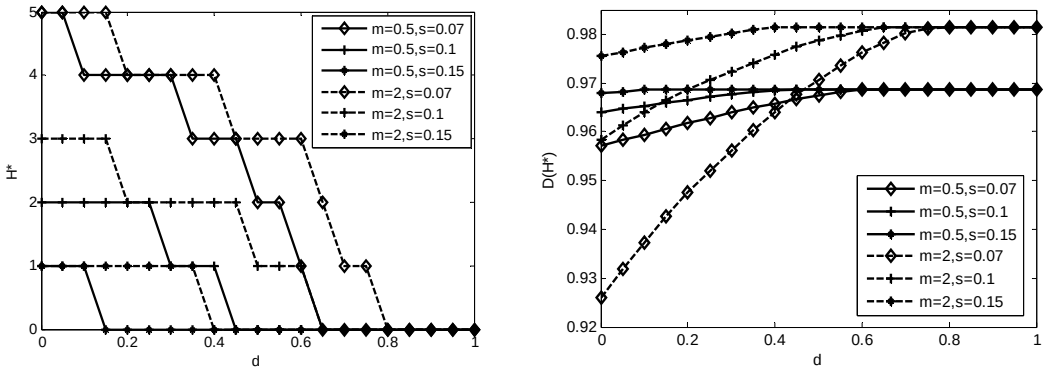


Figure 5.3 $H^*$ and $D(H^*)$ as functions of $d$ for series systems

Figure 5.3 presents the optimal number of FTs $H^*$ and the corresponding $D(H^*)$ as functions of $d$ for $N$=5, $R$=$r$=1 and $\varepsilon = 1$ and different combinations of $s$ and $m$. It can be seen that $H^*$ decreases with the increases of $s$ and $d$. Indeed, it is not cost-effective to build many FTs if they are too expensive or can be rather easily detected by the attacker. $D(H^*)$ increases with the increase of $s$ and $d$.

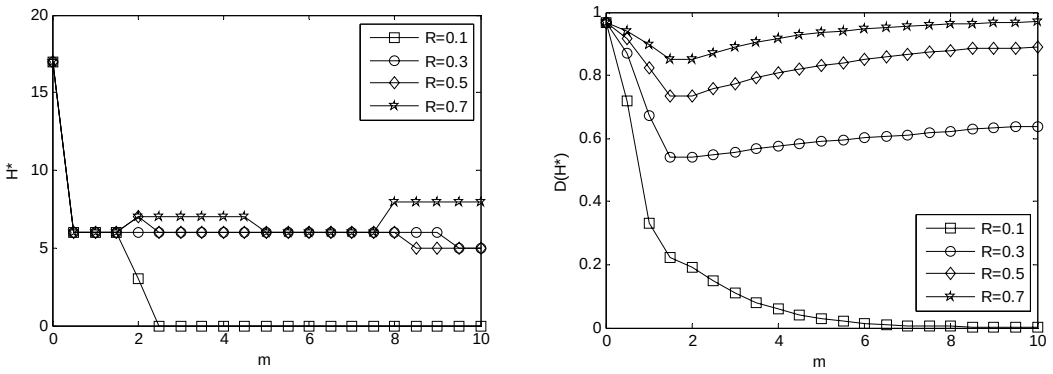


Figure 5.4 $H^*$ and $D(H^*)$ as functions of $m$ for series systems

It can be seen that $H^*$ and $D(H^*)$ are non-monotonic functions of $m$. Figure 5.4 represents the optimal number of FTs $H^*$ and $D(H^*)$ as functions of $m$ for $N$=5, $r$=1, $\varepsilon = 1$, $s$=0.05, $d$=0.4 and different values of $R$. When $R$=0.1 the defense effort is superior, the defender benefits from the increase of the contest intensity and $D(H^*)$ decreases with $m$. When $R$>0.1, $D(H^*)$ as function of $m$ demonstrates non-monotonic behavior. This can be explained by the fact that changes in optimal values of $H^*$ and $Q_k$ can make the defender's object protection effort either inferior or superior. In the former case $D(H^*)$ increases with $m$ whereas in the latter case $D(H^*)$ decreases with $m$.

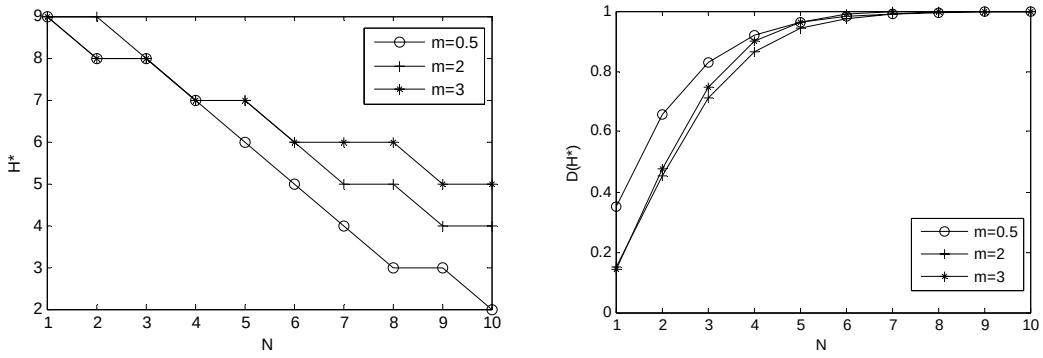


Figure 5.5 $H^*$ and $D(H^*)$ as functions of $N$ for series systems

Figure 5.5 presents the optimal number of FTs $H^*$ and the corresponding $D(H^*)$ as functions of $N$ for $R$=$r$=1, $\varepsilon = 1$, $s$=0.05, $d$=0.4 and different values of $m$. It can be seen

that $D(H^*)$ increases with the increase of $N$. Obviously, the vulnerability of a series system increases with the growth of the number of elements in the system.

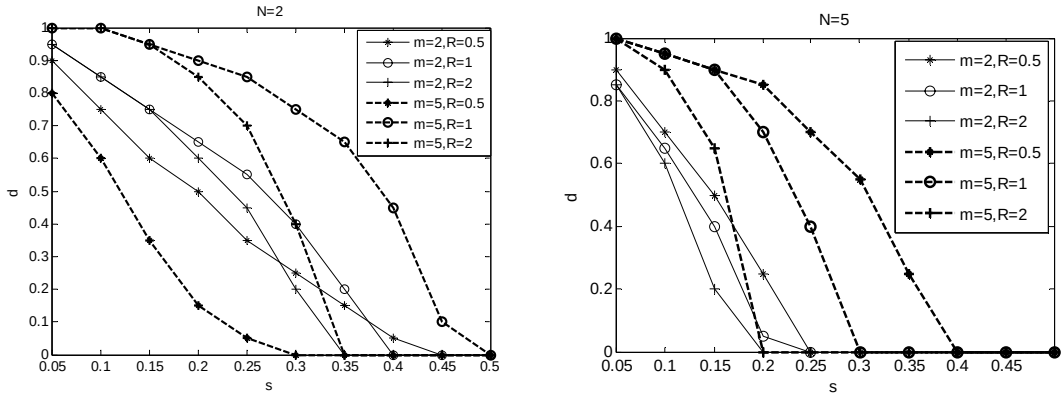


Figure 5.6 Efficiency analysis of deploying false targets for series systems

Figure 5.6 presents the false targets deployment efficiency curves for $r$=1, $\varepsilon = 1$ and different values of $N$, $m$, and $R$. For any pair of ($s$,$d$) above the curve the deployment of any false target is not beneficial for the defender ($H^*$=0). The ($s$,$d$) curves obtained for different combinations of model parameters can be used for making decisions about false targets deployment.

It is interesting that the critical value of $d$ (the maximal value when the deployment of at least one FT is beneficial for the defender for a given $s$) can depend on the attacker's resource $R$ non-monotonically. Figure 5.7 presents the critical value of $d$ as a function of $R$ for $m$=2, $r$=1, $\varepsilon = 1$, $N$=2 and different values of $s$.

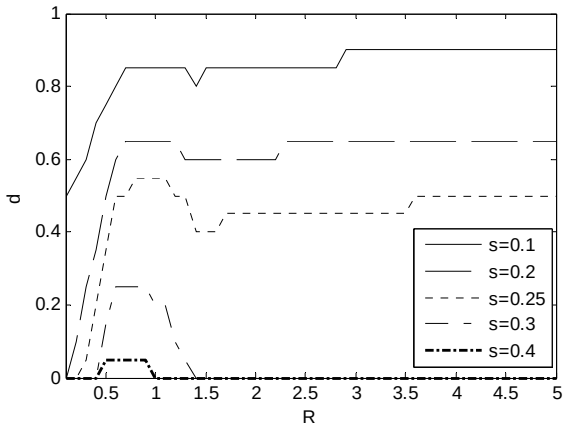


Figure 5.7 The critical value of $d$ as a function of $R$ for series systems

It can be seen that, when $s$=0.25, the deployment of the FTs is not efficient for small values of $R$, then with the growth of $R$, it becomes efficient and then from certain values of $R$ it becomes again inefficient. Indeed, when $R$ is much smaller than $r$ the defender obtains overwhelming superiority in the attack-protection contests and does not need any FT. When $R$ is much greater than $r$ the attacker has enough resources to attack all targets (including the FTs) preserving its superiority in the attack-protection contest and the FTs are not effective. The FTs are most effective when the amounts of the attacker's and the defender's resources are close. In this case the optimal deployment of the FTs can considerably reduce the damage. A numerical comparison of the expected damage is presented below.

***Numerical Comparison***

Here we show a numerical example and compare the efficiency of deploying FTs. Consider the FTs characterized by $s$=0.4 and $d$=0.2 when $N$=$m$=2, $r$= $\varepsilon = 1$ . Since $\lfloor r/s \rfloor = \lfloor 1/0.4 \rfloor = 2$ , $H$ can take only values of 0, 1 and 2.

We have

$$D(H=0) = \max(D(Q_0=1, H=0), D(Q_0=2, H=0)) = \max(\frac{R^5}{R^5+0.5^5}, 1-(R^5+1)^{-2}) \quad ,$$

$$\begin{aligned} D(H=1) &= p_0 D(Q_0^*, H=1) + p_1 D(Q_1^*, H=1) \\ &= 0.8\max(D(Q_0=1, H=1), D(Q_0=2, H=1), D(Q_0=3, H=1)) \\ &\quad + 0.2\max(D(Q_1=1, H=1), D(Q_1=2, H=1)) \quad , \\ &= 0.8\max(\frac{2}{3}\cdot\frac{R^5}{R^5+[0.3]^5}, \frac{2}{3}\cdot\frac{R^5}{R^5+[0.6]^5}+\frac{1}{3}\cdot(1-(\frac{[0.6]^5}{R^m+[0.6]^5})^2), \\ &\quad 1-(\frac{[0.9]^5}{R^m+[0.9]^5})^2) + 0.2\cdot\max(\frac{R^5}{R^5+[0.3]^5}, 1-(\frac{[0.6]^5}{R^5+[0.6]^5})^2) \end{aligned}$$

and

$$
\begin{aligned}
D(H=2) &= p_0 D(Q_0^*, H=2) + p_1 D(Q_1^*, H=2) + p_2 D(Q_2^*, H=2) \\
&= 0.64 \max(D(Q_0=1, H=2), D(Q_0=2, H=2), D(Q_0=3, H=2), D(Q_0=4, H=2)) \\
&\quad + 0.32 \max(D(Q_1=1, H=2), D(Q_1=2, H=2), D(Q_1=3, H=2)) \\
&\quad + 0.04 \max(D(Q_2=1, H=2), D(Q_2=2, H=2)) \\
&= 0.64 \max(0.5 \cdot \frac{R^5}{R^5+[0.1]^5}, \frac{1}{6}(1-(\frac{[0.2]^5}{R^5+[0.2]^5})^2) + \frac{2}{3} \cdot \frac{R^5}{R^5+[0.2]^5}, \\
&\quad \frac{1}{2}(1-(\frac{[0.3]^5}{R^5+[0.3]^5})^2) + \frac{1}{2} \cdot \frac{R^5}{R^5+[0.3]^5}, 1-(\frac{[0.4]^5}{R^5+[0.4]^5})^2) \\
&\quad + 0.32 \max(\frac{2}{3} \cdot \frac{R^5}{R^5+[0.1]^5}, \frac{1}{3}(1-(\frac{[0.2]^5}{R^5+[0.2]^5})^2) + \frac{2}{3} \cdot \frac{R^5}{R^5+[0.2]^5}, 1-(\frac{[0.3]^5}{R^5+[0.3]^5})^2) \\
&\quad + 0.04 \max(\frac{R^5}{R^5+[0.1]^5}, 1-(\frac{[0.2]^5}{R^5+[0.2]^5})^2)
\end{aligned}
$$

For $R$=0.5, as $D(H=0)$=0.5, $D(H=1)$=0.6804, $D(H=2)$=0.9736, deploying FTs is not efficient. For $R$=1, $D(H=0)$=0.9697, $D(H=1)$=0.9596, $D(H=2)$=0.9999. Hence the damage is minimized when one FT is deployed in this case. For $R$=2 we get $D(H=0)$=0.9991, $D(H=1)$=0.9997, $D(H=2)$=1. Hence deploying FTs is not efficient again.

The nonlinear dependence on $R$ makes the intuitive decision about deploying the FTs impossible and emphasizes the importance of using the suggested model in the decision making process.

## 5.3. *N* genuine elements connected in parallel

We consider a system that is built from $N$ identical parallel GEs with the same functionality having performance $g$ each. The system demand is $F$ ($F \le Ng$). The system fails to meet the demand when at least $\lfloor N - F/g \rfloor + 1$ elements are destroyed.

Since $H$ ($H \le r/s$) FTs are deployed, the defense effort exerted on each genuine target is $t=(r-Hs)/Na$. In case that $k$ ($0 \le k \le H$) FTs are detected by the attacker, he chooses $Q_k$ ($\lfloor N - F/g \rfloor + 1 \le Q_k \le N+H-k$) targets out of $N+H-k$ targets to attack and the attack effort allocated onto each target is $T=R/(Q_kA)$. The vulnerability of each GE is determined in (5.2).

For any specific $k$ and $Q_k$ the random number of attacked GEs can vary from $\max(0, Q_k-H+k)$. The probability $\varphi(Q_k, i)$ that among $Q_k$ attacked elements $i$ elements are the genuine ones is determined in (5.3). The probability $\theta(i, j)$ that among the $i$ attacked GEs $j$ elements are destroyed is

$$\theta(i,j) = \binom{i}{j} v^j (1-v)^{i-j}, \ 0 \le j \le i. \tag{5.6}$$

### 5.3.1. Damage proportional to the loss of demand probability

If the system totally fails when the demand is not met, the expected damage is proportional to the loss of demand probability. The demand is not met if the number of

destroyed GEs $j$ is greater than $N$-$F/g$ i.e. $j \geq \lfloor N - F/g \rfloor + 1$. In this case the expected damage to the system can be obtained as

$$D(Q_k, H) = F \cdot \sum_{i=\max(\lfloor N-F/g \rfloor+1, Q_k-H+k)}^{\min(N,Q_k)} \varphi(Q_k, i) \cdot \sum_{j=\lfloor N-F/g \rfloor+1}^{i} \theta(i, j) \quad . \tag{5.7}$$

Thus the most harmful attacker's strategy is $Q_k^* = \arg\max_{1 \leq Q_k \leq N+H-k} D(Q_k, H)$. The total expected damage to the entire system is obtained using (5.5).

The defender chooses the $H$ which minimizes $D(H)$, thus we have the optimal number of deployed FTs $H^* = \arg\min_{0 \leq H \leq \lfloor r/s \rfloor} D(H)$.

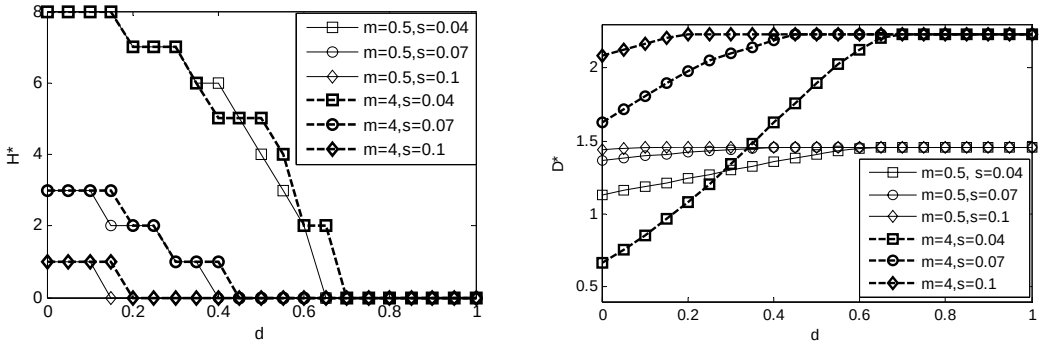


Figure 5.8 $H^*$ and $D(H^*)$ as functions of $d$ for parallel systems with damage proportional to the loss of demand probability

Figure 5.8 presents the optimal number of FTs $H^*$ and the corresponding $D(H^*)$ as functions of $d$ for $N$=8, $R$=$r$=1, $\varepsilon = 1$, $F$=4, $g$=1 and different combinations of $s$ and $m$. It can be seen that, as in the case of series system, $H^*$ decreases with the increase of $s$ and $d$.

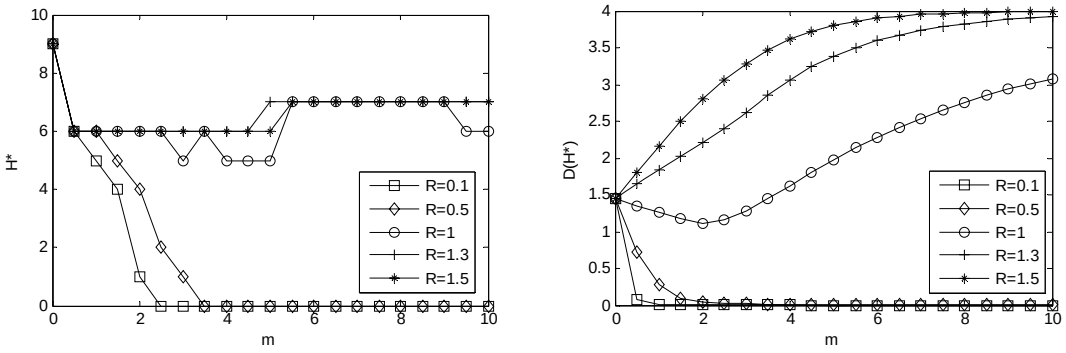


Figure 5.9 $H^*$ and $D(H^*)$ as functions of $m$ for parallel systems with damage proportional to the loss of demand probability

As in the case of series system (Figure 5.4), $H^*$ and $D(H^*)$ are non-monotonic functions of $m$. Figure 5.9 presents the optimal number of FTs $H^*$ and $D(H^*)$ as functions of $m$ for $N$=8, $r$=1, $\varepsilon = 1$, $s$=0.04, $d$=0.4, $F$=4, $g$=1 and different values of $R$. It can be explained in the same way as in the case of series system. When $R$=0.1 and 0.5 the defense effort is superior, the defender benefits from the increase of the contest intensity and $D(H^*)$ decreases with $m$. When $R$=1, $D(H^*)$ as function of $m$ demonstrates non-monotonic behavior. This can be explained by the fact that changes in optimal values of $H^*$ and $Q_k$ can make the defender's object protection effort either inferior or superior. In the former

case $D(H^*)$ increases with $m$ whereas in the latter case $D(H^*)$ decreases with $m$. When $R$=1.3 and 1.5 the defense effort is inferior, the attacker benefits from the increase of the contest intensity and $D(H^*)$ increases with $m$.

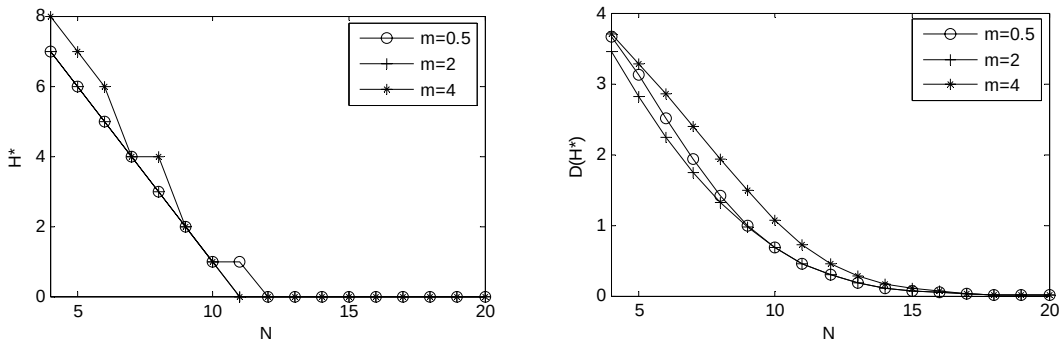


Figure 5.10 $H^*$ and $D(H^*)$ as functions of $N$ for parallel systems with damage proportional to the loss of demand probability

Figure 5.10 presents the optimal number of FTs $H^*$ and the corresponding $D(H^*)$ as functions of $N$ for $R$=$r$=1, $\varepsilon = 1$, $s$=0.05, $d$=0.4, $F$=4, $g$=1 and different values of $m$. It can be seen that $D(H^*)$ decreases with the increase of $N$. Indeed increase of $N$ makes the system less vulnerable because its redundancy increases. In this case the defender spends more resources for protection of the increased number of GEs and deploys fewer FTs for minimizing the system vulnerability.

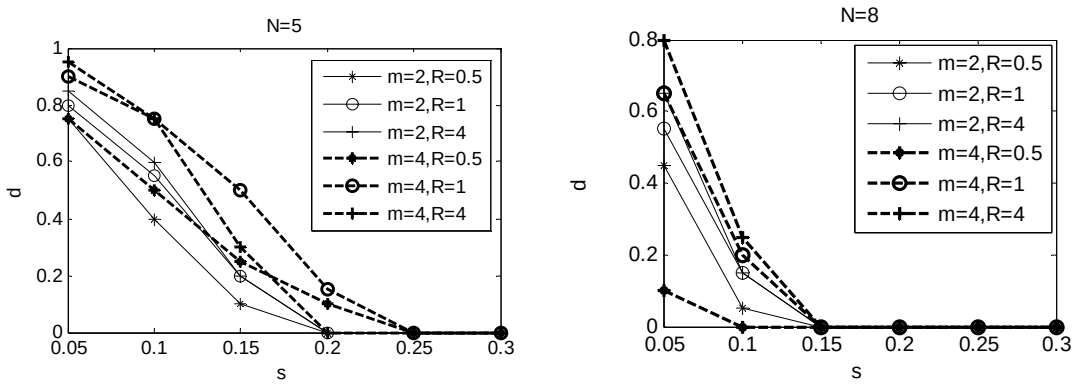


Figure 5.11 Efficiency analysis of deploying false targets for parallel systems with damage proportional to the loss of demand probability

As in the case of series systems, we have plotted the false targets deployment efficiency curves for $F$=4, $g$=1, $r$=1, $\varepsilon = 1$ and different values of $N$, $m$, and $R$, as shown in Figure 5.11. Similar to the case of series system (Figure 5.6), the critical $d$ depends on $R$ non-monotonically**.**

### 5.3.2. Damage proportional to the unsupplied demand

When $j$ GEs are destroyed the amount of unsupplied demand is equal to $\max(0,F-(N-j)g)$. The unsupplied demand becomes positive when $j \geq \lfloor N - F/g \rfloor + 1$ . The expected unsupplied demand can be obtained as

$$D(Q_k, H) = \sum_{i=\max(\lfloor N-F/g \rfloor+1, Q_k-H+k)}^{\min(N, Q_k)} \varphi(Q_k, i) \cdot \sum_{j=\lfloor N-F/g \rfloor+1}^{i} \theta(i, j) \cdot (F - Ng + gj) \qquad (5.8)$$

The attacker chooses $Q_k^* = \arg\max_{1 \le Q_k \le N+H-k} D(Q_k, H)$. The total expected damage to the system is obtained using (5.5). The defender chooses the $H$ which minimizes $D(H)$: $H^* = \arg\min_{0 \le H \le \lfloor r/s \rfloor} D(H)$ .

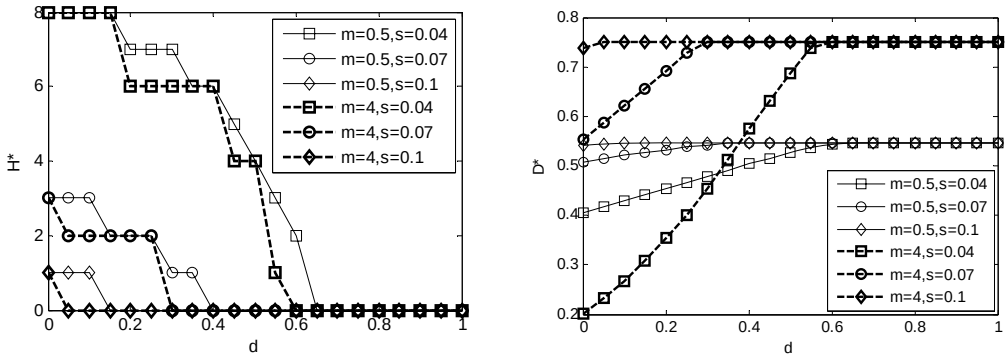


Figure 5.12 $H^*$ and $D(H^*)$ as functions of $d$ for parallel systems with damage proportional to the unsupplied demand

Figure 5.12 presents the optimal number of FTs $H^*$ and the corresponding $D(H^*)$ as functions of $d$ for $N$=8, $R$=$r$=1, $\varepsilon = 1$, $F$=4, $g$=1 and different combinations of $s$ and $m$. It can be seen that the function $H^*(d)$ in Figure 5.12 behaves similarly to that in Figure 5.8. Actually since the loss of demand probability and the unsupplied demand are positively correlated, $H^*$ that minimizes the loss of demand probability (Figure 5.12) is equal or close to $H^*$ that minimizes the unsupplied demand (Figure 5.8). $D(H^*)$ in Figure 5.12 is much lower than that in Figure 5.8. Indeed, the damage proportional to the unsupplied demand is always less than the damage proportional to the loss of demand probability

because when the demand is not met, in the former case the system can still function with reduced performance supplying the part of the demand whereas in the latter case the system totally fails.

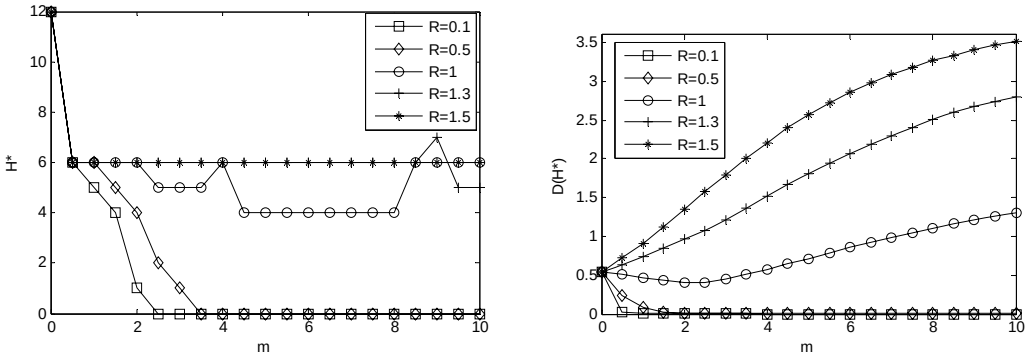


Figure 5.13 $H^*$ and $D(H^*)$ as functions of $m$ for parallel systems with damage proportional to the unsupplied demand

Figure 5.13 presents $H^*$ and $D(H^*)$ as functions of $m$ for $N$=8, $r$=1, $\varepsilon = 1$, $s$=0.04, $d$=0.4, $F$=4, $g$=1 and different values of $R$. Similar to Figure 5.9, when $R$=0.1 and 0.5 $D(H^*)$ decreases with $m$. When $R$=1, $D(H^*)$ as function of $m$ demonstrates non-monotonic behavior. When $R$=1.3 and 1.5, $D(H^*)$ increases with $m$. Since the system doesn't totally fail when the demand is not met, the $D(H^*)$ in Figure 5.13 is lower than that in Figure 5.9.

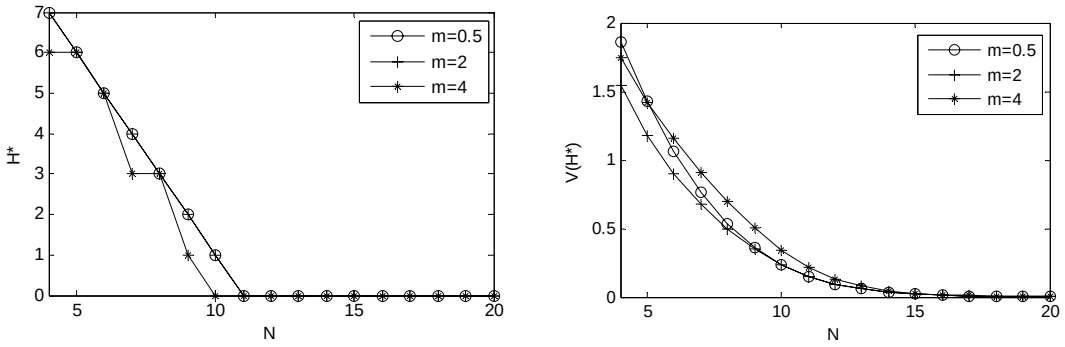


Figure 5.14 $H^*$ and $D(H^*)$ as functions of $N$ for parallel systems with damage proportional to the unsupplied demand

Figure 5.14 presents the optimal number of FTs $H^*$ and the corresponding $D(H^*)$ as functions of $N$ for $R$=$r$=1, $\varepsilon = 1$, $s$=0.05, $d$=0.4, $F$=4, $g$=1 and different values of $m$. Similar to Figure 5.10, $D(H^*)$ decreases with $N$. As the system doesn't totally fail when the demand is not met, the $D(H^*)$ in Figure 5.14 is considerably lower than that in Figure 5.10.

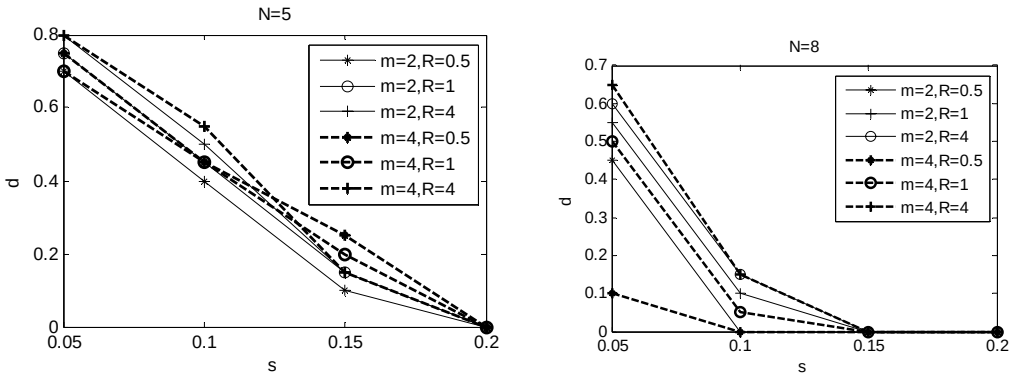


Figure 5.15 Efficiency analysis of deploying false targets for parallel systems with damage proportional to the unsupplied demand

As in subsection 5.3.1, we have plotted the false targets deployment efficiency curves for $F$=4, $g$=1, $r$=1, $\varepsilon = 1$ and different values of $N$, $m$, and $R$, as shown in Figure 5.15. Similar to the cases considered above, the critical $d$ depends on $R$ non-monotonically**.**

## 5.4. Conclusions

This chapter considers defending series and parallel systems against intentional attacks. The defender allocates part of its resource into deploying FTs and uses its remaining resource to protect the genuine system elements. It is assumed that each FT has a nonzero probability to be detected by the attacker and the detections of different FTs are independent. Once the attacker has detected a certain number of FTs, it ignores them and

chooses such number of undetected elements to attack that maximizes the expected damage to the system. The defender decides how many FTs to deploy to minimize the expected damage to the system assuming that the attacker uses the most harmful attack strategy. The expected damage to the series system is proportional to the probability of system destruction. Depending on the type of system the expected damage to the parallel system is either proportional to the loss of demand probability or equal to the unsupplied demand.

The chapter demonstrates the methodology of analysis of optimal defense strategy as the function of different parameters (number of GEs, contest intensity, total attacker's resource). It presents the decision curves that can be used for the making a decision about efficiency of deploying FTs depending on their cost and detection probability.

For any type of considered systems, the optimal number of FTs decreases with the increase of the detection probability or the unit cost of each FT. The number of FTs also decreases with the growth of the number of GEs. For any type of systems the expected damage can be non-monotonic function of the contest intensity.

With the increase of the number of GEs the expected damage to the series system increases, since the defender has to protect all $N$ elements while the attacker can destroy the entire system by destroying any single GE.

With the increase of the number of GEs the expected damage to the parallel system decreases because of the increase of system redundancy.

The expected damage proportional to the unsupplied demand is always much lower than in the expected damage proportional to the loss of demand probability.

The numerical analysis of the presented model shows the complicated interaction of free choice strategic variables and nonlinear dependence of the optimal number of FTs on different parameters. For example, the efficiency of FTs deployment can depend on the attacker's resource non-monotonically. Therefore, the intuitive decisions about the optimal strategy can be misleading and the use of the suggested model can be very helpful for supporting the decisions.

# CHAPTER 6 FURTHER WORK ON SYSTEM DEFENSE WITH FALSE TARGETS

In chapter 5, it is assumed that the detection probability of a false target is constant. This assumption does not address to the case when the attacker can take intelligence actions to detect false targets. In this chapter, we assume that the attacker allocates part of its budget into intelligence actions in order to detect false targets. Analogously, the defender allocates part of its budget into disinformation actions in order to deploy the false targets and prevent them from being detected. The detection probability of a false target is determined by the intelligence and disinformation efforts allocated on the false target by the attacker and the defender. In Levitin and Hausken (2009c) it is assumed that if the attacker's intelligence actions succeed, the attacker can identify and attack the defended object and ignore all false targets. However in many cases (for example, when the attacker can detect only specific features of the FTs) the intelligence actions can result in identifying part of FTs. In this case the attacker has a set of unidentified targets when it launches the attack.

This chapter considers defending a single genuine object including the strategy of deploying false targets that can be detected by the attacker individually and independently. We assume that both the attacker's and the defender's resources are fixed and both of them have full knowledge of each other's efforts. The contest between the defender and the attacker is modeled as a two period game where the defender moves in the first period, and the attacker moves in the second period. The defender builds the system over time and the attacker takes it as given when it chooses its attack strategy. In this chapter we study the defender's most conservative strategy which minimizes the probability of the object destruction assuming that the attacker always chooses the most harmful strategy no matter what the defender's strategy is. It is pointed out in Shier (1991) that the most conservative strategy is "particularly appropriate in the design of robust military systems".

Section 6.1 presents the model. Section 6.2 assumes that the number of false targets is fixed and the attacker tries to detect all the false targets. In Section 6.3 we assume that the defender can choose how many false targets to deploy and the attacker tries to detect all the false targets. In Section 6.4 we assume that the defender can choose how many false targets to deploy and the attacker tries to detect only a subset of false targets.

## 6.1. The model

**Assumptions:**

1. The defender uses identical false targets and allocates the disinformation efforts evenly among them.

2. The attacker allocates the intelligence efforts evenly among the targets it tries to detect.

3. The attacker allocates the attack effort evenly among all the attacked targets.

4. The attacker can successfully identify some targets as false targets (by detecting some features that characterize the FTs), but cannot confidently identify any target as the genuine object (the fact that specific FT features are not detected can mean either that the detection failed or that the target is the genuine object).

The defender has deployed one genuine object and $H$ false targets (FTs). The total attacker's resource is $R$. The attacker can allocate part of its resource $RX$ ($0 \le X \le 1$) into intelligence effort aimed at detecting FTs. The cost of the intelligence effort unit is $B$. The attacker tries to detect FTs among the $H+1$ targets. The intelligence effort allocated on each target is $S=RX/[B(H+1)]$. Once the attacker has detected a certain number $k$ ($0 \le k \le H$) of FTs, it will choose $Q_k$ targets among the $H-k+1$ undetected targets to attack such that $Q_k$ maximizes the probability of the genuine object destruction. The cost of the attack effort unit is $A$. The attack effort allocated on each attacked target is $T=R(1-X)/(Q_kA)$.

The defender's total resource is $r$. It distributes $xr$ ($0 \le x \le 1$) into disinformation actions, which includes deploying $H$ FTs and preventing the FTs from being detected by the attacker, and distributes its remaining resource $r(1-x)$ into protecting the genuine object.

The cost of the protection effort unit is $a$. The cost of the disinformation effort unit is $b$. The effort for protecting the defended object is $t = r(1-x)/a$, whereas the disinformation effort allocated on each FT is $s=rx/(bH)$.

We assume that two contests take place in the considered game: intelligence contest and impact (protection-attack) contest. We here apply the commonly used ratio form of the attacker-defender contest function (Hausken 2005, Tullock 1980, Skaperdas 1996). The probability of the attacker's success in the intelligence contest is

$$w = \frac{S^f}{S^f + s^f} = \frac{\left[\frac{RX}{B(H+1)}\right]^f}{\left[\frac{RX}{B(H+1)}\right]^f + \left[\frac{rx}{bH}\right]^f} = \frac{X^f}{X^f + (1+1/H)^f h^f x^f} \tag{6.1}$$

where $f$ is a parameter that specifies the intensity of the contest, that is how decisively the agents fight or compete in the contest. $f$=0 gives egalitarian distribution. $f$=1 gives proportional distribution. $f$= ∞ gives winner-take-all. $h = \frac{B}{b} \cdot \frac{r}{R}$ is the defender's intelligence superiority parameter that specifies how the intelligence resource ratio $x/X$ is realized into intelligence effort ratio $s/S$.

$h$>1 decreases the probability of detection which gives advantage to the defender, whereas $h$<1 increases the probability of detection which gives advantage to the attacker.

The probability that $k$ ($0 \le k \le H$) FTs are detected by the attacker is given by

$$p_k = \binom{H}{k} w^k (1-w)^{H-k} \tag{6.2}$$

The probability of the attacker's success in the impact contest (object vulnerability) is

$$\omega = \frac{T^m}{T^m + t^m} \tag{6.3}$$

In the case that $k$ FTs are detected by the attacker (with probability $p_k$), the attacker chooses $Q_k$ ($1 \le Q_k \le H\text{-}k+1$) out of $H\text{-}k+1$ undetected targets to attack. Having the probability that the genuine object is attacked $Q_k/(H\text{-}k+1)$ and the probability of attack success (6.3) one can get the genuine object destruction probability as

$$v_k(Q_k) = \frac{Q_k}{H-k+1} \cdot \frac{\left[\frac{R(1-X)}{Q_k A}\right]^m}{\left[\frac{R(1-X)}{Q_k A}\right]^m + \left[\frac{r(1-x)}{a}\right]^m} = \frac{Q_k}{H-k+1} \cdot \frac{(1-X)^m}{(1-X)^m + g^m Q_k^m (1-x)^m} \tag{6.4}$$

where $g = \frac{A}{a} \cdot \frac{r}{R}$ is the defender's impact superiority parameter that specifies how the impact resource ratio $(1\text{-}x)/(1\text{-}X)$ is realized into impact effort ratio $t/T$ when $Q_k=1$.

$g>1$ decreases the probability of target destruction in the case of attack which gives advantage to the defender, whereas $g<1$ increases the probability of target destruction in the case of attack which gives advantage to the attacker.

For each combination of $H$, $x$, $X$ and $k$ the attacker chooses the optimal $Q_k=Q_k^*$ which maximizes $v_k(Q_k)$. For any combination of $H$, $x$, and $X$, the maximal object destruction probability can be expressed as

$$V(H,x,X)=\sum_{k=0}^{H} p_k \cdot v_k(Q_k^*). \tag{6.5}$$

For any defender's strategy ($H$,$x$) the attacker responds with $X$ that maximizes $V(H,x,X)$ obtained in (6.5). We use $X^*$ to denote the optimal $X$ and $V^*(H,x)$ to denote $V(H,x,X^*)$.

The defender must choose the combination of ($H$,$x$), denoted as ($H^*$, $x^*$), which minimizes $V^*(H,x)$.

## 6.2. Fixed number of deployed FTs

In this section we assume that the number of deployed FTs $H$ is fixed. From (6.4) we have

$$\frac{\partial v_k(Q_k)}{\partial Q_k}=\frac{(1-X)^{2m}+(1-m)g^m Q_k^m (1-x)^m (1-X)^m}{(H-k+1)[(1-X)^m+g^m Q_k^m (1-x)^m]^2} \tag{6.6}$$

If $m\le 1$, $\frac{\partial v_k(Q_k)}{\partial Q_k}>0$ and maximal $v_k(Q_k)$ is achieved when $Q_k=H-k+1$. In this case we have

$$v_k(Q_k)=\frac{(1-X)^m}{(1-X)^m+g^m(H-k+1)^m(1-x)^m} \tag{6.7}$$

If $m>1$, $\frac{\partial v_k(Q_k)}{\partial Q_k}=0$ gives $Q_k=\frac{(1-X)}{\sqrt[m]{m-1}g(1-x)}$. For the convenience of later discussion, we denote $F=\left\lfloor \frac{(1-X)}{\sqrt[m]{m-1}g(1-x)} \right\rfloor$, where $\lfloor x \rfloor$ is the maximal integer not greater than $x$. Since $Q_k$ cannot be greater than $H$-$k$+1, we have:

$$Q_k^*=\begin{cases} F & \text{if } F<H\text{-}k+1, v_k(F)\geq v_k(F+1) \\ F+1 & \text{if } F<H\text{-}k+1, v_k(F)<v_k(F+1) \\ H-k+1 & \text{if } F\geq H\text{-}k+1 \end{cases} \tag{6.8}$$

Observe, that when $m>1$ and $F<H$-$k$+1, $Q_k^*$ does not depend on $H$.

Figure 6.1 presents the most harmful attack strategies $Q_k^*$ ($0\leq k\leq 6$) for $H=6$, $g=1$, $x=0.8$ and different combinations of $X$ and $m$. The attacker tends to attack more elements when it has reserved more resource for attacks $R(1-X)$. With the growth of the contest intensity it becomes more important for the attacker to achieve the effort superiority over the defender. Therefore the attacker concentrates greater per-target efforts by attacking fewer targets and $Q_k^*$ tends to decrease with $m$. When $k=H$ the attacker detects all the FTs and attack single defended genuine object: $Q_H^*=1$.

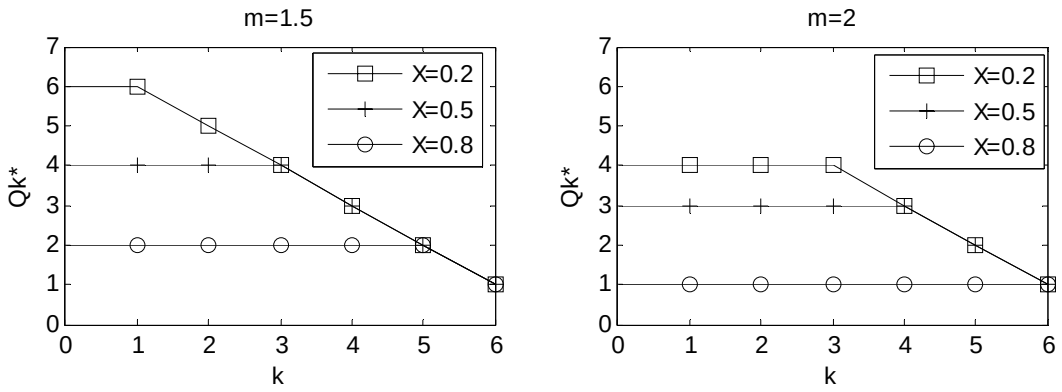


Figure 6.1 Optimal number of attacked targets for different $X$

For $m \leq 1$, the expected object vulnerability is

$$V(x,X)=\sum_{k=0}^{H} p_k v_k(Q_k^*)=\sum_{k=0}^{H}\frac{(1-X)^m\binom{H}{k}w^k(1-w)^{H-k}}{(1-X)^m+g^m(H-k+1)^m(1-x)^m} \tag{6.9}$$

For $m>1$, the expected probability of object destruction is

$$V(x,X)=\sum_{k=0}^{H} p_k v_k(Q_k^*)=\sum_{k=0}^{H}\binom{H}{k}w^k(1-w)^{H-k}v_k(Q_k^*) \tag{6.10}$$

where $v_k(Q_k^*)=\begin{cases}\dfrac{(1-X)^m}{(1-X)^m+g^m(H-k+1)^m(1-x)^m}, & F\geq H-k+1\\ \max(v_k(F),v_k(F+1)), & F<H-k+1\end{cases}$.

The optimal value of $X$ which maximizes $V(x,X)$ can be expressed as $X^*=\text{argmax}_{0\leq X\leq 1}V(x,X)$ and the corresponding object destruction probability is $V^*(x)=V(x,X^*)$. Figure 6.2 shows the attacker's optimal intelligence resource portion $X^*$

and the corresponding object destruction probability $V^*(x)$ as functions of the defender's disinformation resource proportion $x$ for $H$=6, $g$=1, $m$=1.5, $f$=1 and different values of $h$. It can be seen that the behavior of $X^*$ as function of $x$ is irregular, however, in general, it takes an inversed u-shape form. $V^*(x)$ is u-shaped function of $x$. When $x$ is small, the increase of disinformation actions will make the FTs harder to be detected and thus reduce the object destruction probability. Further increase of $x$ can leave the genuine object poorly protected, which increases the possibility of its destruction with low attacker's effort. The u-shaped form of $V^*(x)$ shows that the optimal balance between the protection effort and disinformation effort minimizes the object vulnerability. It can also be seen that $V^*(x)$ decreases with the increase of $h$. Actually from (6.1) we can see that the increase of $h$ will reduce the detection probability of each FT and thus reduce the object destruction probability.

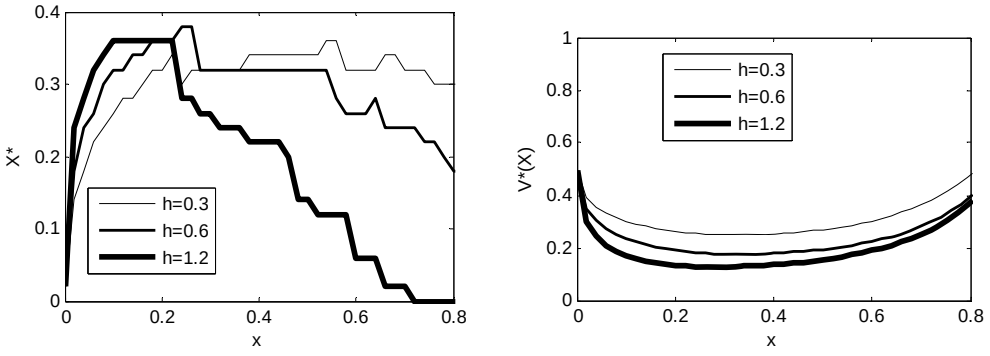


Figure 6.2 $X^*$ and $V^*(x)$ as functions of $x$ for different $h$

In order to understand the irregular behavior of $X^*$ with the increase of $x$, Figure 6.3 shows the object destruction probability $V(x,X)$ as a function of the attacker's intelligence resource proportion $X$ for $H$=6, $g$=1, $m$=1.5, $f$=1, $h$=0.6 and different values of $x$. It can be seen that due to the complexity of $V(x,X)$ the optimal value of $X$ behaves non-monotonically even when the value of $x$ changes in a small range. When $x$=0.2, the optimal value of $X$ is at point A. When the value of $x$ increases to 0.25, the optimal value of $X$ increases to point B. When the value of $x$ increases to 0.3, the optimal value of $X$ decreases to point C. The stepwise changes of $X^*(x)$ are caused by discrete variations of the corresponding $Q_k^*$ in the optimal attacker's strategy.

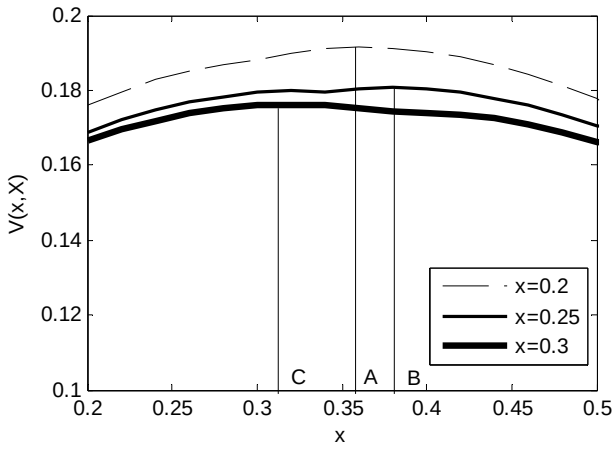


Figure 6.3 $V(x,X)$ as a function of $X$ for different $x$

The optimal value of $x$ (solution of the two-period minmax game) is $x^*=\text{argmin}_{0\le x\le 1} V^*(x)$. The corresponding object vulnerability is $V^*=V^*(x^*)$. Figure 6.4 shows the

defender's optimal disinformation resource proportion $x^*$, the attacker's optimal intelligence resource proportion $X^*$ and the object destruction probability $V^*$ as functions of the defender's disinformation superiority parameter $h$ for $H$=6, $g$=1, $m$=1.5 and different values of $f$. It can be seen that $x^*$ first increases with the increase of $h$ and then decreases with the increase of $h$. When $h$ is small, the increase of $h$ makes FTs harder to be detected thus justifies the allocation of more disinformation resource. When $h$ has reached a certain level the FTs are already very hard to be detected, thus it is more cost-effective to spend more resource on protecting the genuine object. $V^*$ decreases with the increase of $h$, since the increase of $h$ has reduced the detection probability of each FT. When $h$ is low the attacker's intelligence effort is superior, thus $V^*$ increases with the increase of $f$. When $h$ is high the defender's disinformation effort is superior, thus $V^*$ decreases with the increase of $f$.

The oscillations of $X^*$ are similar to those observed in Levitin and Hausken (2009c). The vulnerability function $V(X)$ for fixed $x$ has two maxima: one at some positive value of $X$ and one at $X$=0. These maxima frequently have similar values, which cause the attacker to be indifferent between zero investment into the intelligence effort and a specific positive investment. In practice the oscillating behavior of the attacker's intelligence effort fraction $X^*$ causes the attacker to concentrate all its resources on impact effort sacrificing the intelligence. Indeed, the exact values of the contest intensities are hard to estimate and predict in practice. Small alterations of these intensities may make nonzero intelligence effort beneficial or not beneficial to the attacker. Hence the attacker can never be sure that the nonzero intelligence effort is justified. On the contrary the defender must invest a

nonzero resource fraction into counter-intelligence in order to maintain the optimal solution when the attacker's intelligence investment deviates from the optimal value.

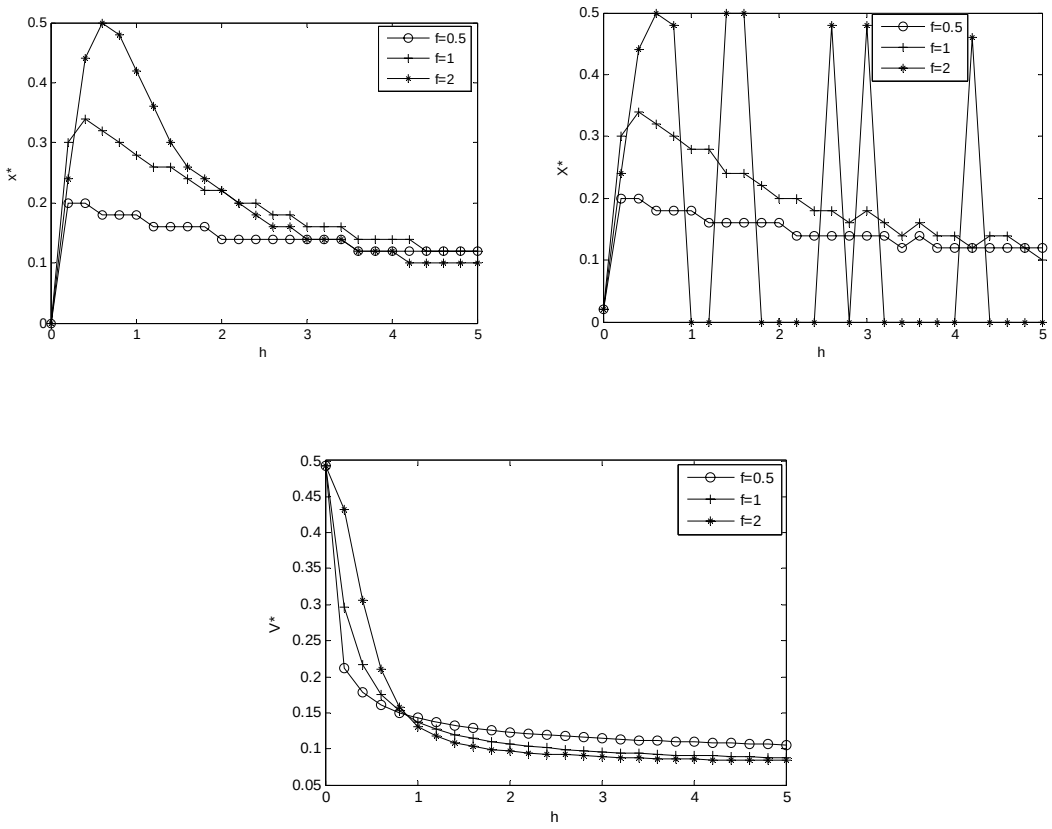


Figure 6.4 $x^*$, $X^*$ and $V^*$ as functions of $h$ for different $f$

Figure 6.5 shows the defender's disinformation resource proportion $x^*$, the attacker's intelligence resource proportion $X^*$, the object destruction probability $V^*$ and $V(x^*,x^*)$ as functions of the defender's impact superiority parameter $g$ for $H$=6, $h$=0.5, $f$=0.5 and different values of $m$. It can be seen that $V^*$ decreases with the increase of the defender's

impact superiority parameter $g$. When $g$ is low the attacker's attack effort is superior, thus $V^*$ increases with the increase of $m$. When $g$ is high the defender's protection effort is superior, thus $V^*$ decreases with the increase of $m$. The attacker's and defender's resource distribution parameters are very close or coincide, which is also consistent with the results of Levitin and Hausken (2009c).

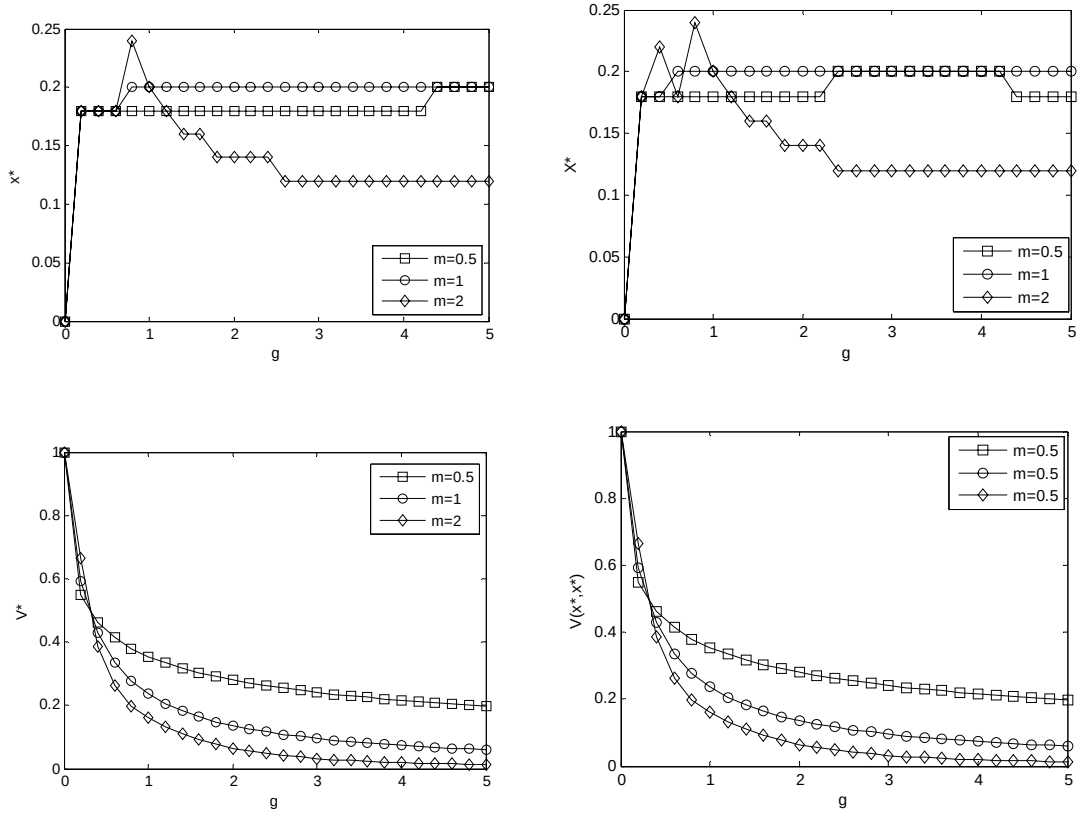


Figure 6.5 $x^*$, $X^*$ and $V^*$ as functions of $g$ for different $m$

Figure 6.6 shows the defender's disinformation resource proportion $x^*$, the attacker's intelligence resource proportion $X^*$, the object destruction probability $V^*$ and $V(x^*,x^*)$ as functions of the number of false targets $H$ for $g$=0.5, $h$=0.5, $f$=0.5 and different values of $m$. $V^*$ decreases with the increase of $H$. The increased number of FTs makes the attacker more difficult to locate the genuine object, which either reduces the probability for the genuine object to be attacked or the attack effort allocated into attacking the genuine object. When $H$ is low the attacker's attack effort on the genuine object is superior, thus $V^*$ increases with the increase of $m$. When $H$ is high much of the attacker's attack effort is distracted by the FTs, thus the attack effort on the genuine object becomes inferior. Hence $V^*$ decreases with the increase of $m$. The attacker's and defender's resource distribution parameters are very close or coincide, which is also consistent with the results of Levitin and Hausken (2009c).

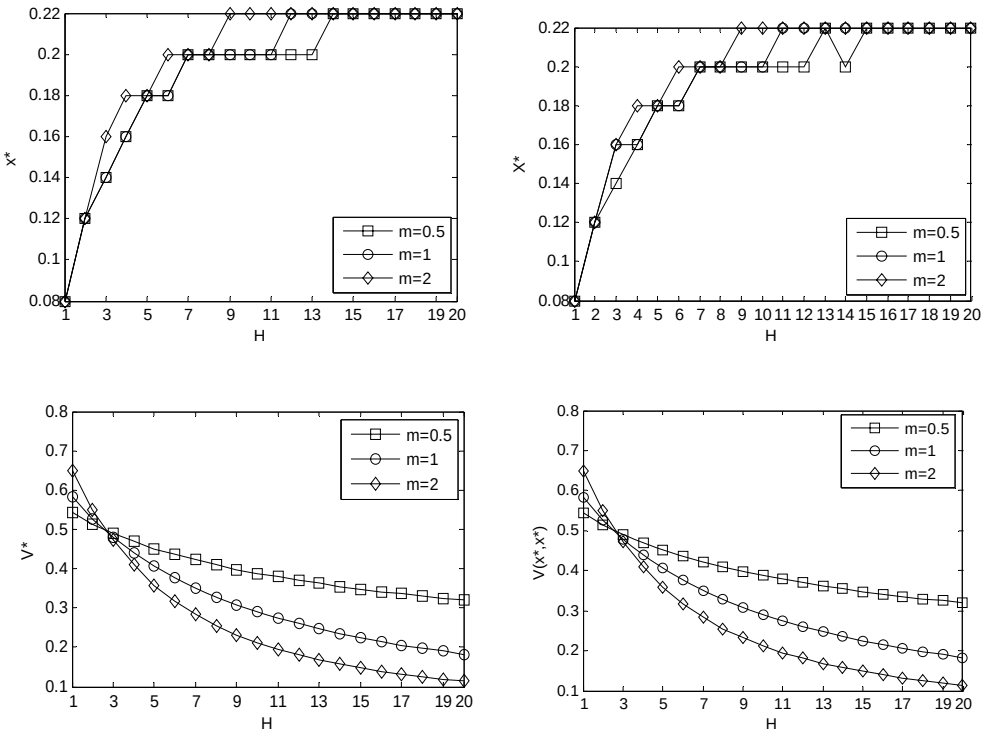


Figure 6.6 $x^*$, $X^*$ and $V^*$ as functions of $H$ for different $m$

## 6.3. Optimal number of FTs

In this section we assume that the cost of each FT cannot be less than $c_{min}$ and the defender can choose how many FTs to deploy. For each fixed $H$ and $x$, the attacker chooses the most harmful $X$ and $(Q_0,\ldots Q_H)$ to maximize the overall destruction probability of the genuine object. The corresponding overall destruction probability of the genuine object is denoted as $V^*(H,x)$, which has the same form as $V^*(x)$ obtained in Section 6.2.

The maximal possible number of FTs the defender can deploy must not exceed $H_{\max}=r/c_{min}$. For any chosen value of $H$, the defender's resource allocated into disinformation actions $rx$ must not be less than $Hc_{min}$, from which follows that $x \geq x_{min}=Hc_{min}/r=H/H_{\max}$. The defender chooses the optimal strategy $(H^*,x^*)$ which minimizes $V^*(H,x)$:

$$(H^*,x^*) = \arg\min_{0\leq H\leq H_{\max};x_{\min}\leq x\leq 1} V^*(H,x).$$

In the following examples we use $V^*$ to denote $V(H^*,x^*)$. Figure 6.7 shows the optimal number of false targets $H^*$, the defender's optimal disinformation resource proportion $x^*$, the attacker's optimal intelligence resource proportion $X^*$ and the object destruction probability $V^*$ as functions of the defender's disinformation superiority parameter $h$ for $H_{\max}=50$, $g=1$, $m=1.5$ and different values of $f$. Similar to Figure 6.4, $V^*$ decreases with the increase of $h$. When $h$ is low the attacker's intelligence effort is superior, thus $V^*$ increases with $f$. When $h$ is high the defender's disinformation effort is superior, thus $V^*$ decreases with $f$. When $h$ reaches a certain level even the cheapest FTs cannot be detected by the attacker, thus $V^*$ doesn't change much with variations of $f$ and $h$.

We can get the asymptotic estimates of $H^*$, $x^*$, $X^*$ and $V^*$ for $h\to\infty$ as follows. When $h$ approaches $\infty$, the detection probability $w$ for any FT approaches zero for any $f$ (see equation (6.1)) even when the defender distributes minimal resource into disinformation actions. In this case the defender always distributes $rx_{min}= rH/H_{\max}=rH/50$ into disinformation actions whereas the attacker does not invest any effort into intelligence contest ($X^*=0$) as he has no chance to win it. In this case

$$V^*(H,x) = V^*(H,H/50) = \sum_{k=0}^{H} p_k v_k(Q_k^*) = v_0(Q_0^*). \tag{6.11}$$

From (6.4) we have

$$v_0(Q_0) = \frac{Q_0}{H+1} \cdot \frac{(1-X)^m}{(1-X)^m + g^m Q_0^m (1-x)^m} = \frac{Q_0}{H+1} \cdot \frac{1}{1+Q_0^{1.5}(1-H/50)^{1.5}} \tag{6.12}$$

For each fixed $H$, $V^*(H,H/50) = v_0(Q_0^*) = \arg\max_{Q_0=0,1,\ldots H+1} \frac{Q_0}{H+1} \cdot \frac{1}{1+Q_0^{1.5}(1-H/50)^{1.5}}$.

Solving the minmax problem $\min_H \max_{Q_o} v_0$ we get $H^*$=25 and $Q_0{}^*$=3. The corresponding values of $x^*$ and $V^*$ are $x^*$=$H^*$/50=0.5 and $V^*$= $V^*$($H^*$,$x^*$)=0.0407.

It can also be seen from Figure 6.7 that $H^*$ and $x^*$ vary with $h$ and $f$ in very similar manner, which means that the fraction of the defenders resource invested into each FT remains almost the same. The cheapest false targets are most favorable in the cases we consider. Actually for fixed intelligence resource $x$, the defender prefers to deploy more false targets with less unit cost.

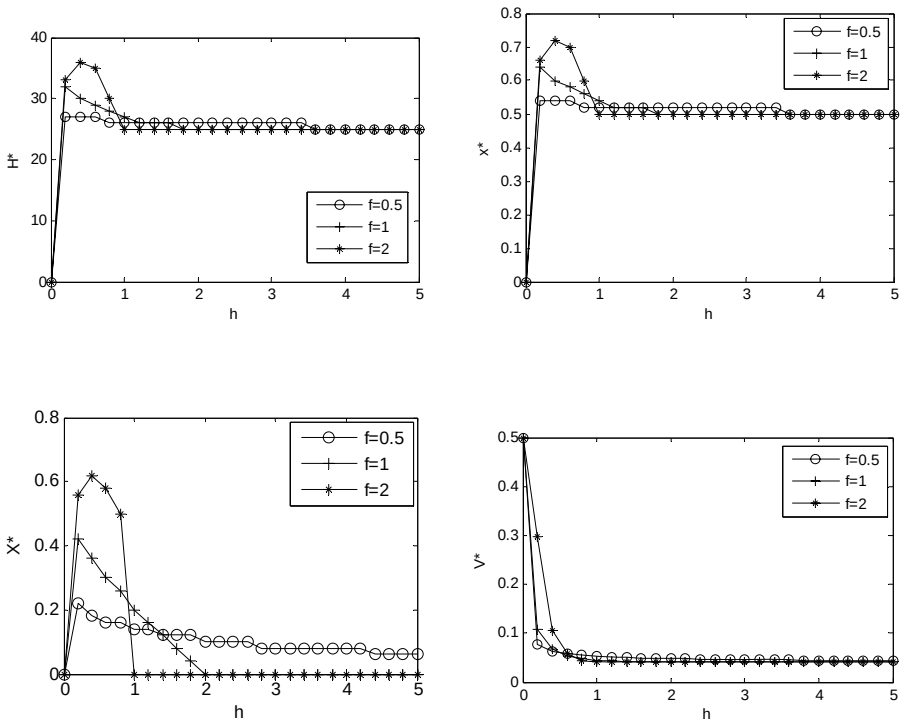


Figure 6.7 $H^*$, $x^*$, $X^*$ and $V^*$ as functions of $h$ for different $f$

Figure 6.8 shows the optimal number of false targets $H^*$, the attacker's optimal intelligence resource proportion $X^*$, and the object destruction probability $V^*$ as functions of the defender's disinformation superiority parameter $h$ for $H_{\max}$=50, $g$=1, $m$=1.5, $f$=2 and different fixed values of $x$. It can be seen that $H^*$ always equals to the maximum possible $H$ ($xH_{max}$) except when $h$=0.

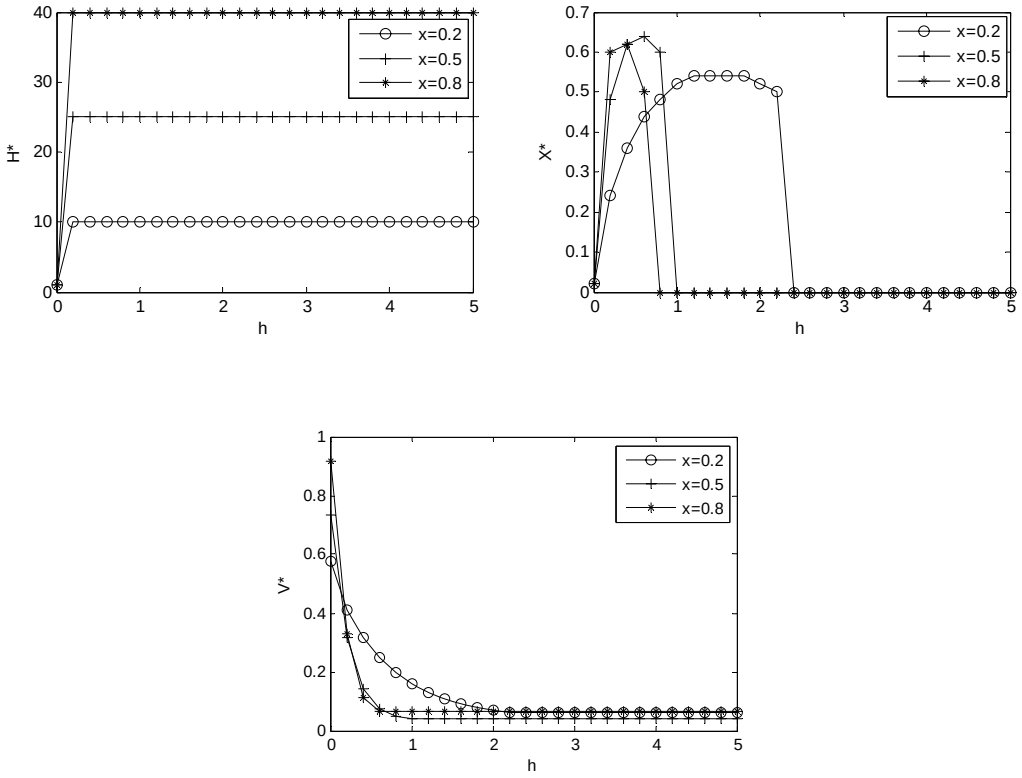


Figure 6.8 $H^*$, $X^*$ and $V^*$ as functions of $h$ for different $x$

Figure 6.9 shows the optimal number of false targets $H^*$, the defender's optimal disinformation resource proportion $x^*$, the attacker's intelligence resource proportion $X^*$ and the object destruction probability $V^*$ as functions of the defender's impact superiority parameter $g$ for $H_{\max}$=50, $h$=1, $f$=1 and different values of $m$. Similar to Figure 6.5, $V^*$ decreases with $g$. $H^*$ and $x^*$ vary with $g$ and $m$ in very similar manner. It can be explained in similar manner as for Figure 6.7.

It can be seen that when $g$ approaches infinity $V^*$ always approaches zero. Indeed, it follows from (6.7) that for $0<m\leq 1$ $\lim_{g\to\infty} v_k(Q_k) = 0$ for any $k$ and $x<1$. For $m>1$ and $g\to\infty$, we have $F = \left\lfloor \frac{(1-X)}{\sqrt[m]{m-1}g(1-x)} \right\rfloor = 0$, which means that the attacker attacks single target. Furthermore from (6.4) and (6.10) we have

$$\lim_{g\to\infty} V^* = \lim_{g\to\infty} \sum_{k=0}^{H} p_k v_k(Q_k^*) = \lim_{g\to\infty} \sum_{k=0}^{H} \frac{p_k}{H-k+1} \cdot \frac{(1-X)^m}{(1-X)^m + g^m(1-x)^m} = 0 \ .$$

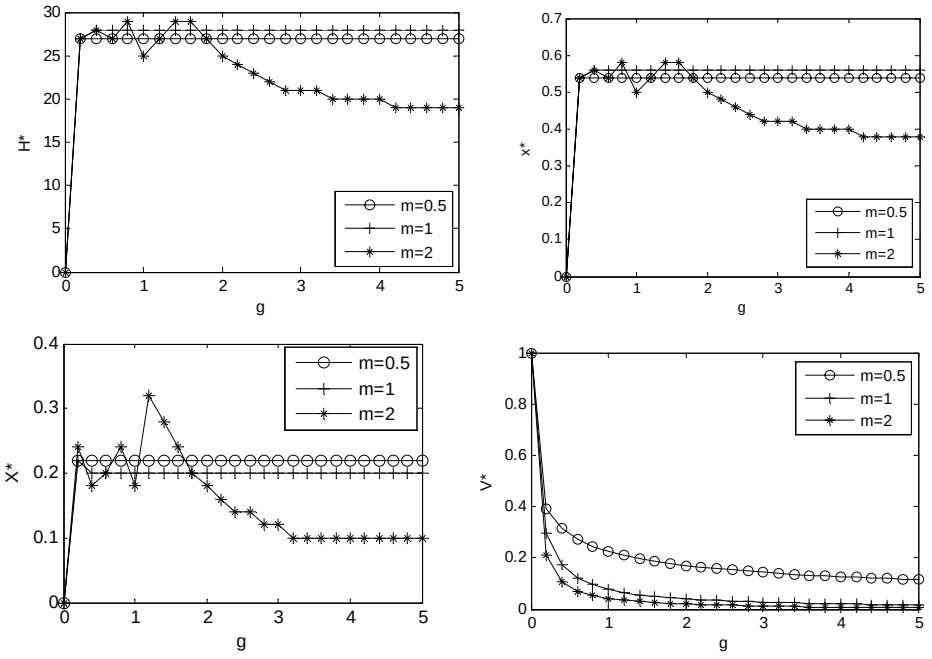


Figure 6.9 $H^*$, $x^*$, $X^*$ and $V^*$ as functions of $g$ for different $m$

In order to show how the impact contest intensity $m$ influences the game solution, Figure 6.10 presents the optimal number of false targets $H^*$, the defender's optimal disinformation resource proportion $x^*$, the attacker's optimal intelligence resource proportion $X^*$ and the object destruction probability $V^*$ as functions of the impact contest intensity $m$ for $H_{max}$=50, $h$=1, $f$=1 and different values of $g$. It can be seen that $H^*$, $x^*$ become more sensitive to variations of $m$ when $g$ increases. $X^*(m)$ oscillates, which means that two different values of $X$ produce very close values of $V$.

The attack effort on the genuine object can be either superior or inferior depending on the optimal values of the free choice variables $H$, $x$ and $X$. Thus $V^*$ as function of $m$ displays non-monotonic behavior. It can be seen from Figure 6.10 that for $H_{max}$=50, $h$=1, $f$=1 the defender benefits from moderate values of the impact contest intensity ($1<m<2$).

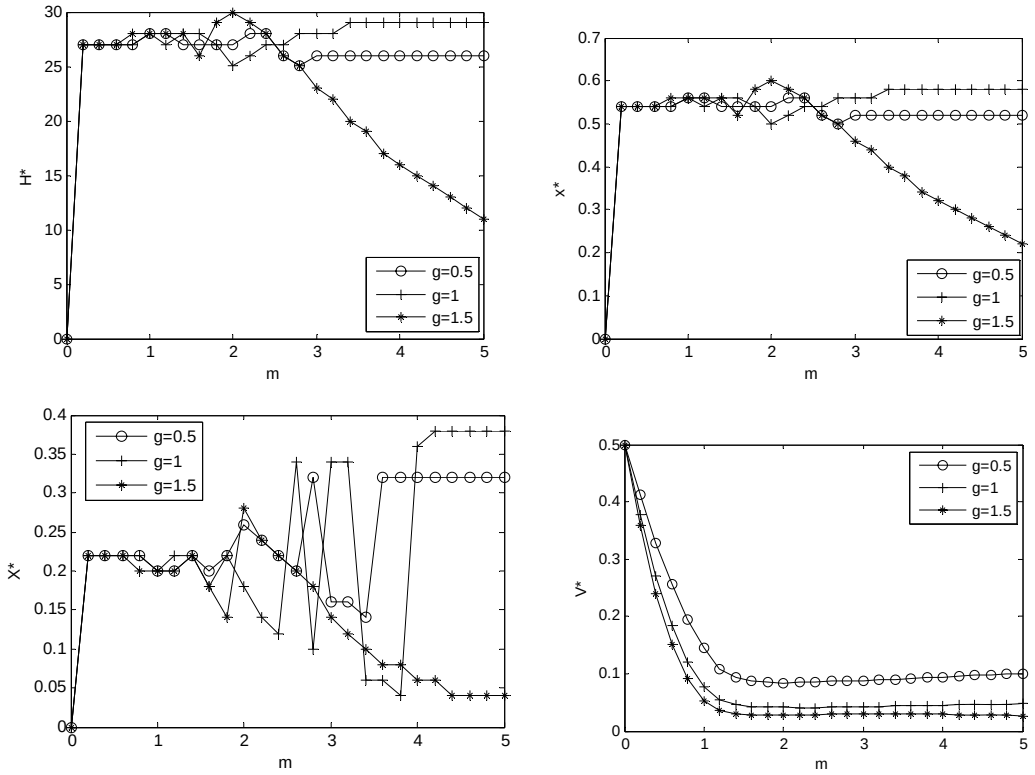


Figure 6.10 $H^*$, $x^*$, $X^*$ and $V^*$ as functions of $m$ for different $g$

## 6.4. The attacker attempts to detect a subset of targets

In this section we assume that the attacker can distribute its intelligence efforts onto a subset of targets. In the case that the attacker chooses $J$ ($0 \leq J \leq H+1$) targets to detect, the intelligence effort allocated on each target is $S=RX/BJ$. The probability of the attacker's success in the intelligence contest is

$$w = \frac{S^f}{S^f + s^f} = \frac{\left[\frac{RX}{BJ}\right]^f}{\left[\frac{RX}{BJ}\right]^f + \left[\frac{rx}{bH}\right]^f} = \frac{X^f}{X^f + (J/H)^f h^f x^f} \tag{6.13}$$

Once the attacker has detected a certain number $k$ ($0 \le k \le J$) of FTs, it will choose $Q_k$ targets among the $H$-$k$+1 undetected targets to attack such that $Q_k$ maximizes the probability of genuine object destruction.

In the case when the genuine object is among the checked targets (with probability $J/(H+1)$), at most $J$-1 FTs can be detected. The probability that $k$ ($0 \le k < J$) FTs are detected by the attacker is given by

$$\tilde{p}_k = \binom{J-1}{k} w^k (1-w)^{J-1-k} \tag{6.14}$$

and the probability that all $J$ FTs are detected is $\tilde{p}_J = 0$.

In the case that the genuine object is not among the checked targets (with probability $1-J/(H+1)$), the probability that $k$ ($0 \le k \le J$) FTs are detected by the attacker is given by

$$\bar{p}_k = \binom{J}{k} w^k (1-w)^{J-k} \tag{6.15}$$

Thus the total probability that exactly $k$ ($0 \le k \le J$) FTs are detected by the attacker is

$$p_k = \frac{J}{H+1}\widetilde{p}_k + \frac{H-J+1}{H+1}\overline{p}_k$$
$$= \begin{cases} \left(\frac{1}{1-w} + \frac{H-J+1}{J-k}\right)\frac{J}{(H+1)}\binom{J-1}{k} w^k (1-w)^{J-k} & \text{for } k < J \\ \frac{H-J+1}{H+1} w^J & \text{for } k = J \end{cases} \quad (6.16)$$

In the case that $k$ FTs are detected by the attacker (with probability $p_k$), the attacker chooses $Q_k$ ($1 \le Q_k \le H\text{-}k\text{+}1$) out of $H\text{-}k\text{+}1$ undetected targets to attack. The genuine object destruction probability is given in (6.4).

For each combination of $H$, $x$, $X$, $J$ and $k$ the attacker chooses the optimal $Q_k=Q_k^*$ which maximizes $v_k(Q_k)$. For any combination of $H$, $x$, $X$ and $J$ the maximal object destruction probability can be expressed as

$$V(H,x,X,J) = \sum_{k=0}^{J} p_k \cdot v_k(Q_k^*) \quad (6.17)$$

For any defender's strategy ($H$,$x$) the attacker responds with the $X$ and $J$ that maximize $V(H,x,X,J)$. We use $X^*$ and $J^*$ to denote the optimal $X$ and $J$ and $V^*(H,x)$ to denote $V(H,x,X^*,J^*)$. Figure 6.11 shows the attacker's optimal intelligence proportion $X^*$, the optimal number of checked targets $J^*$ and the object destruction probability $V^*(H,x)$ as functions of the defender's disinformation resource proportion $x$ for $H$=6, $g$=1, $m$=3, $f$=3 and different values of $h$. Similar to Figure 6.2 the behavior of $X^*$ as function of $x$ is non-monotonic, however, in general, it takes an inversed u-shape form.

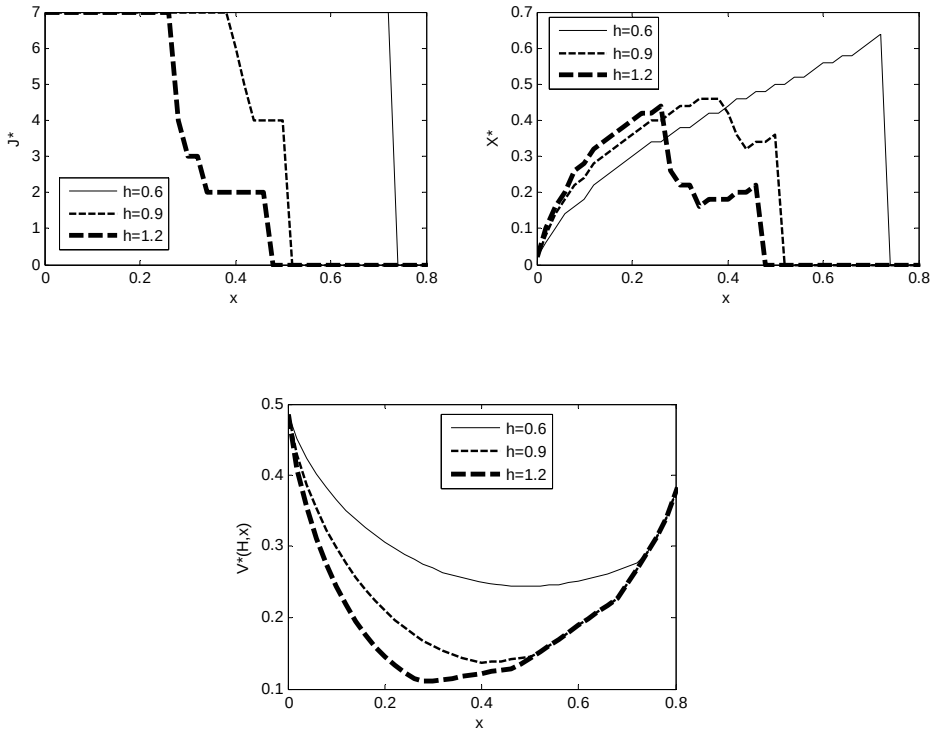


Figure 6.11 $X^*$, $J^*$ and $V^*(H, x)$ as functions of $x$ for different $h$

$V^*(H, x)$ is u-shaped function of $x$ and decreases with the increase of $h$. $J^*$ decreases with the increase of $h$, since the attacker needs to concentrate its intelligence effort on fewer targets in order to gain superiority in intelligence contest when $h$ increases. When $J^*$ doesn't change in some range of $h$, $X^*$ increases to cope with the increased $x^*$. When the attacker prefers to concentrate its intelligence effort, $J^*$ decreases and $X^*$ decreases accordingly.

Similar to Section 6.3, the defender chooses the optimal strategy $(H^*,x^*)$ which minimizes $V(H,x)$: $(H^*,x^*) = \arg\min_{0\le H\le H_{\max};x_{\min}\le x\le 1} V^*(H,x)$. We also use $V^*$ to denote $V(H^*,x^*)$. Figure 6.12 shows the optimal number of false targets $H^*$, the defender's

optimal disinformation resource proportion $x^*$, the attacker's optimal intelligence resource proportion $X^*$, the optimal number of checked targets $J^*$ and the object destruction probability $V^*$ as functions of the defender's disinformation superiority parameter $h$ for $H_{max}$=25, $g$=1, $m$=3 and different values of $f$. Similar to Figure 6.7, $V^*$ decreases with the increase of $h$. $H^*$ and $x^*$ vary with $h$ and $f$ in very similar manner. $H^*$, $x^*$ and $V^*$ converge when $h$ approaches infinity. $X^*$ and $J^*$ converge to 0 when $h$ approaches infinity, since it is not possible for the attacker to detect any false target irrespectively how much resource is allocated into intelligence actions. The oscillations of $X^*$ and $J^*$ have the same nature as in Figure 6.4.

Figure 6.13 shows the optimal number of false targets $H^*$, the defender;s optimal disinformation resource proportion $x^*$, the attacker's optimal intelligence resource proportion $X^*$, the optimal number of checked targets $J^*$ and the object destruction probability $V^*$ as functions of the defender's impact superiority parameter $g$ for $H_{max}$=25, $h$=1, $f$=3 and different values of $m$. $H^*$ and $x^*$ become more sensitive to variations of $g$ as $m$ increases. Similar to Figure 6.9 $V^*$ decreases with $g$ and approaches 0 when $g$ approaches infinity. $H^*$ and $x^*$ vary with $g$ and $m$ in very similar manner. The increase of $g$ makes any intelligence activity ineffective for the attacker ($X^*$=$J^*$=0 for $g$>2).

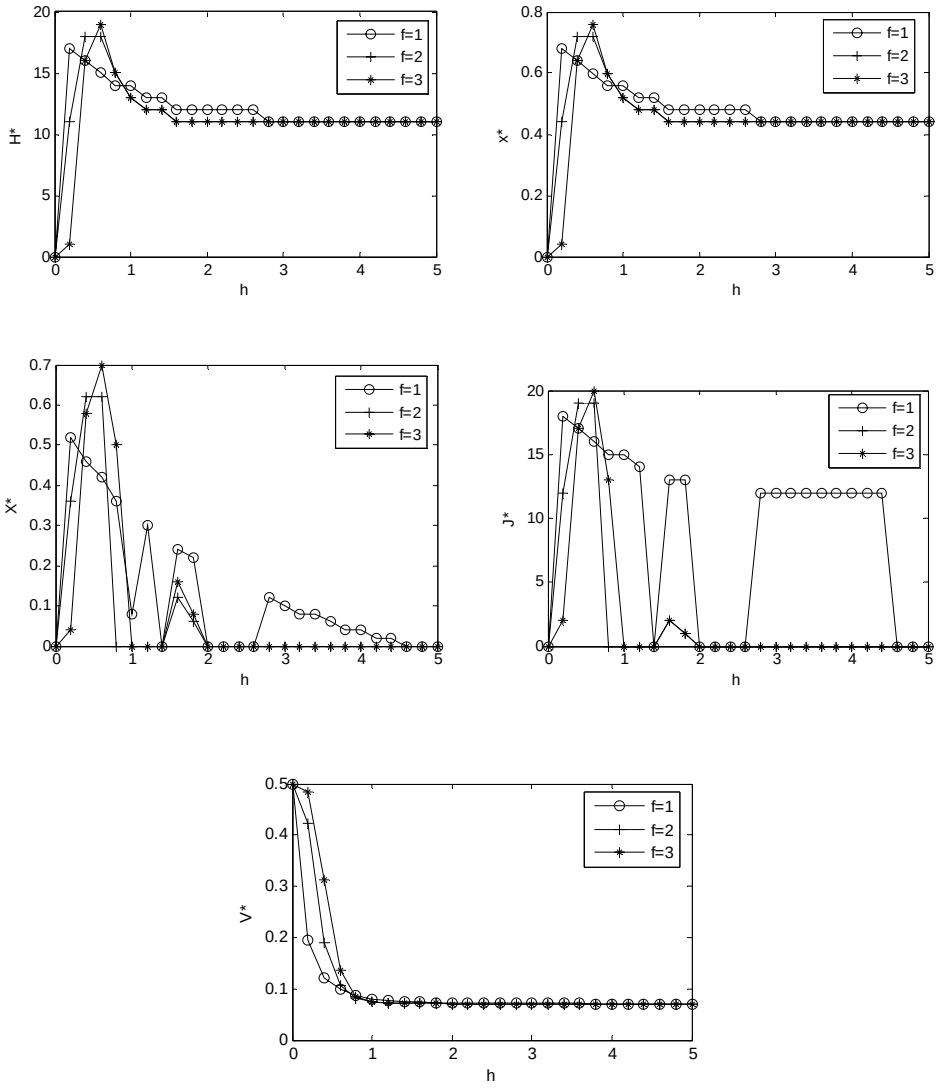


Figure 6.12 $H^*, x^*, X^*, J^*$ and $V^*$ as functions of $h$ for different $f$

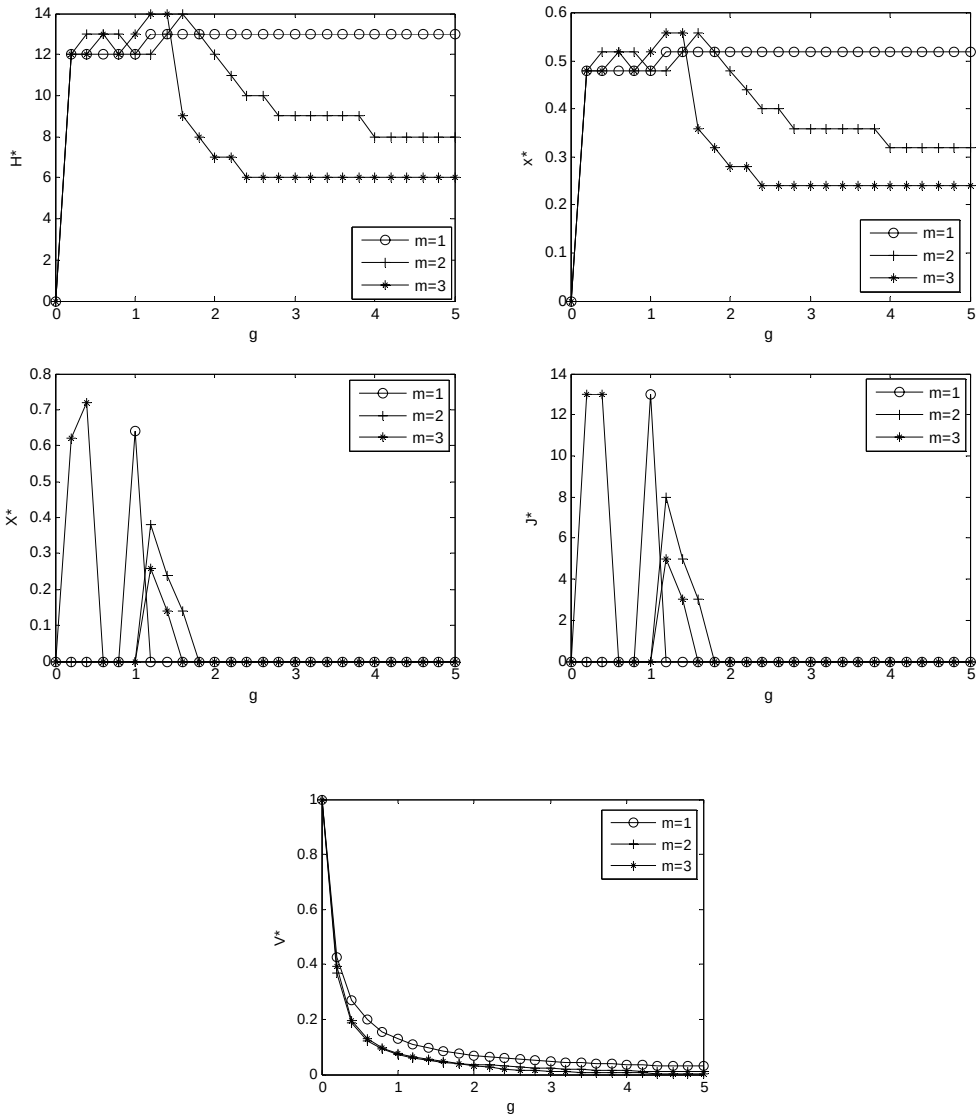


Figure 6.13 $H^*$, $x^*$, $X^*$, $J^*$ and $V^*$ as functions of $g$ for different $m$

## 6.5. Conclusions

This chapter considers the deployment of false targets as a measure to defense systems against intentional attacks. The defender deploys a genuine object and multiple false targets to divert the attacker. The defender allocates its resource between defending the genuine object and investing into disinformation actions to ensure that the attacker cannot distinguish the false targets from the genuine object. The attacker allocates its resource between attacking and investing into intelligence actions trying to detect the false targets. The detection probability of a FT is determined by the attacker's intelligence effort and the defender's disinformation effort allocated on it. Each FT can be detected individually and independently. If the attacker detects a certain number of FTs, it attacks a subset of randomly chosen undetected targets (the attacker chooses such number of targets in the subset that maximizes the object destruction probability). The vulnerability of the genuine object is determined by the attack effort and the protection effort allocated in it. The defender seeks to minimize the object destruction probability. The attacker seeks to maximize the object destruction probability. We consider a minmax two period game in which the defender chooses its strategy in the first period assuming that the attacker responds with the most harmful strategy in the second period.

The complex interaction of the free choice variables and parameters in the game makes its intuitive analysis impossible. The chapter suggests the probabilistic model of the game and presents a methodology based on the analysis of numerically obtained solutions.

The cases when the number of FTs is exogenously given and when this number is optimized by the defender are considered as well as the cases when the attacker tries to detect all FTs or optimally chooses the number of targets he tries to detect.

It is shown that in many cases the resource distribution parameters of both players ($x$ and $X$) behave very similarly, however in some cases the attacker's resource distribution parameter demonstrates oscillating behavior. These effects are similar to those observed in Levitin and Hausken (2009c).

It is demonstrated that for some parameters of the game any intelligence activity is not effective for the attacker. The methodology of numerical analysis that determines the intelligence effort efficiency conditions is presented.

# CHAPTER 7 OPTIMAL SYSTEM REPLACEMENT AND PROTECTION STRATEGY

For systems containing elements with increasing failure rates, preventive replacement of the elements is an efficient measure to increase the system reliability (Levitin and Lisnianski, 1999; Yeh et al., 2010; Chien et al., 2010). Replacing elements that have a high risk of failure, while reducing the chance of failure, can incur significant expenses, especially in systems with high replacement rates. Minimal repair, the less expensive option, enables the system element to resume its work after failure, but does not affect its hazard rate (Beichelt and Fischer, 1980; Beichelt and Franken, 1983; Chang et al., 2010). Since the component replacement reduces its failure rate, the more frequently the component is replaced the higher the availability of the component is. Besides internal failures, a component may also fail due to external impacts, say, natural disasters (Zhuang and Bier, 2007). In order to increase the survivability of a component under external impacts, defensive investments can be made to protect the component. It is reasonable to assume that the external impact frequency is constant over time and that the probability of the component destruction by the external impact decreases with the increase of the protection effort allocated on the component. A tradeoff exists between investments into

system maintenance and its protection. The optimal maintenance and protection strategy needs to take both of these factors into account in order to reach a solution that provides the desired system reliability at minimum cost.

This chapter considers a series-parallel system consisting of components with different characteristics (nominal performances, hazard functions, protection costs etc.). The objective is to minimize the total cost of the damage associated with unsupplied demand and the costs of the system maintenance and protection. A universal generating function (UGF) technique is used to evaluate the system availability for any maintenance and protection policy. A genetic algorithm is used for the optimization. Section 7.1 formulates the problem. Section 7.2 describes the method of calculating the system availability. Section 7.3 provides a description of the genetic algorithm. Numerical examples are shown in Section 7.4.

## 7.1. Problem formulation and description of system model

### 7.1.1. General model and assumptions

**Assumptions:**

1. All the system components are independent.

2. The failures caused by the internal causes and external impacts are independent.

3. The time spent on replacement is negligible.

4. The time spent on a minimal repair is much less than the time between failures.

A system that consists of $M$ subsystems connected in series is considered. Each subsystem $m$ contains $E_m$ elements connected in parallel. The lifetime for the system is denoted as $T_c$. For each component $i$, its nominal performance is denoted as $G_i$ and the expected number of internal failures during time interval $(0,t]$ is denoted as $\lambda_i(t)$, which is an increasing function of $t$. Each component is subjected to internal failures and external impacts. The failures caused by internal failures and external impacts are fixed by minimal repairs. It is assumed that the following two kinds of maintenance actions can be taken (Sheu and Chang, 2009):

1) Preventive replacement. The $i$-th component is replaced when it reaches an age $T_i$. The cost $C_i$ of each replacement is constant. As the preventive replacement is planned action the average time for the replacement is assumed to be negligible.

2) Minimal repair. This action is used after internal failures or destructive external impacts and doesn't affect the hazard function of the component. The average cost for a minimal repair of component $i$ is $\sigma_i$ in the case of internal failure and $\theta_i$ in the case of external impact. The average time for a minimal repair of component $i$ is $t_i$ in the case of internal failure and $\tau_i$ in the case of external impact.

### 7.1.2. The availability of each system element

The average number of internal failures during the period between replacements $\lambda_i(T_i)$ can be obtained by using the replacement interval $T_i$ for each element. Therefore, the total expected number of internal failures of the component $i$ during the system life cycle is

$$(n_i + 1)\lambda_i(T_i) = \frac{\lambda_i(T_i)T_c}{T_i} \tag{7.1}$$

where $n_i = \frac{T_c}{T_i} - 1$ is the number of preventive replacements $n_i$ during the system life cycle.

We use $x_i$ to denote the protection effort allocated on component $i$ and $a_i$ to denote the unit protection effort cost for component $i$. It is assumed that the external impact frequency $q$ is a constant and the expected impact intensity is $d$. The component vulnerability (conditional probability of a component failure caused by an external impact) is evaluated using the contest function model (Hausken 2005, Tullock 1980, Skaperdas 1996) as

$$v(x_i, d) = \frac{d^m}{x_i^m + d^m} \tag{7.2}$$

where $m$ is the contest intensity parameter. The expected number of the failures caused by the external impacts is therefore

$$q \cdot T_c \cdot v(x_i, d) = \frac{q \cdot T_c \cdot d^m}{x_i^m + d^m} \tag{7.3}$$

Hausken and Levitin (2008) discussed the meaning of the contest intensity parameter $m$. A benchmark intermediate value is $m$=1, which means that the investments into protection have proportional impact on the vulnerability reduction. 0<$m$<1 corresponds to the low effective types of protections with component vulnerability less sensitive to variation of the protection effort. $m$>1 corresponds to the highly effective types of protections with component vulnerability very sensitive to variation of the protection effort.

From (7.1) and (7.3) we have the total expected repair time of component $i$ as

$$r_i = \frac{t_i \lambda_i(T_i) T_c}{T_i} + q \cdot T_c \cdot v(x_i, d) \cdot \tau_i \tag{7.4}$$

Furthermore the availability of each element can be obtained as

$$A_i = \frac{T_c - r_i}{T_c} = \frac{T_c - t_i \lambda_i(T_i) T_c / T_i - q T_c v(x_i, d) \tau_i}{T_c}. \tag{7.5}$$

### 7.1.3. The system capacity distribution

The system capacity distribution must be obtained to estimate the entire system availability and the expected unsupplied demand. We use $\boldsymbol{G}=\{G_v\}$ to denote the vector of all the possible total system capacities, which corresponds to its $\boldsymbol{V}$ different possible states; and $\boldsymbol{P}=\{p_v\}$ to denote the vector of probabilities, which corresponds to these states.

The entire system capacity distribution can be defined by using the algorithm presented in Section 7.2 after the availabilities of the system elements are obtained with

(7.5). If we denote the system demand as $W$, the unsupplied demand probability should be calculated as

$$P_{ud} = \sum_{v=1}^{V} p_v \cdot 1(W - G_v > 0) \tag{7.6}$$

The reliability of the entire system requires an availability index $A$=1-$P_{ud}$ that is not less than some preliminary specified level $A^*$.

The total unsupplied demand cost can be estimated with the following expression

$$C_{ud} = \alpha \sum_{v=1}^{V} p_v \cdot (W - G_v) \cdot 1(W - G_v > 0) \tag{7.7}$$

where $\alpha$ is the cost of the unsupplied demand unit.

### 7.1.4. The formulation of the optimization problem

The optimization problem is to find the replacement intervals and protection efforts for system elements $\boldsymbol{T}$=($T_1$,$T_2$,…,$T_N$) and $\boldsymbol{x}$=$\{x_1,x_2,\ldots,x_N\}$ that minimize the sum of costs of the maintenance, protection, and unsupplied demand.

$$\begin{gathered} \boldsymbol{T}, \boldsymbol{x} = \arg\min \{C_{ud}(\boldsymbol{T}, \boldsymbol{x}) + \sum_{i=1}^{N} a_i x_i + \sum_{i=1}^{N} (T_c/T_i\text{-}1) C_i + \sum_{i=1}^{N} l_i\} \\ subject\ to: A \geq A^* \end{gathered} \tag{7.8}$$

where $a_i x_i$ is the protection cost on component $i$, $n_i C_i = \left( \frac{T_c}{T_i} - 1 \right) C_i$ is the expected preventive replacement cost of component $i$, and $l_i = \frac{\sigma_i \lambda_i(T_i) T_c}{T_i} + q \cdot T_c \cdot v(x_i, d) \cdot \theta_i$ is the expected minimal repair cost of component $i$ .

## 7.2. System availability estimation method

The entire system capacity distribution must be obtained in order to evaluate the availability index $A$ and the total unsupplied demand cost $C_{ud}$. The UGF was introduced in Ushakov (1986) and has proven to be extremely effective in evaluating reliability of complex multi-state systems (Liu and Huang, 2010; Yeh and He, 2010). The UGF of a discrete variable $G$ is defined as a polynomial

$$u(z) = \sum_{j=1}^{J} p_j z^{g_j}, \tag{7.9}$$

where the discrete random variable $G$ has $J$ possible values and $p_j$ is the probability that $G$ is equal to $g_j$. In our case, the polynomial $u(z)$ can define capacity distributions, meaning it represents all possible states of the system (or element) by relating the probabilities of each state $p_j$ with capacity $g_j$ of the system in this state.

Since each component $i$ has a nominal performance $g_i$ and its availability is $A_i$, the u-function of component $i$ has only two terms and can be defined as

$$u_i(z) = (1 - A_i)z^0 + A_i z^{g_i}, \tag{7.10}$$

The cumulative performance of parallel elements is equal to the sum of individual performances of these elements. Thus, the u-function of elements connected in parallel can be obtained by using the $\underset{+}{\otimes}$ operator

$$\underset{+}{\otimes}(u_1(z), u_2(z), ..., u_n(z)) = \prod_{i=1}^{n} u_i(z) \tag{7.11}$$

where

$$\underset{+}{\otimes}(u_1(z), u_2(z)) = \sum_{i=1}^{n} p_i z^{x_i} \underset{+}{\otimes} \sum_{j=1}^{J} q_j z^{y_j} = \left( \sum_{i=1}^{n} p_i z^{x_i} \right) \times \left( \sum_{j=1}^{J} q_j z^{y_j} \right) = \sum_{i=1}^{n} \sum_{j=1}^{J} p_i q_j z^{x_i + y_j}$$

The $\underset{+}{\otimes}$ operator is a product of polynomials representing the individual u-functions. Each term of the resulting polynomial is obtained by multiplying the probabilities that correspond to different states of elements and by reaching a summation of the elements' capacities that correspond to these states.

If a system contains subsystem connected in series, the subsystem with the minimal capacity bottlenecks the system. Therefore, this subsystem defines the total system capacity. The $\underset{\min}{\otimes}$ operator should be used to calculate the u-function for a system

containing $M$ subsystems connected in series. This operator for a pair of subsystems connected in series is defined as follows:

$$\underset{\min}{\otimes}(u_1(z),u_2(z)) = \sum_{i=1}^{n} p_i z^{x_i} \underset{\min}{\otimes} \sum_{j=1}^{J} q_j z^{y_j} = \sum_{i=1}^{n}\sum_{j=1}^{J} p_i q_j z^{\min(x_i,y_j)} \tag{7.12}$$

The simple operator should be used to evaluate the probability that the random variable $G$ represented by polynomial $u(z)$ defined in (7.9) does not exceed the value $W$:

$$P_{ud} = p(G \le W) = \delta(u(z),W) = \sum_{g_j \le W} p_j \tag{7.13}$$

Furthermore the availability index $A$ of the entire system can be obtained as

$$A = 1 - P_{ud} = 1 - \sum_{g_j \le W} p_j \,. \tag{7.14}$$

The total unsupplied demand cost can be estimated as

$$C_{ud} = \alpha \sum_{g_j \le W} p_j (W - g_j) \,. \tag{7.15}$$

## 7.3. Optimization technique

Equation (7.8) formulates a complicated combinatorial optimization problem. An exhaustive examination of all possible solutions is not realistic, considering reasonable time limitations. The genetic algorithm (GA) has proven to be an effective optimization

tool for a large number of complicated problems in reliability engineering (Coit and Smith, 1996, Levitin et al., 1998; Lisnianski and Levitin, 2003). To apply the GA to a specific problem the solution representation and the decoding procedures must be defined.

*A. Solution Representation and decoding procedures*

Each solution is represented by string $\boldsymbol{S}=\{s_1,s_2,\ldots,s_N\}$, where $s_i$ corresponds to component $i$ for each $i=1,2,\ldots,N$.

Each number $s_i$ determines both the replacement interval of component $i$ ($T_i$) and the protection effort allocated on component $i$ ($x_i$). To provide this property all the numbers $s_i$ are generated in the range

$$0 \le s_i < (M+1)\cdot\Lambda \tag{7.16}$$

where $M$ is the maximum protection effort allowed to be allocated on a component and $\Lambda$ is the total number of considered replacement frequency alternatives.

The solutions are decoded in the following manner:

$$x_i = [s_i / \Lambda] \tag{7.17}$$

$$v_i = 1 + \mathrm{mod}_\Lambda s_i \tag{7.18}$$

where $v_i$ is the number of replacement frequency alternative for component $i$, $[x]$ is the maximal integer not greater than $x$, and $\mathrm{mod}_x y = y - [y/x]x$. For given $v_i$ and $x_i$ the corresponding $s_i$ is composed as follows:

$$s_i = x_i \cdot \Lambda + v_i - 1 \tag{7.19}$$

Note that all $s_i < \Lambda$ corresponds to the solutions where the component $i$ is not protected.

The possible replacement frequency alternatives are ordered in vector $\boldsymbol{h}=\{h_1, h_2, \ldots, h_\Lambda\}$ so that $h_i < h_{i+1}$, where $h_i$ represents the number of replacements during the operation period that corresponds to alternative $i$. After obtaining $v_i$ from decoding the solution string, the number of replacement for component $i$ can be obtained as

$$n_i = h_{v_i} \tag{7.20}$$

Furthermore replacement interval for component $i$ can be obtained as

$$T_i = \frac{T_c}{n_i + 1} = \frac{T_c}{h_{v_i} + 1} \tag{7.21}$$

For each given pair of vectors $(\boldsymbol{T},\boldsymbol{x})$ the decoding procedure first calculates the availability of each element using (7.5), after which the entire system capacity distribution can be obtained by using (7.10), (7.11) and (7.12). The availability index $A$ and the total unsupplied demand cost $C_{ud}$ can be obtained using (7.13), (7.14) and (7.15).

In order to let the genetic algorithm search for the solution with minimal total cost, when $A$ is not less than the required value $A^*$, the solution quality (fitness) is evaluated as follows:

$$F(\boldsymbol{T}, \boldsymbol{x}) = \omega \cdot (A^* - A) \cdot 1(A^* - A) + C_{ud}(\boldsymbol{T}, \boldsymbol{x}) + \sum_{i=1}^{N} a_i x_i + \sum_{i=1}^{N} (T_c / T_i \text{-} 1) C_i + \sum_{i=1}^{N} l_i \qquad (7.22)$$

where $\omega$ is a sufficiently large penalty.

For solutions that meet the requirements $A \geq A^*$, the fitness of the solution is equal to its total cost.

*B. Crossover and mutation procedures*

The cross operator for given parent strings $\boldsymbol{P1}$, $\boldsymbol{P2}$ and the offspring string $\boldsymbol{O}$ is defined as follows: the $i$-th element ($1 \leq i \leq N$) of the string $\boldsymbol{O}$ is equal to the $i$-th element of either $\boldsymbol{P1}$ or $\boldsymbol{P2}$ both with probability 0.5.

The mutation procedure swaps elements initially located in two randomly chosen positions.

## 7.4. Illustrative examples

The system considered in this example consists of two subsystems. The first subsystem contains 5 components while the second subsystem contains 3 components. The lifetime of the system $T_c$ is 120 months. The system demand $W$ is 45. Figure 7.1 is shown for graphical illustration.

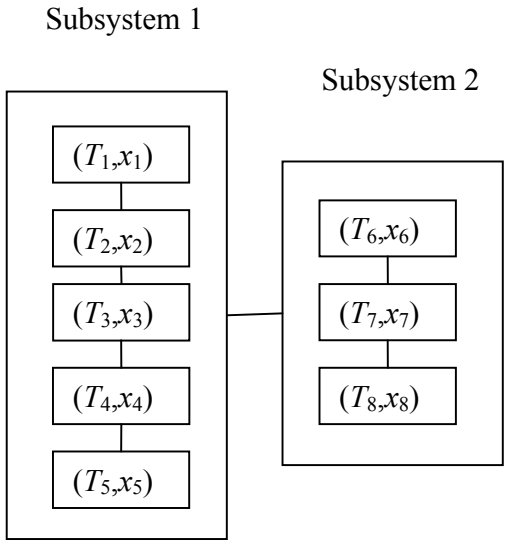


Figure 7.1 Graphical illustration of the considered system

We assume that totally $\Lambda$=6 different replacement frequency alternatives are considered and the alternatives are $\boldsymbol{h}$={4, 9, 14, 19, 24, 29}. The corresponding alternatives for replacement interval are 24 months, 12 months, 8 months, 6 months, 4.8 months and 4 months. The characteristics of the components are presented in Table 7.1.

Table 7.1 The characteristics of the components

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $G_i$ | 10 | 10 | 12 | 12 | 14 | 15 | 15 | 20 |
| $a_i$ | 7 | 8 | 8 | 9 | 10 | 10 | 11 | 12 |
| $C_i$ | 100 | 100 | 110 | 110 | 120 | 140 | 140 | 150 |
| $t_i$(month) | 0.030 | 0.033 | 0.036 | 0.042 | 0.045 | 0.045 | 0.048 | 0.048 |
| $\sigma_i$ | 8 | 8 | 8 | 7 | 8 | 6 | 8 | 7 |
| $\tau_i$(month) | 0.036 | 0.042 | 0.045 | 0.048 | 0.051 | 0.051 | 0.054 | 0.054 |
| $\theta_i$ | 9 | 9 | 9 | 8 | 9 | 7 | 9 | 8 |
| $\lambda_i(4)$ | 0.8 | 0.72 | 0.64 | 0.6 | 0.72 | 0.7 | 0.6 | 0.5 |
| $\lambda_i(4.8)$ | 1.04 | 0.92 | 0.85 | 0.8 | 0.96 | 0.9 | 0.85 | 0.8 |
| $\lambda_i(6)$ | 1.6 | 1.5 | 1.4 | 1.2 | 1.5 | 1.4 | 1.4 | 1.4 |
| $\lambda_i(8)$ | 2.8 | 2.7 | 2.7 | 2.1 | 2.7 | 2.4 | 2.7 | 2.8 |
| $\lambda_i(12)$ | 5.5 | 5.2 | 5.4 | 4.2 | 5.3 | 4.8 | 5.4 | 5.8 |
| $\lambda_i(24)$ | 15 | 15 | 15 | 14 | 15 | 14 | 14 | 16 |

For our example we assume that $q$=0.5, $d$=30, $\alpha$=3000, and the maximum protection effort allowed to be allocated on a component is $M$=50. According to (7.16) we have

$$0 \le s_i < (M+1)\cdot\Lambda = 306$$

For a given solution string, $\boldsymbol{T}$ and $\boldsymbol{x}$ can be decoded using (7.17), (7.18), (7.20) and (7.21). Thereafter (7.22) can be used to obtain the fitness function $\boldsymbol{F}(\boldsymbol{T},\boldsymbol{x})$. For example, the solution string $\boldsymbol{S}$=[195 280 49 218 176 132 270 120] is decoded into $\boldsymbol{T}$= (6 4.8 12 8 8 24 24 24) and $\boldsymbol{x}$ = (32 46 8 36 29 22 45 20). For $m$=1 and $A^*$=0.90, the fitness function for this solution takes the value $F(\boldsymbol{T},\boldsymbol{x})$=18190.

The problem is to find the optimal replacement and protection strategy $(\boldsymbol{T},\boldsymbol{x})$ which minimizes $F(\boldsymbol{T},\boldsymbol{x})$ subject to the availability requirement $A>A^*$.

Table 7.2 contains the optimal solutions obtained for $m$=1 and different values of $A^*$. Each solution was obtained as the optimal one among five different runs of the GA with different randomly generated initial populations. The coefficients of variation among the values of $F(\boldsymbol{T},\boldsymbol{x})$ obtained in the five runs are also presented in Table 7.2. The low values of this coefficient evidence the good consistency of the GA. With the increase of the reliability requirement more resources need to be put into protection actions and the components need to be replaced more frequently, thus the total cost increases. It can also be seen that $C_{ud}$ decreases with the increase of $A^*$. Indeed, when the obtained system availability increases, the unsupplied demand decreases.

Table 7.2 Examples of solutions obtained for $m$=1

| Constraints | $\boldsymbol{T}$ | $\boldsymbol{x}$ | $A$ | $C_{ud}$ | $F(\boldsymbol{T},\boldsymbol{x})$ | variation |
|---|---|---|---|---|---|---|
| *None* | (24 24 24 24 24 24 24 24) | (19 19 19 13 19 13 13 10) | 0.8946 | 1008.9 | 13143 | 0.02% |
| $A^*$=*0.90* | (24 24 24 24 24 24 24 24) | (21 21 21 14 26 14 14 21) | 0.9008 | 955.8333 | 13187 | 0.65% |
| $A^*$=*0.95* | (24 24 24 24 6 24 24 4.8) | (32 32 32 32 42 32 32 50) | 0.9505 | 549.9153 | 17482 | 1.66% |

The maximum availability $A$=0.9645 can be achieved when maximal possible protection and replacement frequency are applied. In this case all the components are replaced every 4 months and the protection effort on each component is 50. The corresponding total cost is 34784.

Table 7.3 contains the optimal solutions obtained for $m$=0.25 and different values of $A^*$. Similar as in the case $m$=1, the total cost increases whereas unsupplied demand cost decreases with the increase of $A^*$. For low intensive contest with $m$=0.25 the total incurred cost becomes lower than in the case of $m$=1 for small $A^*$ and greater than in the case of $m$=1 for high $A^*$. Indeed, when $A^*$ is low the protection effort of an element is generally smaller than the external impact intensity. In this case for smaller $m$ the sensitivity of the element vulnerability to the reduction of the protection effort decreases and the defender

can afford to spend less into the protection. On the contrary, when $A^*$ is high, the protection effort of an element is generally bigger than the external impact intensity. But for smaller $m$ it becomes more difficult to achieve vulnerability reduction by increasing the protection effort and the defender must spend more into the protection.

Table 7.3 Examples of solutions obtained for $m$=0.25

| Constraints | $\boldsymbol{T}$ | $\boldsymbol{x}$ | $A$ | $C_{ud}$ | $F(\boldsymbol{T},\boldsymbol{x})$ | variation |
|---|---|---|---|---|---|---|
| *None* | (24 24 24 24 24 24 24 24) | (6 4 4 4 4 4 4 4) | 0.8981 | 957.7501 | 12112 | 0.44% |
| $A^*$=*0.90* | (24 24 24 24 24 24 24 24) | (6 6 6 6 8 6 6 6) | 0.9000 | 931.1970 | 12147 | 0.23% |
| $A^*$=*0.95* | (24 12 24 12 4 12 24 4.8) | (12 17 12 17 32 17 12 32) | 0.9501 | 486.5937 | 18842 | 1.04% |

The maximum availability $A$=0.9591 can be achieved when maximal possible protection and replacement frequency are applied. The corresponding total cost is 35210, which is greater than the case $m$=1. This is because when $m$=0.25, the protection is not as effective as in the case $m$=1.

Table 7.4 contains the optimal solutions obtained for $m$=4 and different values of $A^*$. Similar as in the cases $m$=1 and $m$=0.25, the total cost increases while the unsupplied demand cost decreases with the increase of $A^*$. The results obtained for the first two cases

are very similar. Actually the near optimal solution obtained without availability constraint has an availability bigger than 0.90. Thus it is also a near optimal solution for the case $A^*$=0.90. As can be seen the protection effort allocated on each element is always very big. This is because the domination of protection effort over external impact is very important when $m$=4.

Table 7.4 Examples of solutions obtained for $m$=4

| Constraints | $\boldsymbol{T}$ | $\boldsymbol{x}$ | $A$ | $C_{ud}$ | $F(\boldsymbol{T},\boldsymbol{x})$ | variation |
|---|---|---|---|---|---|---|
| *None* | (24 24 24 24 24 24 24 24) | (45 43 45 42 43 41 41 41) | 0.9238 | 569.7401 | 12865 | 0.08% |
| $A^*$=*0.90* | (24 24 24 24 24 24 24 24) | (44 44 44 41 44 41 41 36) | 0.9215 | 581.4035 | 12866 | 0.07% |
| $A^*$=*0.95* | (24 24 24 24 8 24 24 8) | (44 44 44 42 50 42 42 49) | 0.9556 | 397.0332 | 14931 | 0.25% |

The maximum availability $A$=0.9792 can be achieved when maximal possible protection and replacement frequency are applied. The corresponding total cost is 33620, which is less than the case $m$=1. This is because the protection is more effective than in the case $m$=1.

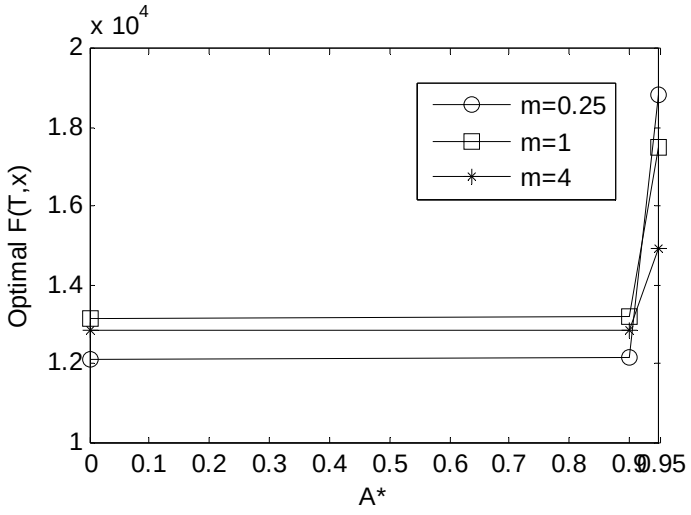


Figure 7.2 Optimal F(***T***,***x***) as functions of $A^*$ for different values of $m$

Figure 7.2 shows the optimal F(***T***,***x***) as functions of $A^*$ for different values of $m$. As can be seen the optimal F(***T***,***x***) curve for $m$=1 is above the curve for $m$=0.25 when $A^*$ is small and below the curve for $m$=0.25 when $A^*$ is big enough. The optimal F(***T***,***x***) curve for $m$=4 is always below that for $m$=1, as the dominant protection effort can considerably reduce the failures caused by external impacts when $m$=4.

## 7.5. Conclusions

In this chapter the optimal resource allocation between replacement and protection of components in a series-parallel system is studied. It is assumed that the failure rate of each component is increasing over time and the failures between replacements are fixed by

minimal repairs. Since the component replacement reduces its failure rate, the more frequently a component is replaced the higher the availability of the component will be. On the other hand, the components may fail due to external impacts. It is assumed that the external impact frequency is constant. The destruction probability of a component in the case of an external impact is determined by the external impact intensity, the protection effort on the component and the contest intensity. The bigger the protection effort allocated on a component is, the lower its destruction probability will be. Thus a tradeoff exists between investments into system maintenance and its protection. In this chapter a framework is proposed to solve the optimal maintenance and protection strategy that provides the desired system reliability at minimum cost, which includes the total cost of the damage associated with unsupplied demand and the costs of the system maintenance and protection. Universal generating function technique is used to obtain the system availability and the total cost of the damage associated with unsupplied demand. Solutions are encoded into strings and a genetic algorithm is employed to search for the string with the best fitness function. Finally the optimal solution is obtained by decoding the optimal string.

Numerical examples are shown in this chapter. With the increase of the reliability requirement more resources need to be put into protection actions and the components need to be replaced more frequently, thus the total cost increases. Meanwhile with the increase of the obtained system availability, the unsupplied demand decreases. The maximum availability can be achieved when maximal possible protection and replacement frequency are applied.

It is shown that whether the total cost increases or decreases with the increase of contest intensity depends on the element protection efforts. When the protection effort on each element is generally superior to the external impact intensity, the total cost decreases with the increase of the contest intensity. Otherwise the total cost increases with the increase of the contest intensity.

# CHAPTER 8 CONLUSIONS AND FUTURE WORKS

## 8.1. Conclusions

In this thesis, the reliability of some networked systems is investigated. These systems include series-parallel systems subject to imperfect fault coverage, linear multi-state consecutively connected systems, series/parallel systems against external intentional attacks and series-parallel systems subjected to both internal failures and external attacks.

Chapter 3 studies the reliability of series-parallel systems with the consideration of imperfect fault coverage. It is assumed that the elements in the same subsystem can be divided into different work sharing groups to perform the same task. Due to imperfect fault coverage, a whole work sharing group can fail if one element fails and the failure is not covered. Different fault coverage models are discussed and the problem of finding the optimal balance between redundancy and task sharing is extended to the cases of multi-fault coverage and performance dependent coverage. Illustrative examples are presented to show that the greatest system reliability (defined as a probability of meeting a certain demand) can be achieved by proper balance between two types of task parallelization.

Chapter 4 studies a linear multi-state consecutively connected system (LMCCS) consisting of elements with increasing failure rates. A framework is proposed to solve the joint element allocation and maintenance optimization problem for LMCCS which minimizes the total system maintenance cost subject to pre-specified system availability requirements. The optimal elements allocation and maintenance strategy are found in the example for three different cases: 1) Fixed element allocation; 2) Even elements distribution among the nodes (no node contains more than one element); 3) Arbitrary allocation of the elements. For all the cases, the minimum maintenance cost increases with the increase of the availability requirement. It is revealed clearly in the results that the flexibility of element allocation enables the system to achieve much higher availability with less maintenance cost.

Chapter 5 and 6 study the defense of systems against external attacks. Chapter 5 considers simple series and parallel systems against external intentional attacks. Different from existing papers which only consider perfect false targets, it is assumed that each false target (FT) has a nonzero probability to be detected by the attacker and the detections of different FTs are independent. The methodology of analysis of optimal defense strategy as function of different parameters (number of GEs, contest intensity, total attacker's resource) is demonstrated. The decision curves are also presented which can be used for the making a decision about efficiency of deploying FTs depending on their cost and detection probability. Chapter 6 considers defending a single genuine object with imperfect false targets. Different from Chapter 5, the detection probability of a false target is assumed to be a function of the attacker’s intelligence effort and the defender’s disinformation effort. The cases when the number of FTs is exogenously given and when this number is

optimized by the defender are considered as well as the cases when the attacker tries to detect all FTs or optimally chooses the number of targets he tries to detect.

Chapter 7 studies the optimal resource allocation between replacement and protection of components in a series-parallel system. It is assumed that the failure rate of each component is increasing over time and the failures between replacements are fixed by minimal repairs. On the other hand, the components may fail due to external impacts. It is assumed that the external impact frequency is constant. A framework is proposed to solve the optimal maintenance and protection strategy that provides the desired system reliability at minimum cost, which includes the total cost of the damage associated with unsupplied demand and the costs of the system maintenance and protection. Numerical examples are shown to illustrate the application. With the increase of the reliability requirement more resources need to be put into protection actions and the components need to be replaced more frequently, thus the total cost increases. Meanwhile with the increase of the obtained system availability, the unsupplied demand decreases.

## 8.2. Future works

This section discusses the limitations of the works contained in this thesis and suggests some directions for future research.

In chapter 3, it is assumed that the same task can be shared by different components in the optimal way. An implicit assumption is that the task can be divided arbitrarily so that a component with greater capacity takes greater amount of task load. In reality, there may be situations where a task can only be divided into discrete number of subtasks. Although the insight of the current research still applies, a framework needs to be proposed to solve the optimal allocation of subtasks into different components. It would be an interesting and challenging issue to incorporate different kinds of fault coverage models with discrete division of tasks. Moreover, in chapter 3 universal generating function is used to calculate the performance distribution of the entire system, it would be interesting to try other methodologies, such as fault tree analysis and ordered binary decision diagram.

In chapter 4, it is assumed that the number of elements that are available is fixed and the allocation and maintenance of these elements are studied. There are situations where different versions of elements are available, say, in the market. In this case, the problem is to decide the number of each version of elements to be allocated into each position and the maintenance actions to be implemented on these elements. The total cost will contain not only the maintenance cost, but also the cost of elements themselves. Another thing that can be done is to use iterative methods instead of universal generating functions to calculate the reliability of linear multi-state consecutively connected systems. The computational complexity of different methods can be compared.

There are a lot of things that can be done on defending system against intentional attacks. Chapter 5 studies the optimal defense of systems with imperfect false targets. It is assumed that only one type of false targets is available. It would be interesting to consider

the case where there are multiple types of false targets with different unit costs and detection probabilities. A framework needs to be proposed to solve the optimal combination of different types of false targets. Another research that can be done is to study the uncertainty that is caused by the contest intensity parameter. The model in chapter 5 uses the contest intensity parameter $m$ that cannot be exactly evaluated in practice. Therefore the study of the influence of this parameter on the optimal and minmax strategies has a qualitative nature. Two ways of handling the uncertainty of the contest intensity can be outlined: first, $m$ can be defined as a fuzzy variable and fuzzy logic model can be studied; second, the range of possible variation of $m$ can be determined and the most conservative "worst case" defense strategy can be obtained under the assumption that $m$ takes the values that are most favorable for the attacker (in this case $m$ can be considered as an additional strategic variable that the attacker can choose within the specified range). The model consider in chapter 5 can also be extended to other systems, say, consecutively connected systems. In consecutively connected systems, some elements are in more important positions than others. Therefore the defender may prefer to allocate more protection efforts on some elements than others and the attacker also prefers to attack the most fragile parts of the system in order to maximize the system destruction probability. It would be very interesting to model the counter-contest between the defender and the attacker. In chapter 5 and 6, the contest between the defender and the attacker is modeled as a two-period game where the defender constructs the system at first period and the attacker attacks the system in second period. Further study can be done to study the case when the defender can take pre-strike to destroy or weaken the attacker's base.

Chapter 7 studies the optimal system maintenance and protection strategy when the system is subjected to internal failures and external attacks. The external attacks considered are limited to unintentional attacks, say, natural disasters. It must be interesting to study the optimal resource allocation strategy when the system is subjected to internal failures and both unintentional and intentional attacks. Say, an attacker who does not have intention to attack may choose to attack when a system is weakened by internal failures or natural disasters. How to model the problem realistically and meanwhile maintain mathematical tractability is an issue to be investigated.

# REFERENCES

Amari, S., 1997. Reliability risk and fault-tolerance of complex systems. Phd Thesis, Indian Institute of Technology, Kharagpur.

Amari, S.V., Dugan, J. B., Misra, R. B., 1999. Optimal reliability of systems subject to imperfect fault-coverage. IEEE Transactions on Reliability 48 (3), 275-284.

Amari, S., McLaughlin, L., Yadlapati, B., 2003. Optimal cost-effective design of parallel systems subject to imperfect fault-coverage. In Proceedings of Reliability and Maintainability Symposium, pp. 29-34.

Amari, S., Pham, H., Dill, G., 2004. Optimal design of k-out-of-n: G subsystems subjected to imperfect fault-coverage. IEEE Transactions on Reliability 53, 567-575.

Ambani, S., Meerkov, S.M., Zhang, L., 2010. Feasibility and optimization of preventive maintenance in exponential machines and serial lines. IIE Transactions 42 (10), 766-777.

Arnold, T.F., 1973. The concept of coverage and its effect on the reliability model of a repairable system. IEEE Transactions on Computers 22: 325–339.

Azaiez, M.N., Bier, V.M., 2007. Optimal resource allocation for security in reliability systems. European Journal of Operational Research 181, 773-786.

Bavuso, S.J., et al., 1994. HiRel: Hybrid Automated Reliability Predictor (HARP) Integrated Reliability Tool System (Version 7.0), 4 vol.s, NASA TP 3452.

Beichelt, F., Fischer, K., 1980. General failure model applied to preventive maintenance policies. IEEE Transactions on Reliability 29, 39–41.

Beichelt, F., Franken, P., 1983. Zuverlassigkeit und Instandhaltung. Berlin: Verlag Technik.

Bier, V., Abhichandani, V., 2002. Optimal allocation of resources for defense of simple series and parallel systems from determined adversaries. Proceedings of the Engineering Foundation Conference on Risk-Based Decision Making in Water Resources X, Santa Barbara, CA, American Society of Civil Engineers.

Bier, V.M., Nagaraj, A., Abhichandani, V., 2005. Protection of simple series and parallel systems with components of different values. Reliability Engineering and System Safety 87, 315–323.

Blanks, K.S., 1994. An effectiveness analysis of the tactical employment of decoys. Master's thesis, Pentagon, Army Command and General Staff Coll Fort Leavenworth KS, Accession Number: ADA284608, http://www.stormingmedia.us/80/8064/A806482.html.

Bouricius, W.G., Carter, V., Schneider, P.R., 1969. Reliability modeling techniques for self-repairing computer systems. In Proceedings of the 24th National Conference, ACM, pp. 295-309.

Chang, C.C., Sheu, S.H., Chen, Y.L., 2010. Optimal number of minimal repairs before replacement based on a cumulative repair-cost limit policy. Computers & Industrial Engineering 59 (4), 603-610.

Chang, Y.R., Amari, S.V., Kuo, S.Y., 2005. OBDD-based evaluation of reliability and importance measures for multistate systems subject to imperfect fault coverage. IEEE Transactions on Dependable and Secure Computing 2 (4), 336-347.

Chakravarthy, S.R., Gómez-Corral, A., 2009. The influence of delivery times on repairable k-out-of-N systems with spares. Applied Mathematical Modelling 33, 2368-2387.

Chien, Y.H., Chang, C.C., Sheu, S.H., 2010. Optimal age-replacement model with age-dependent type of failure and random lead time based on a cumulative repair-cost limit policy. Annals of Operations Research 181 (1), 723-744.

Cluzeau, T., Keller, J., Schneeweiss, W., 2008. An efficient algorithm for computing the reliability of consecutive-*k*-out-of-*n*: F system. IEEE Transactions on Reliability 57 (1), 84-87.

Coit, D., Smith, A., 1996. Reliability optimization of series-parallel systems using genetic algorithm. IEEE Transactions on Reliability 45, 254–266.

Cook., J., Ramirez-Marquez, J.E., 2009. Mobility and reliability modeling for a mobile ad hoc network. IIE Transactions 41 (1), 23-31.

Dighe, N., Zhuang, J., Bier, V.M., 2009. Secrecy in defensive allocations as a strategy for achieving more cost-effective attacker deterrence. International Journal of Performance Engineering 5 (1), 31-43.

Ding, Y., Lisnianski, A., Frenkel, I., Khvatskin, L., 2009. Optimal corrective maintenance contract planning for aging multi-state system. Applied Stochastic Models in Business and Industry 25 (5), 612-631.

Ding, Y., Zuo, M.J., Lisnianski, A., Li, W., 2010. A framework for reliability approximation of multi-state weighted *k*-out-of-*n* systems. IEEE Transactions on Reliability 59 (2), 297-308.

Eryilmaz, S., Tutuncu, G.Y., 2009. Reliability evaluation of linear consecutive-weighted-k-out-of-n: F system. Asia-Pacific Journal of Operational Research 26 (6), 805-816.

Feldman, R.M., Chen, M.C., 1996. Strategic and tactical analyses for optimal replacement policies. IIE Transactions 28 (12), 987-993.

Gen, M., Cheng, R., 1997. Genetic algorithms and engineering design. New York: John Wiley & Sons.

Goldberg, D., 1989. Genetic algorithms in search, optimization and machine learning. Reading, MA: Addison Wesley.

Hauksen, K., 2005. Production and conflict models versus rent seeking models. Public Choice 123, 59-93.

Hausken, K., 2008. Strategic defense and attack for reliability systems. Reliability Engineering and System Safety 93, 1740-1750.

Hausken, K., Levitin, G., 2008. Efficiency of even separation of parallel elements with variable contest intensity. Risk Analysis 28 (5), 1477-1486.

Hausken, K., Levitin, G., 2009. Protection vs. false targets in series systems. Reliability Engineering and System Safety 94 (5), 973-981.

Huang, H.Z., Qu, J., Zuo, M.J., 2009. Genetic-algorithm-based optimal apportionment of reliability and redundancy under multiple objectives. IIE Transactions 41(4), 287-298.

Huang, L., Xu, Q., 2010. Lifetime reliability for load-sharing redundant systems with arbitrary failure distributions. IEEE Transactions on Reliability 59 (2), 319-330.

Hwang, F., Yao, Y., 1989. Multistate consecutively-connected systems. IEEE Transactions on Reliability 38, 472-474.

Kossow, A., Preuss, W., 1995. Reliability of linear consecutively connected systems with multistate components. IEEE Transactions on Reliability 44, 518-522.

Lee, Y.J., Na, M.G., 2009. Design of delay-tolerant controller for remote control of nuclear reactor power. Nuclear Engineering and Technology 41 (1), 71-78.

Levitin, G., 2001. Reliability evaluation for linear consecutively-connected systems with multistate elements and retransmission delays. Quality and Reliability Engineering International 17 (5), 373-378.

Levitin, G., 2003. Optimal allocation of multi-state elements in linear consecutively connected systems with vulnerable nodes. European Journal of Operational Research 150, 406-419.

Levitin, G., 2005. Universal generating function in reliability analysis and optimization. Springer-Verlag, London.

Levitin, G., 2007a. Block diagram method for analyzing multi-state systems with uncovered failures. Reliability Engineering and System Safety 92 (6), 727-734.

Levitin, G., 2007b. Optimal defense strategy against intentional attacks. IEEE Transactions on Reliability 56 (1), 148-157.

Levitin, G., 2008. Optimal structure of multi-state systems with uncovered failures. IEEE Transactions on Reliability 57 (1), 140-148.

Levitin, G., Lisnianski, A., Beh-Haim, H., Elmakis, D., 1998. Redundancy optimization for series-parallel multi-state systems. IEEE Transactions on Reliability 47, 165-172.

Levitin, G., Lisnianski, A., 1999. Joint redundancy and maintenance optimization for multistate series-parallel systems. Reliability Engineering and System Safety 64, 33-42.

Levitin, G., Amari, S.V., 2008a. Multi-state systems with static performance-dependent fault coverage. Journal of Risk and Reliability, Proc. IMechE, PartO: J. Risk and Reliability, 222 (O2), pp. 95-103.

Levitin, G., Amari, S.V., 2008b. Multi-state systems with multi-fault coverage. Reliability Engineering and System Safety 93, 1790-1739.

Levitin, G., Hausken, K., 2008. Protection vs. redundancy in homogeneous parallel system. Reliability Engineering and Systems Safety 93, 1444-1451.

Levitin, G., Hausken, K., 2009a. False targets efficiency in defense strategy. European Journal of Operational Research 194, 155-162.

Levitin, G., Hausken, K., 2009b. Redundancy vs. protection vs. false targets for systems under attack. IEEE Transactions on Reliability 58 (1), 58-68.

Levitin, G., Hausken, K., 2009c. Intelligence and impact contests in systems with fake targets. Defense and Security Analysis 25, 157-173.

Li, C.Y., Chen, X., Yi, X.S., Tao, J.Y., 2010. Heterogeneous redundancy optimization for multi-state series-parallel systems subject to common cause failures. Reliability Engineering and Systems Safety 95 (3), 202-207.

Lisnianski, A., Levitin, G., 2003. Multi-state system reliability. Assessment, optimization and applications. : World Scientific.

Lisnianski, A., Khvatskin, L., Frenkel, I., Ding, Y., 2008. Maintenance contract assessment for aging systems. Quality and Reliability Engineering International 24 (5), 519-531.

Liu, Y., Huang, H.Z., 2010. Optimal replacement policy for multi-state system under imperfect maintenance. IEEE Transactions on Reliability 59 (3), 483-495.

Liu, Y., Li, Y.F., Huang, H.Z., Zuo, M.J., Sun, Z.Q., 2010. Optimal preventive maintenance policy under fuzzy Bayesian reliability assessment environments. IIE Transactions 43 (10), 734-745.

Malinowski, J., Preuss, W., 1996. Reliability increase of consecutive-k-out-of-n: F and related systems through components' rearrangement. Microelectronics and Reliability 36, 1417-1423.

Myers, A., 2008. Achievable limits on the reliability of k-out-of-n: G systems subject to imperfect fault coverage. IEEE Transactions on Reliability 57 (2), 349-354.

Myers, A., Rauzy, A., 2008. Efficient reliability assessment of redundant systems subject to imperfect fault coverage using binary decision diagrams. IEEE Transactions on Reliability 57 (2), 336-348.

Nakagawa, T., Mizutani, S., 2009. A summary of maintenance policies for a finite interval. Reliability Engineering and Systems Safety 94, 89-96.

Patterson, S., Apostolakis, G., 2007. Identification of critical locations across multiple infrastructures for terrorist actions. Reliability Engineering and System Safety 92 (9), 1183-1203.

Pekoz, E.A, and Ross, S.M, 1995. A simple derivation of exact reliability formulas for linear and circular consecutive-*k*-out-of-*n*: F systems. Journal of Applied Probability 32 (2), 554-557.

Perhinschi, M.G., Napolitano, M.R., Campa, G., Seanor, B., Burken, J., Larson, R., 2006. Design of safety monitor schemes for a fault tolerant flight control system. IEEE Transactions on Aerospace and Electronic Systems 42 (2), 562-571.

Powell, R., 2007a. Allocating defensive resources with private information about vulnerability. American Political Science Review 101 (4), 799-809.

Powell, R., 2007b. Defending against terrorist attacks with limited resources. American Political Science Review 101 (3), 527-541.

Ramirez-Marquez, J.E, Rocco S., C.M., Levitin, G., 2009. Optimal protection of general source-sink networks via evolutionary techniques. Reliability Engineering and System Safety 94, 1676-1684.

Rao, P.N.S., Naikan, V.N.A., 2009. An algorithm for simultaneous optimization of parameters of condition-based preventive maintenance. Structural Health Monitoring 8 (1), 83-94.

Shanthikumar, J.G., 1981. A general software reliability model for performance prediction. Microelectronics and Reliability 23, 903-943.

Sheu, S.H., Chang, C.C., 2010. Extended periodic imperfect preventive maintenance model of a system subjected to shocks. International Journal of Systems Science 41 (10), 1145-1153.

Shier, D.R., 1991. Network Reliability and Algebraic Structures. Clarendon Press New York, NY, USA.

Skaperdas, S., 1996. Contest success functions. Economic Theory 7, 283-290.

Tian, Z.G., Zuo, M.J., Huang, H.Z., 2008. Reliability-redundancy allocation for multi-state series-parallel systems. IEEE Transactions on Reliability 57 (2), 303-310.

Tian, Z.G., Zuo, M.J., Yam, R.C.M., 2009. Multi-state k-out-of-n systems and their performance evaluation. IIE Transactions 41 (1), 32-44.

Tullock, G., 1980. Efficient rent-seeking. In: Buchanan JM, Tollison RD, Tulluck G, editors. Toward a theory of the rent-seeking society. College Station, Tx: Texas A&M. University Press; 1980. pp. 97-112.

Ushakov, I., 1987. Optimal standby problems and a universal generating function. Soviet Journal of Computer Systems Science 25, 79-82.

Whitley, D., 1989. The GENITOR algorithm and selective pressure: Why rank-based allocation of reproductive trials is best. In Proceedings of 3rd International Conference on Genetic Algorithms, D. Schaffer, Ed., pp. 116-121, Morgan Kaufmann.

Wu, J., Ng, T.S.A., Xie, M., Huang, H.Z., 2010. Analysis of maintenance policies for finite life-cycle multi-state systems. Computer and Industrial Engineering 59, 638-646.

Xiao, S., Xiao, G.X., Cheng, T.H., 2008. Tolerance of intentional attacks in complex communication networks. IEEE Communications Magazine 46 (1), 146-152.

Xing, L.D., 2007. Reliability evaluation of phased-mission systems with imperfect fault coverage and common-cause failures. IEEE Transactions on Reliability 56 (1), 58-68.

Yeh, R.H., Chang, W.L., Lo, H.C., 2010. Optimal threshold values of age and two-phase maintenance policy for leased equipments using age reduction method. Annals of Operations Research 181 (1), 171-183.

Yeh, W.C., 2009. A convolution universal generating function method for evaluating the symbolic one-to-all-target-subset reliability function of acyclic multi-state information networks. IEEE Transactions on Reliability 58 (3), 476-484.

Yeh, W.C., Lin, C.H., 2009. A squeeze response surface methodology for finding symbolic network reliability functions. IEEE Transactions on Reliability 58 (2), 374-382.

Yeh, W.C., He, X.J., 2010. A new universal generating function method for estimating the novel multiresource multistate information network reliability. IEEE Transactions on Reliability 59(3), 528-538.

Youssef, A.M.A., ElMaraghy, M.A., 2008. Performance analysis of manufacturing systems composed of modular machines using the universal generating function. Journal of Manufacturing Systems 27 (2), 55-69.

Zhang, F., Jardine, A.K.S., 1998. Optimal maintenance models with minimal repair, periodic overhaul and complete renewal. IIE Transactions 30 (12), 1109-1119.

Zhuang, J., Bier, V., 2007. Balancing terrorism and natural disasters-defensive strategy with endogenous attacker effort. Operations Research 55 (5), 976-991.

Zuo, M.J., 1993. Reliability of linear and circular consecutively connected systems. IEEE Transactions on Reliability 42 (3), 484-487.