SHORT TERM SCIENTIFIC MISSION (STSM)
SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

Action number:
STSM title: Advanced optimization methods for multi-objective robust design
STSM start and end date: 09/01/2019 to 15/01/2019
Grantee name: Dimitri Papadimitriou

PURPOSE OF THE STSM:

Contribution Working Group 1. Book chapter

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

The work carried out during the scientific mission can be summarized as follows

1. Generalize the formulation of the reference model such as to cover in addition to (1) content distribution/ information centric systems: (2) storage as a service, (3) computing as a service and their possible combination, in particular (4) computing + storage as a combined Virtual Machine (VM) service. Finally modeling the network as a service paradigm (5) can also be of interest in terms of modeling resource location and allocation for the execution of virtual networks,

2. Extend the applicability of the demand re-routing/protection model to network infrastructure failures For this purpose, use or even enhance advanced optimization methods including adjustable robust optimization and variable decomposition methods to solve such multi-objective reliability problems.

3. The current version of the model covers the situation where congestion is taken into account at the transportation level, for instance, by considering a piecewise linear routing/transportation cost function. However, the proposed model being linear wrt allocation variables it doesn't allow to account for load induced effects such as facility congestion, i.e., when congestion is to be minimized at the facility level even independently of the induced congestion at the transportation level.

Elaborate on the motivation for such optimization model from the operational perspective compared to current practices. This question remains rather open since questioning the relation between optimization model and design/decision support systems.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

The generalization of the model is summarized in the following tables.

1. The first table details the "interpretation" of the different terms that can appear in the objective function of this mixed-integer linear program: facility installation cost (Column 1), facility dimensioning/sizing cost (Column 3), resource production cost - input (Column 4), service supply

cost - output (Column 5), routing/transportation cost – delivery (Column 6). In general, from the modeling perspective, moving from parameterized facility capacity to variable capacity accounts for the main increase in model complexity.

	Installation (nodal)	Sizing	Production	Supply	Transport
(1)	File/Content server	Storage capacity (continuous)	Via content upload from e.g. central repository	Content streaming / download (clients)	Content delivery to clients
(2)	Data storage center Storage rack/chassis Storage slot (node)	Storage capacity (modular if per disk unit)	Via data upload (from clients)	Data storage and download (access time)	Data delivery to clients
(3)	Computing center CPU rack / chassis CPU slot (node)	Number of FLOPs (modular if per processing unit)		CPU load (number of FLOPs/time unit)	Program and I/O
(4)	Computer center Computer rack Computer slot (node)	Number of computers	Number of configured VM (per computer)	Virtual machine	Virtual machine I/O
		Number of VMs (modular since per computer)			
(5)	Physical resources	Physical resources dimensioning	Mapping - Embedding	Virtual N/L states VL capacity	

2. Since the proposed model enables to cover so-called multiple product/service types (compared to the conventional single-product version of the facility location-sizing-allocation-transportation problem), the second table documents the interpretation of the term product for these different applications together with their possible attributes/classes

		Product / Service attributes			Formulation
(1)	Content service	File type	Content type		File/content type indexed by integer Shared capacity model
(2)	Storage service	Access time	Backup - Recovery	RAID level	Service type indexed by integer Dedicated capacity model
(3)	Computing service	Interactive or Batch	Log.level	OS type	Service type indexed by integer Shared capacity model
(4)	VM service	VM type			VM type indexed by integer Shared capacity model (number of VMs per computer)
(5)	VN service	Virtual link/ path	Virtual node/ network		Binary type Shared capacity model

The generalization of the location-sizing-allocation part of the model has no or limited effect on the formulation of the transportation part of the model beside directionality of traffic flows in relation to the expression of demands (i.e., source/client demand point and service/demand description).

- For (1) and (2) currently documented formulation covers them. If data upload is to be included as part of the model, the latter (2) requires nevertheless the formulation of facility incoming flows (from client demand points to facilities) in addition to outgoing flows (from facilities to client demand points).
- For (3), one can assume that either the size of the input (and thus the traffic) is high and output of limited size/negligible or vice-versa. Both cases are covered by the model (due to symmetry principle).
- For (4), the situation is more involved when VMs perform as distributed computers interacting / exchanging data with remote computers. Nevertheless applying a sequential approach would still be possible by first solving the location-sizing-allocation problem and then only solving the transportation problem (as any other mMCF with asymmetric demands). In any case, such modeling yields a worst-case cost estimate on the transportation cost.
- For (5), there is no direct transportation cost involved for service delivery. If the service itself involves traffic exchanges between (virtual) network nodes and client demand points as it would be the case for network service chaining, the proposed model is not directly applicable. The main reason stems because the formulation of the demands would typically include in addition of the source/client demand point and the service/demand description the destination of the demand.

Thus, except if the destination of the demand corresponds to its source, the model is not directly applicable.

The applicability of the demand re-routing model has also been extended to cover multi-link protection scenario (thus network infrastructures failures) beyond un/correlated facility failures. The question remained though open as whether scenario based, uncertainty set-induced parametric robust formulation (with non-adjustable deterministic variables), uncertainty set-induced adjustable/multi-stage robust formulation (with both adjustable and non-adjustable deterministic variables) would provide for best tradeoff between tractability (convexity, decomposability, etc.) vs conservatism.

FUTURE COLLABORATIONS (if applicable)