

Reliable Resource Allocation Models in Network Virtualization

Fujun HE

Abstract

Network virtualization has been introduced as a key role in the next-generation networking paradigm to fend off the ossification of traditional networks. By leveraging the technologies of computer virtualization, network function virtualization, and software-defined networking, a platform with network virtualization provides virtualized resources of computing, functionality, and networking to users in a dynamic manner. While network virtualization leads to a more flexible and efficient network, it brings challenges for network management, one of which is how to efficiently allocate resources with satisfying different requirements. In addition, as the adaptation of network virtualization in different application archetypes is an increasing trend, the reliability of an environment with network virtualization has become a major concern. Resource allocation with protection strategies dealing with the reliability issue is an essential requirement for network virtualization. This thesis studies five specific problems about reliable resource allocation in network virtualization, each of which focuses on a typical application scenario, to achieve a flexible, cost-effective, and dependable network virtualization environment.

Firstly, this thesis proposes a primary and backup resource allocation model that provides a probabilistic protection guarantee for virtual machines against multiple failures of physical machines in a cloud provider to minimize the required total capacity. The probability that the protection provided by a physical machine does not succeed is guaranteed within a given number. Providing the probabilistic protection can reduce the required backup capacity by allowing backup resource sharing,

but it leads to a nonlinear programming problem in a general-capacity case against multiple failures. This work applies robust optimization with extensive mathematical operations to formulate the resource allocation problem as a mixed integer linear programming problem, where capacity fragmentation is suppressed. This work proves the NP-hardness of considered problem. A heuristic is introduced to solve the optimization problem. The results reveal that the proposed model saves about one-third of the total capacity in the examined cases; it outperforms the conventional models in terms of both blocking probability and resource utilization.

Secondly, this thesis proposes a backup computing and transmission resource allocation model for virtual networks with the probabilistic protection against multiple facility node failures. Backup transmission resource allocation is incorporated, where the required backup transmission capacity can affect the required backup computing capacity. This work analyzes backup transmission resource sharing in the case of multiple facility node failures to compute the minimum required backup transmission capacity. A heuristic algorithm is introduced to solve the problem; especially, several techniques based on graph theory are developed to handle the problem with full backup transmission resource sharing. The results observe that the proposed model outperforms a baseline with dedicated protection for computing resource.

Thirdly, this thesis proposes a backup resource allocation model for middleboxes with considering both failure probabilities of network functions and backup servers. This work takes the importance of functions into account by defining a weighted unavailability for each function. This work aims to find an assignment of backup servers to functions where the worst weighted unavailability is minimized. This work formulates the proposed model as a mixed integer linear programming problem. This work proves that the considered problem is NP-complete. This work develops three heuristic algorithms with polynomial time complexity to solve the problem. This work analyzes the approximation performances of different heuristic algorithms with providing several lower and upper bounds. This work presents the competitive evaluation in terms of deviation and computation time among the results obtained by running the heuristic algorithms and by solving the mixed

integer linear programming problem. The results show the pros and cons of different approaches. With the analyses, a network operator can choose an appropriate approach according to the requirements in specific applications.

Fourthly, this thesis proposes an unavailability-aware backup allocation model with the shared protection to minimize the maximum unavailability among functions. The shared protection allows multiple functions to share the backup resources, which leads to a complicated recovery mechanism and makes unavailability estimation difficult. This work develops an analytical approach based on the queueing theory to compute the middlebox unavailability for a given backup allocation. The heterogeneous failure, repair, recovery, and waiting procedures of functions and backup servers, which lead to several different states for each function and for the whole system, are considered in the queueing approach. This work introduces a simulated annealing heuristic to solve the problem based on the developed analytical approach. The results reveal that, compared to a baseline model, the proposed model reduces the maximum unavailability 16% in average in the examined scenarios.

Fifthly, this thesis proposes a master and slave controller assignment model against multiple controller failures in software-defined networks with considering propagation latency between switches and controllers. Given assigned controllers for a switch, the master controller in each failure case is automatically specified based on a low latency first policy. This work defines three objectives to be optimized, which lead to three different problems. This work proves that the adopted policy achieves the optimal objectives for considered problems. This work formulates the proposed model with different goals as three mixed integer linear programming problems. This work proves the NP-completeness for all the three problems. A greedy algorithm with polynomial time complexity is developed; this work shows that it provides a 1/2-approximation for the case without the survivability guarantee constraint. The numerical results observe that the proposed model obtains the optimal objective value with the computation time about 10^2 times shorter than that of a baseline that introduces decision variables to determine the master controllers.