Scientific Report of Short Term Scientific Mission CA15127 RECODIS

Dr. Péter Babarczi
January 23, 2017.

I. Short Term Scientific Mission Details

STSM Title: Alert-Based Reconfiguration of Virtual Networks and Network Functions
STSM Applicant: Dr. Péter Babarczi, Budapest University of Technology and Economics, Hungary
Host: Prof. Massimo Tornatore, Politecnico di Milano, Italy
Period: 2017. 01. 09 - 2017. 01. 20.
Working Group: WG2 - Weather-Based Disruptions

II. Purpose of the Short Term Scientific Mission

Network virtualization is an important technology enabler for those Software-Defined-Networking (SDN) infrastructure providers that intend to share their physical SDN network among multiple tenants. To satisfy the tenants' resilience requirements (e.g., stipulated in a Service Level Agreement), virtual SDNs representing the tenants' network might need to be provisioned exploiting a survivable embedding (i.e., reserving backup resources if the primary nodes/links fail), or virtual nodes and links might be migrated from the current physical resources to new ones as part of disaster preparation upon a hurricane or tornado alert is issued. First goal of the STSM is to investigate and/or develop survivable control and/or data plane embedding strategies (including but not limited to controller and hypervisor placement, the design of switch-to-hypervisor and hypervisor-to-controller communication channels for virtual SDN networks) which considers future (disaster disjoint) embeddings of the virtual network, thus, better supporting the alert-based migration of virtual links. We are especially interested in the trade-off between the required spare capacity/number of required backup rules in the forwarding tables and the degree of migration allowed (e.g., only links can be migrated, number of maximal steps allowed, considering the size of the virtual node in the migration, etc.).

A second goal of the STSM is to investigate different aspects of the problem of protection of virtual network functions and service chains (e.g., in data centres or WAN sized SDNs). One direction might be to investigate the applicability of minimum weight link-disjoint shortest paths with a given number of common nodes in the disaster-resilient Virtual Network Function Placement problem, and try to identify the polynomial-time solvable cases. However, other aspects of resilient service chaining might be also investigated during the STSM.

III. Description of the Work Carried out During the Short Term Scientific Mission

During the first week of the STSM, three topics were discussed in detail with the Host research group (including researchers Prof. Francesco Musumeci and Ali Hmaity) based on the proposed research directions in Section II as potential main lines of the research work carried out during the STSM. In this section we shortly summarize the main findings of these discussions.

A. Minimum-Weight Link-Disjoint Paths for Virtual Network Function and Service Chain Protection

The application of Minimum-Weight Link-Disjoint Paths (MWLD) [1] visiting a specified number of common nodes proposed in one of the previous papers of the Applicant was investigated for the resilient Virtual Network Function (VNF) Placement problem considered by the Host research group in [2]. Polynomial solutions were proposed to the MWLD problem where the number of possible common nodes is upper bounded by a constant k. However, no solutions were provided for the computationally intractable, but practically more relevant cases, i.e., exact k common nodes and an at least k common nodes in the link-disjoint path-pair, which could be beneficial for resilient service chain placement (see motivation in Section IV-A). Furthermore, based on a previous discussion of

the Applicant and Host about availability evaluation of series-parallel structures like link-disjoint path-pairs through pivotal decomposition (or factoring) [3], it was a conclusion of the discussion that extending the current virtual network function provisioning with an availability-aware design is a viable direction which is worth to investigate.

Hence, a recent result on a similar availability-aware connection design, in specific, Quality-of-Service-aware (QoS-aware) route design for a single request [4] was investigated and discussed in detail as a possible framework which is applicable to our problem under consideration. In [4] the authors argue that providing full protection against network failures (against single link failures in specific) through link-disjoint paths is too restrictive and demands excessive redundancy in practice. Thus, tunable survivability is proposed, which provides a quantitative measure for survivability and offers flexibility for the service provider to select paths for the connections by allowing some common links along the two paths. While a given level of survivability has to be satisfied by the path-pair, some bottleneck (e.g., bandwidth) or additive (e.g., delay) metric is minimized in the optimization problem. Note that, with a low survivability requirement even a single path can be sufficient to satisfy it (if it minimizes the additive metric among all solutions). Such an approach can be useful in network parts with higher risk links where disjointness of the two paths is necessary, while they can share some common links (the extent depends on the required survivability level) in the less disaster-prone areas of the topology. Furthermore, sharing common nodes instead of links [1] in the disjoint path-pairs allows the service provider to: 1) deploy the primary and backup VNFs which stores excessive amount of state at the same location, avoiding state-migration (and thus, saving bandwidth and reducing latency) upon an alert is issued; and 2) primary and backup VNFs can be placed closer to each other based on their (software) availability values. Both design objectives can be considered in the path design phase by e.g., requiring exactly k common nodes [1] in the link-disjoint path-pair.

Although [4] tackles the problem of availability-aware design from a similar direction, we concluded our discussion that in VNF placement, where the main objective is to minimize the number of VNF-capable nodes rather than an additive routing cost on the edges makes the problem more complex either we consider the optimization for a set of connections demands (i.e., static routing) as in [2] or we optimize only for the solution cost of individual connection requests (i.e., dynamic routing) reusing as much already deployed VNFs (shared with other service chains) in the network as possible. As the most promising approach, we selected this one as the main research line during the STSM visit. The problem formulation and achieved results are detailed in Section IV.

B. Alert-Based Resilient Connection Design with Diversity Coding

After a short presentation given by the Applicant about the previously achieved theoretical and practical results on General Diversity Coding (GDC) [5], [6], its applicability was discussed for function placement problems in virtual software defined networks and in network function virtualization. The past recent results obtained on QoS-aware GDC routing [7] by the Applicant (before- and after-failure delays are defined on general directed acyclic graphs rather than simple paths) can be applied to alert-based virtual network designs as well (including virtual network function and service chain placement discussed in Section III-A). In fact, one of the main constraints of the service chain model in [2] is that the chains must satisfy an end-to-end latency requirement, exactly what was generalized in GDC for general routing structures [7]. However, important to note that the simplicity of the GDC approach stands only for single edge failures, thus, further elaboration is required before its huge flexibility compared to previous inflexible routing structures (like disjoint path-pairs of $1 + 1$ protection) can be exploited in a disaster-aware connection design.

As a conclusion of the discussions, the flexibility provided by network coding might be further investigated for alert-based reconfiguration of different networking scenarios – including optical transport networks – but points beyond the limits of the current STSM. However, one possible use case where it directly brings benefit for virtual network design is disaster-aware hypervisor placement for virtual SDN networks, detailed in Section III-C.

C. Disaster-Aware Placement of Hypervisors for Virtual Software Defined Networks

One of the most challenging problems of virtual SDN design is the optimal placement of the network hypervisor(s). This problem remains even hard when we assume that the virtual SDN networks and their corresponding controllers are already embedded and fixed in the network [8]. Furthermore, resilient in-band communication channel design from the virtual switches to the hypervisor and from the hypervisor to the corresponding controller might result in excessive resource usage if it is feasible at all.

The previously mentioned flexibility of GDC might be beneficial even in this case, as it provides resilient communication channels even if two disjoint paths for resilience approaches like $1+1$ protection (or three disjoint paths in traditional diversity coding) do not exist between the virtual switches and the hypervisor, and from the hypervisor to the corresponding vSDN controller, and vice versa. Furthermore, if hypervisor failures are also considered in the failure model, the communication paths from the virtual switches to their corresponding controller might/must traverse different hypervisor instances, adding further freedom, and thus, complexity to the optimization problem.

We further note here that this problem is even more complex if we assume that – owing to a weather-based alert – a virtual network (hence, its controller as well) is migrated to different physical resources. One of the main reasons for initiating a reassignment is the latency requirement – e.g., responsiveness to the packet-in messages – which the virtual networks might need to satisfy. However, hypervisor reassignment and novel design of communication channels to the new hypervisor could be a lengthy process. Hence, doing such kind of reassignment for a hypervisor placement which did not consider possibly migrated virtual networks in the network design phase (i.e., hypervisor placement, virtual network assignment and in-band communication channel design) could result in degraded performance and service disruption during the migration.

Owing to the extent of the research work to carry out in this topic exceeds the limits of the STSM, and because it stands the furthest from the Host's and Applicant's currently achieved research results, it was considered as a possible future work after the STSM.

IV. Description of the Main Results Obtained

In this section we detail the main results achieved during the second week of the STSM in the topic of resilient virtual network function and service chain placement (Section III-A). The main idea is to use minimum-weight link-disjoint path-pairs with a specified number of common nodes for a resilient service chain deployment, while availability and quality-of-service requirements of the chain have to be satisfied. As a sound result which provides a reasonable reliability to cost trade-off – i.e., for a significantly decreased cost the connection provides only a slightly worse availability – would help to convince operators to deploy a given method in their network, our goal is to develop and propose such kind of algorithms.

A. Background

As network nodes are typically more reliable than network links, it seems a reasonable assumption in the design that we allow to have some nodes in common in the primary and backup path of the service chain depending on the required availability level of the connection(s), but the two paths must be link-disjoint. The optimal solution of this problem for individual connection demands minimizing the cost of the link-disjoint path-pair was investigated in [1].

In [2] off-site redundancy and on-site redundancy of VNFs were defined. On-site redundancy – where the VNFs of the primary and backup chains are placed at the same VNF-capable nodes to reduce the amount of VNF internal state information that need to be transfered from primary to backup VNF – is beneficial for state-full VNFs, while off-site redundancy – primary and backup VNFs are placed at different VNF-capable nodes – can be used for stateless VNFs. However, in [2] either end-to-end path protection or virtual link-protection design is proposed, which provides off-site or on-site redundancy, respectively, for each VNFs, regardless they are state-full or stateless. With an MWLD-like design, the full design space can be explored between the two extremes (both in availability and capacity), and service chains can be calculated in a more flexible way tailored to the exact requirements of the VNFs. With such a mixed design the placement can be appropriately tuned for protection against weather-based disruption, as upon an alarm is issued, state migration of state-full VNFs does not impose any additional cost (either bandwidth or delay) owing to the fact that it's backup is placed at the same – possibly shielded – node, while for state-less VNFs such migration is not required at all (can be turned off at one place and invoked at the other).

Further note that, the number of common nodes allowed in the MWLD design (tight or lower bound) is a good estimator for the perceived connection availability as well. Thus, availability of the service chain can be evaluated a posteriori (or directly follows from the number of common nodes in the two paths if single link failure model is in place [4]), rather than considered in the design phase, which simplifies the already complicated problem formulation.

B. Problem Formulation

Given a directed graph $G = (V, E)$, where each arc is assigned with positive delay function $d : E \to \mathbb{R}^+$. Furthermore, given a list of service chain (SC) requests $\mathcal{S}$ we want to embed into the network G, each consist of an ordered sequence of one or multiple virtual network functions from the set $\mathcal{F} = \{F_1, F_2, \ldots, F_k\}$. Each service chain request $S_i \in \mathcal{S} : S_i = (s_i, t_i, \mathcal{F}_i, D_i)$ consists of a source node s_i, a target node t_i, the requested VNFs to implement the service chain $\mathcal{F}_i \in \mathcal{F} : \mathcal{F}_i = \{F_{i_1}, F_{i_2}, \ldots, F_{i_j}\}$, and a delay requirement D_i. A valid embedding of a service chain is a simple or non-simple path P_i, which traverses the VNFs in $\mathcal{F}_i$ in the requested order ($s \rightsquigarrow F_{i_1} \rightsquigarrow F_{i_2} \rightsquigarrow \ldots \rightsquigarrow F_{i_j} \rightsquigarrow t$), while the overall delay of the arcs along the chain is less than D_i, formally $\sum_{e \in P_i} d(e) \leq D_i$. Our objective is to embed as many SC requests as possible from the set $\mathcal{S}$, while the number of nodes where VNFs are deployed is minimized. We made the following assumptions:

- Each V in the network can host at most c_v VNFs, or equivalently has c_v CPU cores (we assume a VM for each VNF requires one core).
- A VNF of the same type F_l can be shared between arbitrary SCs without any additional delay (e.g., context switching or upscaling) or any additional cost.

From an operational perspective, the set of VNFs $\mathcal{F}$ can be divided into two disjoint subsets, namely state-full VNFs $\mathcal{F}^f$, which primary and backup instance must be placed at the same node, and state-less VNFs $\mathcal{F}^l$, which primary and backup VNFs must be placed at different locations ($\mathcal{F} = \mathcal{F}^f \cup \mathcal{F}^l$, $\mathcal{F}^f \cap \mathcal{F}^l = \emptyset$). In order to consider *reliable embeddings* surviving single link and single node failures, instead of a single path P_i we require node-disjoint primary and backup paths W_i and B_i between s_i and t_i. However, with the following assumptions we can reduce this unnecessarily strict requirement, and enable the two paths to share some common nodes according to the operational requirements:

- A subset of the network nodes are considered fully reliable (or equivalently, more reliable than other network nodes), and thus, can host the primary and backup VNF for the same service chain, referred to as *common nodes*. As these nodes can be common in W_i and B_i with minor or even without any reduction in the perceived availability, we require node-disjointness only between primary of backup paths of the segments $F_{i_l} \rightsquigarrow F_{i_{l+1}}$ of the service chain, namely a primary segment W_i^l and a backup segment B_i^l.
- Both primary and backup paths W_i and B_i constructed from the different segments have to satisfy the delay bound D_i between s and t. Note that, e.g., W_i^l could share links either with W_i^m or B_i^m, $\forall m \neq l$, while the continuity of the service is still maintained through a combination of primary and backup segments, hence, the delay of these alternating paths between primary and backup segments should also satisfy D_i.
- Successive state-full VNFs in $\mathcal{F}_i^f$ or successive state-less VNFs in $\mathcal{F}_i^l$ in an SC might share a same location (i.e., allocated at the same node).
- The number of state-full VNFs $|\mathcal{F}_i^f|$ defines a k value for each SC, i.e., the number of common nodes allowed along the service chain. As a refinement, k could be defined as the number of state-full VNF sequences in the service chain (owing to the fact that they can share the same physical location).

C. Possible Heuristic Approaches

In order to find a solution for the problem above, as a first option we proposed a greedy heuristic approach which embeds the SC requests one after the other. For this, an auxiliary graph is created following the idea of the upper-bound algorithm in [1]: create a graph $G' = (V, E')$, where each link corresponds to a node-disjoint minimum-delay path-pair in G. Each link E' has two attributes, namely the delay $d'(e)$ of the longest path from the disjoint path pair, and the lower number of available VM cores $f'(e)$ along the nodes of one of the paths. Note that, a path in this graph with $k+1$ links would correspond to a link-disjoint path-pair traversing k common nodes (and probably some other owing to the fact that the paths of different segments can overlap). Further note that, the end-to-end delay of a path in G' (calculated as the sum of $d'(e)$ values of its links) is a worst case upper bound on the worst case delay after multiple link failures affecting the shortest link-disjoint path-pair of each segment, thus, clearly overestimates even the worst case after-failure delay upon any single link failure, which should obey the delay bound D_i. Furthermore, the algorithm might also consider the secondary metric $f'(e)$ on the links either as an optimization objective or constraint for the problem, in order to ensure that a valid embedding of the SC exist, which raises the necessity of the application of a restricted shortest path algorithm to find suitable paths in G'.

As consolidation (i.e., share as many VNFs between SCs as possible while the number of VNF nodes is minimzed) is still the main objective of the algorithm [2], the cost with already placed VNFs of a certain type might be considered as well with different cost in the algorithm, and can be calculated with a modified version of Dijkstra's shortest path finding algorithm. Again, it is important to note that, the corresponding link-disjoint path-pairs of the link is the "best" path found in G' might not be mutually node disjoint in the original graph G, i.e., links can be shared by subsequent segments; however, they do not affect the overall connection availability in the single link and node failure model.

V. Future Collaboration and Forseen Publications Resulting from the STSM

On a short term, we believe that – upon completion – the research results described in Section IV are worth being considered as a submission to IEEE Reliable Network Design and Modeling (RNDM) or another conference. However, in order to publish these results, the next steps to be made are:

- Find the best optimization criteria for the path search algorithm in $G' = (V, E')$, e.g., shortest, widest-shortest, given number of links, etc.
- Implement the proposed heuristical approach and compare it to the baseline results provided by the previously proposed ILP formulation extended with the requirement of k common nodes.
- Evaluate availability of the proposed routing structure under varying node- and link-availability values against single failures, and extend the evaluation against weather-based disasters.

Furthermore, different extension like different reliability values for different VNFs (i.e., software reliability) can be considered as well, besides the availability of the physical nodes and links.

On a longer term, the other research directions identified during the personal discussions, detailed in Section III-B and Sectoin III-C might be further elaborated and would contribute to the success of COST Action CA15127 RECODIS.

References

[1] J. Yallouz, O. Rottenstreich, P. Babarczi, A. Mendelson, and A. Orda, "Optimal link-disjoint node-"somewhat disjoint" paths," in *Proc. IEEE 24th International Conference on Network Protocols (ICNP)*, Nov 2016, pp. 1–10.

[2] A. Hmaity, M. Savi, F. Musumeci, M. Tornatore, and A. Pattavina, "Virtual network function placement for resilient service chain provisioning," in *2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*, Sept 2016, pp. 245–252.

[3] P. Babarczi, J. Tapolcai, and M. Tornatore, "Comments on 'Availability formulations for segment protection'," *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2591–2591, June 2013.

[4] J. Yallouz and A. Orda, "Tunable qos-aware network survivability," *IEEE/ACM Transactions on Networking*, vol. PP, no. 99, pp. 1–11, 2016.

[5] P. Babarczi, J. Tapolcai, L. Rónyai, and M. Medard, "Resilient flow decomposition of unicast connections with network coding," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2014, pp. 116–120.

[6] A. Pašić, J. Tapolcai, P. Babarczi, E. R. Bérczi-Koávcs, Z. Király, and L. Rónyai, "Survivable routing meets diversity coding," in *Proc. IFIP Networking Conference (IFIP Networking)*, May 2015, pp. 1–9.

[7] A. Pašić, P. Babarczi, and A. Kőrösi, "Diversity coding-based survivable routing with qos and differential delay bounds," *Optical Switching and Networking*, vol. 23, Part 2, pp. 118 – 128, 2017, Special Issue on Design and Modeling of Resilient Optical Networks.

[8] M. Tornatore, J. André, P. Babarczi, T. Braun, E. Følstad, P. Heegaard, A. Hmaity, M. Furdek, L. Jorge, W. Kmiecik, C. Mas, L. Martins, C. Medeiros, F. Musumeci, A. Pašić, J. Rak, S. Simpson, R. Travanca, and A. Voyiatzis, "A survey on network resiliency methodologies against weather-based disruptions," in *Proc. 8th Intl. Workshop on Reliable Networks Design and Modeling (RNDM)*, Sept 2016, pp. 23–34.