Joint Allocation of IT and Connectivity Resources for Survivable Services in Geographically Distributed Metro Data Centers

Ajmal Muhammad[(1)], Paolo Monti[(1)], Payman Samadi[(2)], Lena Wosinska[(1)], Keren Bergman[(2)]

[(1)] School of ICT, KTH Royal Institute of Technology, Kista, Sweden, ajmalmu@kth.se
[(2)] Department of Electrical Engineering, Columbia University, New York, NY 10027, USA

Abstract *The paper proposes a survivable and programmable metro-scale converged inter- and intra-datacenter network architecture and exploits its unique features for allocating jointly IT and connectivity resources. The proposed dynamic provisioning strategy offers a substantial reduction of service request blocking.*

Introduction

Metro networks are expected to play a key role in the upcoming 5G network paradigm[1], which is expected to introduce new services with stringent bandwidth, latency, and reliability requirements. In addition, 5G is envisioned to use data center (DC) resources for providing virtualized network services, rather than employing dedicated physical resources. Consequently, enabling new applications and services requires not only innovative transport solutions but also a sophisticated integration of the IT infrastructure. Similarly, to improve the overall network resource utilization, enhanced control plane solutions are needed for the unified orchestration of IT and connectivity resources for these new services. To efficiently address the strict requirements of these services, novel architectural solutions, such as the deployment of small- to medium-sized DCs interconnected through an agile and software-defined optical metro network have been recently proposed[2–5]. Among them, the architecture based on converged inter- and intra-DC network[5] exhibits remarkable flexibility as well as cost effectiveness compared to other alternatives[2–4]. This converged architecture combines the benefits of electronic packet switching in conventional metro DC network with the flexibility of optical switching by providing two types of connectivity services among the DCs, i.e., dynamic and background connections. Dynamic connections support on-demand lightpaths between racks in different metro DCs facilitating bulk data transfer, while background connections set up connectivity among the DCs serving low data-rate traffic flows.

Some services (e.g., media distribution and machine-type communication (MTC)[1]) employ redundancy to ensure high service availability. As a result, new traffic patterns and huge traffic volumes are generated requiring tailored resource-efficient provisioning solutions. For example, a considerable amount of connectivity and IT resources are required for instantiating data replication and backup, which needs additional bandwidth for the synchronization between the source and backup content. A converged architecture offers the opportunity to jointly manage IT and connectivity resources, which can potentially improve the overall network resource utilization. To this end, our study aims at investigating the benefits of applying a converged architecture for survivable optical data center networks by *(i)* supporting the overall resource orchestration and *(ii)* proposing a dynamic provisioning strategy that seeks to balance the load over both the DCs and the transport network segments to minimize the service request blocking. The proposed strategy is based on a multi-objective function, which is properly tuned for optimization of IT as well as connectivity resources needed for service instantiation and synchronization.

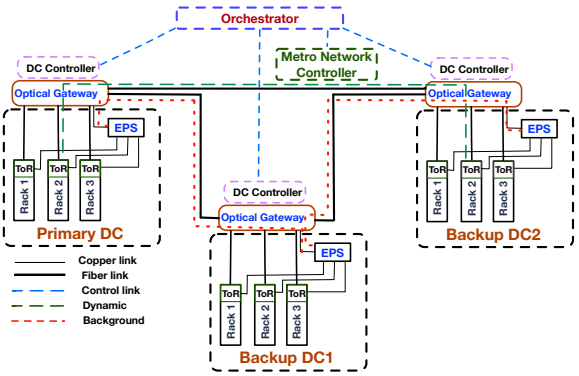


Fig. 1: Proposed control plane for joint resource allocation in converged metro network architecture.

Control Plane for Joint Resource Allocation

The schematic diagram of the proposed control plane for joint resource allocation in a converged architecture is shown in Fig. 1. The electronic packet switch (EPS) network provides the intra-DC connectivity among the racks by aggregating the top-of-rack (ToR) traffic. The optical gateway (OG) combines both the EPS and ToR switches and facilitates the inter-DC communication over a metro network. More details on the OG and the data plane of the converged architecture can be found in[5,6]. The control plane is based on the Software Defined Networking (SDN) concept. It comprises three main modules, i.e., DC controller, metro controller, and orchestrator, which are arranged in a hierarchical manner. The DC controller configures the OGs and EPSs while the metro network controller coordinate the resource allocation in the transport network. These controllers have full knowledge of resources in their respective domains and provide this information to the orchestrator. The orchestrator is responsible for accommodating new service requests by assigning the required IT and connectivity resources. The selection

of the backup DC for a given service request depends on the availability of *(i)* the IT resources (servers) required by the service request and *(ii)* connectivity resources for transferring and subsequently updating the stored data. A dynamic connection (green dashed lines in Fig. 1) is employed for large data transfer to the backup DC, while a background connection (red dotted lines in Fig. 1) is used for exchange of synchronization information between the primary and the backup DC. The orchestrator runs the resource allocation algorithm to efficiently select the required IT and connectivity resources. The resources occupied by the dynamic connection are automatically released once the data is transferred. The bandwidth allocated for the update traffic and the IT resources in the primary and the backup DC become free when the service holding time expires. A provisioning strategy executed by the orchestrator for assigning resources to dynamic survivable service requests is presented next.

Dynamic Provisioning of Survivable Service Request (DP-SSR)

The working principle of Dynamic Provisioning of Survivable Service Request (DP-SSR) follows the steps presented in Algorithm1. Given a service request R, if there are enough IT resources available at source DC s and at least one candidate backup DC d, then the connectivity resources for transferring the data DA and exchanging the synchronization information from s to d are computed. For data transfer via dynamic connection, free wavelengths on k shortest paths (π^k_{sd}) from s to d are explored while active background connections (BG_{sd}) are checked to find free bandwidth for the update traffic. The algorithm prefers existing background connections (BG_{sd}) over establishing new ones by assigning more weight (Ω_j) to them in the selection process. Besides, preference is given to the background connection with free bandwidth ($freeBW_j$) exactly equal to the required value (BW_{syn}), or, if such connection does not exist, to the connection with the largest free bandwidth. If connectivity resources are available for dynamic connection as well as for synchronization traffic, then the triplet SR_d, p^i_{sd}, and j is a viable solution and the value of objective function $O^s_{d,i,j}$ is computed. The objective function $O^s_{d,i,j}$, is a linear combination of these three quantities with coefficient α for the IT resources, coefficient β for the connectivity resources for transferring data, and coefficient γ for the connectivity resource for synchronization traffic. Among all the feasible solutions, the one with the largest value of the objective $O^s_{d,i,j}$ is selected. However, if a feasible solution is not found, then the service request R is blocked. Finally, the wavelength for the dynamic connection is selected from the set of available wavelengths ($w_{p^i_{sd}}$) using a first-fit approach.

Numerical Results

The performance of the DP-SSR strategy is evaluated using a custom-built event-driven simulator. Simulations are performed on metro network topology[6] with 38 nodes and 59 bidirectional fiber links, each one supporting 80 wavelengths. The number of DC locations (|D|) is 15, uniformly distributed among the network nodes. Each metro DC hosts 320 servers assuming 40 servers per rack, while each ToR switch has one tunable 10G transceiver connected to the OG and one grey 10G transceiver connected to EPS. Service requests are assumed to arrive in the network following a Poisson process, each one requiring a number of servers (SR) chosen uniformly between 1 and 10 in both source and backup DC. The size of data transfer (DA) for each request is uniformly distributed between 1 and 250 GB while the value of BW_{syn} is selected between 1 and 5 GB with equal distribution. The request holding time t_h is assumed to be exponentially distributed with mean equal to 1 time unit. Moreover, the value of k for $\pi_{sb,k}$ is set to 3. Simulation results are

Algorithm 1: Dynamic Provisioning of Survivable Service Request (DP-SSR)

$\mathcal{G}(\mathcal{V},\mathcal{E})$: a directed graph where $\mathcal{V}$ is the set of vertices representing the network nodes, $\mathcal{E}$ is the set of edges representing the network links;
D $\in \mathcal{V}$: set of datacenter locations; $\boldsymbol{W}$: maximum number of wavelengths on each link; **BW** : wavelength bandwidth; $\boldsymbol{SR_d}$: available servers at DC $d \in$ D;
$\boldsymbol{R = \{s, SR, DA, t_h, BW_{syn}\}}$: service request with source DC $s \in$ D, required number of servers SR, data size to be transferred DA, holding time t_h, and required bandwidth for updating BW_{syn};
$\boldsymbol{\pi^k_{sb}}$: the set of k shortest paths from source DC s to candidate backup DC $b \in$ D;

Initialization: $\widetilde{O} = -1$;
if $(SR \leq SR_s)$ **then**
  for each DC $d \in$ D $\setminus s$ **do**
    while $(SR \leq SR_d)$ **do**
      for each path $p^i_{sd} \in \pi^k_{sd}$ **do**
        Let $w_{p^i_{sd}}$ be the set of free wavelengths for p^i_{sd};
        if $w_{p^i_{sd}} \neq \{\emptyset\}$ **then**
          Let BG_{sd} be the set of background connections from s to d with free bandwidth larger than BW_{syn};
          if $(BG_{sd} \neq \{\emptyset\})$ **then**
            if $j \in BG_{sd}$ with free bandwidth $freeBW_j$ equal to BW_{syn} **then**
              $\Omega_j =$ BW $\cdot$ BW;
            else
              Select $j \in BG_{sd}$ with largest free bandwidth $freeBW_j$;
              $\Omega_j =$ BW $\cdot\, freeBW_j$;
            end if
          end if
        else
          Reserve $j \in w_{p^i_{sd}}$ for new background connection BG_{sd} and set $\Omega_j =$ BW;
        end if
        Compute $O^s_{d,i,j} = \alpha \cdot SR_d + \beta \cdot |w_{p^i_{sd}}| + \gamma \cdot \Omega_j$;
      end for
    end while
  end for
end if
Select d, p^i_{sd}, and j: $\widetilde{O}$= max $\{O^s_{d,i,j}\}$;
if $\widetilde{O} \neq -1$ **then**
  Return d, p^i_{sd}, j;
else
  Block R;
end if

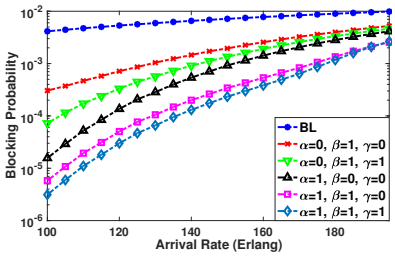


Fig. 2: Total BP.

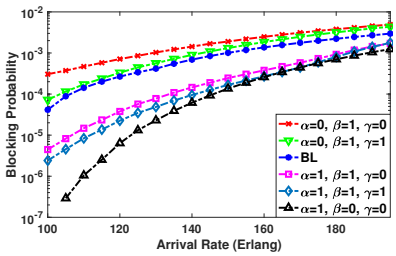


Fig. 3: BP due to lack of IT resources.

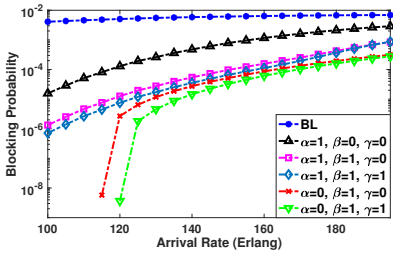
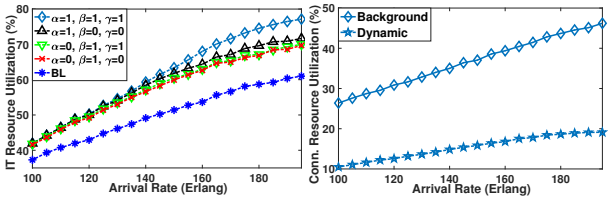


Fig. 4: BP due to lack of connectivity resources.



(a) **(b)**

Fig. 5: (a) Average IT resource utilization; (b) Average connectivity resource utilization for $\alpha = 1, \beta = 1, \gamma = 1$

compared against a baseline (BL) strategy that connects the source DC to the closest backup DC (with sufficient IT resources) using the shortest path with required connectivity resources.

Fig. 2 displays the total blocking probability (BP) as a function of the network load. Several observations can be made from the results shown in Fig. 2. First, the objective function tuned for overall resource optimization ($\alpha = 1, \beta = 1, \gamma = 1$) outperforms all the other options. Specifically, it reduces the total BP by 90% on average compared to BL strategy, while it exhibits an average decrease of the total BP by 67% and 58% with respect to the case when the objective function is tuned only for IT resources ($\alpha = 1, \beta = 0, \gamma = 0$) and connectivity resources ($\alpha = 0, \beta = 1, \gamma = 1$), respectively. Secondly, all the other combinations of the tuning coefficients demonstrate significant decrease in BP with respect to BL strategy. Third, appropriate selection of the background connection for synchronization traffic by considering it in objective function (scenarios with $\gamma = 1$) leads to an average decrease of BP by 15% compared to the case when it is excluded from the objective function ($\gamma = 0$). Note that in the latter situation (i.e., $\gamma = 0$), DP-SSR just picks the first candidate background connection for BW_{syn}. To obtain insight into the BP caused by lack of each type of resources, Fig. 3 and Fig. 4 highlight the BP due to unavailability of IT and connectivity resources, respectively. As expected, the objective function aimed at IT resources optimization shows better performance for BP caused by lack of IT resources, while BP due to insufficient connectivity resources is higher. Nevertheless, the objective function tuned for overall resource optimization achieves a beneficial trade-off between these two components of the total BP.

Fig. 5a shows the average IT resource utilization in each DC for different optimization objectives. The figure reveals that by accounting also for connectivity resources ($\alpha = 1, \beta = 1, \gamma = 1$), the IT resource utilization is improved notably (for higher loads) compared to the results for IT resource optimization only ($\alpha = 1, \beta = 0, \gamma = 0$). Similarly, Fig. 5b depicts the average connectivity resource utilization for the two types of optical connections when the objective function aims at maximizing both IT and connectivity resources. Dynamic connections that perform large data transfer in a fast way are short-lived and occupy less connectivity resources than background connections which provide bandwidth for synchronization traffic during the entire service time.

Conclusions

The paper presents software-defined architecture capable of managing both IT and connectivity resources for metro-scale inter- and intra-DC networks. To exploit this capability of the architecture, a dynamic provisioning algorithm that jointly allocate the IT and connectivity resources for survivable service requests is proposed. This algorithm provides a framework that can be flexibly adapted to optimize the utilization of a specific type of resource depending on its availability or cost. Simulation results show the benefit of the proposed algorithm, which offers reduction of the request blocking probability by more than 55% when aptly tuned for a realistic network scenario. These results emphasize the importance of joint orchestration of resources for converged metro-scale DC networks, in order to avoid poor blocking performance.

Acknowledgment

This work was jointly supported by the Swedish Research Council (VR) framework grant No. 2014-6230, Celtic-Plus sub-project SENDATE-EXTEND funded by Vinnova, H2020-ICT-2014 project 5GEx (Grant Agreement no. 671636), CIAN NSF ERC (EEC-0812072), NSF NeTS (CNS-1423105), and DoE ASCR under Turbo Project DE-SC0015867.

References

[1] 5G PPP Architecture Working Group, “View on 5G architecture,” white paper, (2016).
[2] M. Schiano et al., “Flexible node architectures for metro networks,” IEEE/OSA JOCN, (2015).
[3] G. Chen et al., “First demonstration of holistically-organized metro-embedded cloud platform with all-optical interconnections for virtual datacenter provisioning,” Proc. OECC, (2015).
[4] S. Yan et al., “Archon: A function programmable optical interconnect architecture for transparent intra and inter data center SDM/TDM/WDM networking,” IEEE/OSA JLT, (2015).
[5] P. Samadi et al., “Software-defined optical network for metro-scale geographically distributed data centers,” OSA Optics Express, (2016).
[6] M. Fiorani et al., “Flexible network architecture and provisioning strategy for geographically distributed metro data centers,” IEEE/OSA JOCN, (2017).