ESWA 10038
23 May 2015
ARTICLE IN PRESS
No. of Pages 10, Model 5G

Expert Systems with Applications xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Reliability and topology based network design using pattern mining guided genetic algorithm

Nasim Nezamoddin, Sarah S. Lam *

Department of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY 13902, United States

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Genetic algorithm
Network design
Network reliability
Pattern mining

ABSTRACT

This research proposes a new reliable network design methodology that is based on a pattern mining guided genetic algorithm (GA). The proposed method can be applied for a variety of applications including telecommunication, ad hoc, and power systems. In these networks, failures in certain parts of a network make it necessary for other parts to tolerate a higher traffic load in order to maintain adequate network connections. In addition, path changes due to dynamic routing of traffic can cause a time delay of communications in the network. To understand and reduce the connection failures costs, vigorous investigations are required to select the best design option under budget constraints. Given that many options for network topology and reliability allocation exist, a GA guided with pattern mining is proposed as an effective optimization method to design reliable network while considering link and node failures. Experimental designs under various assumptions have concluded that the guided GA approach is effective in identifying a network solution within a short period of time.

1. Introduction

The design of reliable networks is an important research area, particularly because of the serious problems in telecommunications, ad hoc, and power system networks that result from connection losses between the network sites. Connection failures in these networks often cause delays in communication, which degrade the overall system performance. Given that resources and budgets are limited, it is unrealistic to have an absolutely reliable network design. Therefore, there exists a tradeoff between network performance and the costs of designing and using a less reliable network. While network design varies according to the specific area of application, its basic concepts are consistent in all applications. An optimal design of network systems determines the best architecture with proper reliability indexes. Other variables including the network cost, load, reliability allocation limitations, and the inherent possibility of failure must be considered in the network design (Johnson, Lenstra, & Rinnooy Kan, 1978).

A significant amount of research has been conducted to address the challenges in reliable network designs. Network reliability can be estimated using minimal cuts, Monte Carlo simulation, and artificial neural networks (Altiparmak, Dengiz, & Smith, 2009; Srivaree-ratana, Konak, & Smith, 2002; Yeh, Lin, & Yeh, 1994). One of the issues in network design problem is computation burden of testing the large number of possible designs. In general, network design problems are reported as NP-Hard problems (Kiu & McAllister, 1998). Meta-heuristics have received considerable attention and are considered as effective ways to search the design space and find the most reliable design (Dengiz, Altiparmak, & Belhin, 2010; Konak, 2012). For instance, a genetic algorithm (GA) -based reliable network system design combined with Monte Carlo simulation was studied in Altiparmak, Dengiz, and Smith (1998). GA was used for multiple-objective design optimization problems (Marseguerra, Zio, Podofillini, & Coit, 2005). Their research tried to determine the best design option when the components in the design could cause performance uncertainties. It was indicated that encoding strategies have the most significant effect on quality of the solution in meta-heuristic approaches (Chou, Premkumar, & Chao-Hsien, 2001). They tested the relationship between the effect of encoding, mutation, and crossover strategies on the performance of GA in finding the optimal solution for designing communication networks. GA was applied in variety of application specific network design problems including power systems (Leary, Srinivasan, Mehta, & Chatha, 2009), optical transport networks (Morais, Pavan, Pinto, & Requejo, 2011), and transportation systems (Gen, Kumar, & Kim, 2005). Due to GA's flexibility and success in finding the best or near optimal in most of the network design problems, it has been considered as a main method for solving similar cases (Chen, Kim, Lee, & Kim, 2010).

* Corresponding author. Tel.: +1 607 777 4754; fax: +1 607 777 4094.
E-mail address: sarahlam@binghamton.edu (S.S. Lam).

http://dx.doi.org/10.1016/j.eswa.2015.05.019
0957-4174/

Other heuristic methods, such as simulated annealing (SA), tabu search and particle swarm, have been used to address network design problems (Chelouah & Siarray, 2000; Dengiz & Alabas, 2001; Papagianni, Papadopoulos, & Pappas, 2008). Combination of GA with other heuristics also showed promising results in network design problems (Won, Malisia, & Karray, 2010). Although many meta-heuristics have been proposed to solve the complex design optimization problems, these methods can suffer from time consuming searches while moving through different areas of the design space. Combining data mining techniques with meta-heuristics may be a promising approach that can overcome the large computational expenses for network design problems (Jourdan, Dhaenens, & Talbi, 2006).

In recent years, mathematical models have been used to solve simplified versions of network design problems. Some examples include mixed integer programming (Desai & Sen, 2010), dynamic programming (Elshqeirat, Soh, Rai, & Lazarescu, 2013; Elshqeirat, Soh, Rai, & Lazarescu, 2014; Elshqeirat, Soh, Rai, & Lazarescu, 2015), and branch and bound techniques (Song & Luedtke, 2013). These methods attempted to incorporate more features of reliable network design problems, including multiple terminals (Song & Zhang, 2014), and dependency between failures (Barrera, Cancela, & Moreno, 2015). Nonetheless, these approaches remain incomplete in incorporating the main features of reliable network design problems. Examples include the served loads (or traffics) and nodes failure. Furthermore, they are not efficient in designing complex networks with a large number of nodes and links.

This research proposed an integrated approach of pattern mining, meta-heuristics and simulation for reliable network design. The main contributions of this research are as follows:

- Introduce a reliable network design model where failures may happen to both links and nodes. Design decisions concern with selecting the appropriate topology and the elements' reliability indexes while considering the operation costs.
- Incorporate node-to-node traffic in the model where failures can force path changes of traffic and can cause delays. Dynamic routing is considered as the main routing scheme where the traffic chooses the shortest available path when failures occur in the network.
- Design a pattern-guided genetic algorithm that efficiently searches the solution space.

The rest of the paper is organized as follows: Section 2 discusses the basic concepts of network design areas and different reliability measurement tools. Section 3 presents the design assumptions and the simulation model. Section 4 discusses the proposed method including pattern mining and genetic algorithm. Section 5 summarizes the results and sensitivity analysis findings and the last section provides conclusions and suggestions for future work.

2. Problem statement

This research addresses reliable network problems where the network is presented by an undirected graph UG (N,A) with a defined transmission rate between the nodes (t). Nodes and links in this network are designed based on the determined reliability. Reliability indexes (between 0 and 1) indicate an independent probability that node and link can operate without failures. If a specific link fails, the traffic will need to be guided onto another path and will result in possible delays. If the origin and/or destination nodes and all possible paths fail, a transaction failure will follow. These delays and failures are presented in terms of the delay cost and the transmission lost penalties. The traffic for that transaction affects the network's tolerated penalties. Decisions on network design are concerned about the selection of the best topology and reliability allocation in order to minimize possible penalties. Each link incurs different design cost that can be used as an indicator of the distance or any variable showing link interestingness for selection. The model for the suggested network design problem is shown below:

$$\min\ C_o = \sum_{t=1}^{T} u_t (c_d \bar{d}_t + c_l \bar{l}_t) \tag{1}$$

$$s.t.\quad \sum_{k=1}^{N} c_k + \sum_{k'=1}^{A} x_{k'} c_{k'} \leqslant B \tag{2}$$

$$c_k = f(y_k) \quad \forall k \tag{3}$$

$$c_{k'} = f'(z_{k'}) \quad \forall k' \tag{4}$$

$$\bar{d}_t = g(x_1, \ldots x_{k'} \ldots, x_A, y_1, \ldots y_k \ldots, y_N, z_1, \ldots z_{k'} \ldots, z_A) \quad \forall t \tag{5}$$

$$\bar{l}_t = g'(x_1, \ldots x_{k'} \ldots, x_A, y_1, \ldots y_k \ldots, y_N, z_1, \ldots z_{k'} \ldots, z_A) \quad \forall t \tag{6}$$

$$x_{k'} \in (0, 1)$$

$$y_k \in (R_1, R_2, \ldots, R_{m'})$$

$$z_{k'} \in (R_1, R_2, \ldots, R_m)$$

k: Node index.
k': Link index.
t: Transmission index.
C_o: Total expected cost of network operation.
T: Total number of transactions.
N: Total number of nodes.
A: Total number of links.
u_t: Traffic rate in transmission t.
c_d: Penalty cost of transmission delay.
c_l: Penalty cost of losing transmission.
$\bar{d}_t$: Expected delay for transmission t.
$\bar{l}_t$: Expected transmission lost for transmission t.
$x_{k'}$: Indicator of link k' existance.
B: Total available budget for network design.
y_k: Reliability assigned to node k.
$z_{k'}$: Reliability assigned to node k'.
R_i: Possible reliability index.
c_k: Cost associated with assigning reliability to node k.
$c_{k'}$: Cost associated with assigning reliability to link k'.
m': Number of possible reliability allocations for links.
m: Number of possible reliability allocations for nodes.

The objective function is the cost of delays and transaction losses, which is a subsequent indicator of network performance. Eq. (2) is the budget constraint for design and reliability allocations. Designing more reliable nodes and links costs more, and costs are determined based on the reliability, as shown in Eqs. (3) and (4). Eqs. (5) and (6) show that delays and transmission losses will be affected by the decisions related to topology and reliability allocations for nodes and links. The following formula illustrates the number of possible design options for a network with N nodes (Srivaree-ratana et al., 2002):

$$design\ options = (m+1)^{\frac{N(N-1)}{2}} (m')^{N} \tag{7}$$

For instance, designing a network that has 20 nodes and 5 possible reliability allocations for nodes and links can have a total of 6.73×10^{161} possible options for the network. Therefore, it is

Please cite this article in press as: Nezamoddin, N., & Lam, S. S. Reliability and topology based network design using pattern mining guided genetic algorithm. *Expert Systems with Applications* (2015), http://dx.doi.org/10.1016/j.eswa.2015.05.019