

Design of a routing scheme for bulk data transfers in multi-domain SnF OCS networks

Qian Liu*, Xiao Lin[†], Shengnan Yue*, Weiqiang Sun*, Weisheng Hu*

*State Key Laboratory of Advanced Optical Communication Systems and Networks
Shanghai Jiao Tong University, Shanghai, China

[†]College of Physics and Information Engineering, Fuzhou University, Fuzhou, China
Email: qianliu4@sjtu.edu.cn, linxiaocer@fzu.edu.cn, sunwq@sjtu.edu.cn

Abstract—Rapid growth of overloaded traffic leads to bottlenecks on inter-domain links. There may not be sufficient bandwidth for end-to-end transmission upon the arrival of transfer bulk data requests. Storage can be introduced into the forwarding path so that the delay tolerant data can be temporarily stored and forwarded when network links become less congested. In this paper, we study the *Store-and-Forward* (SnF) problem in a multi-domain environment. We present a routing scheme based on a framework named TS-MLG (time-shifted multilayer graph) to perform inter-domain bulk data transfer by taking both spatial and temporal aspects into consideration. With this scheme, the cross-domain routing path with storage and transmission information, can be determined through conventional routing computations. An updating strategy is proposed to realize temporal bandwidth information exchange between physical and logical layers. Numerical simulations reveal that SnF can improve the inter-domain bandwidth utilization and optimize the network performance at the practical cost of maintenance overhead and computational complexity.

Keywords: multi-domain, bulk data, bandwidth, store-and-forward, routing

I. INTRODUCTION

The rapid growth of inter-domain traffic causes bottleneck links, and optimizing these links are critical[1–4]. As the residual bandwidth of inter-domain links exhibit temporal and spatial varying characteristics, there may not be sufficient bandwidth resource for end-to-end reservation and transmission on the arrival of request. If the request can be temporarily stored until there is sufficient bandwidth for forwarding, blocking can be avoided. Although the internet is inherently a multi domain network, existing researches on bandwidth optimization with *store-and-forward* (SnF) are still carried out in single domain network perspective[5–7].

Multi domain lightpath provisioning is a challenging problem for limited information exchanged between domain, hierarchical routing is a promising approach as it can compute routes in a skeleton or incremental manner and expanding them to a complete end-to-end lightpath via inter-domain signaling [8–12]. The conventional routing process only consider the current state and cannot find such a spatial-temporal path. But for SnF, it should be able to schedule current and future bandwidth resources for request, facing whether it should be stored, how long it will be stored, and when it will be transferred.

Besides, the use of storage introduces additional complexities into routing process. TS-MLG (time-shifted multilayer graph) [7] is a routing framework for bulk data SnF in optical circuit switch (OCS) networks that greatly simplifying the spatial assignments and temporal arrangements process. The framework is applicable to the scenario of complete "global" information. Considering that the abstraction layer on hierarchical routing is a single domain view, we can adopt to the principle of TS-MLG to compute a skeleton path over time and space. The TS-MLG update is only related to the routing results. As for hierarchical routing, the logical layer update not only depends on the routing results of the physical domain, but also the abstract scheme. So, the question sets on how to update the TS-MLG on the abstraction layer according to dynamics of the physical domain.

In this paper, after we extend TS-MLG to traditional hierarchical routing to handle inter-domain spatial-temporal schedule, we also propose a change-based update policy consider both the abstraction layers skeleton path routing results, the physical layers segment signaling and the topology aggregation scheme, to adapt our routing architecture. Moreover, simulation results show that introduce SnF can optimize the inter-link bandwidth utilization and improve the network transfer performance at the practical cost of maintenance overhead and computational complexity.

The rest of the paper is organized as follows. In section II, we review the relevant works on SnF and hierarchical routing. The description of routing and updating schemes are given in Section III. In Section IV, we present simulation results with the numerical analysis, the conclusion and future work directions are drawn in Section V.

II. RELATED WORK

A. Provisioning Process of SnF

Introducing storage to network brings an additional complexity into the routing process, for we need to arrange spatial resource allocation with temporal arrangements. The work in [7] proposes a routing framework named TS-MLG that greatly simplifies the provisioning problems. A TS-MLG is a multilayer graph comprised of a series of snapshots, i.e., layers. Each layer depicts the network dynamics at a certain time. As is shown in Fig.1(b), solid lines connect different nodes with the same layer represent the bandwidth capacity.

Dash lines between the same node pairs in different layers refer to the storage capacity of these nodes. The red arrow lines represent a routing path; traversing a spatial link (the red solid line) denotes physical delivery of data transfers from one node to another; traversing a temporal link (the red dash line) denotes storing the request on the particular node for a certain period of time. As illustrated, when a request from node3 to node6 arrives at time t_0 , an end-to-end path is not possible at the current moment, as the network has insufficient bandwidth as shown by Snapshot at t_0 . By performing shortest path routing on the TS-MLG, an end-to-end path, i.e., 3-5-5'-6', is determined, and the request stored in intermediate storage node5 from t_0 to t_1 .

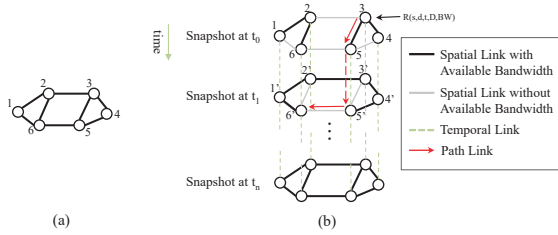


Fig. 1: Example of the routing process with TS-MLG. (a) Original physical topology. (b) Schematic of a TS-MLG.

B. Multi Domain Hierarchical Routing

For confidentiality issue, the intra-domain will provide limited information, and topology aggregation is proposed for information compression by condensing the original topology to an abstract topology with fewer virtual links and nodes. Hierarchical routing is in a top-down manner [9–11, 13, 14] by first computing a skeleton path upon the upper abstract topology, then performing expansion signaling based on the link state and resource database in the underlying layer at present. At last, when all domains complete their establishment successively, the intra-domain subpath will be merged into one integrated end-to-end lightpath. Noted that, no matter what update policy is used, timer- or change-based, the upper layer can only access to the current network state. So the skeleton path and the detailed end-to-end lightpath are all with the current timestamp.

III. ROUTING AND UPDATING SCHEME

In this paper, we extend TS-MLG to traditional hierarchical routing and propose a two-layer routing scheme to handle spatial-temporal with inter-domain schedule and update. The upper layer is on an abstract topology view and refers to the principle of TS-MLG for spatial-temporal network maintenance and routing, while the underlying recording the intra and inter domain dynamics responsible for skeleton path signaling.

A. Design Rationale and Assumption

We take full mesh topology aggregation to implement internal topology reduction. The border optical switching nodes are assumed to be capable of wavelength conversion and equipped with unlimited storage capacity. Assuming the tuple $R(s,d,t,D,BW)$ defines a request arrives at t , where D is the

duration and BW is the required bandwidth, s and d indicates the source and destination. Noted that requests are delay tolerant bulk transfers with large volume, i.e., the duration is range from hours to days. So, the processing overhead and the propagation delays between two nodes can be ignored compared with the duration and transmission delays. Once an integrated end-to-end lightpath is admitted, the related storage and wavelength resource are reserved for serving. These resources can be used for other requests prior to the reservation time.

B. Overview of the Proposed Solution

Fig.2 depicts the TS-MLG-based Hierarchical routing scheme. As the key issues of routing process, the upper layer LTS-MLG (Logical TS-MLG) records the bandwidth dynamics. Topology Aggregation Policy rules the abstract node/link and physical node/link mapping relationships. The local TED (Traffic Engineering Database) of the physical domain maintains the intra domain and inter-link time-related dynamics. If no skeleton path exists in the upper LTS-MLG or can't find physical path mappings, the request will be blocked and lost.

C. Principles of Hierarchical SnF Routing

As shown in Fig.2, The blue arrow lines depict the routing scheduling process, while the red denote the updating process:

1) Principles of Routing:

- *Step 1:* We only use busy/idle mode to establish the availability model based on the aggregated bandwidth information. When a new request arrives(1), analyze the source and destination nodes location, if intra domain business, go to the next step; otherwise, go to *Step3*;
- *Step 2:* Extract the related portion of (s,d) from the whole LTS-MLG as Fig.3(a) shows. Compute a logical skeleton path over the partly LTS-MLG (2), block R if fails; otherwise, go to *Step4*;
- *Step 3:* Routing over the LTS-MLG to get a skeleton path as Fig.3(b) shows(2). Block R if fails; otherwise, go to *Step4*;
- *Step 4:* Choose the proper physical mappings for each logical segment according to the Local and Inter Topology TED. If any physical mapping operation fails, reject R ; otherwise, combine all physical mappings forming an integrated end-to-end lightpath(3)(4).

As showed in Fig.4, *1-6-4* and *1-4* are two alternative physical mappings of skeleton path segment 4-5. Noted that, italics denote the physical mappings and the normal fonts are abstract node pairs. When *1-6-4* or *1-4* has available bandwidth during time t_1 to t_1+D , 4-5 can find its physical mapping. Otherwise, if neither *1-6-4* nor *1-4* have available bandwidth within a certain time interval, signaling process would be failed.

2) *Principles of Updating:* We illustrate how physical end-to-end lightpath establishment and release influence the bilevel updating through a specific example below. As described above in *Step4*, the physical mapping of the skeleton path

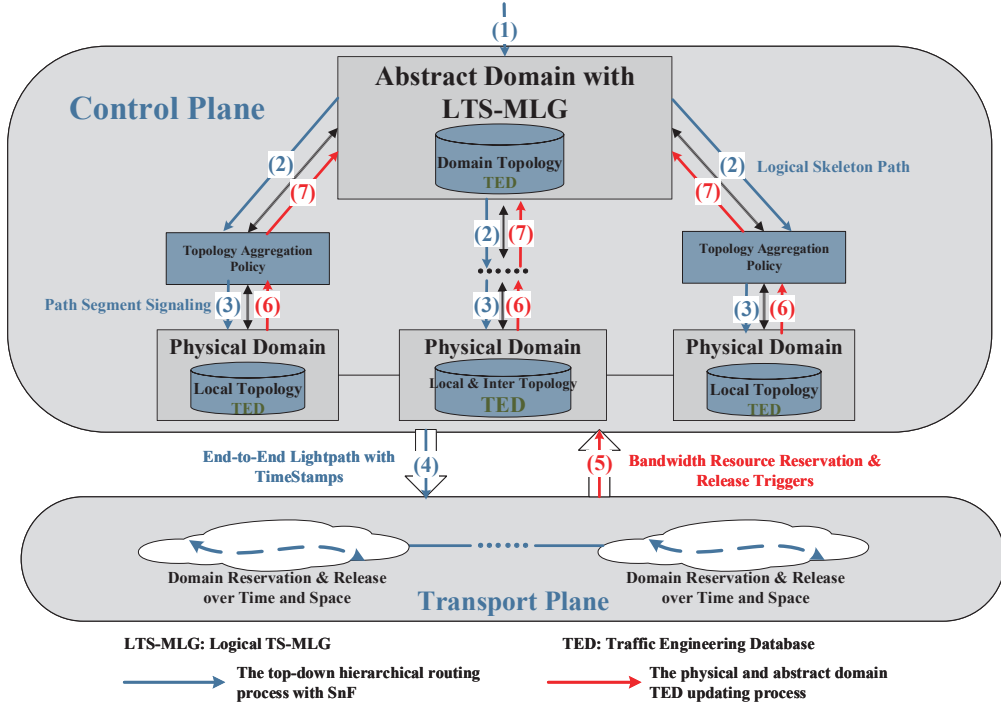


Fig. 2: Overview of the TS-MLG-based Hierarchical routing architecture.

segment 4-5 is 1-6-4 at time t_1 . 4-6 (1-6-5 and 1-2-5) and 6-5 (5-6-4 and 5-4) are relevant to 1-6 and 6-4 respectively according to Topology Aggregation.

- *Step 1(Physical updating process)*: All snapshots between time interval $[t_1, t_1+D]$, i.e., t_1 and t_1' in Fig.4, will be updated by removing the bandwidth resource along subpath 1-6-4. As for the snapshot at the moment t_1+D , release the bandwidth used by R (5);
- *Step 2(LTS-MLG updating process)*: We combine physical link dynamics and topology aggregation policy to determine how to update the logical links(6)(7). As stated, the dynamic of 1-6-4 affects 1-6-5 and 5-6-4 at the same time. In addition to logical segment 4-5, 4-6 and 6-5 in LTS-MLG are also need to be updated.

Fig.5 depicts an example of the dynamics of LTS-MLG. When one logical segment establishes its physical mapping shown in Fig.4, the LTS-MLG in Fig.5(a) lack of a snapshot of t_1' due to network information compression caused by topology aggregation. So, the updating process of LTS-MLG shown in Fig.5(b) should add a new

snapshot of t_1' to reflect the network dynamics. The blue block emphasizes the part that needs to be updated. When some logical links removed or released, the remaining unrelated parts of the snapshot just keeps constant.

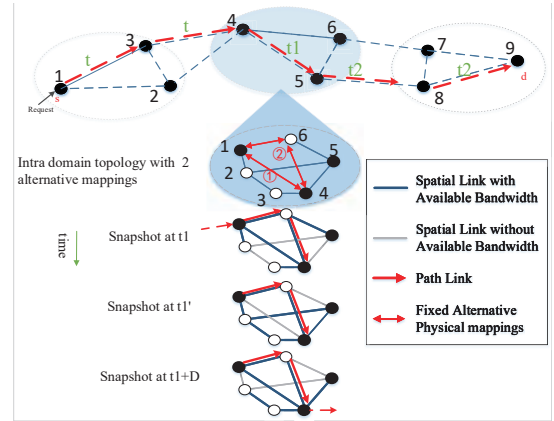


Fig. 4: Physical domain signaling.

IV. PERFORMANCE EVALUATION

We adopt a 3-domain 12-node multi domain network model depicted in Fig.6. Request arrivals are assumed to be independent and uniformly distributed among all abstract nodes with a Poisson process λ , the D is assumed to be exponentially distributed with μ . The load $\rho = \lambda/\mu = 5$, and μ is 0.2. Herein, one bidirectional wavelength per intra-domain link and two in the inter-domain.

The blocking performances without storage are given in Table I, as the baseline. We define a value K to represent

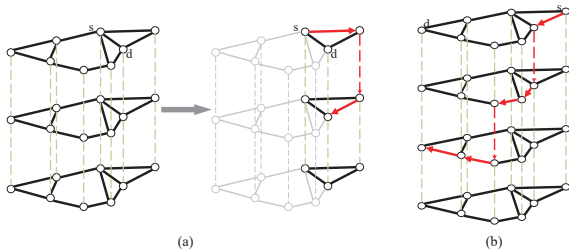


Fig. 3: LTS-MLG skeleton path routing. (a) Intra domain skeleton routing. (b) Inter domain skeleton routing.

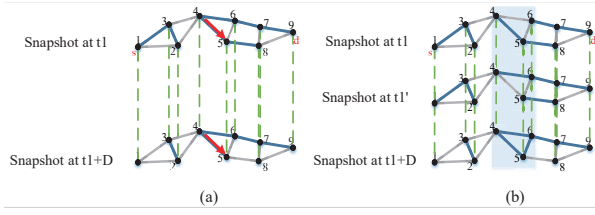


Fig. 5: Example of the dynamics of LTS-MLGs. (a) LTS-MLG responding to request R at t1. (b) LTS-MLG updating according to physical segment signaling results and topology aggregation scheme.

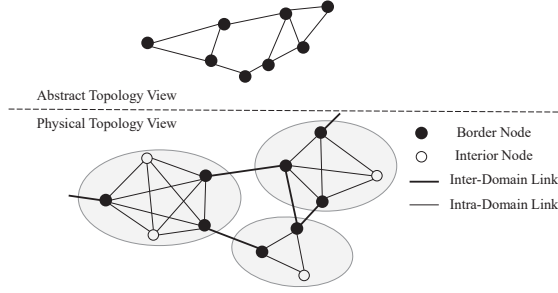


Fig. 6: Simulation model.

the number of physical mappings of the same (s, d) node pair. Layer in the table records the maximum number of layers that the end of the requests transfer in LTS-MLG. SnF Ratio is defined as the ratio of stored requests to totally succeed requests. Delay is defined as the time between the request arrival and the end of data transfer. Because μ is 0.2, the average duration of requests is $1/0.2$ time units. We averaged the results over 10 simulation runs, each with 100,000 requests.

We assume a request can be stored at any abstract node and defer its transmission indefinitely. We investigate the blocking performance with different values of K , and compare with the traditional hierarchical routing under the same network

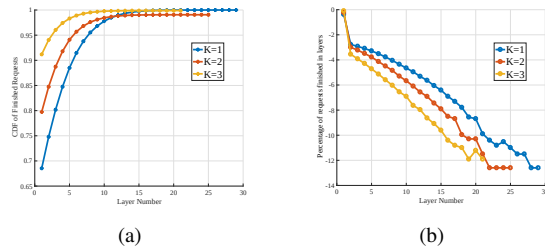


Fig. 7: PDF and CDF of requests finished on different layers with different K .

TABLE I: PERFORMANCE OF THE NETWORK WITHOUT STORAGE

K	Blocking Probability	Average Delay	Inter-Domain Link Utilization	Simulation Time
1	0.1662	5.0000	0.2474	4829
2	0.1182	5.0000	0.2504	5022
3	0.0545	5.0000	0.2565	4679

settings. The statistical results in Table I and Table II reveal that the success rate of SnF outperforms the control group, and the results show that the blocking performance can be almost avoided via SnF.

Fig.7 shows the Cumulative Distribution Function (CDF) and the Probability Density Function (PDF) of requests finished on different layers. Fig.7(a) shows that the routing process can be completed at 27 layers, and as the number of layers increase, the requests finished rate progressively increases. Fig.7(b) shows that the majority of requests could be served immediately, and only a few requests required multiple SnFs. In detail, when K is set to 3, more than 92% of requests finished on the first layer, which means less than 8% of the requests required SnF. Moreover, percentages of requests finished on the same layer increases as K increases.

We can observe that a larger K corresponds to better network performance. This is because more edge-disjoint shortest pairs means greater load balancing capabilities. What's more, it's a trade-off between network performance and layers, i.e., the computation complexity.

Routing with limited layers, i.e., limited number of SnF, is considered to mitigate these issues. Fig.8(a) indicates the blocking probability decreases as the layer number increases. When the limited layer is 15, the blocking probability of $K = 3$ can be down close to 0. As the curvilinear trend in Fig.8(b) shows, a larger limited layer also means better use of the inter domain bandwidth resource.

V. CONCLUSIONS

In this paper, we have studied introduce assistive storage into multi domain OCS networks for delay tolerant data transfers. We propose a routing and updating scheme to cope with the spatial-temporal scheduling problems. simulation results suggest SnF can optimize the network performance and improve the inter-link bandwidth utilization at the cost

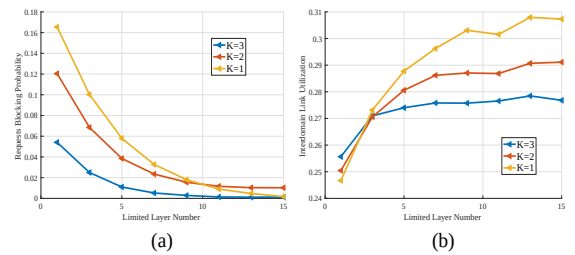


Fig. 8: Network performance under a given layer. (a)Blocking probability. (b)Inter-domain link utilization.

TABLE II: PERFORMANCE OF THE NETWORK WITH STORE AND FORWARD

K	Layer	Blocking Probability	Average Delay	Avg SnF	Inter-Domain Link Utilization	Simulation Time
1	29	0	6.3142	0.3146	0.3057	7304
2	25	0.0094	5.6478	0.1950	0.2891	6479
3	21	0.0011	5.2058	0.0871	0.2785	6273

of maintenance and computation overhead. Since the multi-domain routing is sensitive to the burden of the control plane, we intend to optimize the skeleton routing algorithm in combination with LTS-MLG features in the future to reduce the computation complexity and signaling overhead caused by temporary storage.

ACKNOWLEDGMENT

The work is supported by the National Natural Science Foundation of China (NSFC) (61433009, 61671286, and 61471238).

REFERENCES

- [1] M. Zhu, J. Li, Y. Liu, D. Li, and J. Wu, "Ted: Inter-domain traffic engineering via deflection," *IEEE 22nd International Symposium of Quality of Service (IWQoS)*, pp. 117–122, 2014.
- [2] T. Bressoud, R. Rastogi, and M. Smith, "Optimal configuration for bgp route selection," *IEEE INFOCOM*, pp. 916–926, 2003.
- [3] C. David D and *et al.*, "Measurement and analysis of internet interconnection and congestion," *Social Science Electronic Publishing*, 2014.
- [4] C. Zhang and *et al.*, "Field trial of bandwidth on demand services based on hierarchical control over multi-domain on networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 11, pp. 1057–1063.
- [5] L. Tong, X. Zheng, Y. Xia, and M. Li, "Delay tolerant bulk transfers on inter-datacenter networks," *IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2016.
- [6] N. Laoutaris, M. Sirivianos, X. Yang, and P. Rodriguez, "Inter-datacenter bulk transfers with netstitcher," *Proceedings of SIGCOMM*, pp. 74–85, August 2011.
- [7] X. Lin, W. Sun, M. Veeraraghavan, and W. Hu, "Time-shifted multilayer graph: a routing framework for bulk data transfer in optical circuit-switched networks with assistive storage," *J. Opt. Commun. Netw.*, vol. 8, no. 3, pp. 162–174, Mar. 2016.
- [8] I. Nishioka, Y. Iizawa, and S. Araki, "Multi-domain ason/gmpls network operation: Current status and future evolution," *Proc. SPIE*, vol. 6784, 67840T, 2007.
- [9] M. Chamanian and A. Jukan, "A survey of inter-domain peering and provisioning solutions for the next generation optical networks," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 33–51, First Quarter 2009.
- [10] K. Liang, M. Rahnamay-Naemi, H. M. K. Alazemi, N. MinAllah, M. Peng, and N. Ghani, "Optimization approach for multi-domain optical network provisioning," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 5, no. 12, pp. 1413–1424, 2013.
- [11] J. Zhang, Y. Zhao, X. Cao, W. Gu, and Y. Ji, "Issues on routing in multi-layer and multi-domain optical networks," *9th International Conference on Optical Communications and Networks (ICOON 2010)*, pp. 125–127, 2010.
- [12] S. Yan and *et al.*, "First demonstration of all-optical programmable sdm/tbm intra data centre and wdm inter-dcn communication," *2014 The European Conference on Optical Communication (ECOC)*, pp. 1–3, 2014.
- [13] Q. Liu, M. Kok, N. Ghani, V. Muthalaly, and M. Wang, "Hierarchical inter-domain routing in optical dwdm networks," *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pp. 1–5, 2006.
- [14] X. Yang and *et al.*, "Policy-based resource management and service provisioning in gmpls networks," *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pp. 1–12, 2006.



Qian Liu is currently working towards a M.S. degree in the school of electronic information and electrical engineering in Shanghai Jiao Tong University, China. Her research interests include bulk data transfer in large scale optical networks.



Xiao Lin is an assistant professor in the Department of Internet of Things Engineering, College of Physics and Information Engineering at Fuzhou University, China. He received his Ph.D. degree in information and communication engineering at Shanghai Jiao Tong University. His research interests include intelligent edge computing, optical switches, reliability, and bulk data transfer in optical networks.



Shengnan Yue is a Ph.D. Candidate from electronic information and electrical engineering in Shanghai Jiao Tong University in China. Her research interests include bulk data transfer and delay-tolerant networks.



Weiqliang Sun is a professor in the State Key Laboratory of Advanced Optical Communication Systems and Networks at Shanghai Jiao Tong University, China. His primary research interests include dynamically configured optical networks, hybrid switching systems, and pricing in next-generation networks.



Weisheng Hu is the Director of the State Key Laboratory of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University, China. He serves on panels for four journals and 12 conferences, including IEEE JLT, OFC, APOC, LEOS/PS, CLEO/PS, ICICS, INFOCOM and ICC workshop. His interests are in optical transport networks with GMPLS control, and optical packet switching. He is the co-author of over 200 peer-reviewed journal and conference papers.