Scientific Report: Short Term Scientific Mission

COST Action CA15127

Sasko Ristov

July 23, 2017

1 STSM Details

STSM Title: Cost/Robustness trade-off optimization for VM image placement

STSM Applicant: Dr. Sasko Ristov, Distributed and Parallel Systems Group, Faculty of Computer Science, University of Innsbruck, Innsbruck, Austria

Host: Dr.-Ing. Carmen Mas Machuca, Chair of Communication Networks, Technical University of Munich (TUM), Munich, Germany

Start Date: 06.07.2017

End Date: 17.07.2017

COST Action CA15127 - Resilient communication services protecting end-user applications from disaster-based failures (RECODIS):

Working group: WG3

2 Purpose of the STSM

The purpose of this STSM was to start a collaboration in order to analyze and join two replica placement models with different purposes: one multi-objective cost/performance VM placement and single-objective controller placement for a resilient SDN control plane. Our plan was to define a model for VM placement in a core network of data centers, that will optimize the network utilization, but at the same time it will optimize the robustness of the network and minimize the impact of a possible disaster with maximal impact. Improvement of the robustness requires creating several (at least two) disjoint paths to the required content. Creating such paths that will guarantee required protection level will increase the cost and network utilization. Therefore, the goal was to optimize these two conflicting criteria:

- Maximize the robustness of the network, i.e. maximize the survivability of already created network paths in case of a disaster, and
- Minimize the cost of the recovery in case of a possible disaster, which will have maximized impact.

3 Description of the work carried out during the STSM

3.1 Problem to solve

During the STSM, we defined a model to address the following problem. In a normal network state, before a disaster, we are determining several issues. First, we optimize how many and where to place all replicas of all contents in order to minimize the total network utilization. At the same time, the algorithm tries to find a placement that will maximize the robustness of the system, after which a protected level of quality of service (bandwidth per user request) will still be retained. This means that for each user request, the algorithm will create several disjoint end-to-content paths. The placement also considers, that, in case of a disaster, the recovery procedure to recreate the failed end-to-content paths and potential migrations of content data from one to another data center will be minimized.

To summarize, by using multi-objective optimization, we try to optimize three conflicting objectives:

- Minimize network resource utilization represented with bandwidth and total delay,
- Maximize the network's robustness in case of a disaster with multiple failures of links and data centers, and
- Minimize the recovery cost in case of a disaster, represented as the additional bandwidth and delay that is necessary to recover the network back to the normal state, which is the same as before the disaster.

3.2 Work carried out

During the STSM, a working day was organized in two parts: discussion and research. At the end of each working day, through discussion we summarized and confirmed the work that was done from the last discussion, as well as we were planning the work for the next day.

The whole work was divided into three phases: *i*) defining the disaster, its impact and accordingly the network robustness, *ii*) defining in details multiple objectives and modeling their constraints and *iii*) planning the implementation and evaluation of the model with various parameter values and different real core networks.

The summary of our results for all three activities are described in the following section.

4 Description of the main research results obtained

This section present the assumption that were considered before developing the model in order to cover core networks. Further on, it presents the model for a disaster with multiple failures and its maximized impact.

4.1 Assumptions

Without loosing the generality, in order to simplify the model, we considered several assumptions for the network:

- *One replica per data center.* We considered this assumption since we are not interested in providing better performance for a faster access to the content, but we observe a data center as a single point of failure, so adding more replicas to the data center will not provide better robustness. Additionally, the link to a data center is limited.
- We consider that the network is designed such that exist a placement in which at least two link-disjoint paths exist from each source (node) to each content.
- *Disaster with huge impact.* In case a data center is failed, we assume that the node (intermediary network devices) will also fail. This means that all links connected to the failed data center will also fail.

- *Single link between two nodes in core networks.* We are assuming that maximum one direct link exist between each two nodes, since, in general, core networks usually have have such design. Still, a network can have even several redundant different paths between two nodes, out of which maximum one is direct.

- *Bi-directional traffic*, which means that we will consider (i, j) and (j, i) as separate links for traffic, each with separate link capacity. Nevertheless, if a link (i, j) fails, we consider that the link (j, i) fails, as well.

- *Static random failures.* For the basic model, we considered static random failures, while cascaded impact of a disaster, such as epidemic, were left for future collaboration.

4.2 Modeling a disaster, robustness and maximize the impact of a disaster

Many researchers have considered the failures of network elements (links or network nodes) and then tried to improve the network's robustness. We also consider multiple type and impact of failures, by extending the failures of a link, node, or a data center.

4.2.1 Modeling a (l_f, d_f) disaster

We introduced (l_f, d_f) ***Disaster*** as a situation in which l_f links and d_f data centers fail. In order to simplify the model, the node failure, is modeled as a failure of all links connected to that node.

4.2.2 Modeling a (l_f, d_f) robustness

Further on, we defined the (l_f, d_f) ***Robustness*** of a network. That is, the network is (l_f, d_f) *robust* if l_f and d_f are maximal such that the system is (l_f, d_f) disaster tolerant, while it is neither $(l_f + 1, d_f)$ nor $(l_f, d_f + 1)$ disaster tolerant. This means that the network will protect all user requests with some protection level in case of (l_f, d_f) disaster, but neither can protect in case of $(l_f + 1, d_f)$, nor $(l_f, d_f + 1)$ disaster.

We must note that a system can have several values for the robustness. For example, a system's robustness can be (8, 2) or (10, 1). In reality, the latter case is more probable, since the link failure probability is much higher than the data center failures.

Also, we must mention that the system robustness is changing during the time, since the number of replicas will be changing, as well as the number of links, nodes and data centers.

4.2.3 Consider the worst case: Model a disaster with the maximal total impact (over links and data centers)

In order to maximize the (l_d, d_d) **Disaster Impact**, we will assume the worst case, that is, failure of those l_d links (i, j) that are allocated with maximum bandwidth (DIL) and failure of those d_d data centers d that store maximum number of replicas (DID), as defined in (1).

$$DI = \sum_{top\ l_d\ (i,j)} \frac{DIL_{(i,j)}}{\sum_{\forall(i,j)\in E} DIL_{(i,j)}} + \sum_{top\ d_d\ d} \frac{DID_d}{\sum_{\forall d\in D} DID_d} \tag{1}$$

We considered the static random failures, since the failures caused by natural or technical disasters are unintentional failures.

4.3 Planning the implementation and evaluation

The model implementation was started by introducing the constraints and objectives in the jMetal framework [DNA10, DN11]. Its evaluation will be done after completing the implementation by using different networks.

Since the model is very complex and NP-hard, we need to introduce some heuristics that will be done as a follow up.

5 Future collaboration with the Host institution

Since the model was so complex, its implementation will be finished as a follow up, and discussions will be organized remotely. Nevertheless, after finishing the implementation and evaluation, the collaboration will be even extended with several other researchers of both universities, that will work in extending the model in several directions:

- Considering Node-disjoint end-to-content path, instead of link-disjoint. Just to note that node-disjoint end-to-content paths are at the same time link-disjoint, but the opposite is not always true.
- Model the network in order to allow multiple direct links between two nodes in order to increase the network robustness.
- Assumption is that the end time of a request is much greater than the recovery of a disaster. We would like to generalize our model in order to add that some requests will finish and release some resources, which can be used in the recovery phase.
- Consider other disaster models, such as *Random Dynamic Epidemic Disaster Model* - *SID Susceptible-Infected-Disabled*, or a disaster with different impact / probability in different regions, or *SIDR - Susceptible-Infected-Detected-Prevent*, or *SIRS - Susceptible-Infected-Removed-Susceptible*.

6 Foreseen publications/articles resulting from the STSM

The plan is after completing the implementation and evaluation, to submit this complete work as a conference paper. Later on, by extending the model and introducing heuristic, we plan to publish the work as a journal article.

References

[DN11] Juan J. Durillo and Antonio J. Nebro. jmetal: A java framework for multi-objective optimization. *Advances in Engineering Software*, 42:760–771, 2011.

[DNA10] J.J. Durillo, A.J. Nebro, and E. Alba. The jmetal framework for multi-objective optimization: Design and architecture. In *CEC 2010*, pages 4138–4325, Barcelona, Spain, July 2010.