공학박사 학위논문

Multi-Criteria Decision Support Frameworks for Disaster Risk Mitigation of Community and Lifeline Networks

커뮤니티와 라이프라인 네트워크의 재난 리스크 저감을 위한 다중 기준 의사결정지원 방법론 개발

2020 년 2 월

서울대학교 대학원

건설환경공학부

최 유 정

Abstract

Community and lifeline networks are exposed to natural disasters such as earthquakes, floods, and typhoons. While the financial resources for disaster management are limited, the risks from those disastrous events are increasing due to high population density, a large number of network components, complex network topology, and network interdependency. In these circumstances, identifying critical disaster scenarios and effectively mitigate the risk of those scenarios within decision constraints are among the most important tasks of infrastructure network and communities. Therefore, in this dissertation, multi-criteria decision support frameworks are developed for disaster risk mitigation planning of infrastructure networks and communities.

This work contributes to the disaster risk mitigation of the lifeline network and community in several aspects. For lifeline networks, first, a novel Multi-Group Non-Dominated Sorting Genetic Algorithm (MG-NSGA) is proposed to identify critical post-disaster scenarios of large-size lifeline networks. The MG-NSGA enables us to identify the solutions of large-size network optimization problems with better optimality and robustness when compared to original NSGA-II. Besides, to select critical' post-disaster scenarios from the scenarios archive, which is collected during the sampling process, the concept of 'critical zone' is also presented. By setting boundaries within the solution space in terms of decision variables (e.g. network performance, number of components which failed at the initial post-disaster stage), 'critical zone' facilitates the disaster risk mitigation planning focused on critical scenarios. Second, by combining the MG-NSGA and flow-based overload cascading failure model (OCM), a critical cascading failure scenario identification framework

is developed. Using the cascading failure results of the OCM, this framework successfully identifies critical cascading failure scenarios of the power grid, which were induced by the multi-component failure. Third, the 'elite-set updating' method is developed. The method identifies the cost-effective retrofit combinations against the set of critical cascading failure scenarios, which are identified using MG-NSGA and OCM, with a relatively small computational cost. When compared with the all-candidate method, which enumerates all possible combinations of candidate components, the proposed method significantly reduces the computational cost while showing a similar level of excellent performance in identifying optimal retrofit combinations. Finally, for community-level disaster risk mitigation planning, a novel community disaster resilience clustering (CDRC) method is proposed. By clustering the structural damage and the socio-economic recoverability of sub-communities, which are evaluated separately, CDRC enables the categorization of the neighborhoods within the community using the similar disaster resilience characteristics without modeling the intertwined dependency between the physical and socio-economic aspects of the community. The novel methods proposed in this work are applied to the virtual and real lifeline networks and communities to demonstrate their accuracy and effectiveness. The methods proposed in this dissertation are expected to support the decision-making process for disaster risk mitigation of both lifeline networks and community scale.

Table of contents

List of Tables

List of Figures

1. Introduction

1.1. Motivation

Building “disaster-resilience,” which is defined as the ‘ability to adapt to and recover from hazards, shocks or stresses without compromising long-term prospects for development’ (Hyogo Framework, UNISDR, 2005), is one of the essential tasks for lifeline network and community. Since both systems are including numerous components, retrofit priorities should be decided for effective disaster risk management. In those decision-making processes, it is essential to consider the multiple criteria comprehensively (e.g. the importance of components in system functionality and monetary constraints for disaster risk management). Therefore, for the lifeline network and community, multi-criteria decision support frameworks are required to facilitate informed decision-making process for disaster risk mitigation.

Besides, infrastructure networks, the backbone of communities, are prone to manmade- and natural disasters and often suffer from un-prepared disasters. For instance, severe power blackouts in Italy in 2003, North-Eastern U.S and Canada in 2003, and Eastern India 2011 were the results of cascading failures triggered by unexpected, low-likelihood damage of network components. To reduce the disaster risk of the infrastructure networks, in particular, appropriate retrofits and emergency recovery strategies should be prepared to ensure proper post-disaster network performance. In general, however, the available budget is quite limited, while infrastructure network features numerous components. In these circumstances, identifying critical disaster scenarios of infrastructure networks and planning the

cost-effective risk mitigation strategy against those events are among the most critical issues that need to be addressed in efforts to reduce the disaster risk.

For community-level disaster risk mitigation planning, not only structural systems but also socioeconomic systems of a community should be comprehensively considered in the decision-making process for enhancing the disaster resilience of the community. Since various systems jointly influence the initial post-disaster condition and the recovery process after the disastrous events, focusing on the only physical or socioeconomic system could misestimate the disaster resilience of the community. On this account, multiple-aspects of community (e.g. structural, social, and economic) should be evaluated quantitatively and comparably during community-level disaster resilience assessment.

1.2. Technical Challenges and Research Objectives

The dissertation focuses on the following tasks to overcome technical challenges in multi-criteria decision support frameworks for disaster risk mitigation of lifeline networks and community:

First, it is noted that multi-objective genetic algorithms (MOGAs) are often employed for identifying the optimal solutions for network-related problems and known for their efficiency. However, a large size of a network may hamper the multi-objective genetic algorithm from obtaining final solutions that are accurate and robust against variability. This issue is caused by losing sample population diversity during the evolutionary process. The diversity of the sample population thus required to be preserved throughout the generations to improve the search performance of the MOGA. Therefore, this work aims to developed a multi-group non-dominated

sorting genetic algorithm (MG-NSGA) which can identify critical scenarios of large-size network with improved robustness and optimality.

Second, for the power supply network, in particular, it is essential to consider the cascading failure while identifying critical post-disaster scenarios since the consequence of an initial failure and that after occurrence of cascading failure could differ significantly. However, critical node detection problems (CNDP), which consider cascading failure of a power grid has been rarely addressed in the literature due to the complex mechanism of the cascading failure and its high computational cost. In general, to overcome this challenge, simplified cascading failure models based on topology were adopted, and only the scenarios each of which is induced by a single component failure are considered. However, differences between the topology-based modeling and flow-based modeling in cascading failure simulation are reported in multiple works of literature, and multi-component failures could induce cascading failure under natural or man-made disasters. Therefore, in this dissertation, to identify critical cascading failure scenario of the power grid with a better representation of the cascading mechanism, a flow-based overload cascading model (OCM) is combined with MG-NSGA, with consideration of multi-component failure at the initial post-disaster stage.

Third, to identify the cost-effective retrofit combinations under budget constraints, examining the all possible retrofit combination would be ideal. However, for the large size network, such an approach is computationally intractable. Therefore, to reduce this computational cost, an effective search method termed ‘elite-set updating’ method is developed for cost-effective network retrofit combination identification.

Lastly, in community-level disaster resilience assessment, there is general agreement that a holistic and multidisciplinary approach is required to address the disaster resilience of communities. However, integrated models themselves and data sets for those models yet remain insufficient to describe the multiple aspects of disaster resilience of the community. It means that existing community disaster resilience assessment frameworks may not be applicable to decision-making practice. Hence, there is a need for an integrated framework which could be realized with currently available community data sets to understand the multiple-dimensional disaster resilience of urban society. To this end, a community disaster resilience clustering (CDRC) method, which is compatible with publicly available community data sets, is developed to integrate information regarding the structural and socioeconomic systems of the community.

1.3. Organization of Dissertation

With the research objectives listed above, this dissertation consists of six chapters. Following this chapter, in Chapter 2, a new multi-group non-dominated sorting genetic algorithm (MG-NSGA) is developed to increase the robustness and optimality of the final non-dominated solutions of large-size network optimization problems. In Chapter 3, the critical cascading failure scenario identification framework of power supply networks, which combines the flow-based overload cascading failure model and proposed MG-NSGA, is illustrated through two power supply network examples. In Chapter 4, the ‘elite-set updating’ method is presented to identify the cost-effective network retrofit combination. In Chapter 5, the community disaster resilience clustering (CDRC) method is developed and

illustrated through a case study of two communities. Lastly, in Chapter 6, major findings of this dissertation are summarized, and a list of future research topics is also provided.

2. Critical Post-Disaster Scenario Identification using Multi-Group Non-Dominated Sorting Genetic Algorithm (MG-NSGA)

2.1. Introduction

In critical scenario identification, either probability-based or consequence-based search could be applied. When using the probability-based approach, network risk analysis methodologies often utilize disaster scenarios that are sampled based on the failure probabilities of the network components (Bruneau *et al*., 2003; Chang *et al*., 2000; Jayaram and Baker, 2010). The consequences of the sampled post-disaster scenarios are estimated by computing the average or probabilistic distribution of the socioeconomic loss. Such approach is often used due to an exceedingly large number of component failure combinations and high computational costs required for network analysis (Bruneau *et al*., 2003). This probability-based sampling approach, however, may not be able to identify critical post-disaster scenarios with high impact due to their relatively low probabilities of occurrence and imperfect probabilistic information used for sampling. As a result, the community may lose the opportunity to prepare against truly catastrophic situations and thus the network could remain vulnerable to such scenarios. It should be noted that, in the past, devastating consequences were often triggered by low-likelihood failure combinations in a variety of infrastructure networks.

To mitigate such unexpected but critical losses in infrastructure networks, besides the aforementioned probability-based sampling approach, consequence-based search of critical post-disaster scenarios needs to be performed. To date, in the areas of computer science and operational research, the so-called Critical Node

Detection Problem (CNDP) has been studied to find critical scenarios. To detect critical node failure combinations that degrade the network functionality most, single- (Arulselvan *et al*., 2009), bi- (Ventresca *et al*., 2018; Rocco *et al*., 2009), and multi-objective (Ventresca *et al*., 2015) optimization problems were defined and studied. In particular, the efficiency of using evolutionary algorithms in this type of optimization problems has been reported. However, application of multi-objective genetic algorithms to large-size lifeline network for the purpose of identifying critical post-disaster scenarios has not been studied extensively. This may be due to the large computational cost and difficulties in solving large size network problems as described in recent publications. For instance, Ventresca *et al*. (2018) noted that convergence issue, which is often observed as the number of the components increases, may occur due to the lack of population diversity. When the sample population loses the diversity during the evolution process, the final non-dominated solutions may converge to suboptimal points.

To overcome this challenge reported for large-size networks, a Multi-Group Non-dominated Sorting Genetic Algorithm (MG-NSGA) is developed in this chapter. The main feature of MG-NSGA is dividing the search space into multiple groups in the Pareto rank assignment stage to preserve the sample diversity. In addition, the concept of 'critical zone' is proposed to identify critical scenarios in the solution space, which would demand further investigations with a priority. Among sample populations that MG-NSGA investigated as it approaches the Pareto surface, post-disaster scenarios belonging to the pre-specified zone are identified as critical ones. The non-dominated solutions obtained by the proposed MG-NSGA are compared with those by NSGA-II (Deb *et al*., 2002), which is known to provide superior results among existing Multi Objective Genetic Algorithms (MOGAs) (Ventresca *et al*.,

2018) to investigate the performance of MG-NSGA. Benchmark examples featuring large-size networks are introduced to further demonstrate the performance of the MG-NSGA and its applicability to large-size real lifeline networks.

2.2. Review of Multi-Objective Genetic Algorithm (MOGA)

The method proposed in this section facilitates finding critical post-disaster scenarios that lead to out-of-proportion degradation of system performance using one of the MOGAs, termed NSGA-II. To provide background information for the proposed MG-NSGA, the following sections summarize basic ideas of MOGA, NSGA-II, and recent applications of MOGAs to network problems reported in the literature.

2.2.1. Basic ideas of MOGA

Genetic algorithm (GA) is a bio-inspired population-based search algorithm, which pursues a balance between exploitation and exploration within the search space (Gen *et al*., 2008). During the evolution process, the algorithm explores the search space through genetic operators such as crossover combinations and random mutations. At the same time, GA exploits the previous solutions whose objective functions gave good fitness values. The algorithm evolves the initial sample population to better fitted sample populations based on the fitness of the samples, which are repeatedly evaluated in terms of the objective functions.

Genetic algorithm can be applied to multi-objective optimization problems with more than one decision variable. Using MOGA, a group of non-dominated solutions, i.e. Pareto optimal solutions, can be successfully obtained. During the last few decades, various MOGA methods have been developed with different fitness

assignment mechanisms. MOGA methods became popular in diverse research fields since the most of the decision making problems in practice involve multiple decision variables.

2.2.2. Non-dominated Sorting Genetic Algorithm II (NSGA-II)

Among many MOGAs, NSGA-II (Deb *et al*., 2002) is one of the most widely used algorithms, which can efficiently solve multi-objective optimization/search problems (Gen *et al*., 2008). Therefore, to address critical post-disaster scenario identification problem, NSGA-II is adopted as the base algorithm of the proposed search method (Chapter 2.3.).

The evolution process of NSGA-II over two adjacent generations is illustrated in Figure 2.1. In the beginning of the *i*th generation, the fitness of the population is evaluated in terms of the objective functions. Then, the offspring population is generated using genetic operators, i.e. crossover and mutation. In this stage, the size of population is temporarily doubled.

Two main features of NSGA-II during the Pareto rank assignment process are (1) fast non-dominated sorting, and (2) crowding distance sorting. Among the increased population of a given generation, i.e. the union of the parent population and offspring population, Pareto solutions on the non-dominated front are assigned to the first Pareto rank. Next, the Pareto front of the population except for first rank solutions are assigned to the second Pareto rank. Such process is repeated until every sample in the increased population receives a Pareto rank. In addition to this fast non-dominated sorting, the crowding distance sorting is performed to provide a secondary criterion that is used after the replacement process.

At the replacement process, i.e. the last stage of evolution between generations, the solutions with the lower Pareto rank and longer inter-distance are selected as the population of i+1th generation. For instance, the distributions of ith and i+1th populations are plotted in Figure 2.1 It is shown that the solutions with better fitness values, i.e. those close to the first Pareto surface, are preferred by the algorithm. Eventually, throughout the evolution process of NSGA-II, better-fitted and more-spread-out solutions are identified to achieve the final Pareto solutions.

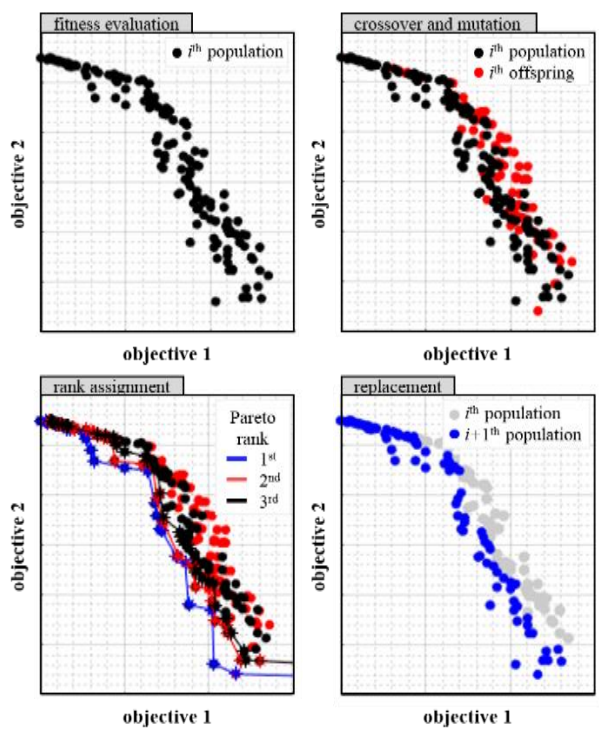


Figure 2.1 The evolution process between two adjacent generations by NSGA-II

2.2.3. Application of MOGA to Network-related Problems

Because of advantages in adaptability, robustness, and flexibility, MOGA methods have been applied to diverse network related problems. In particular, MOGA has been utilized to solve a wide array of infrastructure network optimization problems such as post-disaster restoration optimization (Rocco *et al*., 2018), topology/design optimization (Pasad and Park, 2004; Cadini *et al*, 2010), and interdiction problems (Rocco *et al*., 2009). By selecting various combinations of objective functions, critical post-disaster scenarios were identified for the lifeline networks. However, while previous MOGA methods including NSGA-II efficiently deliver the non-dominated solutions to the network examples with relatively small number of components, the accuracy and robustness of the approach decreased as the number of the network components increased (Ventresca *et al*., 2015). One of the causes of such issue is the lack of population diversity. To apply the MOGA to lifeline networks, which often consist of numerous components, therefore, it is essential to overcome the population diversity issue by developing an MOGA algorithm that can improve the accuracy and robustness against variability of final Pareto solutions for large size networks.

2.3. Critical Post-Disaster Scenarios Identification Frameworks

For effective identification of critical post-disaster scenarios of large size network using MOGA, a Multi-Group NSGA is developed in this section. To this end, GA-based representation of post-disaster scenarios of network components, and definitions of the multi-objective functions are introduced. Lastly, the concept of the critical zone is also proposed to support the decision-making process using the post-disaster scenarios identified by MG-NSGA.

2.3.1. MG-NSGA

While applying NSGA-II to large-size network examples, it is found that MOGA may lose meaningful solutions at early stages of the evolution process, which hampers the search process from converging to the final Pareto surface. As reported in the literature (Ventresca *et al*., 2015), this is because the fast-non-dominated-sorting algorithm in NSGA-II inevitably sacrifices the population diversity while pursuing early convergence at each step of evolution. For instance, as shown in Figure 2.1, samples with lower rank have high chance to survive at the next generation, while samples with higher rank quickly disappear. Therefore, meaningful samples, which are not positioned in the front yet, are less likely to survive in the next steps of evolution. Such tendency of losing population diversity at an early stage leads to convergence toward suboptimal solutions. When the population loses the diversity, some critical scenarios whose failure combination is totally different from quickly converged solutions could be ignored.

In order to sustain diversity in the population, solutions positioned in the dominant space during the evolution should be given higher chances to survive in the next generation. To this end, the multi-objective search space is horizontally divided into multiple groups in terms of the performance measure axis and non-dominated front is assigned to each group as shown in Figure 2.2(a). It is important to note that the sample populations are divided into the multiple groups temporarily during the Pareto rank assignment stage and combined together to progress toward the replacement stage. Therefore, samples are not restricted in a certain group during the evolution process. With this “Multi Group” NSGA method, the front of each group receives the first rank to increase their chance of survival in the following

evolution process. For example, Figure 2.2 shows that, although the given populations are identical, the scenarios selected for first and second Pareto using MG-NSGA includes more diverse samples than those by NSGA-II. With the proposed approach, the population can preserve the diversity until the convergence to the final Pareto surface is achieved. Eventually, in terms of the quality of the final solutions, the proposed multi-group algorithm is expected to have advantages in accuracy and robustness against variability.

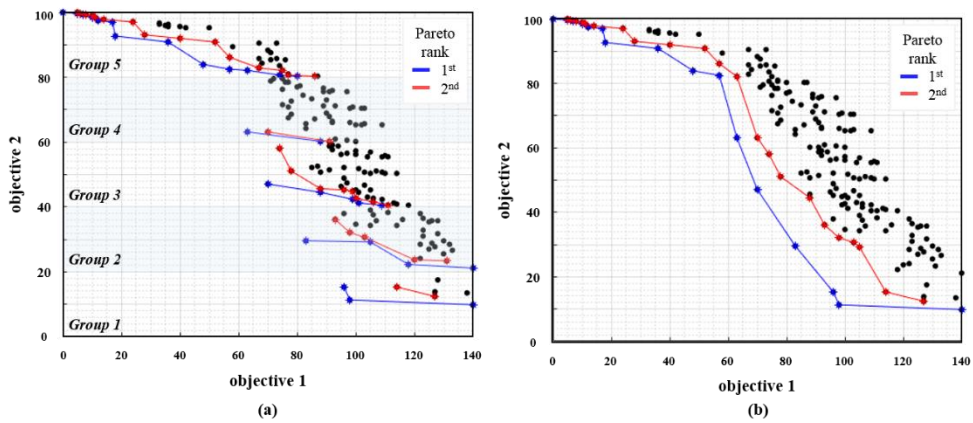


Figure 2.2 Comparison of Pareto rank assignment results using (a) MG-NSGA; and (b) NSGA-II

It is noted that the more groups are introduced, the more diverse post-disaster scenarios will be preserved during the evolutions. Although a large number of groups should slow down the convergence toward the final Pareto surface at early generations of the evolution, from the experience, a similar total number of generations were required for the final convergence. It was also found that, an increase in the number of groups may not necessarily improve the identification

results after exceeding a certain number of groups. Therefore, it is important to develop a systematic procedure to determine an appropriate number of groups for a given problem to gain optimal performance of MG-NSGA. Figure 2.3 shows the flowchart of MG-NSGA. While following the classical steps of MOGA, MG-NSGA shows difference in terms of Pareto rank assignment.

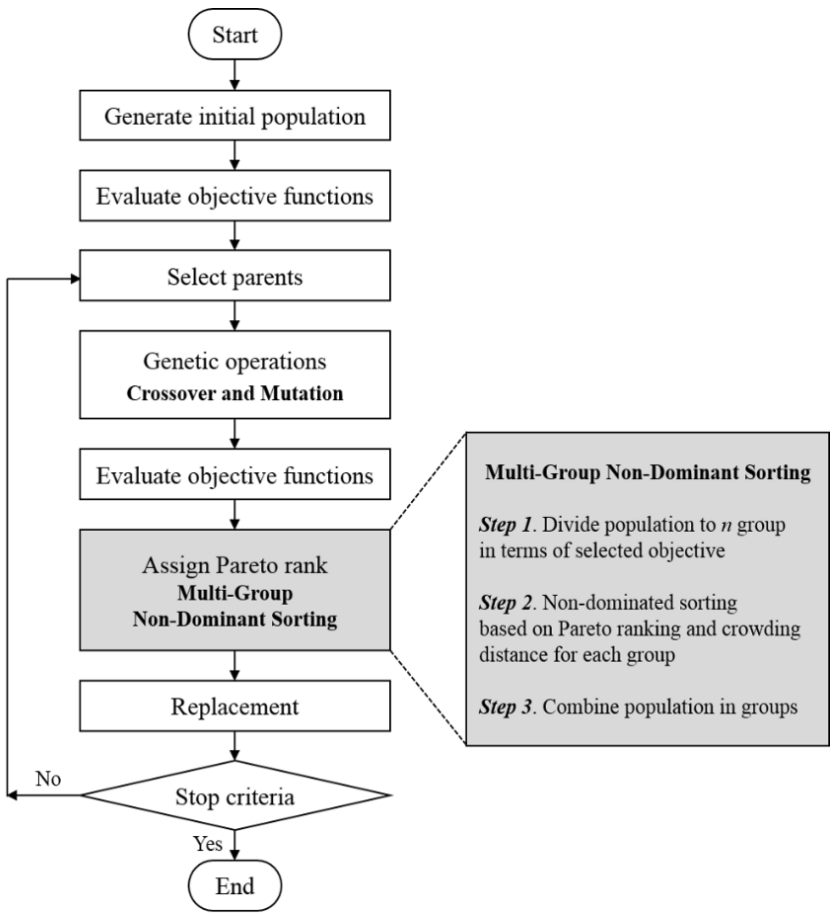


Figure 2.3 Flowchart of MG-NSGA

2.3.2. Modeling Post-disaster Scenarios of Lifeline Networks

To model the conditions of the lifeline network in GA-based methods, a binary matrix is generally used. This approach is adopted in this research to describe post-disaster scenarios of lifeline network in the proposed MG-NSGA-based search. As illustrated in Figure 2.4(a), each chromosome represents a single post-disaster scenario in which the values 0 and 1 respectively indicate the failure and survival states of the corresponding network component. The example in Figure 2.4(a) shows that if the post-disaster scenarios are described by binary states, i.e. failures or survivals, of k components, and a population consists of m chromosomes, the population can be expressed by an $m \times k$ binary matrix. Figure 2.4(b) shows all possible link failure scenarios of a three-link network and the corresponding genetic representations.

In addition, the string size k is determined by the type of disaster and how the infrastructure network is modeled, e.g. node-based or link-based approaches. Therefore, the number of the links, nodes, or their sum determine the string size. Since this study considers only post-disaster survivals and failures of the existing network components, i.e. addition of new components is not considered, the size of the binary matrix does not change over the evolution process. On the other hand, it is important to note that only initial population is generated by the random sample generation. The sample populations, after first population, are generated and evolve using genetic operators, crossover and mutation, and finally determined during the selection process.

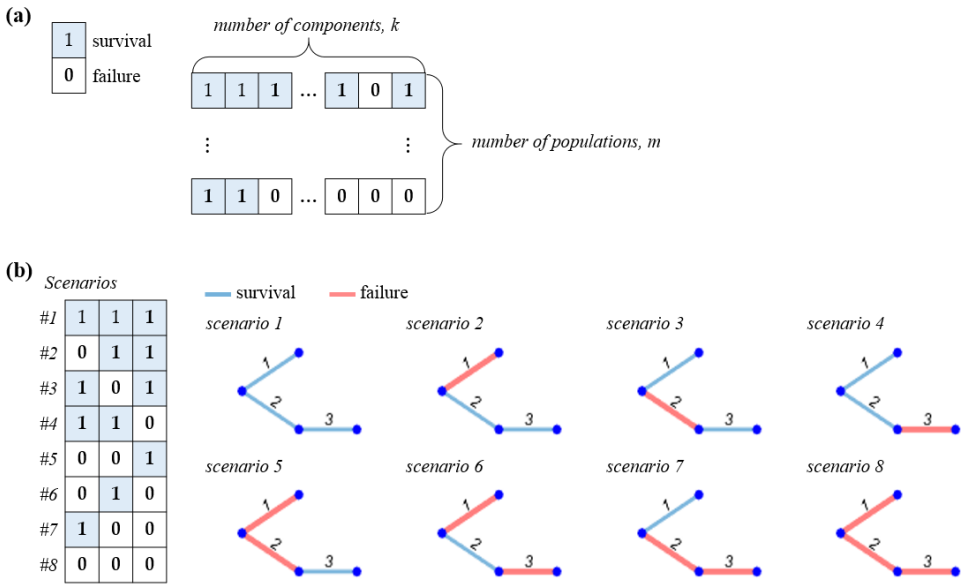


Figure 2.4 Modeling structure adopted for MG-NSGA: (a) binary matrix for genetic representation of post-disaster scenarios; and (b) example of link failure scenarios and corresponding genetic representations

2.3.3. Defining Multi-Objective Functions to Identify Critical Post-Disaster Scenarios

The definition of 'critical' post-disaster scenarios is incorporated into the proposed identification method via multiple objective functions. Once a sample population is generated in binary form, each sample's fitness as a critical scenario is quantified in terms of the objective functions. Therefore, it is important to define and use appropriate objective functions. It is noted that flexibility of genetic algorithms facilitates applications of the proposed search algorithm to general objective functions.

To identify critical post-disaster scenarios entailing out-of-proportion consequences, the following two objective functions are used: (1) the post-disaster performance measure of the network, and (2) the number of the malfunctioning network components. At the Pareto front identified for each of the numerical examples in this chapter, the post-disaster performance measure of a network (i.e. number of nodes connected to the nodes of interests) decreases as the number of the failed component increases. This shows that these two objective functions facilitate effective identification of critical post-disaster scenarios by multi-objective optimization.

2.3.3.1. Objective function 1: post-disaster network performance measure

One of the essential tasks of the disaster risk management is to identify serious post-disaster functional interruption scenarios so that their physical or socioeconomic loss can be reduced by a proper pre-disaster decision making process. Minimizing a selected post-disaster performance measure naturally leads to identifying scenarios with more critical consequences. Various performance measures can be considered as candidates based on the decision maker's understanding of the particular disaster risk, characteristics of the disaster of concern, and circumstances of the community. The proposed framework can incorporate such diverse definitions of post-disaster impact, simply by using an objective function that quantifies the post-disaster performance of the network. For each type of lifeline network, different kinds of network performance measures are considered most appropriate. Therefore, it is important to select the measure which could represent the specific performance of the lifeline network. The numerical examples of this chapter uses performance

measures representing the post-disaster network connectivity between the node of interest and the other nodes of the network, described by

$$f_{i1} = n(\boldsymbol{A}_i) \tag{2-1}$$

where f_{i1} is the number of the nodes still connected to the node of interest under the *i*th scenario, and $n(\cdot)$ denotes the cardinality of the set, and $\boldsymbol{A}_i$ is the set of the visited nodes that can be visited from the node of interest under the *i*th scenario according to the connectivity analysis results. It is noted that other kinds of performance measures, e.g. network flow capacity, travel delay time, could be used as alternative objective functions without losing generality of the proposed framework.

2.3.3.2. Objective function 2: number of failed components

This study proposes to use the number of failed components as one of the objective functions that can be used together with the post-disaster performance measure. The corresponding objective function is described as

$$f_{i2} = k - \sum_{j=1}^{k} x_{ij} \tag{2-2}$$

where f_{i2} is the number of the failed components under the *i*th scenario, k is the length of the scenario representation (see Figure 2.4), and x_{ij} is the element of the binary matrix in Figure 2.4 which corresponds to the *j*th component in the *i*th scenario. This objective function is chosen because, first, minimizing the number of failed components helps identify scenarios with serious consequence which is induced

from relatively small number of component failures, i.e. scenarios leading to out-of-proportion consequences. Second, minimizing the number of failed components helps identify more likely scenarios without calculating their probabilities because the probability of joint failures of many components is relatively low. As a result, the GA search tends to give preferences to more likely scenarios but without evaluating the probabilities of all joint failure scenarios necessarily. Third, if the probability is used as an objective function instead, one needs to have a reliable probabilistic model of the hazard and fragility of the network components. This is useful especially when such information is not available.

2.3.4. The Concept of Critical Zone

MG-NSGA enables us to identify numerous post-disaster scenarios throughout the evolution process. However, not all the scenarios are necessarily critical scenarios in terms of the multiple objective functions. Therefore, a critical zone needs to be introduced in the solution space to provide boundaries. The definition of the critical zone will significantly affect the decision-making process because the priority list of critical components will be determined by the boundaries of the zone. The definition of the critical zone is case-specific, and should reflect the socioeconomic condition and risk-averseness of the decision makers. In the numerical examples of this paper, critical zones were defined arbitrarily to demonstrate the concept. For example, critical zone boundary constraint could be expressed as the following equation:

$$f_{i1} = n(\boldsymbol{A}_i) \leq N \tag{2-3}$$

$$f_{i2} = k - \sum_{j=1}^{k} x_{ij} \leq K \tag{2-4}$$

where *N* and *K* are case-specific constraints on the 'number of components failed at the post-disaster stage' and the 'number of still connected nodes to the node of interest' respectively. Since the critical zone should be determined in the context of network management and socioeconomic impact, it is desirable to perform further research in these contexts to make optimal risk management decisions using the critical scenarios identified by the proposed approach. This concept of the 'critical zone' can be applied to various fields, where the decision-making process does not aim to identify all non-dominated solutions necessarily, and rather, finding critical solutions satisfying certain criteria (e.g. disaster impact) is actually important.

2.4. Case studies

2.4.1. Transportation Networks: EMA Highway Network

2.4.1.1. Problem setting and critical scenario searching by MG-NSGA

EMA highway network is the transportation network consisting of 74 nodes and 129 links. The network is modeled based on data from Zhang (2016). To represent the link failure scenarios, binary strings with length 129 are used. The topology of the EMA highway network in Figure 2.5(a) shows that the network has a small number of highly connected nodes such as #22, #48, and #60. In this example, the node #60, i.e. the red-colored node in Figure 2.5(a), is considered as the community of interest and check the connectivity between the node #60 and other nodes after disruption.

Since the loss of the connectivity with other nodes would lead to direct and indirect losses in regional economy, the number of nodes maintaining post-disaster connection with the node #60 is selected as 'post-disaster network performance measure' (Objective function 1) discussed above. For the connectivity evaluation, Breadth First Search algorithm is used. Together with the other objective function 'the number of the failed links' (Objective function 2), MG-NSGA effectively identifies critical post-disaster scenarios.

In addition, to apply the MG-NSGA approach to the test case, two MG-NSGA parameters, i.e. the number of the groups in the solution space, and the boundaries of each group, are determined. The number of groups from 2 to 20 with uniform length are introduced to the Objective function 1. For instance, 8 groups with uniform length are shown in Figure 2.5(b).

For the MG-NSGA with the various number of groups and for NSGA-II, the same set of the initial population are applied for comparison. For numerical experiments, the following options are used: the size of the initial population 150, mutation ratio 1/129, which is $1/k$ where k is string size (Ochoa, 2002), and stopping criteria 5,000th generation, which is a conservative choice according to several test runs. To the best of experience, the algorithm delivered the optimal convergence with these decision variables with the EMA network.

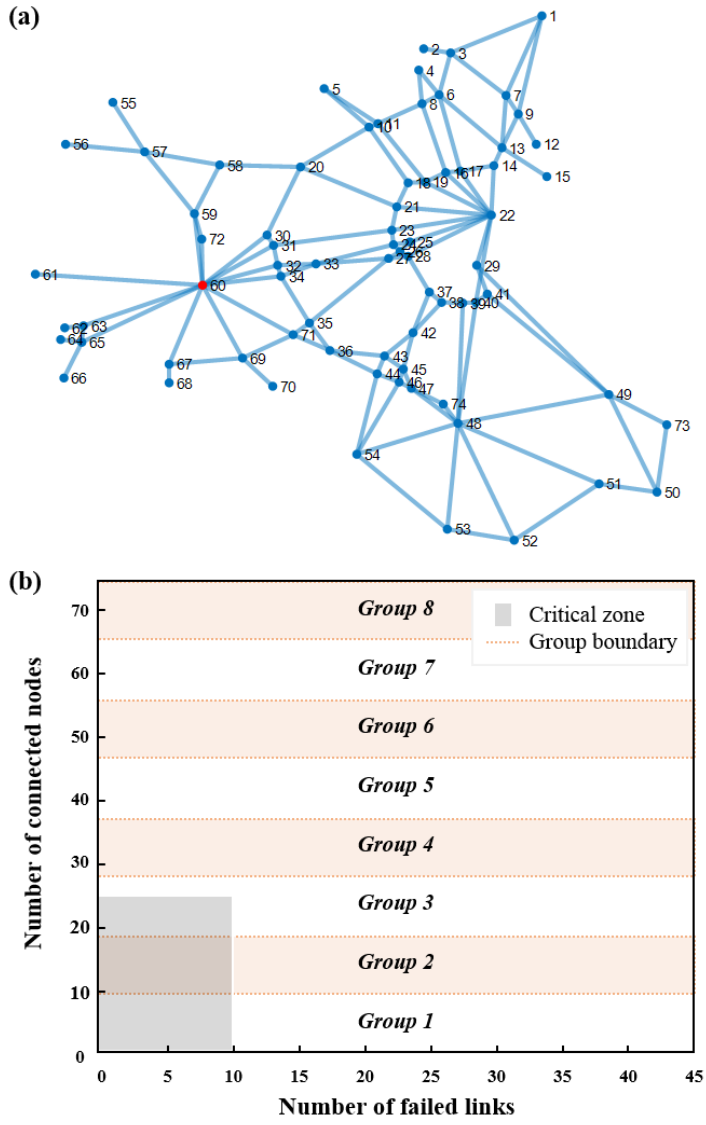


Figure 2.5 Easter Massachusetts (EMA) highway network example: (a) topology of the highway network; and (b) boundaries of the 8 groups selected for MG-NSGA and critical zone

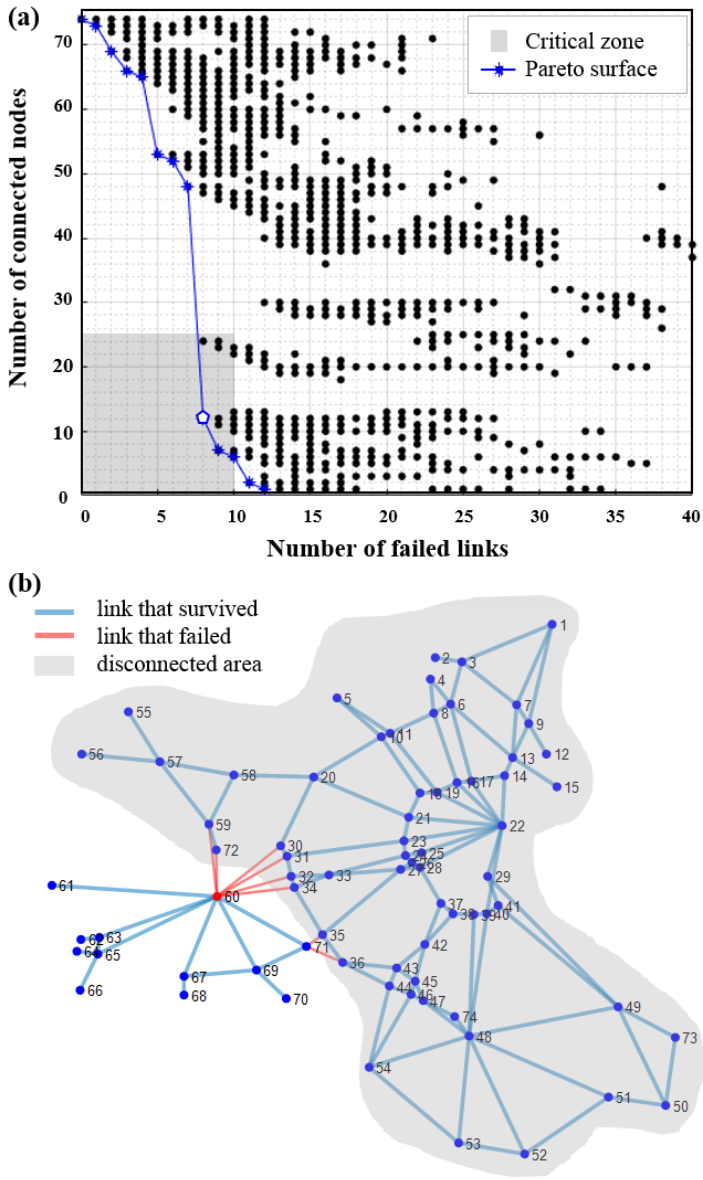


Figure 2.6 Identification of critical scenarios for EMA highway network by a single run of MG-NSGA: (a) Post-disaster scenarios archive and final Pareto solutions; and (b) one of the identified critical post-disaster scenarios ("⬠" marker in (a))

2.4.1.2. Results and discussions

Figure 2.6(a) shows all scenarios investigated from a single searching process using MG-NSGA when 8 groups are introduced. Each point in Figure 2.6(a) represents a post-disaster scenario, i.e. a combination of network component states, and it is seen that the proposed approach identifies not only critical scenarios but also many others in the feasible domain. The Pareto surface in Figure 2.6(a) (blue line) shows that the number of nodes maintaining connection with the node #60 tends to decrease as the number of failed links increases. It is noteworthy that, the relationship between the number of failed components and the disaster impact is far from being linearly proportional. The Pareto surface also shows that the node #60 in the EMA network could lose most of the connectivity by the failures of 8 components.

Suppose the decision maker defines 'critical' scenarios as the cases in which "only less than 25 nodes, i.e. less than one third of the total network nodes, are connected with the node #60 even though at most 10 links fail due to a disastrous event." This is an example of the critical zone definition reflecting the significance of EMA highway network and budget available for the disk management efforts. In other words, by the critical zone definition, the decision maker aims to address consequences in which less than 25 nodes are connected using the available budget that can cover retrofits of up to 10 links. For instance, a critical scenario induced by the failures of eight links (represented by "⏢" marker in Figure 2.6(a)) is illustrated in Figure 2.6(b). It is shown that the critical scenario is induced by the failures of links which connect the node #60 to the eastern part of the network, which coincides with the insights based on the network topology. Since the proposed method does not require additional information or a procedure to identify the topological effect,

MG-NSGA is expected to provide a valuable insight for highly complex and/or large-size networks for which a manual identification of critical scenarios based on insights or component importance measures would be intractable.

To investigate the optimality and the variability of the MOGA-based search results caused by random selection of the initial population and heuristic nature of MOGA, the final Pareto solutions and post-disaster scenarios archive, such as those in Figure 2.6(a) are repeatedly obtained using 50 different sets of initial populations. Figure 2.7 visualizes the performance of NSGA-II and the MG-NSGA with the different number of group between 2 and 20. To compare the quality of the Pareto solutions achieved from different numbers of the group, the sum of the 'number of connected nodes' plotted in boxplot and its median, minimum, and optimal values are compared with each other. The minimum value is evaluated from the single run while the optimal value is the minimum value which could be evaluated from the 50 runs. Although it may depend on the given problem, it is observed from numerical experiments that the use of 10 to 15 groups generally provides stable outcomes. As an example, one of the groups which deliver the most optimal Pareto surface, MG-NSGA with 15 groups is illustrated in Figure 2.8(a). However, as the graph indicates, the parameter (the number of groups) does not have a significant effect on the analysis result as long as the number is not particularly small or large.

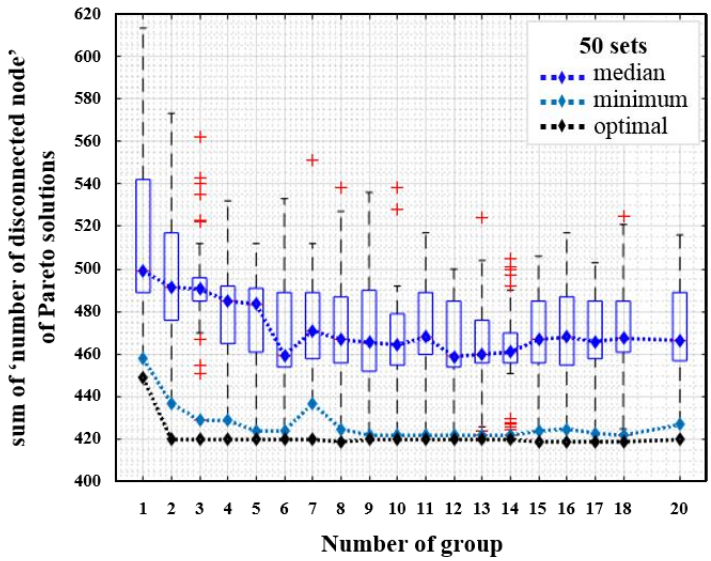


Figure 2.7 Performance comparison using 50 Pareto solutions of EMA highway network

Figure 2.8 visualizes the variability of the Pareto surfaces using box-plots at each number of failed links for the MG-NSGA with 14 groups and NSGA-II. In general, the median values are close to the optimal points, i.e. minimum values among 50 Pareto solutions at each number of failed links, and show robust outcomes regardless of the number of failed link. When compared to the results by NSGA-II and the MG-NSGA with 15 groups in Figure 2.7 and 2.8(a) and (b), the optimality of the solutions by MG-NSGA is better than NSGA-II. The optimal solutions by MG-NSGA (blue line) and NSGA-II (red dotted line) are distinct when 7 links are failed for the EMA highway example. Moreover, improvement in the efficiency by MG-NSGA is significant. For most of the numbers of the failed links, median values from the NSGA-II results show a large difference from the Pareto solutions (red

dotted line). Furthermore, the smaller variability of the MG-NSGA results indicate that a smaller number of independent runs would be needed for reliable identification. Especially, in results for between 5 and 9 failed components, the bias and variability by NSGA-II are significant. It is expected that the proposed MG-NSGA will deliver more accurate identification results than NSGA-II even with fewer runs.

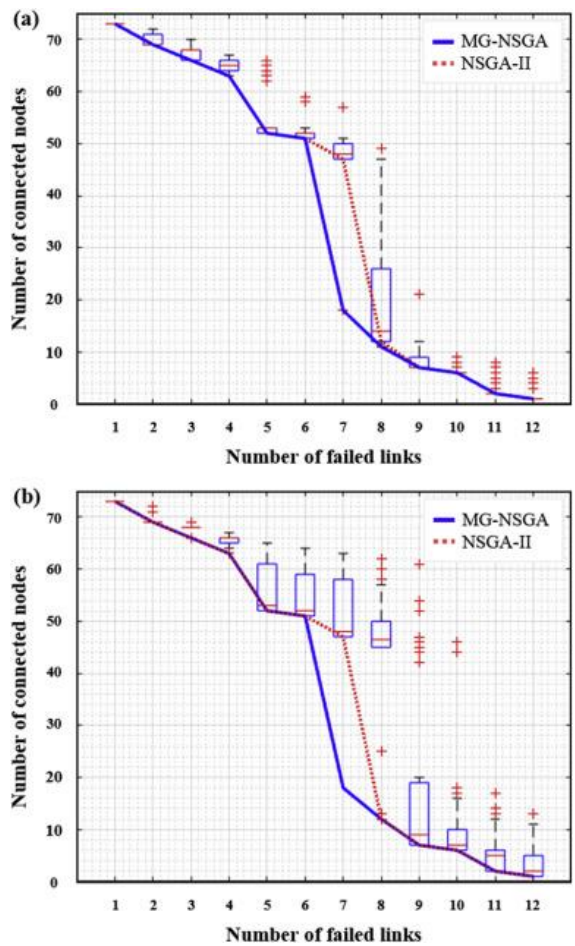


Figure 2.8 Box-plot of 50 Pareto solutions of EMA highway network by (a) MG-NSGA with number of 15 group and (b) NSGA-II

2.4.2. Transportation Networks: Jeju Island Transportation Network

2.4.2.1. Problem setting and critical scenario searching by MG-NSGA

To further investigate advantage of using MG-NSGA for large-size network problem, a larger-size network example, Jeju Island transport network in Figure 2.9, is investigated. The network model is developed for this study based on the network topology and the coordinates of nodes of the Jeju Island transportation network. As shown in Figure 2.9, the network consists of 273 nodes and 422 links. To represent link failure post-disaster scenarios, therefore, binary stings of length 422 are used.

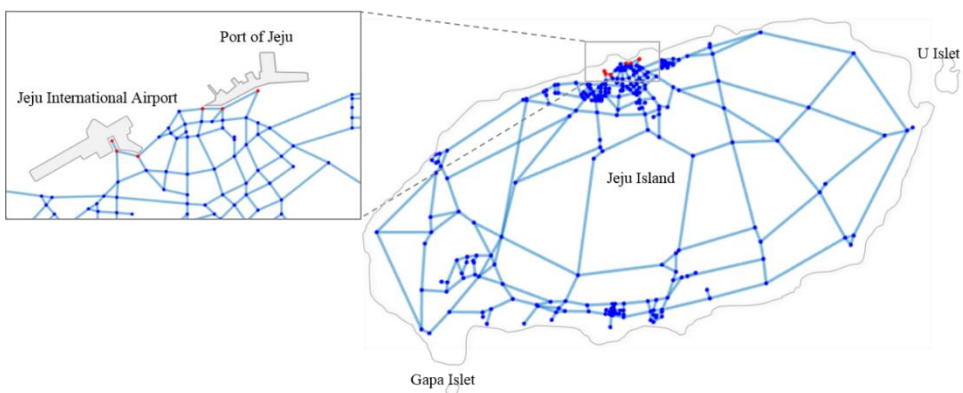


Figure 2.9 Jeju Island transportation network

In this example, the six red-colored nodes directly connected to the airport and the port are set as the terminal nodes. Since Jeju Island is accessible only via flights or maritime transports, disconnection with the terminal nodes means isolation and significant loss in the regional economy, which is heavily reliant to tourism. The

number of nodes maintaining connection with the six nodes and the number of failed components are used as the objective functions.

To perform MG-NSGA, sample space is divided into several number of groups between 2 to 20 with uniform spans. The performance of MG-NSGA with multiple groups and NSGA-II are compared using the sum of the 'disconnected nodes' of Pareto solutions from 30 sets initial populations. In addition, the same options, i.e. population size 150, mutation ratio 1/422, and stopping criteria 50,000th generation, which is conservatively selected based on the results of several test runs, were used for both algorithms. According to the experience, the algorithm delivered the optimal convergence with these decision variables with the Jeju Island transportation network.

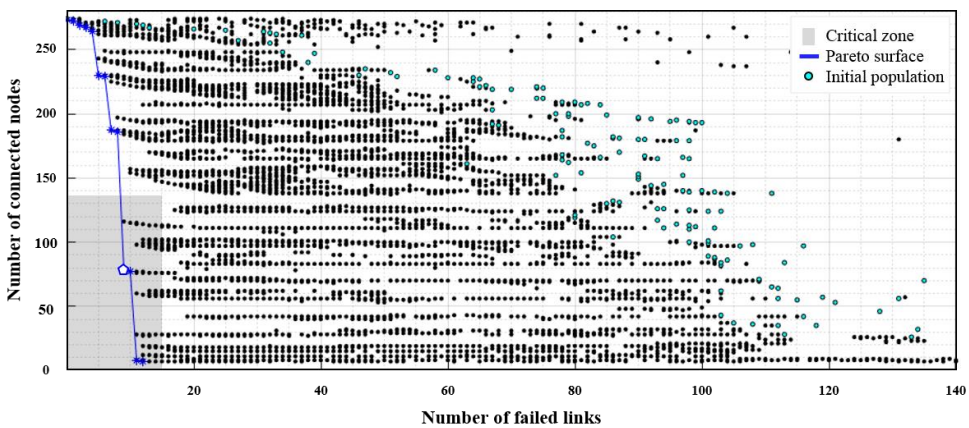


Figure 2.10 Post-disaster scenarios archive and final Pareto solutions for Jeju Island transportation network collected from single run of MG-NSGA with 20 groups

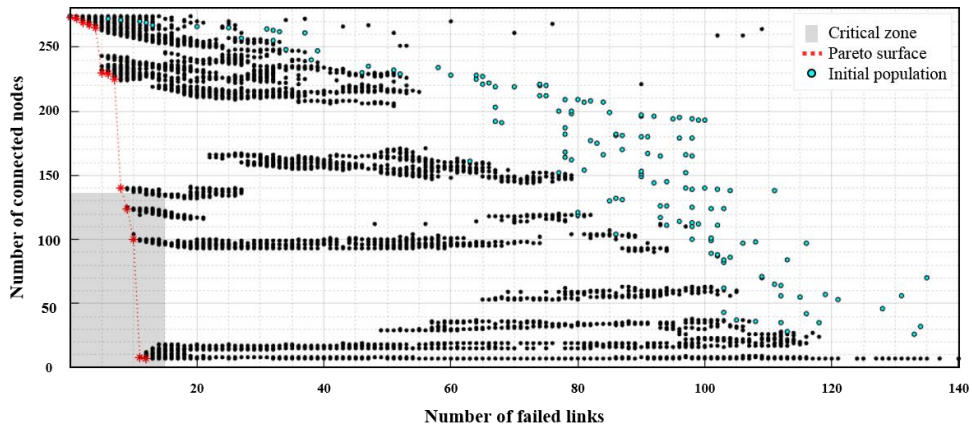


Figure 2.11 Post-disaster scenarios archive and final Pareto solutions for Jeju Island transportation network collected from single run of NSGA-II

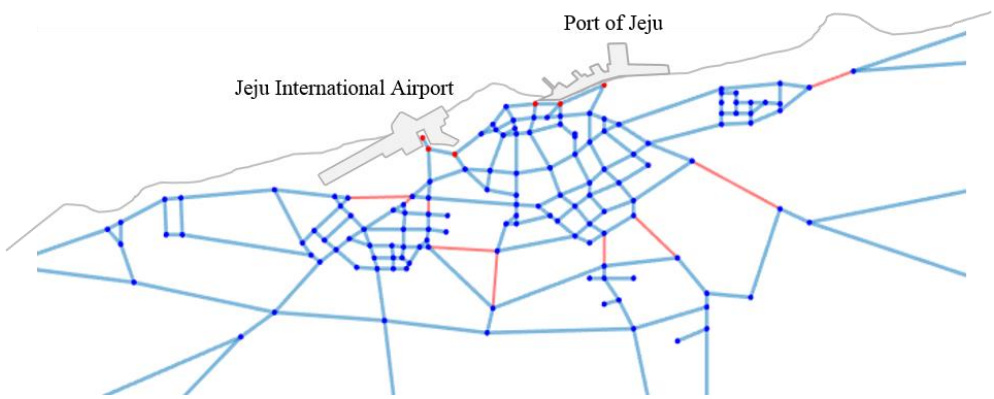


Figure 2.12 Critical post-disaster scenario identified by MG-NSGA ("⌂" marker in Figure 2.10)

2.4.2.2. Results and discussions

Figures 2.10 and 2.11 shows post-disaster scenarios and the Pareto solutions achieved by a single run of the proposed MG-NSGA with 20 groups and NSGA-II,

respectively. Both results are achieved from the same initial population represented by light green dots. The two approaches produce different outcomes especially in terms of the diversity of the solutions. The samples explored by MG-NSGA are significantly diverse as compared to those by NSGA-II. The MG-NSGA successfully preserves the diversity of sample while NSGA-II loses the diversity of population during the evolution process. As a result, the final non-dominated solutions also show a significant level of difference.

The functionality and disaster-robustness of a network heavily rely on its topology (Strogatz, 2001). Therefore, the evolutions and final Pareto surface in the multi-objective function plots are also influenced by the network topology. Most of the single-component-failure cases do not make impact on system connectivity because of the redundancy of the networks. However, a certain combination of a small number of component failures – especially those in the critical post-disaster scenarios identified by MG-NSGA – could create significant loss in the network connectivity. Moreover, as shown in Figure 2.10, the connectivity loss between the terminal nodes and the other nodes are not linearly proportional to the number of failures necessarily. The Pareto surface indicates that the increase of the network-level impact by additional failed link is not constant. For example, in Figure 2.10, the post-disaster network performance on the Pareto surface shows the largest decrease as the number of failed links increases from 8 to 9. For this reason, the failed components included in critical scenarios may not be detected by investigating network components one by one. Therefore, comprehensive demonstration and investigation using scenario-wise consequence-based sampling proposed in this dissertation is considered an adequate approach to detect critical post-disaster scenarios in terms of disaster impact measure.

After the search is completed, critical scenarios are identified using the 'critical zone' which is visualized by a gray-shaded area in Figure 2.10. Suppose the decision maker defines a critical scenario as cases in which "less than a half of the total network nodes are connected with the terminal nodes after less than 15 links fails." This means that the decision maker utilizes budget covering 15 links, and disconnection of a half of the network nodes from the airport and the port of the Jeju Island is considered catastrophic. For instance, one of the critical post-disaster scenarios (represented by "⬠" marker in Figure 2.10) identified by MG-NSGA (which NSGA-II fails to identify) is illustrated in Figure 2.12. It is shown that the critical scenario induced by the failures of 9 links located in the downtown area, near the airport and the port, isolate most of the neighbors of the Islands from the terminals. In the disaster risk management perspective, the identified critical scenarios represent worst case situations which need to be addressed with a priority during the disaster mitigation planning. To avoid or reduce the negative impact of such critical scenario which may occur in the infrastructure network, stakeholders need to pay special attention to retrofits, regular maintenance, and recovery resource allocation for the network components appearing in the identified critical scenarios.

Furthermore, by comparing the results produced by 30 different initial populations, the accuracy and variability of the post-disaster scenarios identified by MG-NSGA with different numbers of groups and NSGA-II are examined. In Figure 2.13, the quality of the Pareto solutions achieved from different numbers of the group, the sum of the 'number of connected nodes' plotted in boxplot and its median, minimum, and optimal values are compared with each other. Similar to the results from the EMA network, it is again observed from numerical experiments that the result is not sensitive to the number of groups unless it is too small or large.

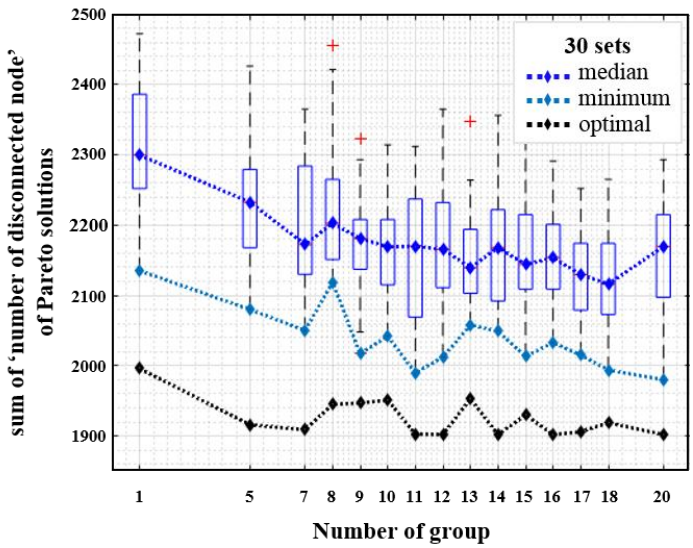


Figure 2.13 Performance comparison using 30 Pareto solutions of Jeju example

Figure 2.14(a) and (b) show the box-plots of the network performance measure at each number of the failed links by MG-NSGA with 16 groups and NSGA-II, respectively. MG-NSGA with 16 groups is chosen among the groups which deliver the optimal Pareto surface. In each plot, the optimal solutions by MG-NSGA (blue line) and those by NSGA-II (red dotted line) are also shown for comparison. For the Jeju transportation network example, the Pareto solutions by MG-NSGA and NSGA-II show significant difference. As compared to the proposed approach, NSGA-II identifies suboptimal and less significant solutions as the non-dominated solutions. Again, this is due to the lack of population diversity of NSGA-II, which makes it difficult for the search algorithm to get out of local minima. The results from the boxplots in Figure 2.13 and 2.14 show that MG-NSGA has a clear advantage in robustness against variability since gap between the 1st quintile and the median value is closer to the optimal solution than NSGA-II. Based on these results, MG-NSGA

is expected to show improved search performance in terms of both accuracy and variability in real large-size networks.

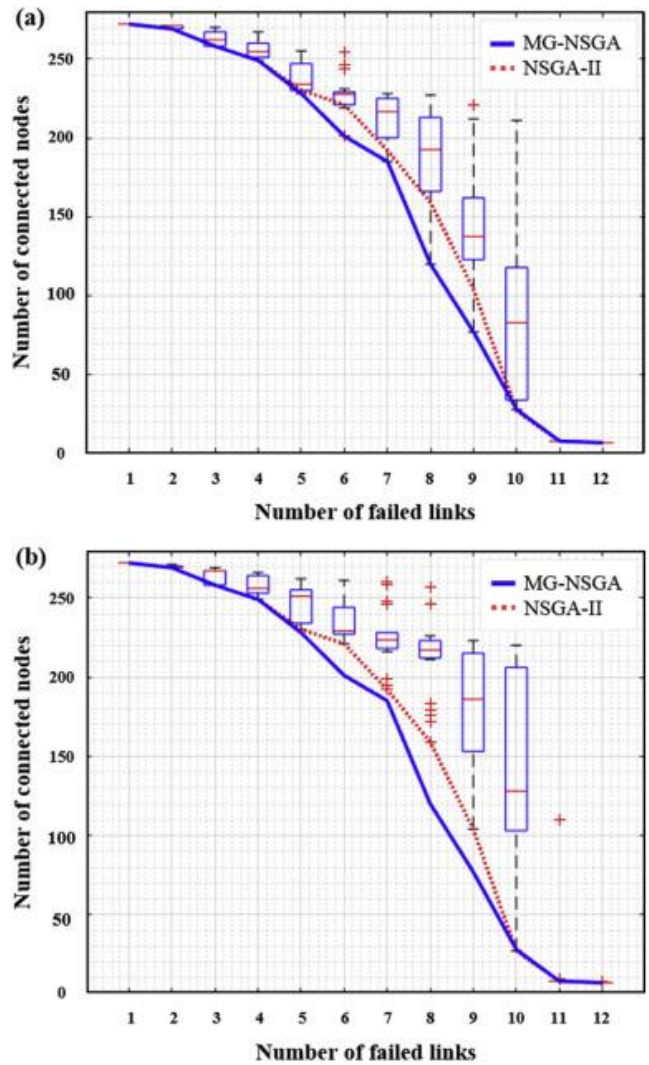


Figure 14. Box-plot of 30 Pareto solutions of Jeju transportation network by (a) MG-NSGA and (b) NSGA-II

2.5. Summary

A Multi-Group NSGA is developed to resolve an underlying issue of using NSGA-II algorithm for large-size network problems, i.e. losing the population diversity at an earlier stage of the evolution process. By dividing the search space during the Pareto ranking assignment process, MG-NSGA successfully identifies critical post-disaster scenarios for large-size infrastructure networks. Final non-dominated solutions from the proposed method show significant improvement in accuracy and robustness against the variability when compared to NSGA-II. In other words, post-disaster scenarios with more critical consequences were identified by using MG-NSGA with the less computational cost. In addition, the concept of 'critical zone,' which allows the decision makers to define critical scenarios in the solution space, is also proposed. It is expected that the critical scenarios identified by MG-NSGA and the critical zone concept will facilitate decision-makers' efforts to reduce the disaster risk of complex infrastructure networks. Through comprehensive demonstration and investigation of critical post-disaster scenarios, decision-makers can gain useful insights on retrofits and maintenance of network components.

3. Critical Cascading Failure Scenarios Identification of Power Grid using MG-NSGA

3.1. Introduction

Power supply networks are critical infrastructure systems, which support core functionalities of urban communities. However, these are often prone to risk of cascading failures due to a high level of interconnectivity and dependency between the network components. For instance, serious power blackouts in Italy (2003), North-Eastern U.S and Canada (2003), and Eastern India (2012) were the results of cascading failures triggered by unexpected, low-likelihood damage of network components (Anderson *et al.*, 2005; Makarov *et al.*, 2005). The devastating outcomes from these historic blackouts highlighted needs for identifying critical cascading failure scenarios so that the network management authorities can make optimal decisions regarding retrofits of network components.

To date, cascading failure of power supply networks has been studied extensively in the fields of reliability engineering and electrical engineering. As a result, various cascading failure models (Guo *et al*., 2017) have been proposed and network optimization problems against the cascading failures (Guo *et al*., 2017; Fang *et al*., 2014 and 2015) have been also investigated. In these studies, 1) initial failure setting and 2) cascading failure modeling were the key elements in determining the ultimate outcome of the cascading failures. First, before modeling the cascading failure, initial component failure conditions of the power supply network need to be described. In this procedure, the number of the components that failed at the initial stage has been limited to one (Guo *et al*., 2017; Fang *et al*., 2014 and 2015; Wang *et*

al., 2009) whereas the cascading failures induced by multiple components were rarely addressed in the literature. This is because computational complexity and costs required for simulating cascading failures increase rapidly as the number of components failed at initial stage increases. However, natural disasters such as earthquake, typhoon, and flood tend to cause multiple components distributed in a large region to fail simultaneously. Therefore, it is important to search critical cascading scenarios without imposing constraints on the number of components that failed at initial stage.

Once the initial conditions are assumed, the corresponding cascading failures are investigated using a model that can simulate cascading effects. To this end, various simulation approaches using topological model, stochastic simulation model, and high-level static and dynamic simulation models have been developed (Guo *et al*., 2017). Especially, topology-based simulation approaches were often adopted for design, management, and protection optimization problems of power supply networks because of computational simplicity. However, despite meaningful results and insights provided by the approach, it is noted that the topology-based model has a significant limitation in modeling complex phenomenon as compared to the flow-based models (Hines *et al*., 2010; Pagani and Aiello, 2013; Cauadra *et al*., 2015; Guo *et al*., 2017; Bompard *et al*., 2019). To facilitate disaster risk reduction based on better understanding of complex cascading failure mechanism, it is essential to adopt a power-flow based model.

To overcome these limitations in the literature, this dissertation proposes a method to identify critical cascading failure scenarios of power supply networks using a flow-based cascading model without imposing constraints on the number of components that failed at initial stage. In particular, the flow-based model termed

"overload cascading model" (OCM) (Koc *et al*., 2013; Pahwa *et al*.,2014) is employed during the search by the multi-group non-dominated sorting genetic algorithm (MG-NSGA) (Chapter 2). The concept of 'critical zone' is also utilized to identify critical cascading scenarios based on the given definitions of criticality.

3.2. Flow-Based Cascading Failure Modeling of Power Grid

Several issues of using a topology-based cascading failure model were extensively discussed in the literature (Wang *et al*., 2009 and 2011; Hines *et al*., 2010; Pagani and Aiello, 2013; Fang *et al*., 2014 and 2015; Cauadra *et al*., 2015; Guo *et al*., 2017; Bompard *et al*., 2019). For instance, Guo *et al*. (2017) noted that topology-based cascading models are relatively simple but not successful in representing the physical features of the power supply network. Such approaches may result in erroneous conclusions which lead to misjudgment in disaster risk reduction planning. This issue can be addressed by using a flow-based model that can simulate the complex mechanism of the cascading failure of power supply network. Therefore, a flow-based cascading model, termed overload cascading model (OCM) (Koc *et al*., 2013; Pahwa *et al*.,2014) is adopted. The algorithm of OCM illustrated in Figure 3.1 can simulate the sequential overload line trip mechanism. First, the load flow demands are estimated for the initial post-disaster topology of the power grid. Next, the load flow demand at each power transmission line is compared with its capacity, and the overloaded transmission lines are removed from the initial network topology. These processes, i.e. load flow estimation, overload line check, and network topology updating, are repeated until the load flows are completely stabilized, i.e. no more cascading failures occur. In the following sub-sections, the main modules of OCM,

i.e. DC power flow analysis, and the power supply network topology updating procedure, are summarized respectively.

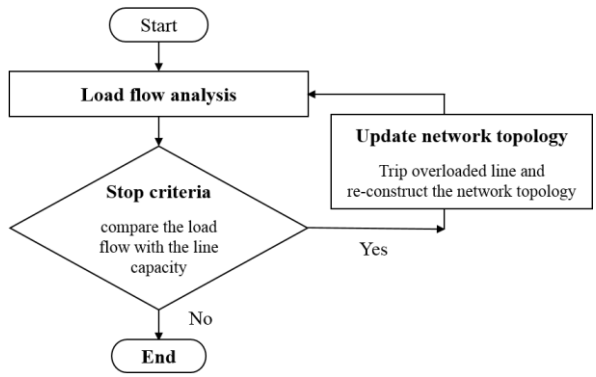


Figure 3.1 Flowchart of the overload cascading model (Pahwa *et al*., 2013)

3.2.1. Load Flow Analysis

In general, power network flow analysis is classified into two types: active current (AC) and direct current (DC) load flow analysis, depending on whether the reactive power is considered or not. The DC load flow analysis is often used in the power supply network optimization problem because of its simplicity and effectiveness (Koh *et al*., 2003). This study also adopts the DC load flow analysis to estimate line power flows in OCM. This is because both critical scenario search process using MG-NSGA and the simulation of cascading failures using OCM are iterative and thus AC flow analysis would make the search computationally intractable. The DC load flow analysis is considered a simplified version of AC load flow analysis under the following assumptions: (1) line resistance is negligible; (2) voltage angle difference is assumed to be small; and (3) magnitudes of bus voltage are set to the

flat voltage profile. Based on these assumptions, the DC flow equations are expressed as:

$$f_{ij} = b_{ij}\theta_{ij} \tag{3-1}$$

where f_{ij} and b_{ij} are respectively the active power flow and susceptance of the line l_{ij} connecting node *i* and *j*; and θ_{ij} is the voltage phase difference between the nodes. Thus, the power grid can be expressed in terms of the active power flows at the nodes, i.e.

$$P_i = \sum_{j=1}^{d} f_{ij} = \sum_{j=1}^{d} b_{ij}\theta_{ij} \tag{3-2}$$

where P_i is the active power flow at node *i*, and *d* is the degree of node *i*. As a result, the active power flows in a certain transmission line *i*, which connects the bus *s* and *r* can be calculated as

$$P_{Li} = \frac{\theta_{sr}}{X_{Li}} \tag{3-3}$$

where X_{Li} is the reactance of line *i*. The load flows through branches are then expressed using a matrix form

$$\mathbf{\Theta} = \boldsymbol{B}^{-1}\mathbf{P} \tag{3-4}$$

$$\mathbf{P}_L = (\mathbf{b} \times \mathbf{A})\mathbf{\Theta} \tag{3-5}$$

where **P** denotes the vector of bus active power injections, ***B*** is the admittance matrix, **Θ** is the vector for bus voltage angles, $\mathbf{P}_L$ is the vector for branch flows, **b** is the susceptance matrix, and **A** is the bus-branch incidence matrix (Grainger and Stevenson, 1994).

In summary, when the network topology and electric properties (i.e. generated power and/or power demand at each node, reactance of each line) of the network components are given, the power flow of each line can be evaluated using the DC load flow equations given above.

3.2.2. Updating Power Grid Topology

While simulating cascading failures by OCM, if the load flow in a transmission line exceeds its own capacity, the line is removed from the network topology. The removal process of the overload line is implemented by setting the admittance value of the removed line to zero (Koc *et al*., 2013). In this procedure, the power grid may be separated into multiple sub-networks, so-called 'islands.' As illustrated in Figure 3.2, an island could have one of three states: 1) intermediate, 2) stabilized state with generation, and 3) no generation. If an island still has overloaded line(s) even after the isolation, the island is considered to be at the intermediate stage which requires further evaluations. On the other hand, if an island reaches the other two states, no additional evaluations are required. When the conditions of all islands are either stabilized or isolated from generation, the cascading failure analysis is terminated. Lastly, the impact of the cascading failure is estimated based on the final conditions of the stabilized islands.

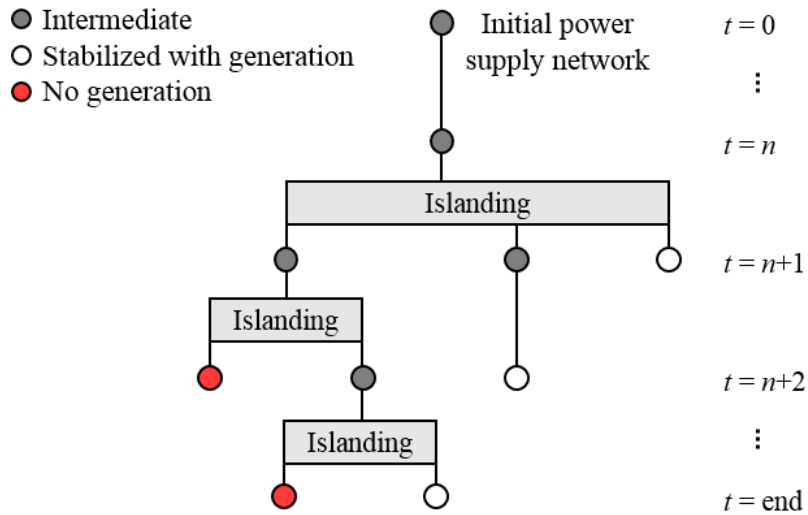


Figure 3.2 A tree structure of cascading failures and islanding process (Koc *et al*., 2013)

3.3. Cascading Failure Scenarios Identification Frameworks

Multi-objective optimization can be used to obtain a set of critical failure scenarios, for example, initial component failure scenarios in which a small number of failures lead to disproportionate system-level failure or impacts (Zio *et al*., 2012). In particular, MG-NSGA and the concept of 'critical zone' were proposed in Chapter 2 (Choi and Song, 2019). In this chapter, this approach is now applied to network cascading failures by adopting a measure of cascading failure impact as one of the objective functions. In the following sections, GA-based representation of initial post-disaster scenarios and objective functions to identify critical cascading failure scenarios are illustrated respectively in the context of cascading failure of power supply network.

3.3.1. GA-based representation of initial post-disaster scenarios

It should be noted that, the evolution process searches for the initial post-disaster failure states which will eventually result in significant or disproportionate damage through cascading failures. Therefore, the scenarios described by the GA-based form during the Multi-Objective genetic algorithm (MOGA) procedure represent the initial component failure condition, not the final cascading failure condition. As shown in Figure 3.3(a), each scenario can be represented by a binary string, where the values 1 and 0 respectively indicate the survival and failure of the network component at the initial stage. For example, Figure 3.3(b) shows the initial component failure scenarios of a simple power supply system consisting of 3 nodes and 2 links.

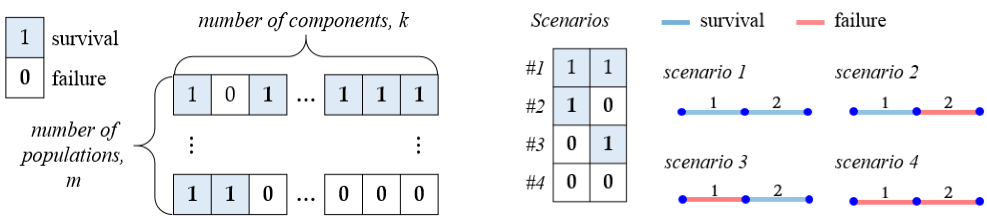


Figure 3.3 (a) GA-based representation of post-disaster scenarios of initial failures; and (b) an example of initial link failure of a two-link power grid and corresponding GA-based representations

3.3.2. Multiple objective functions to identify critical cascading failure scenarios

To identify critical cascading failure scenarios using MOGA, two objective functions having a trade-off relationship are selected in this research: (1) number of components that failed at the initial stage, and (2) network functionality at the final cascading stage. The Pareto surface obtained by minimizing the objective functions, i.e. non-dominated solutions would show that the post-cascading-failure network functionality decreases as the number of initially failed components increases. Therefore, the two objective functions are expected to identify a small number of initial failures that result in serious impacts through cascading failures in the power supply network (Choi and Song, 2019; Zio *et al*., 2012).

3.3.2.1. Objective function 1: number of components that failed at initial stage

To identify scenarios entailing devastating consequence even with a relatively small number of component failures, i.e. scenarios leading to out-of-proportion consequence, the 'number of components that failed at initial stage' is introduced as one of the objective functions. The objective function is described as

$$f_{i1} = k - \sum_{j=1}^{k} x_{ij} \tag{3-6}$$

where f_{i1} is the 'number of components that failed at initial stage' in the *i*th scenario, k is the length of the GA-based representation in Figure 3.3(a), and x_{ij} is the element of the binary matrix in Figure 3.3(b) corresponding to the *j*th component in the *i*th scenario.

3.3.2.2. Objective function 2: measure of network functionality at final cascading stage

The second objective function represents the functionality of the network at the final cascading stage to identify scenarios leading to serious loss in terms of network functionality. The flexibility of the genetic algorithm facilitates applications of various network functionality measures (e.g. number of still connected nodes to the nodes of interest (Choi and Song, 2019)) as the second objective functions. In this study, 'total active link capacity' is adopted as an example of power grid functionality measure. The 'active link' is defined as the link that withstands the power flow demand at the final cascading failure stage while connected with at least one single generator node (Koc *et al*., 2013b). The corresponding objective function is

$$f_{i2} = \sum_{j=1}^{k} act_{ij} \times cap_j \tag{3-7}$$

where f_{i2} is the 'total active link capacity at final cascading stage' of the *i*th scenario, act_{ij} denotes a binary variable which is 1 if the jth link is 'active' in the ith scenario and 0 otherwise, and cap_j is the capacity of the *j*th link. The matrix of act_{ij} could be obtained by performing cascading failure analysis for each scenario using OCM described in Section 3.2.

3.4. Case studies

3.4.1. Power Supply Networks: IEEE 30-Bus Example

The IEEE 30-bus system example (Alsac and Stott, 1974) is selected to demonstrate the applicability and efficiency in using 1) OCM and MG-NSGA method together for identifying critical cascading failure scenarios and 2) the elite set updating method for evaluating cost-effective retrofit combinations. As illustrated in Figure 3.4, the power supply network is modeled by the graph consisting of 30 nodes (6 generators and 24 substations) and 41 edges (transmission lines).

Critical cascading failure scenarios, defined as those entailing significant losses in terms of 'total active link capacity' (Objective function 2) by a small 'number of the links that failed at the initial stage' (Objective function 1), are searched. To evaluate the second objective function value using OCM, the open source Matlab code MATCAC (Koc *et al*., 2013) was adopted after slight modifications. Unlike the existing research, this work does not restrict initial post-disaster conditions to single component failure scenarios in generating and selecting initial conditions by MOGA. After identifying cascading scenarios by MG-NSGA, those in the critical zone are selected for the purpose of decision-making.

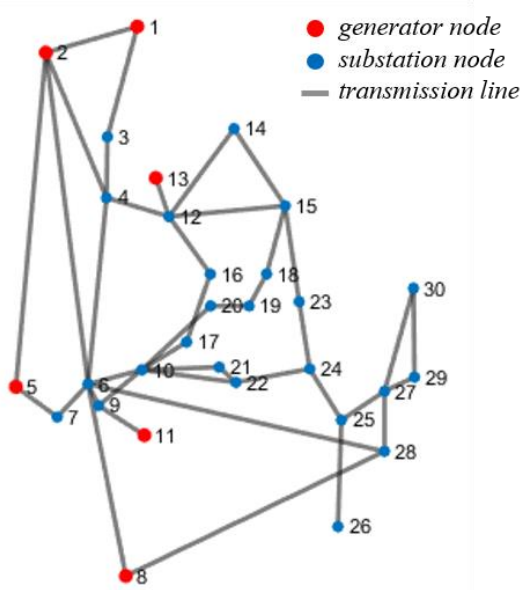


Figure 3.4 Network topology of the IEEE 30-bus power supply system

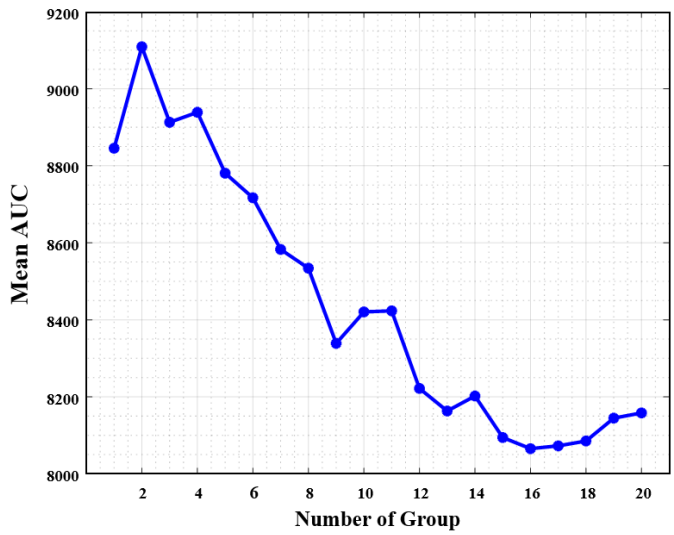


Figure 3.5 Mean AUCs of 50 initial populations for NSGA-II (number of group = 1) and MG-NSGA (number of groups varied from 2 to 20)

To investigate the effectiveness of MG-NSGA in identifying critical cascading scenarios, the results by NSGA-II are compared with those by MG-NSGA while varying the number of groups from 2 to 20. For comparison between the performances of the two MOGAs, the area under curve (AUC) of the Pareto curves (Bradley, 1997) is obtained for each of 50 different initial populations, and the mean of the 50 AUCs is computed for each MOGA. For numerical experiments, the following options were used: the size of the initial population 150, mutation ratio 1/41 (which is 1/ k where k is the string size (Ochoa, 2002)), and stopping criteria 2,000th generation, which is a conservative choice determined based on several test runs. According to the experience gained in this research, the algorithm delivered the optimal convergence with these options for the 30-bus power system. Figure 3.5 compares the mean AUC by NSGA-II with those by MG-NSGA and confirms that MG-NSGA provides better performance unless the number of group is too small. Especially, it is indicated in the figure that MG-NSGA with 16 groups deliver the best performance. Hence, the results by MG-NSGA with 16 groups are used henceforth. For further investigation, the boxplots of the 50 Pareto solutions by MG-NSGA (Figure 3.6a) and those by NSGA-II (Figure 3.6b) are compared with each other. The results confirm that MG-NSGA outperforms NSGA-II in terms of variability of optimal solutions as well.

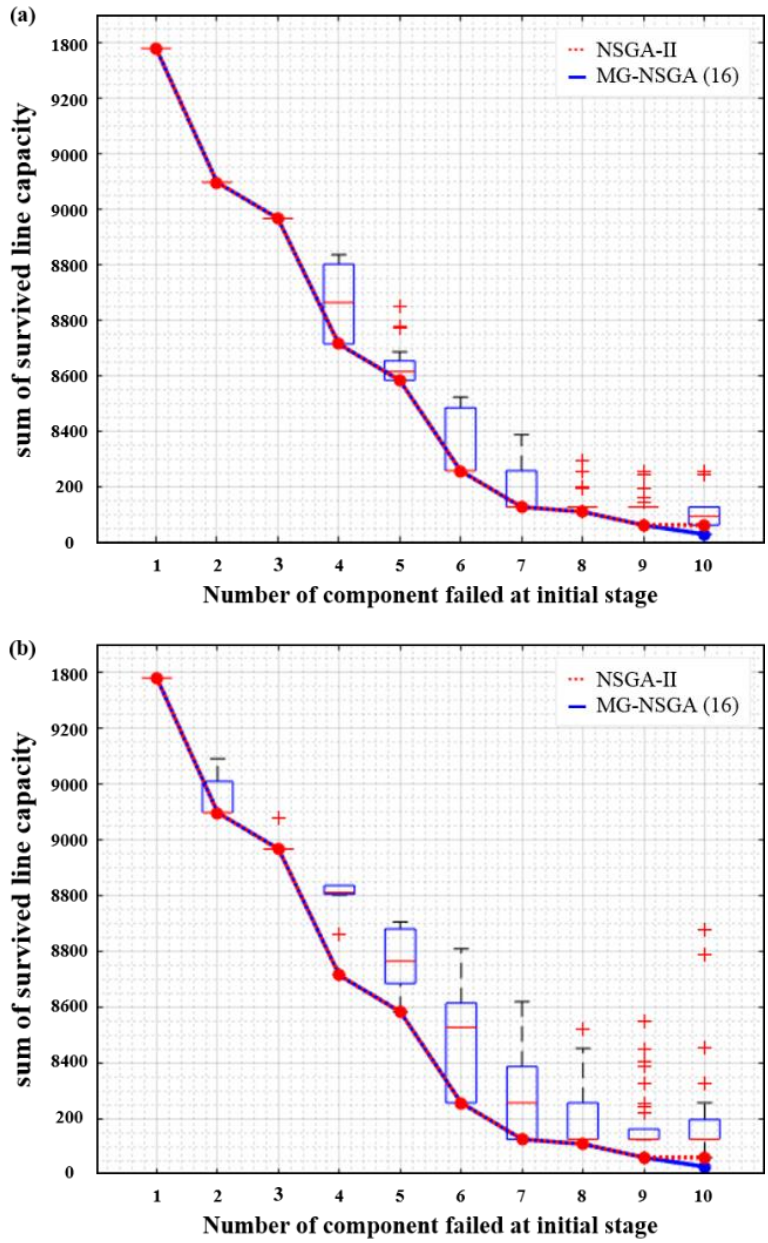


Figure 3.6 Box-plots of 50 Pareto solutions of IEEE 30-bus system by (a) MG-NSGA with 16 groups; and (b) NSGA-II

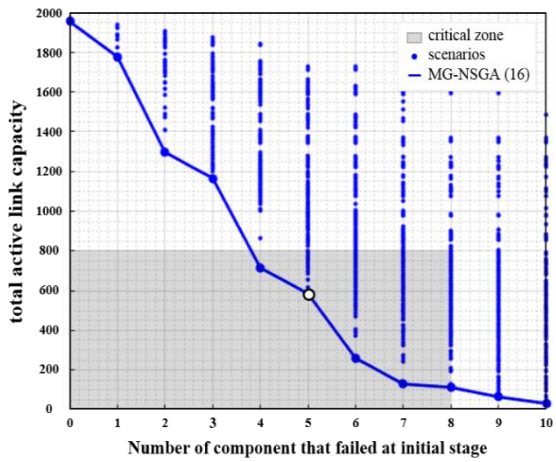


Figure 3.7 Post-disaster scenarios archive and final Pareto solutions collected from 50 runs of MG-NSGA using 16 groups

Using the proposed combination of OCM and MG-NSGA, critical cascading failure scenarios of 30-bus system are identified. The post-disaster scenario archive and the final Pareto solutions collected from 50 runs of MG-NSGA with 16 groups, are shown in Figure 3.7. Each point in the plot represents the cascading failure initiated by a different initial failure scenario. On the Pareto surface, the 'total active link capacity' at the final stage decreases as the 'number of components that failed at the initial stage' increases. This confirms that cascading failure scenarios induced by multi-component failures at the initial stage, which were rarely addressed in the literature, need to be considered in efforts for critical scenario identification and decision-making process.

In this example, it is assumed that the decision maker defines 'critical' scenarios as the cases in which "total active link capacity is less than 800 (MVA) at final cascading stage, even though at most 8 links fail initially." Based on this critical zone

definition, a total of 1,904 critical cascading failure scenarios were identified. For instance, the progress of the cascading scenario represented by "o" marker in Figure 3.7 is illustrated in Figure 3.8. Although a relatively small number of components experience initial failures, the system loses two thirds of the total link capacity at the final cascading stage due to the sequential load re-distributions and the overload trips. It is noteworthy that the number of active links drops quickly as progressing from t=3 to t=4. Such progress of the cascading failure at each step and the relationship between the initial component failures and the final cascading results cannot be detected by insight. Critical scenario identification by MG-NSGA with OCM and the concept of the 'critical zone' would help provide meaningful insights for the decision makers.

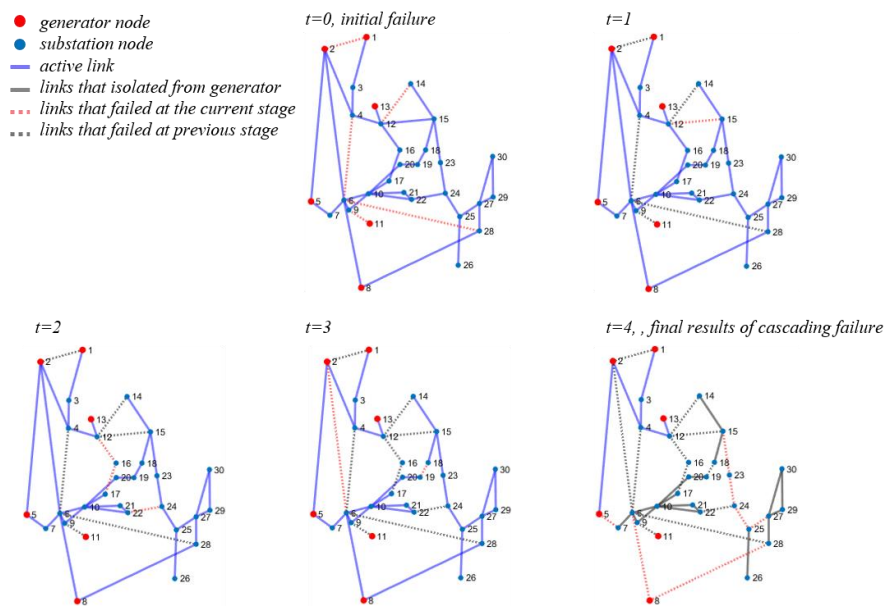


Figure 3.8 Progress of critical cascading failure scenario induced by initial failures of five components ("o" marker in Figure 3.7)

3.4.2. Power Supply Networks: IEEE 118-Bus Example

This section demonstrates the applicability of the proposed methods to large-size power supply networks by the IEEE 118-bus example. The network topology and the electric properties of the power system are from (Christie, 1993) and (Illinois Institute of Technology, 2019). As illustrated in Figure 3.9, the power supply network is modeled as the graph of total 118 nodes (generators and substations) and 186 edges (transmission lines).

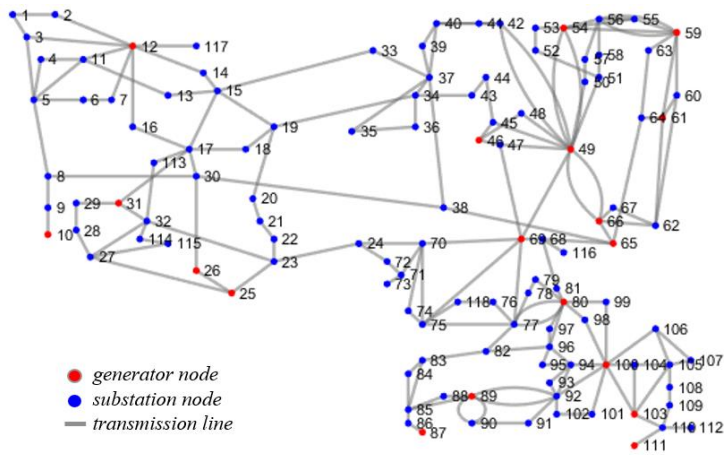


Figure 3.9 Network topology of the 118-bus power supply system

Using the same objective functions as IEEE 30-bus example, both NSGA-II and MG-NSGA with 10 groups, are used to search critical cascading failure scenarios. To consider the variability of the results caused by initial populations, 60 different initial populations were examined for both MOGAs. For numerical experiments, following options were used: the size of the initial population 150, mutation ratio 1/186 (which is $1/k$ where k is the string size (Ochoa, 2002)), and stopping criteria

20,000th generation, which is a conservative choice made based on several test runs. According to the experience, the algorithm delivered the optimal convergence with these decision variables for the 118-bus power system.

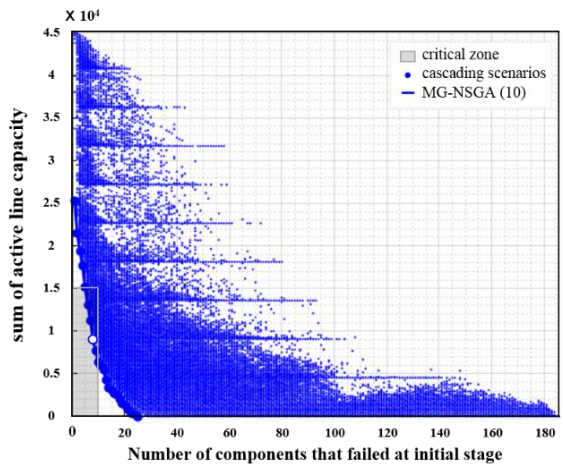


Figure 3.10 Archive of post-disaster scenarios archive and final Pareto solutions collected from 60 run of MG-NSGA with 10 groups

Using MG-NSGA with OCM, critical cascading failure scenarios of large-size network are successfully identified. In Figure 3.10, the post-disaster scenarios archive and final Pareto solutions collected from 60 runs of MG-NSGA are plotted. Each point in the figure represents a certain cascading scenario. Figure 3.10 indicates that the entire network links could be deactivated by the cascading failure induced by initial failures of 25 components. Among the large number of scenarios examined during the evolution process, those in the 'critical zone' determined by the decision makers are selected: the cases in which "total active link capacity is only less than 15,000(MVA) even though at most 10 links fail." By this critical zone definition, a

total of 2,015 critical cascading failure scenarios were identified for the 118-bus system. As one of the critical cascading failure examples, the progress of cascading failures for the scenario marked as the "o" marker in Figure 3.10 is illustrated in Figure 3.11. With the failures of only 8 links at initial stage, which is less than 5% of total number of links, most of the links are deactivated at the final cascading stage due to series of the load re-distribution and disconnection. In this larger-size power grid, more complex cascading failures scenarios were observed. Since obtaining the set of critical cascading scenarios by insight is extremely difficult, applying the proposed frameworks to large-size power supply network would facilitate the decision making process. Lastly, Figures 3.12(a) and (b) show the boxplots of the Pareto solutions achieved by MG-NSGA and NSGA-II respectively. MG-NSGA outperforms NSGA-II in terms of both optimality and variability.

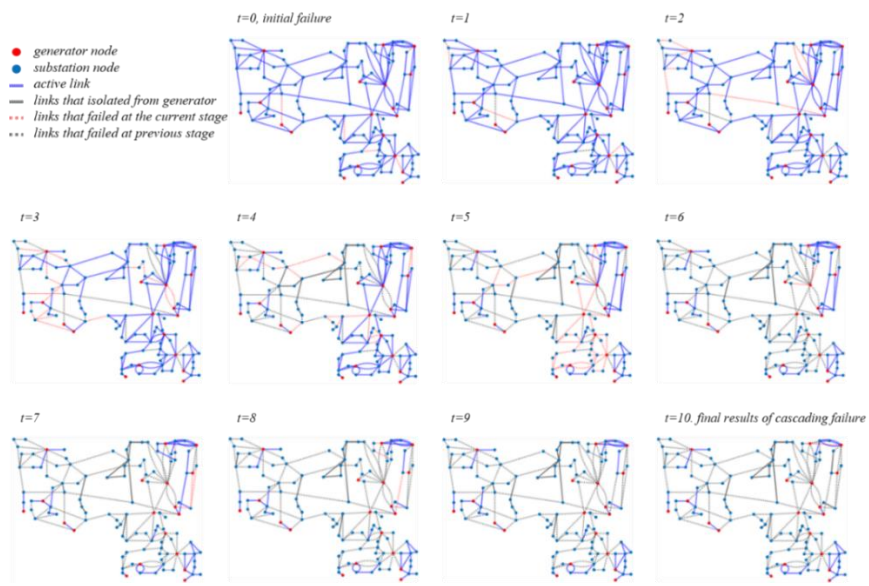


Figure 3.11 Critical cascading failure scenario induced by eight initial component failures ("o" marker in Figure 3.10)

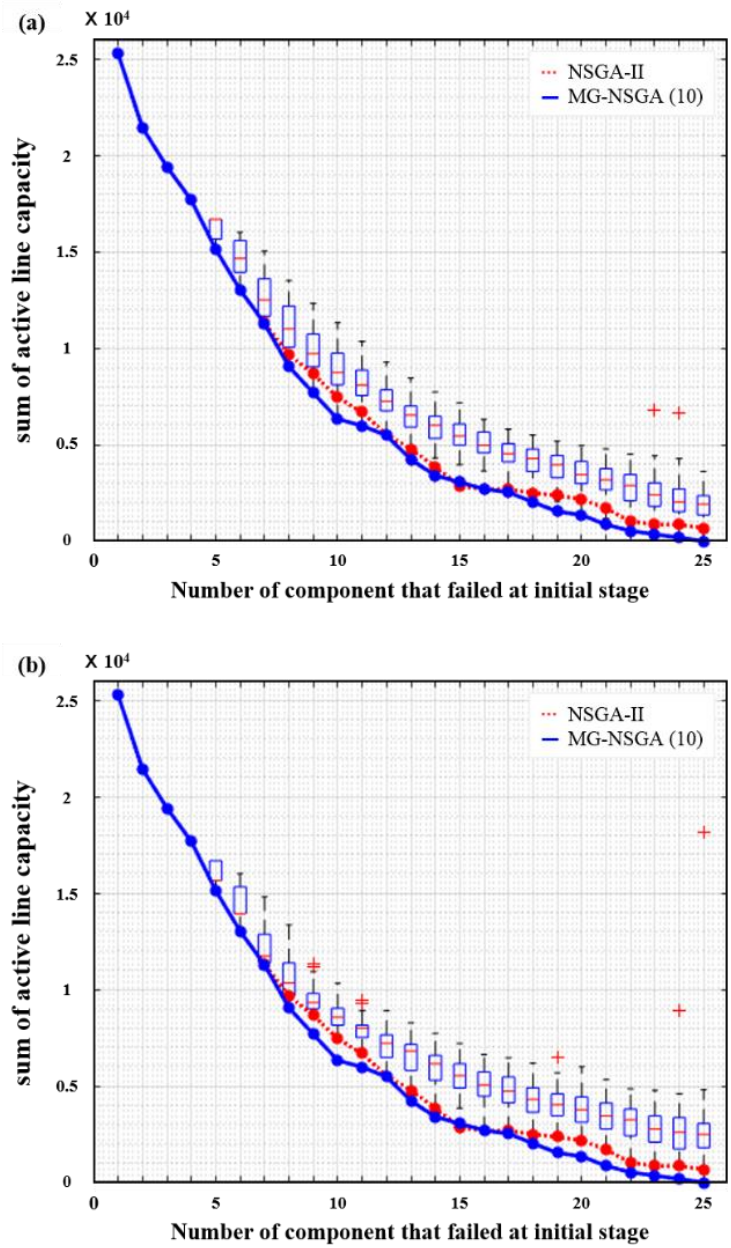


Figure 3.12 Box-plots of 60 Pareto solutions of IEEE 118 bus system by (a) MG-NSGA with 10 groups; and (b) NSGA-II

3.5. Summary

By combining the flow-based overload cascading model and the MG-NSGA, critical cascading failure scenarios are successfully identified. The results from MG-NSGA and NSGA-II are also compared with each other. Results indicate that MG-NSGA delivers the solutions with better accuracy and less variability than those of NSGA-II. Using the proposed framework, critical cascading scenarios, which are induced by multi-component failure at the initial post-disaster stage, are identified in terms of "total active link capacity" at the final cascading stage. By using the network performance measure, which is evaluated from the result of the flow-based cascading failure modeling, the socio-economic outcome of each scenario is well-represented. Furthermore, the results of the critical cascading failure scenario search show that the risk of cascading failure increases as the number of components failed at the initial stage increases at the final Pareto surface. Such results indicate that cascading failure scenarios induced by multiple component failures at the initial stage should not be neglected in power grid disaster risk mitigation planning. Through comprehensive demonstration and investigation of critical cascading failure scenarios, decision-makers can gain useful insights on retrofits and maintenance of the power grid. Moreover, the set of the identified critical cascading failure scenarios can be used to optimize various power network management decisions, e.g. optimal retrofit combination under budget, retrofit sequence, capacity allocation, intentional islanding.

4. Elite-set Updating to Identify Cost-Effective Network Retrofit Combinations

4.1. Introduction

The purpose of identifying critical cascading scenarios is to identify optimal countermeasures against the identified scenarios using the available resources. To find the most cost-effective retrofit combinations, examining all possible retrofit combinations would be ideal, but often computationally intractable. Therefore, an "elite set updating method" is proposed to effectively evaluate the cost-benefit trade-off relationship of the potential retrofit combinations. In this chapter, the proposed "elite set updating" method is described and applied to two power grid examples.

4.2. Elite-set Updating Method

4.2.1. Selecting 'Candidates' and 'Elite' Components of Network

To reduce the number of retrofit combinations that will be considered during the decision-making process, 'candidate' components are first selected by investigating impact and cost by potential retrofit of each component. The components whose retrofit will make greater impact (according to the procedure explained below) are identified as 'impact' set. To take into account the budget limitation as well, the components with lower retrofit cost are identified as 'cost-effective' set as shown in Figure 4.1 (left). The non-dominated components in 'cost' and 'impact' data space are selected as 'cost-effective' components. Finally, the network components in the union of the impact set and cost-effective set are identified as the 'candidates' as

illustrated in Figure 4.1 (right). Among the identified candidates, components belonging to the cost-effective set, i.e. yellow areas in Figure 4.1, are termed 'elite' components in this work.

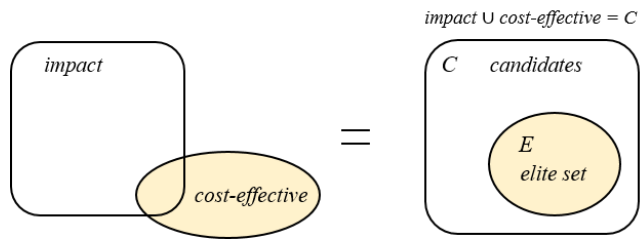


Figure 4.1 Identification of candidate and elite set

During the selection procedure, the 'impact' of retrofitting each single component is measured by re-simulating the identified cascading failure scenarios under the assumption that the component of interest has been retrofitted and thus can withstand at the initial post-disaster stage. Numerically, to represent the condition that *j*th component of the power supply network has been retrofitted, the default value of *j*th row of the critical cascading scenarios matrix be set as 1, which indicates the survival of the *j*th component at the initial stage because of the increased robustness against the disastrous event (see Figure 3.3(a)). The benefit of retrofitting the *j*th component is measured by the increase of the average 'total active link capacity' of the power supply network, which is defined as

$$Impact_{[j]} = \frac{\sum_{i=1}^{N_{cs}} (f_{i2} - f_{i2}^{[j]})}{N_{cs}} \tag{4-1}$$

where f_{i2} is the 'total active link capacity at the final cascading stage' of the originally identified *i*th critical cascading scenario, and $f_{i2}^{[j]}$ denotes the 'total active link capacity at the final cascading stage' of the *i*th critical cascading scenario with *j*th component withstanding at initial cascading stage. In addition, N_{cs} is the number of total critical cascading failure scenarios identified by the method presented in Chapter 3. After evaluating the 'impact' of every component, the components with bigger impact than a prescribed threshold value are selected as the impact set. For example, in Figure 4.2 (left), components #1, 3, 8, 9, and 16 are selected as 'impact' set. The threshold value, used in this candidate selection, is a case-specific parameter which could be determined by the network management authorities. Second, the 'cost-effective' set is selected from the scatter plot of impact and the cost of retrofitting single component. As illustrated in Figure 4.2 (right), components #8 and 21, two non-dominated are selected as 'cost-effective' solutions.

The final candidate set is obtained as the union of impact and cost-effective sets, which includes component #1, 3, 8, 9, 16 and 21. Among the components in the candidate set, the 'cost-effective' components #8 and 21, are identified as the 'elite' set. The 'elite set,' a subset of the candidate set showing cost-effectiveness, will be continuously updated through the 'elite set updating' process which will be described below. The elite set initially includes cost-effective but less impactful components, e.g. component #21 because it is likely to be part of optimal solutions of retrofit combinations with the larger size. If those kinds of components are neglected, suboptimal cost-benefit solutions will be identified.

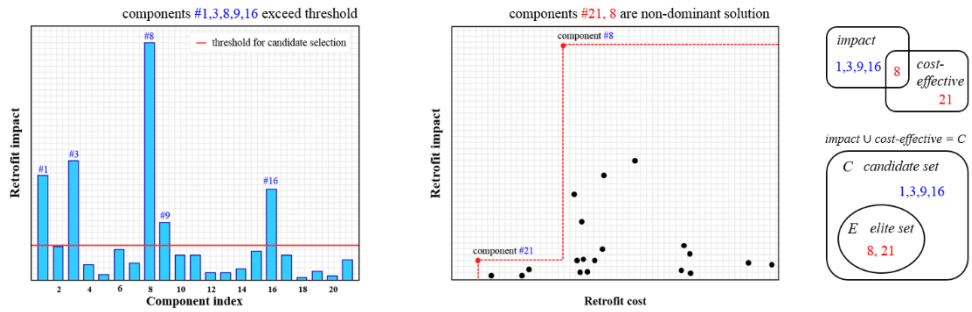


Figure 4.2 Conceptual demonstration of selecting 'candidates' and 'elites'

4.2.2. Generating Retrofit Combinations using Candidates

Next, retrofit combinations are generated by using the candidate components selected by the procedure in section 4.2.1. Exploring retrofit combinations only among the candidates would significantly reduce both the number of combinations and computational time cost when compared to those by the complete enumeration (termed "all-candidates" method henceforth). The 'elite set updating' method proposed in this chapter gradually updates the elite set as generating and exploring more retrofit combinations in the candidate set. At each step, under the assumption that the elite components are more likely to be a part of the cost-effective optimal solutions, the following two groups of combinations with size *n* are generated. For the first group, *n* components are selected from the elite set only while the second group chooses $(n-1)$ components from the elite set and one non-elite component in the candidate set. The second group with $(n-1)$ elite components and 1 non-elite candidate component is proposed by the rule of thumb. While evaluating the optimal retrofit combinations, the non-dominated solutions which include more than one non-elite candidate are not identified. Figure 4.3 shows an example with $n = 2$,

using the components selected in Figure 4.2. In this case, one and eight combinations are generated for the first and second group respectively. On the other hand, if the all-candidates method is used, a total of 15 combinations would be generated. Although only 6 retrofit combinations are reduced by using the elite set updating method in this example, it should be noted that computational cost of $6 \times N_{cs}$ cascading failure assessments are being saved.

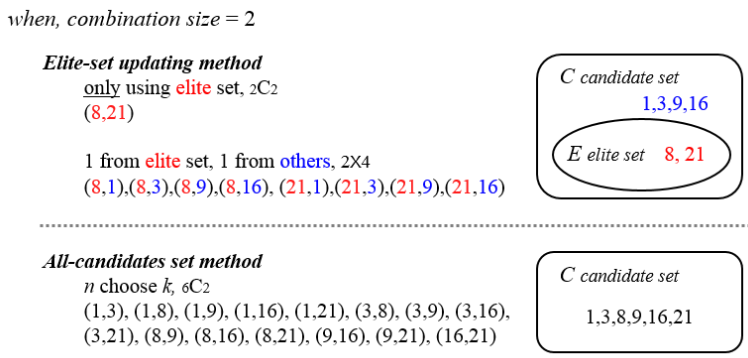


Figure 4.3 Example of generating retrofit combinations using elite set updating method and all-candidates method

Again, the proposed retrofit combination generating approach has the underlying assumption that cost-effective combinations will not include more than one candidate component which is not elite components. Therefore, theoretically, the elite-set updating method has the potential to converge to local minima if the non-elite component appears the same time for the non-dominated retrofit combination. Besides, if the cost-effective retrofit combination includes the components outside the candidate set, results also may converge to the suboptimal points. However, such possibilities could be reduced by checking assumptions by

generating a 'third group' with the $(n-2)$ elite component and 2 other candidates for the combination size larger than two and also increasing the number of candidates.

4.2.3. Evaluating the Cost and Benefit of Generated Retrofit Combinations

Once retrofit combinations are generated, the cost of each retrofit combination is estimated and compared with the given budget. If the cost exceeds the budget, the corresponding combinations are removed from the retrofit combination sets. Otherwise, the improvement of the post-disaster network functionality by protecting the retrofit combinations are evaluated. In particular, the increase in the mean 'total active link capacity' in Equation 4-2 is measured for each combination. For example, to evaluate the impact of retrofitting components #8 and 21 together, the default value of the 8th and 21th rows of the critical cascading scenarios matrix are set as 1. Lastly, as shown in Figure 4.4, the evaluation results are added to the cost-benefit (impact) curve.

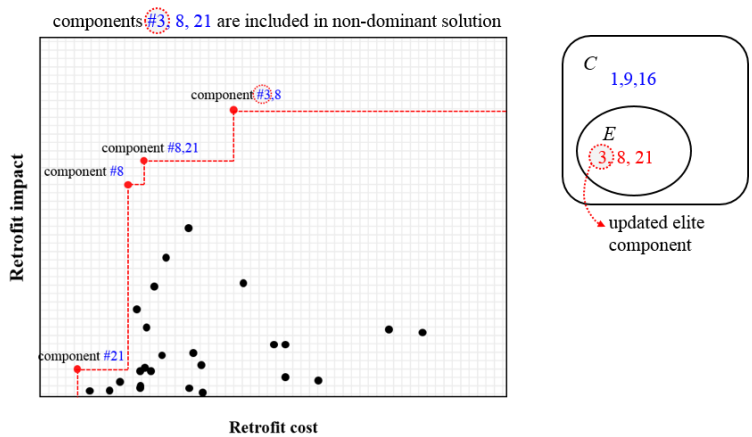


Figure 4.4 Example of cost and benefit evaluation of retrofit combinations (size $n = 2$) and corresponding updating of the elite set

4.2.4. Updating Elite Set

After performing cost and benefit evaluation for retrofit combinations with size *n*, the non-dominated solutions are checked to identify new component(s) appearing in the Pareto solution set. If new elite components are identified, the elite set is updated. For example, in Figure 4.4, up to retrofit combination size 2, combinations (21), (8), (8, 21), and (3, 8) are identified as non-dominated solutions. Since the component #3 was not included in the elite set at the previous size step, i.e. $n = 1$, the elite set is now updated to include #3. It is important to note that only the elite set is updated through the process while the size of candidate set remains the same, i.e. exploring within the candidate set. Figure 4.5 shows the flowchart of the proposed elite set updating method in which each step is illustrated by the example with combination size $n = 3$. After evaluating the retrofit combination size up to 3, the component #1 is newly identified as an elite among the candidates. These procedures are repeated until the algorithm meets one of the prescribed stopping criteria such as retrofit budget, size of the combinations.

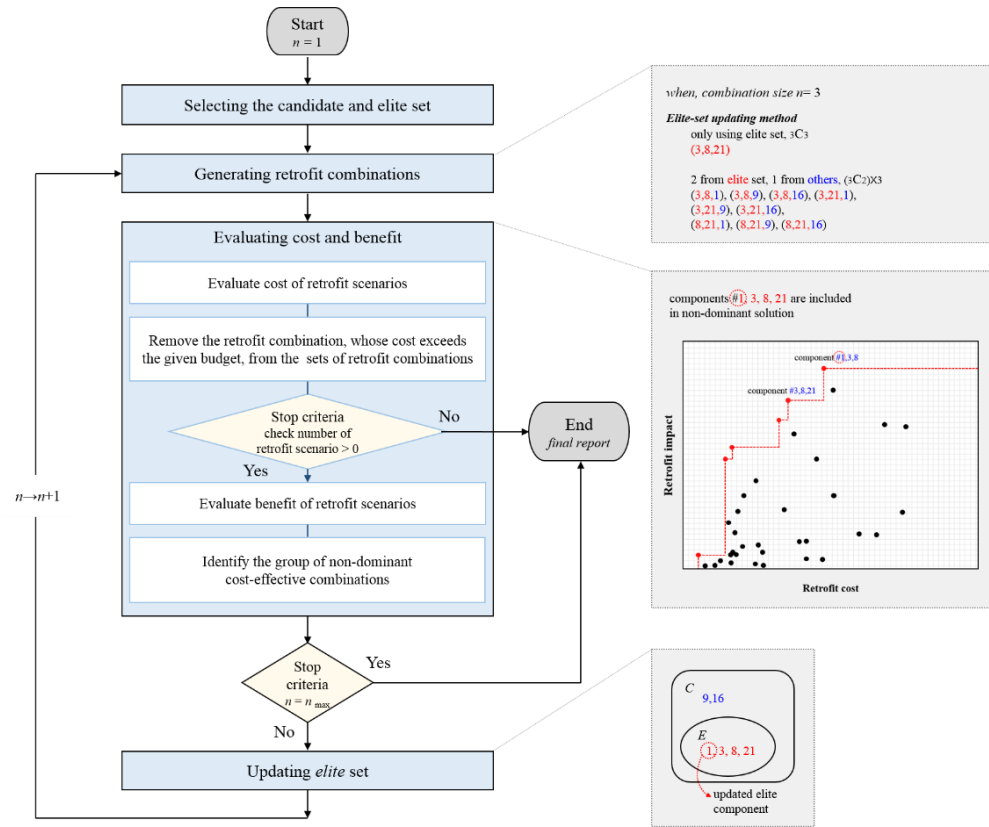


Figure 4.5 Flow chart of the elite set updating method and illustrations of the procedures using the example with combination size 3

4.3. Case Studies

4.3.1. Power Supply Networks: IEEE 30-Bus Network

Based on 1,904 critical cascading scenarios identified by MG-NSGA (Results of Chapter 3), the most cost-effective retrofit combinations are identified by the proposed elite set updating method. In this example, the cost of retrofitting each network component is assumed to be proportional to the length of the transmission

line. Under this assumption, the retrofit cost is scaled such that the cost for the shortest transmission line is 1. In addition, while selecting candidate components in terms of impact, 56.125 (MVA) is used as the threshold, which is 10% of the mean 'total active link' over 1,904 scenarios. As for the stopping criteria, only 'maximum combination length' is used.

To test the validity and the effectiveness of the proposed method, the results by the elite set updating and all-candidates methods were compared with each other. Those were compared with the exact solutions, i.e. complete enumeration without candidate selections. Since this exhaustive search is computationally intractable as the combination size or network size increases, complete numerations were performed for combinations with up to three components only in this example.

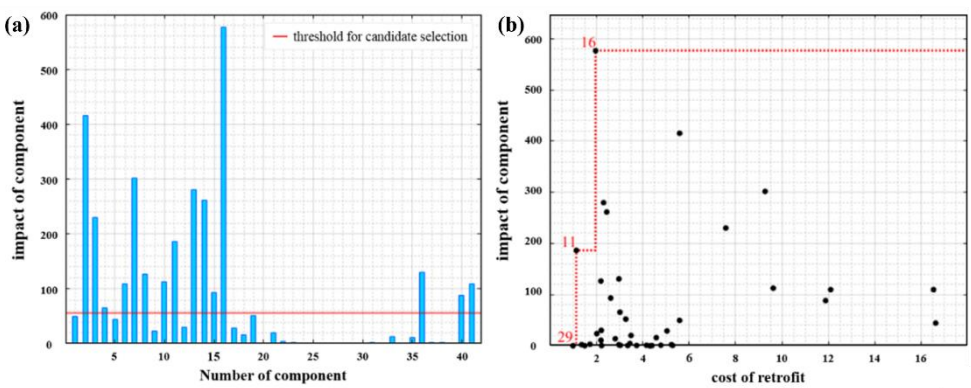


Figure 4.6 Candidates selected for the 30-bus example based on (a) impact and (b) cost-effectiveness

The candidate and elite components are selected as illustrated in Figure 4.6 The candidates selected in terms of impact criteria are shown in Figure 4.6(a). In Figure 4.6(b), by considering cost-effectiveness, three components #11, 16, and 29 are

selected as the elite components. A total of 16 candidates selected for the 30-bus system are illustrated in Figure 4.7.

After evaluating the cost and the benefit of the retrofit combinations with length up to 8, the final cost-benefit stair-step graph is plotted as shown in Figure 4.8. Accordingly, in Table 2, the non-dominated solutions are listed for each size of retrofit combinations along with the elite components updated at each iteration. Until the combinations with length up to three, the three retrofit combination generation methods delivered the same results. In addition, for the retrofit combinations size up to 8, the elite set updating and all-candidates methods deliver the same Pareto solutions. According to this result, the main assumption introduced for the elite set updating method, i.e. the assumption that the solutions included in (i-1) size combination are more likely to be included in i size combinations, seem reasonable.

Besides, the number of the combinations generated at each size of the retrofit combination using 1) the proposed method 2) the all-candidates method, and 3) complete enumeration without candidate selection are also compared in Table 4.1. While the same cost-effective solutions are achieved by using the three methods, the elite set updating method requires much less number of combinations. Because reducing one retrofit combination could save the computational cost of *Ncs* cascading simulations, the elite set updating method is significantly effective than other methods. For this example, the total number of retrofit combinations reduced from 39,227 to 4,457, which indicates that only 11% of the computational expense required by adopting the elite-set updating method.

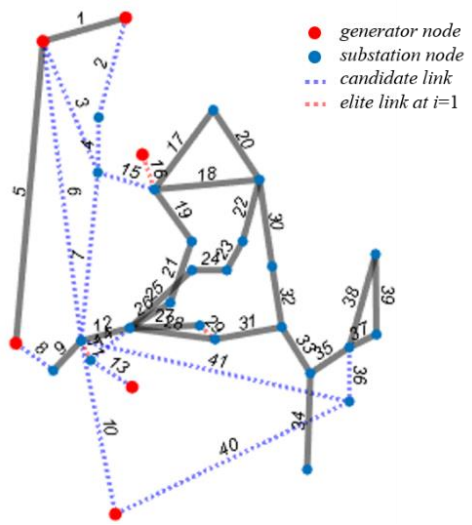


Figure 4.7 Topological distributions of candidates and elites after first iteration

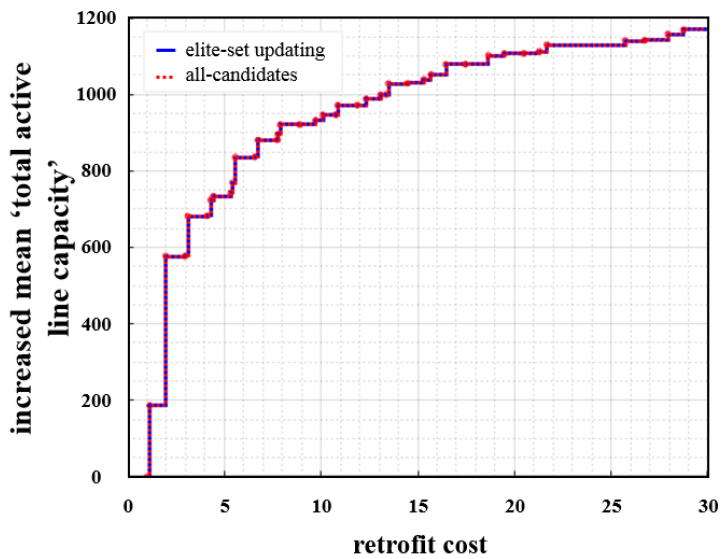


Figure 4.8 Stair-step plot showing cost-benefit relationship of retrofit combinations of 30-bys system

Table 4.1 Pareto solutions of cost-effective retrofit combinations for 30-bus system

Size	Number of combinations*			Pareto solutions	Updated Elite set
	(1)**	(2)**	(3)**		
1	41	41	41	(29), (11), (16)	11, 16, 29
2	42	120	820	(16, 29), (11, 16), (13, 16), (14, 16), (7, 16)	7, 13, 14,
3	170	560	10,660	(11, 16, 29), (8, 11, 16), (11, 13, 16), (11, 14, 16), (13, 14, 16)	8
4	385	1,820	-	(16, 14, 11, 29), (16, 13, 14, 29), (16, 14, 11, 8), (16, 13, 14, 11), (16, 13, 14, 36), (16, 2, 14, 11), (16, 2, 13, 14)	2, 36
5	1,134	4,368	-	(16, 13 ,14 , 11, 29), (16, 13, 14, 11, 8), (16, 14, 11, 36, 8), (16, 13, 14, 11, 36), (16, 2, 13, 14, 29), (16, 2, 14, 11, 8), (16, 2, 13, 14, 11), (16, 2, 13, 14, 36), (16, 2, 7, 14, 11), (16, 2, 7, 13, 14)	-
6	1,092	8,008	-	(16, 13, 14, 11, 36, 29), (16, 13, 14, 11, 36, 8), (16, 2, 13, 14, 11, 29), (16, 2, 13, 14, 11, 8), (16, 2, 13, 14, 11, 36), (16, 2, 7, 13, 14, 11), (16, 2, 7, 14, 11, 36), (16, 2, 7, 13, 14, 36)	-
7	708	11,440	-	(16, 2, 13, 14, 11, 36, 29), (16, 2, 13, 14, 11, 36, 8), (16, 2, 13, 14, 11, 36, 4), (16, 2, 7, 14, 11, 36), (16, 2, 7, 14, 11, 36, 29), (16, 2, 7, 13, 14, 36, 29), (16, 2, 7, 14, 11, 36, 8), (16, 2, 7, 13, 14, 11, 36), (16, 2, 7, 13, 14, 36, 4)	4
8	885	12,870	-	(16, 2, 13, 14, 11, 36, 4, 29), (16, 2, 13, 14, 11, 36, 8, 15), (16, 2, 13, 14, 11, 36, 8, 4), (16, 2, 7, 13, 14, 11, 36, 29), (16, 2, 7, 13, 14, 11, 36, 8), (16, 2, 7, 13, 14, 11, 36, 4)	15
total	4,457	39,227	-	-	-

* Number of retrofit combinations which are generated by each method; ** (1) elite-set updating method; (2) all-candidates method; (3) complete enumeration without candidate selection

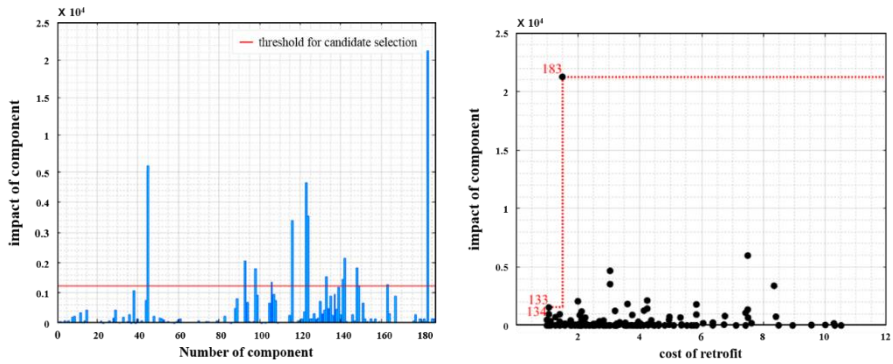
4.3.2. Power Supply Networks: IEEE 118-Bus Network

Based on 2,015 critical scenarios identified in Chapter 4, the most cost-effective retrofit combinations are obtained by the proposed 'elite set updating' method. The cost of retrofitting is assumed to be proportional to the length of the transmission line. While selecting candidates in terms of impact criteria, 122.75 (MVA), which is 10% of the mean 'total active link' over 2,015 scenarios, is used as the threshold. For the stopping criteria, only 'maximum combination length' is used. To verify the validity and the effectiveness of the proposed method, results from elite set updating method and all-candidates method were compared with each other.

Using the 'impact' criteria, candidate components are selected as shown in Figure 4.9. Most of the components in the 118-bus system have minor impacts on reducing cascading failure risks, while a small number of components can improve the robustness of the system. In Figure 4.9(b), using the additional criteria 'cost-effectiveness,' three components #133, 134, and 183 are selected as elite components.

The final cost-benefit stair-step graph, evaluated from the retrofit combinations with length up to 10, is plotted in Figure 4.10. According to these results, In Table 4.2, the non-dominated solutions are listed for each size of retrofit combinations along with the elite components at each iteration. The elite set updating and the all-candidates method provide the same Pareto solutions. Up to retrofit combination size 10, a total of 2,502 and 30,997 combinations were explored by the elite set updating method and all-candidates method respectively. For this example, only 8% of the computational expense was required by adopting the elite-set updating method. Considering the computational cost of simulating the cascading failure of large size power grid and the size of the critical scenarios set, the proposed elite set updating

method is considered helpful as a process to find optimal retrofit combinations, especially for large-size networks.



(a) (b)

Figure 4.9 Candidates selected for the 118-bus example based on (a) impact and (b) cost-effectiveness

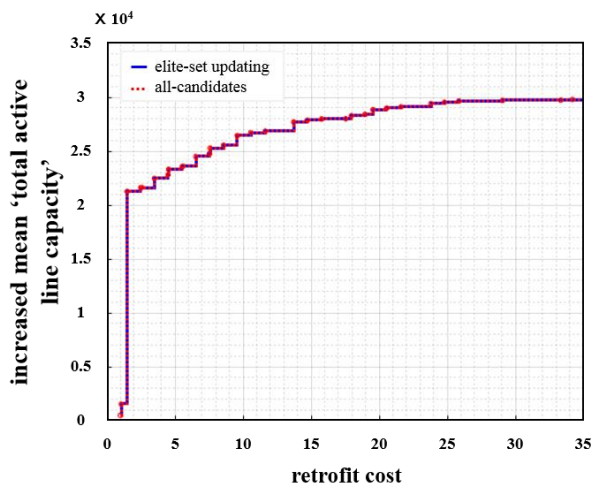


Figure 4.10 Stair-step plot showing cost-benefit relationship of the retrofit combinations of 118-bus system

Table 4.2 Pareto solutions of cost-effective retrofit combinations for 118-bus system

Size	Number of combinations*		Pareto solutions	Updated Elite set
	(1)**	(2)**		
1	186	186	(133), (134), (183)	133, 134, 183
2	39	105	(183, 134), (183, 133), (183, 93), (186, 124)	93, 124
3	110	455	(183, 134, 93), (183, 134, 124), (183, 133, 123), (183, 124, 93), (183, 124, 123)	123
4	150	1,365	(183, 134, 124, 93), (183, 134, 124, 123), (183, 124, 93, 123), (183, 124, 123, 163), (183, 124, 123, 98)	98, 163
5	546	3,003	(183,134,124, 93,123),(183,124, 93,123,163), (183,124,123,163, 98)	-
6	420	5,005	(183, 134, 124, 93, 123), (183, 133, 134, 124, 93), (183, 124, 93, 123, 163), (183, 133, 124, 163, 98), (183, 124, 123, 163, 98), (183, 124, 93, 163, 142), (183, 134, 124, 93, 123)	142
7	540	6,435	(183, 133, 134, 124, 93, 123, 163), (183, 134, 124, 93, 123, 163, 142), (183, 134, 124, 93,123, 163, 98), (183, 124, 93, 123, 163, 98, 142)	-
8	225	6,435	(183, 133, 134, 124, 93, 123, 163, 98), (183, 134, 124, 93, 123, 163, 98, 142), (183, 124, 93, 123, 163, 98, 142, 116)	116
9	235	5,005	(183, 133, 134, 124, 93, 123, 163, 98, 142), (183, 134, 124, 93, 123, 163, 98, 142, 116)	-
10	51	3,003	(183, 133, 134, 124, 93, 123, 163, 98, 142, 149), (183, 133, 134, 124, 93, 123, 163, 98, 142, 45), (183, 133, 134, 124, 93, 123, 163, 98, 142, 116)	45
total	2,502	30,997		

* Number of retrofit combinations which are generated by each method

** (1) elite-set updating method; (2) all-candidates method

4.4. Summary

Using the set of critical cascading failure scenarios, cost-effective retrofit combinations are identified using the ‘elite set updating’ method. Elite set updating method effectively reduces the computational cost optimization by selecting candidates and elite components among all network components. When compared to the ‘all-candidate’ method, which enumerates all possible combinations using candidate components, the ‘elite set updating’ method requires less than 10% computational cost while both methods deliver the identical cost-benefit curve results. Therefore, the ‘elite set updating’ method is expected to support the disaster risk mitigation planning of the power grid by providing cost-effective retrofit combinations within the budget.

5. Clustering-based Assessment of Disaster Resilience of Communities

5.1. Introduction

Building a ‘disaster-resilient’ community is becoming increasingly crucial to disaster management authorities. Since community disaster resilience inherently has the multidimensional aspects, it has been studied in the fields of engineering, sociology, phycology, economics, and policy.

Especially in the field of engineering, the concept of the ‘disaster resilience triangle’ is widely adopted, and various methods have been developed for modeling the initial post-disaster conditions and recovery trajectory. For example, post-disaster housing recovery models, which incorporate not only initial structural damage state but also socio-economic variables, were developed. However, because such recovery models are developed for the specific community under a particular disaster scenario, applicability to the other communities remains a challenge.

On the other hand, in the field of social science, several framework and models have been developed to identify the sub-community which may be suffering more from the lack of disaster resilience than other communities. For example, the resilience score of the communities was estimated using the comparable disaster resilience index, which could be included in the resilience category such as social, economic, institutional, community capital, infrastructural, and ecological. While these approaches allow for understanding the overall disaster resilience capacity of the community, the value of the disaster resilience score may not represent the condition of the community under specific disaster scenarios since the purpose of

those methods focuses on the general disaster resilience of the community rather than certain disaster scenario.

Therefore, for the community which does not have the comprehensive recovery model yet, community disaster resilience clustering (CDRC) method is proposed to facilitate a comparative assessment of disaster resilience of community under specific disaster scenario. The framework of existing regional structural damage loss estimation and the community disaster resilience index are adopted to measure the structural damage and socio-economic recoverability. After a brief review of existing engineering and social science-based approaches and the proposal of the clustering method, three methods are compared through a community example. The required input data, analysis and results are presented for each method for the comparative study.

5.2. Literature Review: Community Disaster Resilience Assessment

A variety of frameworks has been proposed to assess community disaster resilience. However, a community disaster resilience assessment framework that captures all aspects (e.g. engineering, environmental, social, economic, etc.) of the disaster resilience is yet developed. While more investigation is needed regarding the overall mechanism of disaster resilience, each framework has its own theoretical background and advantages. This section provides a brief review and summary of community resilience assessment frameworks that are built on two different theoretical backgrounds, i.e. engineering and social science.

5.2.2. Engineering Approaches

In the engineering field, to quantify the community disaster resilience, the so-called 'resilience triangle' concept was proposed by Bruneau *et al*. (2003). In this quantitative assessment framework for seismic resilience of community, earthquake loss of resilience, *R*, is quantified by the expected degradation of the quality of the system over time (Figure 5.1). Numerically, the loss of resilience, *R* could be expressed as follow:

$$R = \int_{t_o}^{t_1} [100 - Q(t)]\, dt \qquad (5\text{-}1)$$

where *Q(t)* denotes a quality of the infrastructure of a community; and t_0 and t_1 are respectively the time points when an earthquake occurs, and infrastructure is completely repaired. The major advantage of this approach is its general applicability (Hosseini *et al*., 2016). The resilience triangle approach has the potential to be applied to all scales of resilience from individual structures and up to community-level.

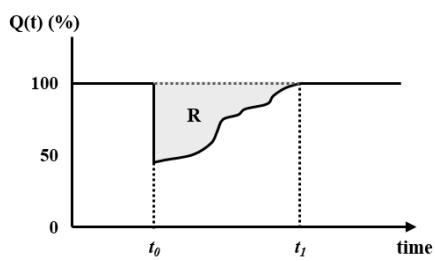


Figure 5.1 Measure of seismic resilience using resilience triangle (adapted from Bruneau *et al*. (2003))

Despite the potential of the ‘disaster resilience triangle’ method, it could not the applied in ‘community-level’ at the time when it proposed, due to the absence of the recovery model which can consider both physical and socioeconomic attributes of the community. Besides, Chang and Shinozuka (2004) proposed a method to evaluate the technical, organizational, social, and economic dimensions of community disaster resilience alternatively using the system failure probabilities at the initial post-disaster stage. Although the socioeconomic dimensions of disaster resilience are examined, social and economic variables are yet introduced in the frameworks, and those resilience measures are evaluated from the system failure probability of the lifeline network, which only considers the condition of the system at initial post-disaster state.

The post-disaster recovery models of housing and infrastructure inventories are necessary for modeling the post-disaster processes. Such recovery models were first developed with consideration of the type of structure and its initial damage states only. For example, building a recovery time table in the HAZUS-MH earthquake technical manual considers the structural damage state and the occupancy classes of the buildings. The influence of the socioeconomic aspects of the community was yet introduced in the models.

However, recent development of the recovery model, which considers the socioeconomic status of the community, facilitates the application of a resilience triangle approach to the community-level. Lin and Wang (2017) proposed a simulation-based building portfolio recovery model (BPRM) to estimate the recovery trajectory of the functionality overtime and recovery trajectory of a community building portfolio. In this method, economic variables reflected using

four different types of finance-related waiting time (e.g. insurance, private loan, SBA-backed loan, and not covered).

Later, housing recovery models were further investigated with the relationship between the various resilience indicators using longitudinal collection of recovery data. Using empirical data from the city of Napa, Despotaki *et al*. (2018) attempted to link the community resilience indicator and the recovery process. In this research, initial damage state, age of building construction, homeownership, employment, and etc. are relevant variables of the housing recovery. However, application of proposed model to the other community with different socioeconomic status yet remains challenging because the model was fitted to the case-specific data set. On the other hand, housing recovery model which consider physical, economic, and social condition were proposed by Sutley *et al.* (2019) using the empirical data collected overtime after a flood disaster, recovery-based fragility functions were built. The investigation in the relationship between the recovery trajectory and the socioeconomic variables showed that these variables can be strong predictors of the recovery process.

By the recent development of the comprehensive recovery models, modeling the condition of the community at the initial post-disaster stage and following the recovery process are theoretically possible. However, recent studies have only dealt with the situation where a hypothetical community with full data sets and matrix or recovery model fitted with the longitudinally collected recovery data. Whereas the proposed study focuses on the situation where the recovery model is yet developed while the building inventory and census data are publically available. There is a considerable gap between the ideal framework and the data sets available for many communities.

To overcome this practical challenge, the clustering-based disaster resilience assessment, which does not require the data-driven model that links the recovery process with the socioeconomic variables, is proposed in this chapter. As an alternative to the comprehensive recovery model, socioeconomic recoverability is measured to understand the recovery process of the community.

5.2.3. Social Science Approaches

Community disaster resilience is also extensively addressed in the field of social science. The most distinct features between the engineering field and the social science can be found in terms of the variables and methods. Social science-based disaster resilience assessments at the community-level are mainly semi-quantitative assessments, which selects community disaster resilience index based on the literature review, experts' opinions, analytic hierarchy process (AHP), to measure the resilience of the places.

Cutter *et al*. (2008) proposed a place-based model for understanding community resilience to a natural disaster. The disaster resilience of place (DROP) model is designed to prompt comparative assessment of disaster resilience at local or community level with the multiple variables in six dimensions including ecological, social, economic, institutional, infrastructure, and community competence. As a first step implementation, the method is yet proposed in the conceptual level. Later, this framework was applied to the US (Cutter *et al*., 2010 and 2014) to enable decision-makers to compare the disaster resilience of the inter- and intra-metro area.

Besides, community-specific models that use a weighted average of disaster resilience indexes were also developed. For example, Yoon *et al*. (2014) developed a method that constructs a set of indicators measuring Community Disaster Resilience Index (CDRI) in terms of five humans, social, economic, environmental, and institutional factors. By examining of the relationship between the aggregated CDRI and disaster losses (economic damage and human loss from 2001 and 2010 in South Korea), using an ordinary least squares (OLS) regression method and a geographically weighted regression (GWR) method, the disaster resilience of the communities was assessed.

Despite the recent developments of the various community disaster resilience assessment frameworks in the social science approach, the relationship between the socioeconomic variables and the infrastructure system is beyond the scope of those methods and remain neglected. Moreover, the physical condition of the building and the infrastructure systems at post-disaster state are ambiguously considered using proxy measures such as the percentage of certain housing types, built year, etc.

In addition, most of disaster resilience index for the purpose of evaluating the general distaste resilience rather than case-specific disaster resilience. Considering the fact that a disaster risk mitigation plan cannot be not optimal to every type of the disaster, some disaster resilience index may be not adequate for a certain type of disaster.

Therefore, to provide the community disaster resilience with a better measure for hazard and damage of the building system, without requiring the complex interdependency model between the physical and socioeconomic variables, the community disaster resilience clustering (CDRC) method is proposed in the following section. The CDRC separately evaluates the infrastructural resilience

dimension and other resilience dimensions to overcome the lack of a recovery model. By adapting the conventional structural damage analysis within the community, the physical condition of the structural systems after a specific disaster could be understood with better resolution.

5.3. Community Disaster Resilience Clustering (CDRC) Method

Disaster resilience of community is multi-dimensional capacity, which could be estimated by the initial post-disaster condition and the recovery trajectory until community recovers its original functionality. In this section, to quantitatively estimate the resilience properties while overcoming the absence of recovery data and model, the CDRC method is proposed. The CDRC method categorizes the communities with similar resilience characteristics (e.g. structural damage at the initial post-disaster stage and socio-economic recoverability measure) instead of quantifying disaster resilience in certain unit. In the CDRC methods, two resilience measures are quantified separately and used for clustering the communities in terms of disaster resilience. By separately evaluatingthese two measures, each measure of resilience can adopt well-defined methods in each field. Moreover, correlation models between physical and socio-economic elements are not required while the effects of both sides on disaster resilience are reflected using clustering analysis. In Figure 5.2, the concept of the CDRC method is illustrated. In the following sub-sections, structural damage estimation, socio-economic recoverability quantification, and disaster resilience clustering steps of CDRC are addressed, respectively. By adopting structural damage or loss estimation models for various types of disasters, the CDRC method can be applied to diverse disaster scenarios. As one of the

examples, CDRC based resilience analysis is performed for an earthquake disaster scenario.

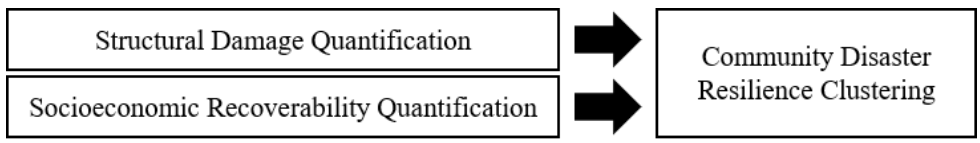


Figure 5.2 Conceptual framework of CDRC method

5.3.1. Quantifying Structural Damage

As a measure of structural damage, average 'mean expected structural damage loss' within the community is being used. The flowchart of structural damage quantification is illustrated in Figure 5.3.

First, a data inventory is required for structural damage estimation. However, often building data sets which are adequate for the structural damage analysis are not available at the community-level. Therefore, data selection, cleansing and transforming steps are required before further evaluation. For example, communities examined in chapter 5.4.2, 'national GIS building standard data set' (Open data portal, South Korea) is selected as baseline data of building inventory. At the data cleaning step after data selection, building data with false information (e.g. have 50 below ground level, have construction permit beyond current date, etc.) and incompatible code are removed from the initial data set. As the last step of data processing, data information is transformed to the format which could be later used in fragility mapping.

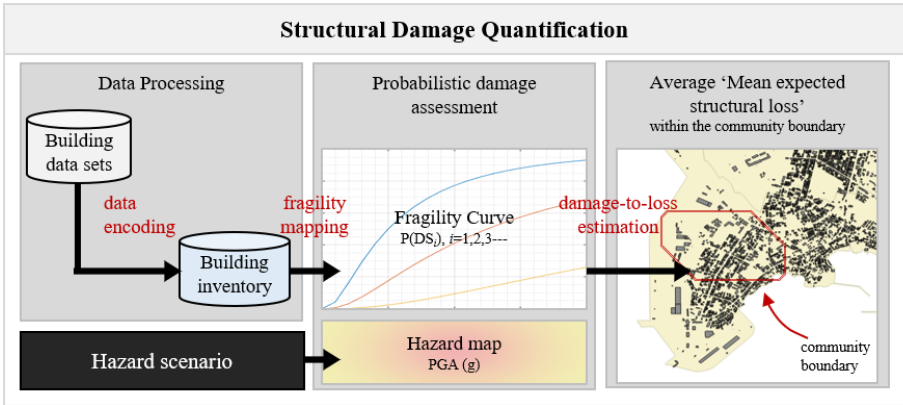


Figure 5.3 Flowchart of structural damage estimation

After, using fragility mapping, each building is matched with the fragility model. The probability of the structural damage states (e.g. slight, moderate, extensive, and complete) of each building is determined using the PGA (g) value at building location and matched with fragility model.

Once the probability of each damage state is evaluated the mean structural damage state of each building is estimated using the damage-to-loss correlation (Bai *et al.*, 2007). Numerically, the 'mean expected structural damage loss' of each building could be expressed as follows:

$$\mathrm{E} = \sum_{i=1}^{4} \left(P(DS = DS_i) \times \mu_{DF_i} \right) \tag{5-2}$$

where $P(*)$ denotes the probability of occurrence of * and DS is the damage state (i=1,2,3,4 corresponds to slight, moderate, extensive, and complete damage state respectively). In addition, μ_{DF_i} indicates the mean damage factor for the damage state i (Table 5.1). As the last step of structural damage assessment, the average

'mean expected structural damage loss' of building structures within the community boundary is evaluated using QGIS, an open-source GIS application.

Table 5.1 Damage states and factor (Mid-America Earthquake Center report, September, 2008)

Damage State	Damage Factor
Insignificant	0.005
Moderate	0.155
Heavy	0.550
Complete	0.900

The value of the average 'mean expected structural damage loss' indicates the average structural damage condition of the community at initial post-disaster stage. The values 0 and 1 represent the 'none' and 'total' structural damage state. Therefore, communities with higher structural damage have lower level of disaster resilience.

5.3.2. Quantifying Socioeconomic Recoverability

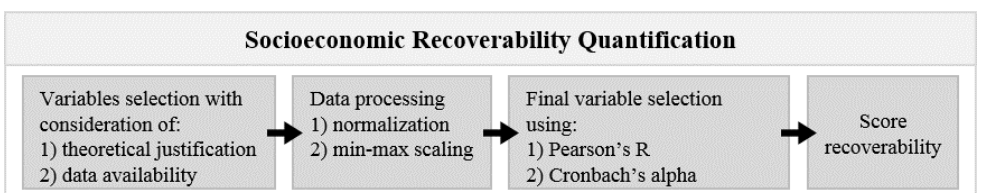


Figure 5.4 Flowchart of socioeconomic recoverability quantification

As a measure of socioeconomic recoverability, baseline resilience index for community (BRIC) is adopted (Cutter, Burton, and Emrich, 2010; Cutter, Ash, and Emrich, 2014). Besides the BRIC approach, recovery modeling and recovery regression were not applicable since the recovery data is absent for most of the

communities. The flowchart of socioeconomic recoverability quantification is illustrated in Figure 5.4.

Selecting community disaster recoverability index is the most important and challenging part of the socioeconomic recoverability measure assessment. In a community-level assessment, a great amount of data is required. Yet often publically available data sets are limited and not in the data format which could be directly used in the analysis. Therefore, while selecting a community disaster resilience index, both 1) data availability and 2) justification based on literature were considered. During recoverability index selection, variables of social, economic, institutional, community capital, and infrastructural dimensions are considered. Besides, in infrastructural dimension, the variables that are related and used in the structural damage estimation (e.g. housing age, housing type, etc.) are intentionally excluded.

If both criteria of the variable selection are satisfied, the value of collected variables are transformed into scale which could be compared to each other (e.g. percentage, per capita, etc.). In addition, the transformed values are further scaled using min-max rescaling. This procedure is performed to set all variables in a similar scale. By subtracting the minimum value of the variable and divided by the range of variable (difference between maximum and minimum value), best value and worst value of each variable are set as 0 and 1 respectively.

Before the final selection on recoverability index selection, correlations between the variables and internal consistency are considered using Pearson's R and Cronbach's alpha respectively. The variables with high correlation (e.g. Pearson's $R>0.70$) are excluded for further evaluation. In addition, to achieve the internal consistency and acceptable level of reliability among the recovery variables, several variables are excluded. Table 5.2 is an example of final recoverability index for

Ulsan South Korea. The detail of these indicator selection process are further discussed in case study (Section 5.4.2.2).

As the last step, to equally weight each category of the recoverability, the average value of the variables within the same category are used. The final recoverability score is evaluated as the average of multiple-categories of recoverability. The value of the average 'disaster recoverability score' indicates the relative socio-economic recoverability of the community. Since those values are 'relative' value among the subject communities, it is not valid besides the subject group. The values 0 and 1 represent the weakest and strongest relative recoverability within communities. Therefore, communities with higher socio-economic recoverability have higher disaster resilience.

Table 5.2 Disaster recoverability indicator sets of Ulsan, South Korea

Category	Index
Social	Population under 65 (%)
Social	Non-foreign population (%)
Social	Number of doctor per 1000
Economic	Homeownership (%)
Economic	Female labor participation (%)
Economic	Per capital income
Economic	Labor employed by government (%)
Community Capital	Voter participation (%)
Community Capital	Number of religious, community organization per population
Institutional	Population stability (%)

5.3.3. Clustering Community Disaster Resilience

As the last step of CDRC, communities are clustered in terms of structural damage and socioeconomic recoverability which are evaluated in previous steps of CDRC. As illustrated in the figure, communities with similar resilience characteristics are clustered (Figure 5.5). By these categorizations, spatially distributed disaster resilience patterns of the community could be examined and the communities in the same resilience cluster could require similar disaster resilience planning.

To cluster the communities with similar disaster resilience properties, *k*-mean clustering method is applied. First, using the elbow-method, the number of clusters is chosen. Second, the disaster resilience cluster with the number, which exceeds the number chosen by the elbow method, is investigated. Third, by clustering each disaster resilience criteria separately, disaster resilience clusters determined by the threshold of each criterion is examined. For those cases which select 3 group for each measure as an example, each cluster of disaster resilience is assigned one of the low-moderate-high rank for each disaster resilience measure. Since each cluster could be later incorporated with the disaster resilience planning, both distribution of the community data set and the following decision process should be considered comprehensively in the clustering stage.

Furthermore, by grouping the disaster-resilience-alike communities, a group of communities which may be suffering from the disaster longer than other community due to lack of disaster resilience could be identified. For instance, the red cluster in Figure 5.5 is expected to require a longer recovery period than other communities due to high structural damage measure and relatively low recoverability.

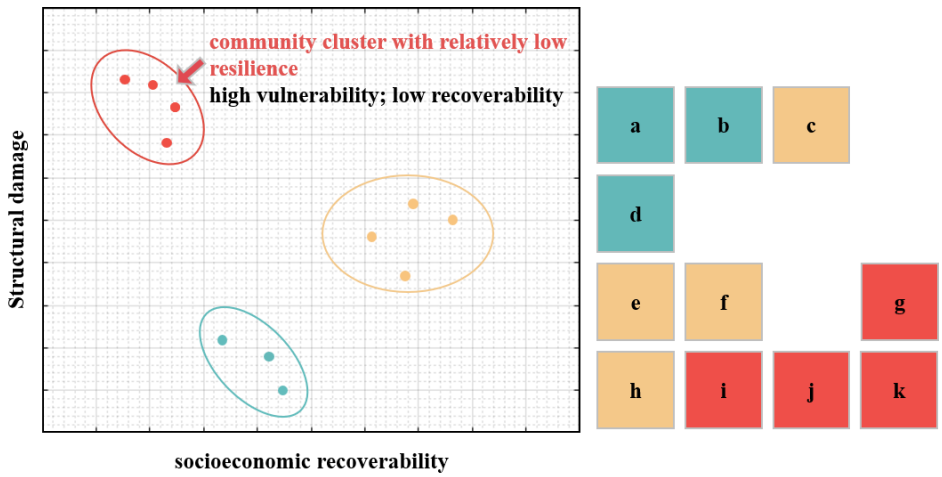


Figure 5.5 Expected outcome of CDRC framework – disaster resilience cluster in resilience measure space and spatial space

5.4. Case Studies

In this section, the applicability of the proposed CDRC method is demonstrated by two community examples; a virtual community, and real community - Ulsan, South Korea.

5.4.1. Virtual Community Example

In this section, a virtual community is examined to compare three community disaster resilience assessment methods: 1) building portfolio loss estimation (BPLE) and building portfolio recovery model (BPRM) methods (Lin and Wang, 2017a, 2017b); 2) baseline disaster resilience for community (BRIC) method (Cutter *et al.*,2010 and 2014); and 3) the proposed CDRC method. Three methods are compared in terms of required input data, analysis, and results.

The presented virtual community is assumed to have a full data set that are required for all three methods. The virtual community in size of approximately 15km by 15km. Also, the scenario earthquake is assumed to have the Moment magnitude 7.0 and epicenter located approximately 15km southeast of the virtual community. As illustrated in Figure 5.6, the community includes 8 zones that show different distributions of building type and financing states. The detailed description of building type and financing resource of each zone are summarized in Table 5.3. Resilience of sub categorizes and the assumed resilience score are given in Table 5.4.

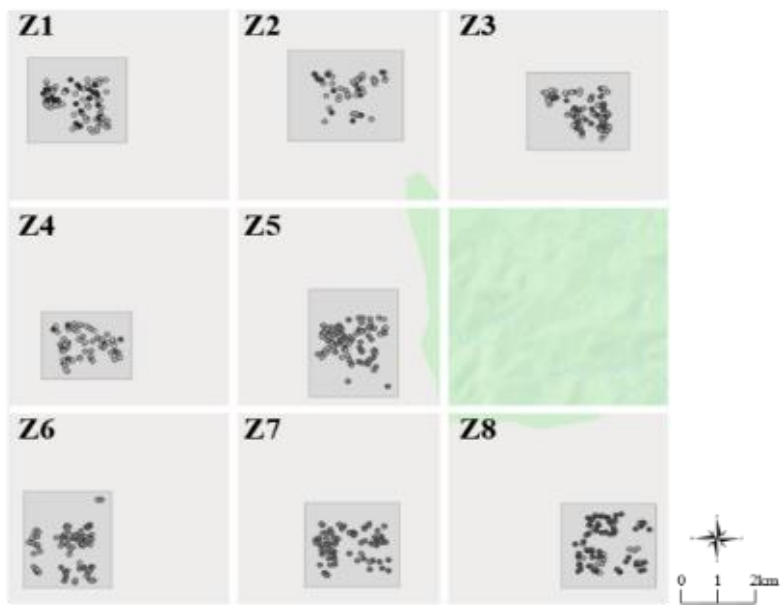


Figure 5.6 Zone and the buildings of the virtual community

Table 5.3 Household characteristics and financing resource for building restoration of residential zones (Zone 1-8)

Zone ID		Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8
Number of Household		93	48	96	71	103	98	82	80
Type and built Year of buildings (%)	W1*(1970)	50	50	50	30	-	40	100	100
	W1*(1985)	50	50	25	30	50	30	-	-
	W2**(2000)	-	-	25	40	50	30	-	-
Financing Resources (%)	Insurance	25	20	30	50	60	40	5	5
	Private Loan	15	20	35	40	35	40	10	-
	SBA-backed	30	30	15	10	5	10	5	5
	Other	30	30	20	-	-	10	50	90

Table 5.4 Resilience score of residential zones (Zone 1-8)

Zone	Social	Economic	Infrastructural	Community Capital	Institutional	Environment	Resilience score
1	0.495	0.468	0.198	0.187	0.458	0.388	0.366
2	0.477	0.457	0.151	0.151	0.365	0.021	0.271
3	0.780	0.538	0.653	0.339	0.623	0.495	0.571
4	0.747	0.641	0.719	0.451	0.525	0.824	0.651
5	0.853	0.643	0.675	0.311	0.576	0.572	0.605
6	0.786	0.622	0.726	0.451	0.529	0.495	0.601
7	0.482	0.384	0.286	0.251	0.380	0.477	0.377
8	0.606	0.382	0.295	0.135	0.316	0.422	0.359

5.4.1.1. Comparison of input data

The input data required to perform each method is summarized in Table 5.5. As described in the table, both BPLE and CDRC methods require the hazard scenario

data and building inventory for structural damage estimation at initial post-disaster state. However, while CDRC uses socioeconomic data set to measure the recoverability, BPRM models recovery trajectory. Therefore, for recovery trajectory modeling, the community-specific recovery data sets (e.g. delay time model, repair time model, and restoration financing resource statics) are required. For the community which has longitudinal collection of housing recovery data in community-scale, BPRM method could be applicable and will deliver trajectory of building portfolio functionality improvement over time. Despite the usefulness of BPRM method, however, for the most of community those data sets are not available.

Besides, the BRIC method and the proposed CDRC method require the data of community-level resilience variables in multiple-dimensions such as social, economic, institutional, and community capital. Although collecting these data sets are challenging due to limited data set adequate for community-scale analysis, both BRIC and the proposed method have flexibility in choosing variables. Due to this flexibility in variable selection, the applicability of the BRIC and CDRC methods increase. Although CDRC is adopting the BRIC method to measure the recoverability measure of the community, it is important to note that the CDRC method evaluates the 'recoverability' score while the BRIC method deals with the 'resilience' score. Since the variables of infrastructural and ecological resilience measure are related to the structural damage estimation, those input data sets are not required as input data in the socio-economic recoverability quantification.

Table 5.5 Comparison of input data set require for community disaster resilience assessment frameworks

Method	Input Data
BPLE BPRE	Hazard scenario: type, intensity, location
	Building inventory: structure type, built year, # of floor, occupancy type, area, location, etc.
	Recovery data: delay time, repair time, and restoration financing resource statistic
BRIC	Census data: variables to construct disaster resilience index (e.g. social, economic, institutional, etc.)
CDRC	Hazard scenario: type, intensity, location
	Building inventory: structure type, built year, # of floor, occupancy type, area, location, etc.
	Census data: variables to construct disaster recoverability index (e.g. social, economic, institutional, etc.)

5.4.1.2. Comparison of analysis

Comparison of analysis of the BPLE-BPRM, BRIC, and the proposed CDRC method is illustrated in Figure 5.7. As illustrated in the figure, BPLE and CRDC share the modeling procedure of structural damage estimation at the initial post-disaster stage. However, while BPLE further models the recovery trajectory after the disaster stage using discrete state, continuous time Markov Chain (CTMC), CDRC evaluates the recoverability measure using a resilience index.

Instead of performing recovery modeling using the recovery data set, CDRC adopts the social science methods to estimate the recoverability. Therefore, the BRIC method and socio-economic recoverability quantification process of CDRC are similar to each other. In terms of purpose, however, BRIC aims to measure the general disaster resilience score while CDRC measures socio-economic recoverability against the specific disaster scenario. Therefore, several resilience

variables (e.g. housing and hazard related variables) considered in the BRIC model are excluded in measuring the recoverability process of the CDRC method.

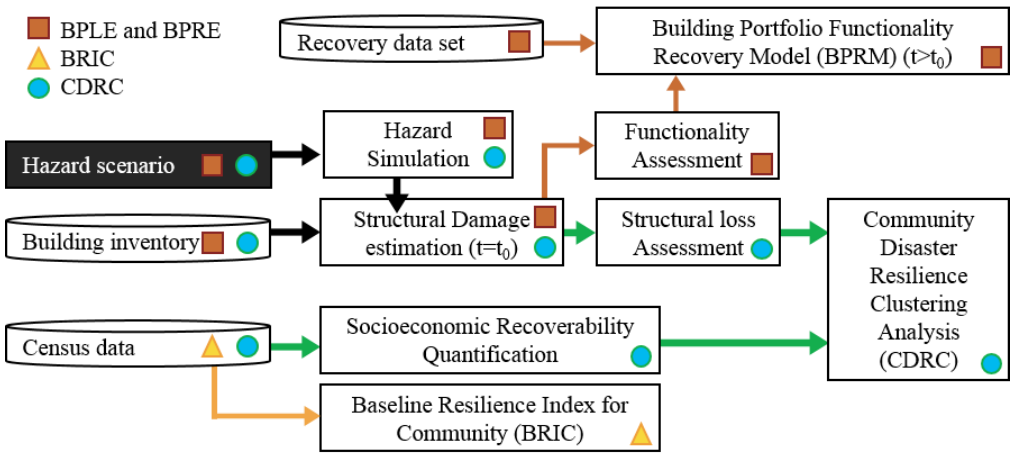


Figure 5.7 Comparison of analysis used in BPLE-BPRM, BRIC, and CDRC

5.4.1.3. Comparison of results

In this section, the results of the resilience assessment using three methods are compared with each other. First, the building portfolio functionality recovery trajectory is achieved using BPLE and BPRM (Figure 5.8). By modeling the recovery process of each building CTMC, the percentage of the building portfolio, which reaches the full function state, is measured for each zone. Secondly, using the BRIC model, the resilience score of each zone is evaluated (Figure 5.9). It is shown in the figure that Zone 3-6 have a higher resilience score than others. Lastly, the disaster resilience cluster of community and its spatial distribution are identified using the CDRC method (Figures 5.10-11).

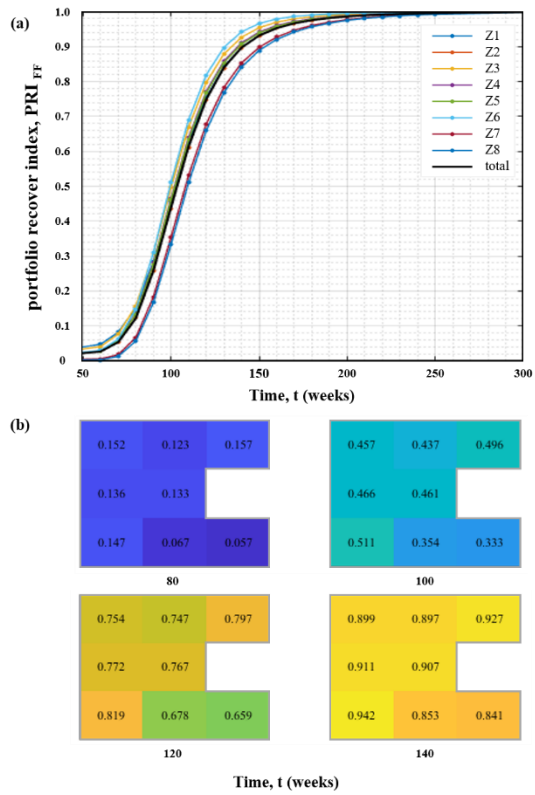


Figure 5.8 An illustration of (a) mean recovery trajectory of 8 zones over time and (b) at week 80, 100, 120, and 140

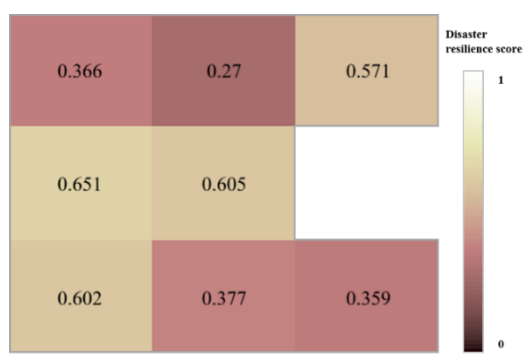


Figure 5.9 Disaster resilience score of virtual community by BRIC

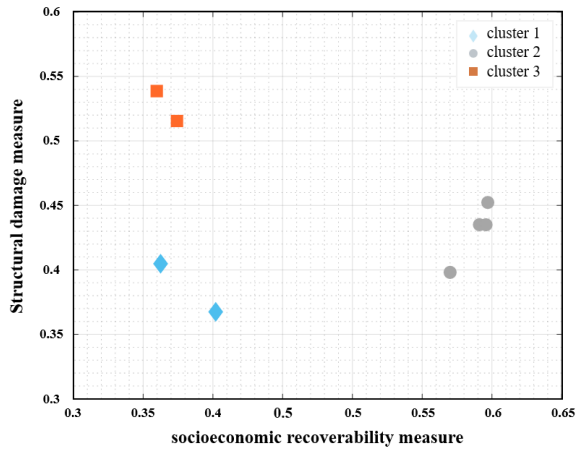


Figure 5.10 Disaster resilience cluster of 8 zones by CDRC

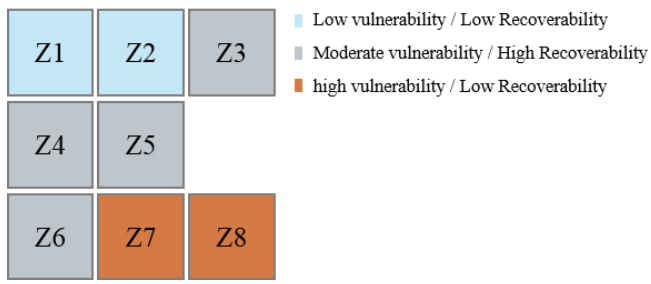


Figure 5.11 Spatial distribution of disaster resilience pattern of 8 zones by CDRC

Since the virtual community is intentionally designed to have different resilience characteristics over the region, three clusters with clear distinct are identified in the figures (Figures 5.10-11). Figure 5.11 shows that Zone 7-8, colored in orange, have high structural damage measures and relatively low recoverability when compared to the other zones. Due to its resilience characteristics, Zone 7-8 are expected to experience a longer recovery process. This lack of resilience is also observed in the results of BPLE and BPRM. In Figure 5.8 (b) at week 120, the difference between Zones 7-8 with other residential areas is notable. Furthermore, the difference between Zones 1-2 and Zones 3-6, which is identified as different

resilience cluster using CDRC, are shown in the results of the BPLE and BPRM. (Figure 5.8). Although initial damage is more significant in Zones 3-6 than Zones 1-2, less recovery time is needed to regain the full-function state because of better financial states.

Besides, the results from BRIC and the resilience cluster identified using CDCR are slightly different from each other. For instance, in the results of BRIC analysis, Zones 1-2 and Zones 7-8 have little difference in the resilience score while CDRC categorizes those zones in different disaster resilience clusters. This gap is coming from the structural damage dimension used in CDRC. While the resilience score is measured for disaster in general, CDRC is focused on a specific disaster scenario. Considering the fact that general resilience planning may not be adequate to a certain disaster scenario, CDRC has strength in categorizing the communities under a specific disaster situation.

5.4.2. Communities in Ulsan, South Korea

5.4.2.1. Problem setting and data sources

In this section, applicability of CDRC to real community example is demonstrated with the communities of Ulsan, South Korea. The 56 zones, smallest administrative units (Eup, Myeon, and Dong, EMD), and the buildings of Ulsan communities are illustrated in Figure 5.12. To perform CDRC, the scenario event was chosen to have a moment magnitude of 6.5 and an epicenter located at East Sea which approximately 10km apart from Dong-Gu, Ulsan, South Korea. The building inventory, total 101,037 buildings, are constructed from ‘national GIS building standard data set’ (Open Data Portal, South Korea). Lastly, socioeconomic recovery

variable data sets (Table 5.2) are collected from multiple open sources (e.g. statically Geographic Information Service (SGIS), South Korea; Statistics Korea (KOSTAT), South Korea; Open Data Portal, South Korea).

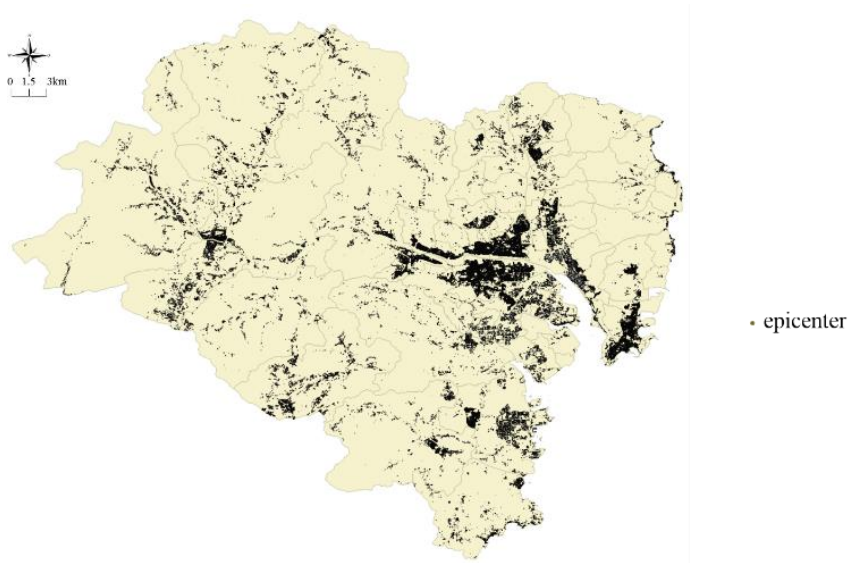


Figure 5.12 Zone and the buildings of the Ulsan, South Korea

5.4.2.2. Results and Discussions

Structural Damage Quantification

The structural damage measure is evaluated using the average 'mean expected structural losses' of buildings within each community boundary (Figure 5.13). It is shown in the figure that the physical vulnerabilities of communities are generally linear to the distance from the epicenter. However, some of the communities (colored in red in Figure 5.13) located in central Ulsan show the highest structural damage measure under the given earthquake scenario due to its soft ground condition, which is prone to the earthquake hazard.

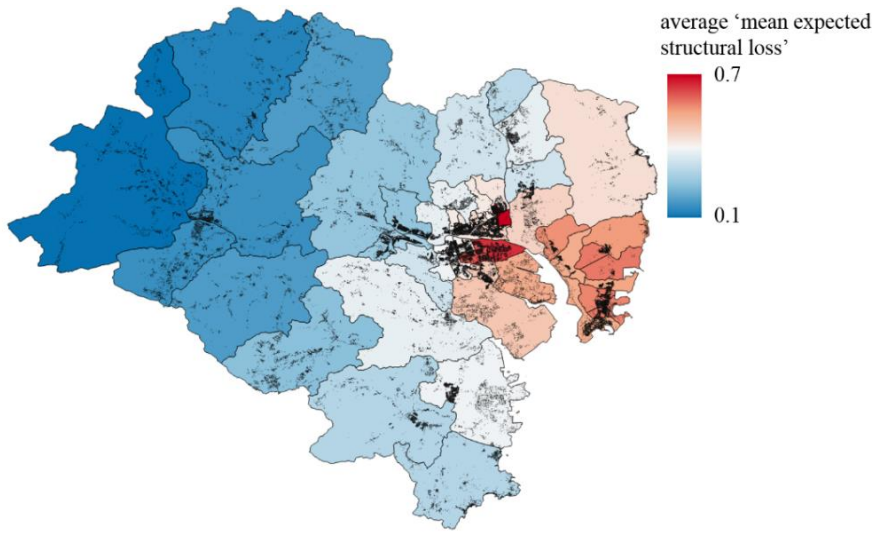


Figure 5.13 Structural damage measures of Ulsan communities

Socio-economic Recoverability Quantification

For the socioeconomic recoverability, the recoverability score is measured using baseline indicators. Since the index values required the statistics of 56 zones, the limited number of variables are used for comparative analysis. In the first variable selection, data sets of 15 variables were collected.

For the variables which are selected by validity and availability, the correlation coefficients between the variables are evaluated. As a result, three sets of variables (i.e. population under 65 and population have high school education; employment rate and per capita income; non-vulnerable social group and Number of religious, community organization per population) have a correlation coefficient which is larger than 0.8. Also, to ensure internal consistency, modified Cronbach's alpha values, α^*, of all possible variable combinations are evaluated.

$$\alpha^* = \alpha \times x \tag{5-3}$$

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} var(X_i)}{var(\sum_{i=1}^{n} X_i)}\right) \tag{5-4}$$

$$x = \begin{cases} 1 \text{ if all variables have a correlation coefficient lower than 0.6} \\ 0 \text{ if variables with a correlation coefficient higher than 0.6 exist} \end{cases} \tag{5-5}$$

where α^* is modified Cronbach's alpha value of index set, $var(X_i)$ denotes variance of the ith index value and $var(\sum_{i=1}^{n} X_i)$ denotes the variance of sum of index values. As a result, five variables, i.e. the percentage of population with high school education, employment rate, the percentage of population in non-vulnerable social group, the percentage of population without special needs, and number of hospital bed per 1000 are excluded and 10 variables in 4 categories of recoverability (Table 5.2) are selected as final group of recoverability index. Those variables do not have correlation coefficients higher than 0.6 and as group the Cronbach's alpha value is higher than 0.6, which is an acceptable level of internal consistency (Figure 5.14).

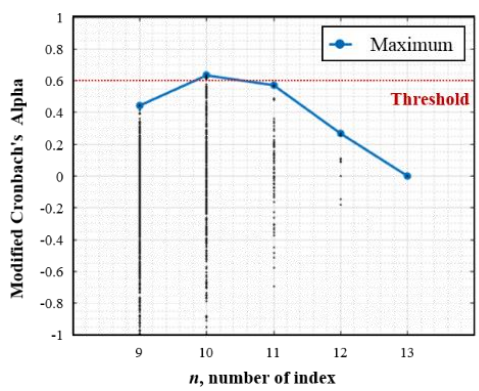


Figure 5.14 Modified Cronbach's alpha by number of index

First, three indicators in social recoverability (i.e. population under 65, non-foreign population, and number of doctor per 1000) are intended to represent the demographic properties of the community incorporated with comprehensive wellness which increase recoverability.

Population under 65. Effective response and recover against the disaster is often correlated with age factor. The elderly population tends to be less adaptive to a disastrous situation than others (Morrow *et al*, 2008; Peek *et al*, 2010; Cutter *et al*, 2014; Yoon *et al*, 2016).

Non-foreign population. At the disaster recovery stage, being a foreigner tends to be found a disadvantage due to a language barrier and being neglected in government support (Norris *et al.*, 2008; Sherrieb *et al.*, 2010; CARRI Report).

Number of doctors per 1000. The greater density of the doctor is correlated with higher health access which is the important recovery resource (Norris *et al.*, 2008; Chandra *et al.*, 2011).

Second, four indicators in economic recoverability (i.e. homeownership, per capital income, female labor participation, and labor employed by government) are used to represent the economic vitality and equality which could be advantage in recovery context.

Home-ownership and per capita income. Both indexes are relate with the general economic vitality. Home-ownership is often correlated with better economic conditions and a higher motivation for housing restoration. Also, it is reported in multiple research efforts that the poorest individual and communities are often suffering from the longest recovery period (Comerio *et al.*, 1997; Haveman *et al.*, 2005; Pendall *et al.*, 2012; Burton *et al.*, 2014; Despotaki *et al*., 2018, CARRI Report).

Female labor participation. It is the index which represents the income equality by gender. The quality in income equality is incorporated with a fair distribution of recovery resources which strengthen the disaster resilience of community (Enarson *et al*., 2012; Sherrieb *et al*., 2010).

Labor employed by government. The employment by the government is a beneficial factor correlated with garnering economic supports (Cutter *et al*.,2014). The experience in dealing with authorities is also helpful in the complex recovery process (CARRI Report).

Third, three indicators in the community capital category (i.e. voter participation and the number of religious, community organization per population) are representing the connection between the individual and the community. The citizens of the community with the strong community capita expected higher tendency to assist their neighborhoods.

Voter participation. The political engagement, which represents the level of community engagement, can be measured by voter participation (Peacock *et al*., 2010; Sherrieb *et al*., 2010).

The number of religious, community organization per population. The membership of those organizations provides the social connections which facilitate the disaster recovery process. On the other hand, the community with less social connections may be prone to receive such mutual supports (Sherrieb *et al*., 2010; CARRI Report).

Lastly, one indicator (i.e. population stability), which represents the policies and governance of the disaster recoverability, in the institutional category. Due to the

absence of the indicators such as ‘mitigation spending’ and ‘disaster aid experience’ in the EMD unit, ‘population stability’ is selected.

Population stability. The fluctuation of the population could be a burden to local authorities. Both decrease and increase of the population could have negative impact respectively. On the other hand, population stability could be the measure of the disaster preparedness (Sherrieb *et al*., 2010; USNRC, 2012; Cutter *et al*., 2014).

Although some of the variables are not directly indicating the recoverability of the community, theoretically valid, available, and quantitative indexes are selected with consideration of correlation and internal consistency. Therefore, using the recoverability score measured by those chosen composite indexes, the ‘broad brush’ of recoverability capacity of the communities could be estimated. As a result, Figure 5.15 shows that communities of central and east-side of Ulsan (colored in green in Figure 5.15) have relatively high recoverability scores than other communities.

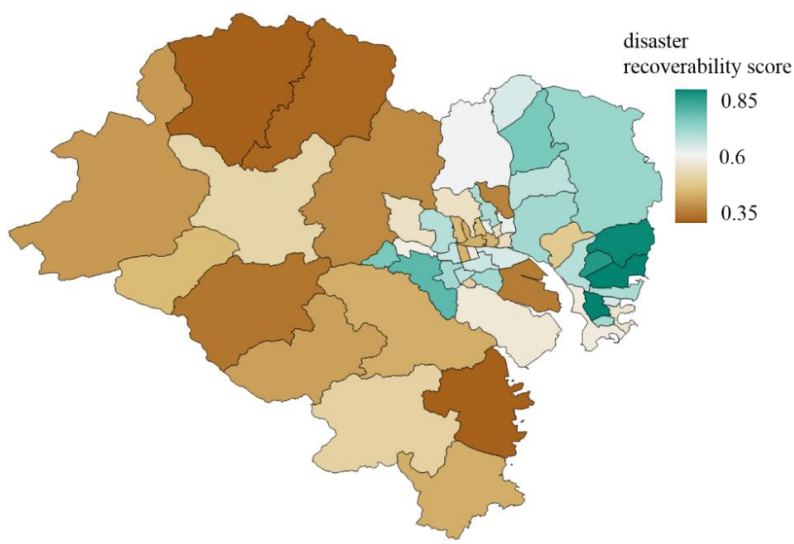


Figure 5.15 Socioeconomic recoverability measures of Ulsan communities

Disaster Resilience Clustering

Using the two disaster resilience measures, disaster resilience of the communities is investigated. For the Ulsan communities, disaster resilience measures are distributed in data space as illustrated in Figure 5.16. To choose the number of groups for the Ulsan communities, the mean 'sum of point-to-centroid' is evaluated for the number of the cluster from 1 to 12 are evaluated. As indicated in Figure 5.17, the Ulsan communities grouped into 5 clusters are expected to have a reasonable level of similarity, whose in-cluster points have a relatively short distance from the cluster centroid. With the number chosen by the elbow method, 5 different clusters in Ulsan are identified in Figure 5.18 (a) and (b). Besides, to investigate the groups identified by more than five cluster, communities clustered in 9 different groups are also illustrated in Figures 5.18 (c) and (d). Lastly, as a result of separate clustering by each measure, 9 disaster resilience clusters with low, moderate, and high physical vulnerability and socio-economic recoverability are identified in Figure 5.18 (e) and (f).

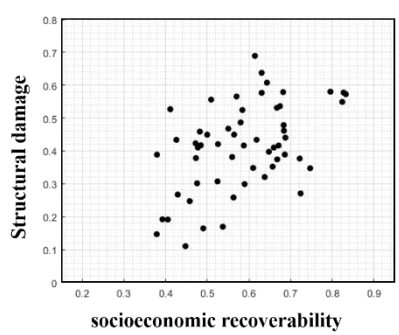


Figure 5.16 Disaster resilience distribution of Ulsan community

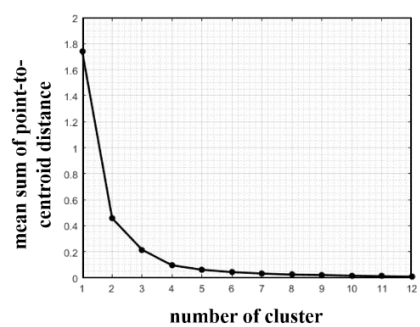


Figure 5.17 Results of elbow method

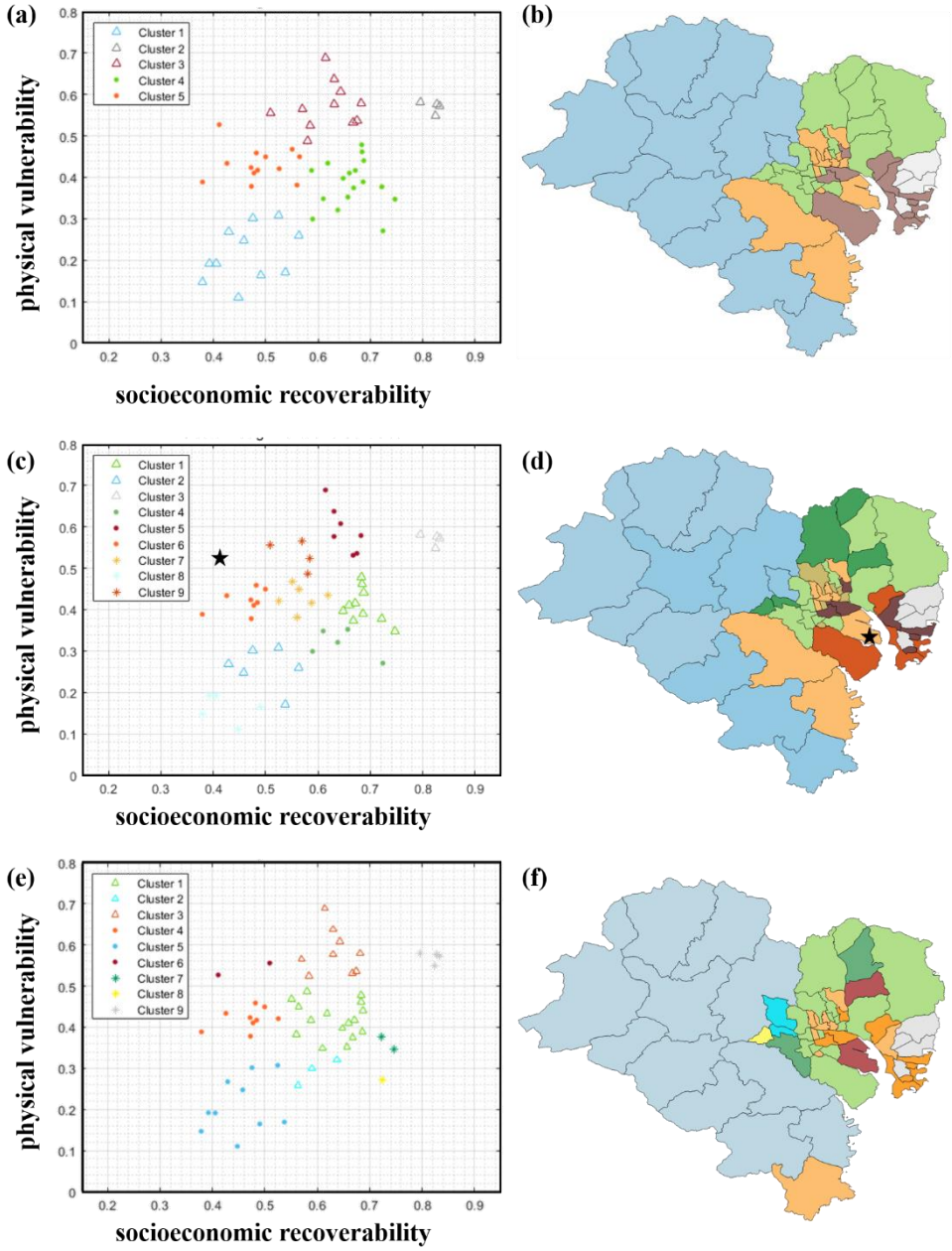


Figure 5.18 Disaster resilience clusters of Ulsan by CDRC in data and geographic space

Disaster Resilience Planning

It is noted from Figure 5.18 that communities of Ulsan have different resilience properties, which could be categorized into multiple groups and thus require different plans and approaches to improve disaster resilience for each cluster. Using the identified resilience clusters, appropriate natural disaster resilience plans that are adequate for each community could be developed.

For instance, one of the communities colored in light-orange color ('★' mark in Figure 5.18 (c) and (d)) may suffer with a longer period time of recovery due to its moderate to high vulnerability and yet low recoverability. Hence, retrofitting of buildings with higher seismic code is encouraged for about 660 buildings (33% of building set), which have the 'mean expected structural damage' value that exceeds the 0.6 (Figure 5.19). In terms of recoverability, contracting with disaster-insurance are encouraged to compensate for the relatively low economic recoverability of the community.

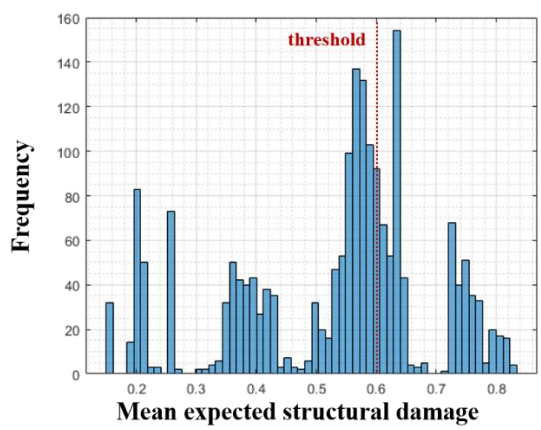


Figure 5.19 Histogram of 'mean expected structural damage' of community which marked as '★' in figure 5.18

5.5. Summary

In this chapter, the clustering-based community disaster resilience assessment framework termed CDRC is proposed. For the virtual community example, data, analysis, and results of the CDRC, BPLE-BPRM, and BRIC methods are compared with each other. It is demonstrated from the example that the CDRC method is useful for the communities which do not have a collection of the recovery data. To further illustrate the applicability of the CDRC, the proposed method is applied to the real communities of Ulsan, South Korea. Using publicly available data sets, CDRC effectively identifies the pattern of the disaster resilience of the communities. For the communities which yet have the sufficient recovery data collection, it is believed that the proposed method is providing a useful and practical analysis framework for assessing urban resilience to natural disasters despite the assumptions made in both structural damage and socioeconomic recoverability analysis for simplification of assessments. If more accurate data are available from reliable sources, the CDRC method is expected to provide a more accurate assessment of disaster resilience since the large-scale assessment relies on the data processed at smaller-scale analyses.

6. Conclusions and Future Research

6.1. Summary and Major Findings

The goal of this dissertation is to support the disaster risk mitigation planning of both the lifeline network and the community by developing the multi-criteria frameworks.

In order to identify critical post-disaster scenarios of the large-size lifeline networks, MG-NSGA is proposed. When compared with the NSGA-II method, the developed MG-NSGA method improves the robustness and optimality of the results. Especially, for the large-size lifeline networks, the performance gap between the original NSGSA-II and MG-NSGA-II algorithm significantly increases. Considering the fact that improvements of MG-NSGA are achieved without further increase in computational cost for evaluation, MG-NSGA enhances the applicability of MOGA to large-size network optimization problems. Also, the concept of 'critical zone' is presented. By proving the boundary with the sample space in terms of decision variables, the critical scenarios which may be a great concern for network management authorities are selected. By providing the selected set of scenarios that significantly threaten the network functionality with relatively small damage at the initial post-disaster state, the concept of 'critical zone' is expected to support the decision-makers' efforts to reduce the disaster risk of complex infrastructure networks.

To identify critical cascading failure scenarios of the power grid, the combination of OCM and MG-NSGA is presented. By introducing the flow-based overload cascading failure model without imposing constraints on the number of components that failed at the initial post-disaster condition, the cascading risk inherent in the power supply network is investigated. As a result, as the number of

components that failed at the initial stage increases, the negative impact at the final cascading stage gradually increases. By using the assumption that only a single component damaged at the initial post-disaster scenario, disaster risk could be underestimated greatly. Therefore, it is noticed by the results that critical cascading failure scenarios induced by the multiple components failure at the initial stage should be considered for disaster risk mitigation planning of the power grid.

Using the set of the identified critical cascading failures scenarios, optimal network retrofit combinations are identified with consideration of both cost and the benefit of the retrofit. Since the examining all possible retrofit combinations is computationally infeasible, 'elite set updating method' is proposed. When compared to the other method termed 'all-candidates' method, enumeration of candidates without considering elite-set, the proposed method significantly reduces the computational time without losing optimality of the final solutions. The proposed 'elite set updating' method is expected to support the decision-makers by suggesting the optimal retrofit combinations that enhance the robustness of power gird against the cascading failure within the budget.

Finally, to provide disaster resilience-driven decision support in a community-level, clustering-based community disaster resilience assessment framework CDRC is developed. Using the structural damage and socioeconomic recoverability measures quantified by 'mean expected structural losses' and average 'recoverability score', communities with similar disaster resilience capacity are clustered. By comparing CDRC with other engineering approaches and social science approach validity of the CDRC is also investigated. Unlike others, CDRC only requires publically available data sources and could be performed without sophisticated recovery models. Therefore, the proposed approach is expected to be applicable to a

wide range of communities for the purpose of understanding the distribution of different disaster resilience characteristics within the communities.

In summary, the developed approaches are expected to contribute to disaster risk mitigation planning by 1) searching critical post-disaster scenarios with better optimality and robustness, 2) providing critical cascading scenarios with flow-based cascading simulation, 3) identifying optimal retrofit within budget, and 4) categorizing communities with the similar disaster resilience characteristics.

6.2. Future Research Topics

The methods proposed in this dissertation yet have several limitations, and further improvements are desired to support the disaster risk mitigation planning in practice.

- Development of adaptive sample space dividing technique for MG-NSGA: In this dissertation, MG-NSGA divides the sample space into multiple groups with uniform space. However, not all groups are necessarily contributing to the improvement of the sample diversity and the quality of the final Pareto solutions during the evolution process. Therefore, the development of the sample space dividing technique which is adaptive to the problems and may further enhance the efficiency of the MG-NSGA.

- Investigation on a network performance measure, which facilitates the proposed MG-NSGA method to disaster risk mitigation planning: MG-NSGA could be performed with various network performance measures. Therefore, network system performance measures, which could represent

the post-disaster functionality of the lifeline network, should be further investigated.

- Development of sample re-using methods for the evolution process near the Pareto surface: In this thesis, cascading failure scenarios are identified using a combination of MG-NSGA and OCM. Near the Pareto surface, the computational cost of this analysis increased since samples generated near the Pareto surface are mostly critical cascading failure scenarios, which have a series of dependent failures that are modeled from repeated flow analysis. Because the sample population converges slowly near the Pareto surface, re-using the sample's fitness value generated at a previous generation might reduce the computational cost greatly.

- Investigation on other disaster risk mitigation plan against critical post-disaster scenarios: In this dissertation, optimal retrofit combinations are identified using the ‘elite-set updating’ method. Besides, other disaster risk mitigation strategies such as intentional islanding could be optimized using the identified critical post-disaster scenarios.

- Apply CDRC to other types of disaster scenarios such as floods and typhoons: the CDRC framework was applied to the community under earthquakes scenario. By introducing flood damage and loss model, various hazard risk of a community could be further investigated.

- Longitudinal recovery data collection: for most of the community, the recovery process at the post-disaster stage is yet collected systemically.

However, for better modeling and measuring of the recovery process, the recovery data in the regular time interval is necessary. Since the recovery process could be varied by communities and local authorities, systemic recovery data collecting is encouraged as a future research topic.

- Development of the recovery process estimation, which can consider the post-disaster condition of both lifeline network and community building set: In this dissertation, the interdependency of the lifeline network recovery process and the building recovery process is yet introduced. However, for the better estimation of the community recovery process, it is important to model the intertwined relationship between the lifeline network and buildings in the community.

Reference

Andersson, G., *et al*. "Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance." *IEEE transactions on Power Systems* 20.4 (2005): 1922-1928.

Alsac, O. and Stott, B. “Optimal Load Flow with Steady State Security" *IEEE Trans. on Power Apparatus and Systems*, 93.3 (1974), 745-751.

Arulselvan, A., Commander, C. W., Elefteriadou, L., Pardalos, P. M. "Detecting critical nodes in sparse graphs." *Computers & Operations Research* 36.7 (2009): 2193-2200.

Bompard, E. and Joint Research Center of the European Commission. "A Perspective Overview of Topological Approaches for Vulnerability." *networks* 66 (2019): 65102.

Bradley, A. P. "The use of the area under the ROC curve in the evaluation of machine learning algorithms." *Pattern recognition* 30.7 (1997): 1145-1159.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O’Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., von Winterfeldt, D. “A framework to quantitatively assess and enhance the seismic resilience of communities.” *Earthquake Spectra* 19.4 (2003): 733–752.

Cadini, F., Zio, E., Petrescu, C. A. "Optimal expansion of an existing electrical power transmission network by multi-objective genetic algorithms." *Reliability Engineering & System Safety* 95.3 (2010): 173-181.

Chang S.E., Shinozuka M., Moore J.E. "Probabilistic earthquake scenarios: extending risk analysis methodologies to spatially distributed systems." *Earthquake Spectra* 16.3 (2000): 557-572.

Chang, S. E., & Shinozuka, M "Measuring improvements in the disaster resilience of communities." *Earthquake spectra* 20.3 (2004): 739-755.

Choi, E. and Song, J. “Development of Multi-Group Non-dominated Sorting Genetic Algorithm for Identifying Critical Post-Disaster Scenarios of Lifeline Networks.” *International Journal of Disaster Risk Reduction* 41 (2019): 101299.

Christie, R. "118 bus power flow test case." (1993).

Cuadra, L., Salcedo-Sanz, S., Del Ser, J., Jiménez-Fernández, S., and Geem, Z. "A critical review of robustness in power grids using complex networks concepts." *Energies* 8.9 (2015): 9211-9265.

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. "A place-based model for understanding community resilience to natural disasters." *Global environmental change* 18.4 (2008): 598-606.

Cutter, S. L., Burton, C. G., & Emrich, C. T. "Disaster resilience indicators for benchmarking baseline conditions." *Journal of homeland security and emergency management* 7.1 (2010).

Cutter, S. L., Ash, K. D., & Emrich, C. T."The geographies of community disaster resilience." *Global environmental change* 29 (2014): 65-77.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. A. M. T. "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE transactions on evolutionary computation* 6.2 (2002): 182-197.

Despotaki, V., Sousa, L., & Burton, C. G "Using resilience indicators in the prediction of earthquake recovery." *Earthquake Spectra* 34.1 (2018): 265-282.

Fang, Y., Pedroni, N., and Zio, E. "Optimization of cascade-resilient electrical infrastructures and its validation by power flow modeling." *Risk Analysis* 35.4 (2015): 594-607.

Fang, Y. P., Pedroni, N., and Zio, E. "Comparing network-centric and power flow models for the optimal allocation of link capacities in a cascade-resilient power transmission network." *IEEE Systems Journal* 11.3 (2014): 1632-1643.

Fema, Hazus-MH. "Mr3 technical manual." Multi-hazard Loss Estimation Methodology Earthquake Model (2003).

Gen, M., Cheng, R., Lin, L. Network models and optimization: Multiobjective genetic algorithm approach. Springer Science & Business Media, 2008.

Grainger, J. J., and Stevenson Jr, W. D. "Power System Analysis, Mc GrawHill." New York (1994).

Guo, H., *et al*. "A critical review of cascading failure analysis and modeling of power system." *Renewable and Sustainable Energy Reviews* 80 (2017), 9-22.

Heinz Center. "Human links to coastal disasters." The H. John Heinz III Center for Science (2002).

Hines, P., Cotilla-Sanchez, E., and Blumsack, S. "Do topological models provide good information about electricity infrastructure vulnerability?." *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20.3 (2010): 033122.

Holmgren, Å. J. "Using graph models to analyze the vulnerability of electric power networks." Risk analysis 26.4 (2006): 955-969.

IEEE 118 bus system, Available: http://motor.ece.iit.edu/data/JEAS_IEEE118.doc

Jayaram, N., Baker, J. W. "Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment." *Earthquake Engineering & Structural Dynamics* 39.10 (2010): 1109-1131.

Koç, Y., Verma, T., Araujo, N. A., and Warnier, M. "Matcasc: A tool to analyse cascading line outages in power grids." *2013 IEEE International Workshop on Inteligent Energy Systems (IWIES)*. IEEE, 2013.

Koç, Y., Warnier, M., Kooij, R. E., and Brazier, F. M "A robustness metric for cascading failures by targeted attacks in power networks." *2013 10th IEEE International Conference on Networking, Sensing and Control (ICNSC)*. IEEE, 2013.

Koh, H. M., Kim, Y. H., and Park, W. S. "A Model for Seismic Reliability Assessment of Electric Power Transmission Network System." *Journal of Earthquake Engineering Society of Korea* 7.2 (2003): 93-102.

Korea Open Data Portal, Available: https://www. data. go. kr

Lin, Y. H., Tsai, K. T., Lin, M. D., Yang, M. D. “Design optimization of office building envelope configurations for energy conservation.” *Applied energy* 171 (2016): 336-346.

Lin, P., & Wang, N."Stochastic post-disaster functionality recovery of community building portfolios I: Modeling." *Structural Safety* 69 (2017): 96-105.

Lin, P., & Wang, N. "Stochastic post-disaster functionality recovery of community building portfolios II: Application." *Structural Safety* 69 (2017): 106-117.

Makarov, Y. V., Reshetov, V. I., Stroev, A., and Voropai, I. "Blackout prevention in the united states, Europe, and Russia." *Proceedings of the IEEE* 93.11 (2005): 1942-1955.

Morrow, B. H. Community resilience: A social justice perspective. Vol. 4. Oak Ridge, TN: *CARRI Research Report*, 2008.

National Research Council. Facing hazards and disasters: Understanding human dimensions. National Academies Press, 2006.

Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F., & Pfefferbaum, R. L. "Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness." *American journal of community psychology* 41.1-2 (2008): 127-150.

Ochoa, G. "Setting the mutation rate: Scope and limitations of the 1/L heuristic." Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computation. Morgan Kaufmann Publishers Inc., 2002.

Pagani, G. A., and Aiello, M. "The power grid as a complex network: a survey." Physica A: Statistical Mechanics and its Applications 392.11 (2013): 2688-2700.

Pahwa, S., Scoglio, C., and Scala, A. "Abruptness of cascade failures in power grids." *Scientific reports* 4 (2014): 3694.

Prasad, T. D., Park, N. S. "Multiobjective genetic algorithms for design of water distribution networks." *Journal of Water Resources Planning and Management* 130.1 (2004): 73-82.

Rocco, C. M., Barker, K., Moronta, J., Ramirez-Marquez, J. E. "Community detection and resilience in multi-source, multi-terminal networks." Proceedings of the Institution of Mechanical Engineers, Part O: *Journal of Risk and Reliability* (2018): 1748006X17751516.

Rocco, C. M. S., Salazar, D. E. A., Ramirez-Marquez, J. E. "Multi-objective network interdiction using evolutionary algorithms." 2009 Annual Reliability and Maintainability Symposium. 2009.

SGIS (Stastical Geographic Information System), Available: https://sgis.kostat.go.kr

Statistics Korea, Available: http://kostat.go.kr/portal/korea/

Song, J. “Multiple Criteria and Scales of Resilience Analysis” Risk Seminar, ETH Risk Center, Oct 3, 2017

Strogatz, S. H. "Exploring complex networks." *Nature* 410.6825 (2001): 268.

Sutley, E. J., Hamideh, S., Dillard, M. K., Gu, D., Seong, K., & van de Lindt, J. W. "Integrative Modeling of Housing Recovery as a Physical, Economic, and Social Process." (2019). *13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP)*, Seoul, Korea

Ventresca, M., Harrison, K. R., Ombuki-Berman, B. M. "The bi-objective critical node detection problem." *European Journal of Operational Research* 265.3 (2018): 895-908.

Ventresca, M., Harrison, K. R., Ombuki-Berman, B. M. "An experimental evaluation of multi-objective evolutionary algorithms for detecting critical nodes in complex networks." European Conference on the Applications of Evolutionary Computation. Springer, Cham, 2015.

Wang, J. W. and Rong, L. L. "Cascade-based attack vulnerability on the US power grid." *Safety Science* 47.10 (2009): 1332-1336.

Wang, J. W. and Rong, L. L. "Robustness of the western United States power grid under edge attack strategies due to cascading failures." *Safety science* 49.6 (2011): 807-812.

Yang, M. D., Chen, Y. P., Lin, Y. H., Ho, Y. F., Lin, J. Y. "Multiobjective optimization using nondominated sorting genetic algorithm-II for allocation of energy conservation and renewable energy facilities in a campus." *Energy and Buildings* 122 (2016): 120-130.

Yoon, D. K., Kang, J. E., & Brody, S. D. "A measurement of community disaster resilience in Korea." *Journal of Environmental Planning and Management* 59.3 (2016): 436-460.

Zhang, J. "InverseVIsTraffic," https:// github.com/jingzbu/InverseVIsTraffic (2016)

Zio, E. and Golea, L. R. "Identifying groups of critical edges in a realistic electrical network by multi-objective genetic algorithms." Reliability Engineering & System Safety 99 (2012): 172-177.

국문 초록

최유정

건설환경공학부

서울대학교 대학원

커뮤니티와 라이프라인 네트워크는 지진, 홍수, 대형사고 등과 같은 재난으로 인한 리스크에 지속적으로 노출되어 있으며, 인구 밀도와 하위 구성 요소들 간의 상호 연관성이 높아짐에 따라 재난으로부터 예상되는 리스크가 점차적으로 증가하고 있다. 반면에 재난 리스크 저감을 위한 예산과 자원은 한정되어 있으므로 커뮤니티와 라이프라인 네트워크 구성요소들의 보수 보강 우선순위를 결정하는 것은 효과적인 재난 리스크 관리를 위한 핵심 과제이다. 따라서 본 논문에서는 도시와 라이프라인 네트워크 각각의 재난으로부터 예상되는 피해 뿐 아니라 복구과정을 종합적으로 고려한 다중 기준 의사결정 지원 방법론들을 개발하였다. 라이프라인 네트워크와 커뮤니티의 재난 위험 저감 의사결정을 위해 개발된 방법론들은 다음과 같다.

첫 번째로, 효과적인 재난 리스크 저감 의사결정을 위해서 네트워크에 큰 피해를 야기하는 중대 재난시나리오를 도출하는 연구를 수행하였다. 기존의 관련 연구들에서는 문제 해결을 위해서 주로 다목적 유전 알고리즘이 활용하였는데, 특히 Non-dominated Sorting Genetic Algorithm-II (NSGA-II)가 가장 대표적으로 활용되는 알고리듬이다. 하지만 네트워크 사이즈가 증가할 때 NSGA-II 를 활용한 전역해 탐색 과정에서 모집단의 다양성이 탐색 초기 단계에서 사라지고 최종적으로

국소 최적해에 수렴하는 문제가 발생하였다. 이러한 문제 해결을 위해서 본 연구에서는 탐색 성능을 개선된 새로운 Multi-Group Non-dominated Sorting Genetic Algorithm (MG-NSGA)이 제안되었다. 제안된 MG-NSGA 방법은 파레토 순위 (Pareto rank) 평가 과정에서 일시적으로 해 공간(sample space)을 복수의 구역으로 나눈 후 새롭게 정의된 복수의 공간마다 파레토 순위를 매김으로써 NSGA-II 와 비교하여 보다 안정적으로 우수한 비지배해 집단을 도출하였다. 또한, 재난 시나리오를 도출한 이후 '중대'한 재난 시나리오를 선택하기 위하여 해공간에서 일부 구역을 '중대 구역 (Critical Zone)'으로 정의하는 개념을 제안하였다. 제안된 '중대 구역' 개념은 의사 결정에 영향을 미치는 변수를 기준으로 해 공간상의 경계를 결정하여 중대 구역 내에 있는 시나리오들 만을 중대 재난 시나리오로 정의하는 개념으로서 재난 위험 대응 과정에서 집중적으로 대응해야 재난 시나리오를 선택할 수 있는 기준을 마련하였다.

두 번째 연구는 앞서 개발된 MG-NSGA 를 활용하여 전력망 네트워크에 발생 가능한 중대 종속 고장 (Critical Cascading Failure) 시나리오를 효과적으로 도출하는 연구를 수행하였다. 기존의 관련 연구들을 살펴보면 종속 고장 해석의 높은 계산 비용으로 인해서 네트워크의 형상을 기반으로 종속 고장을 모델링하며 단일 구성 요소의 파괴로 유발되는 시나리오만을 고려하는 등 단순화하여 종속 고장을 모사하는 연구가 일반적이다. 하지만 네트워크 형상만을 고려한 종속 고장 모사는 과전류로 발생하는 일련의 연쇄 고장 현상을 구현하는데는 한계가 있으며 또한 지진과 태풍과 같은 자연재해는 동시에 넓은 지역에 분포한 구성요소를 파괴시킬 수 있기 때문에 단일 구성 요소 파괴만을 고려한 해석은 전력망에 잠재된

재난 위험을 과소평가하는 문제가 있다. 따라서 이를 극복하기 위해서, 직류 조류 해석을 기반으로 한 과전류 연쇄 고장 모델(Overload Cascading Model, OCM)과 MG-NSGA 를 결합함으로써 효과적으로 전력망의 중대 종속 고장 시나리오를 도출하는 방법론을 제안하였다.

이어서 세 번째 연구는 앞서 개발된 방법론으로 찾은 중대 시나리오들에 효과적으로 대응하기 위해 최적의 보수 보강 전략을 탐색하였다. 최적의 비지배해 집합을 도출해야 하는 과정에는 보수 보강에 소요되는 비용과 재난 리스크 저감 효과를 모두 고려되어야 한다. 이때 네트워크 구성요소 수가 증가함에 따라 최적화에 높은 계산 비용이 요구되는 문제가 발생한다. 이 문제를 해결하기 위해 적은 수의 조합에서 도출된 최적해를 중심으로 점차적으로 해 공간상의 탐색 범위를 늘려가는 Elite-Set Updating 방법론을 제안하였다. 제안된 방법론은 동일한 조합 탐색 범위에서 우수한 결과를 보이며 계산량을 크게 줄여 효과적으로 최적의 보수 보강 안을 제시하였다.

마지막으로 도시 커뮤니티의 재난 복원력의 평가를 위해서 새로운 도시 재난 복원력 클러스터링 방법론(Community Disaster Resilience Clustering, CDRC)을 제안하고 이를 기존의 재난 복원력 평가 방법론들과 비교하는 연구를 수행하였다. 재난 대응의 방법론으로서 회복력의 개념이 관심을 받게 된 이후, 공학분야에서는 재난 발생 직후 구조물의 피해와 그 이후의 기능의 복구 과정을 모델링하는 방법론이 발전되었다. 이때 구조물 복구 모델이 주요한 역할을 하는데 모델 구성에 요구되는 데이터의 부족으로 인해 일반적인 커뮤니티 적용에 제약이 있었다. 따라서 본 연구에서는 복구 데이터와 모델의 부재를 극복하기 위하여 사회경제

지표들을 통해 간접적으로 회복력(Recoverability)을 평가한 도시 재난 복원력 클러스터링 방법론을 제안하였다. 제안된 CDRC 방법론은 시나리오 재난재해 상에서 예상되는 물리적 피해와 회복력을 각각 평가함으로써 유사한 재난 복원력 특징을 가지는 지역들을 찾아 효과적으로 도시 커뮤니티의 재난 피해 및 회복 양상을 유추하였다.

본 논문에서는 네 가지 연구주제를 통해서 각각 네트워크 중대 시나리오 탐색의 성능을 높인 MG-NSGA, OCM 과 MG-NSGA 의 결합을 통해 효과적으로 전력망의 연쇄 고장 시나리오를 도출하는 방법론, 효과적으로 네트워크 보수 보강 조합을 도출하는 Elite-Set Updating 방법론, 그리고 도시 커뮤니티의 재난 회복력 클러스터를 찾는 CDRC 방법론을 제안하였다. 제안된 방법론들은 가상 그리고 실제 라이프라인 네트워크와 도시 커뮤니티 예제들에 적용하여 정확성과 효율성을 보였으며 각각 네트워크와 도시 커뮤니티 단위에서 재난 리스크 저감을 위한 최적화된 의사결정에 기여할 것으로 기대된다.

주요어: 라이프라인 네트워크, 도시 커뮤니티, 재난 리스크, 의사결정 지원, 다목적 유전 알고리즘, 전력망, 연쇄 고장, 보수 보강 전략, 재난회복력.

학번: 2015-30286