Low latency network topology using multiple links at each host

Ryuta Kawano†, Ikki Fujiwara††, Hiroki Matsutani†, Hideharu Amano†, Michihiro Koibuchi††

†Keio Univ. JAPAN, ††National Institute of Informatics / JST JAPAN

Email: blackbus@am.ics.keio.ac.jp

Introduction

Target: Low latency network topology for HPC interconnects

•Network latency between hosts (computation nodes) is a pressing concern for **parallel application** on large-scale High Performance Computing (HPC) systems.

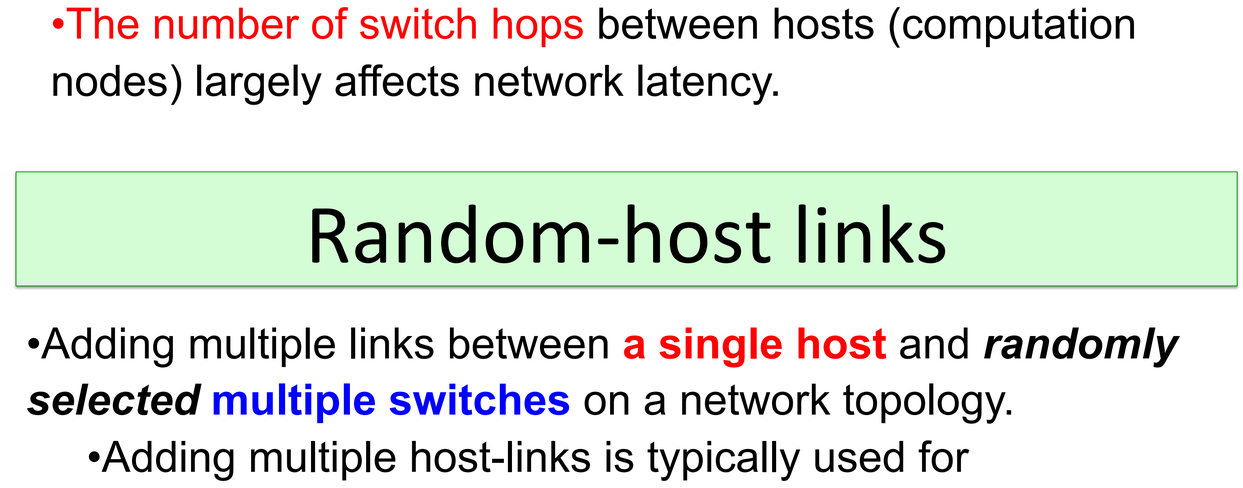
•The number of switch hops between hosts (computation nodes) largely affects network latency.

Random-host links

•Adding multiple links between **a single host** and ***randomly selected*** **multiple switches** on a network topology.

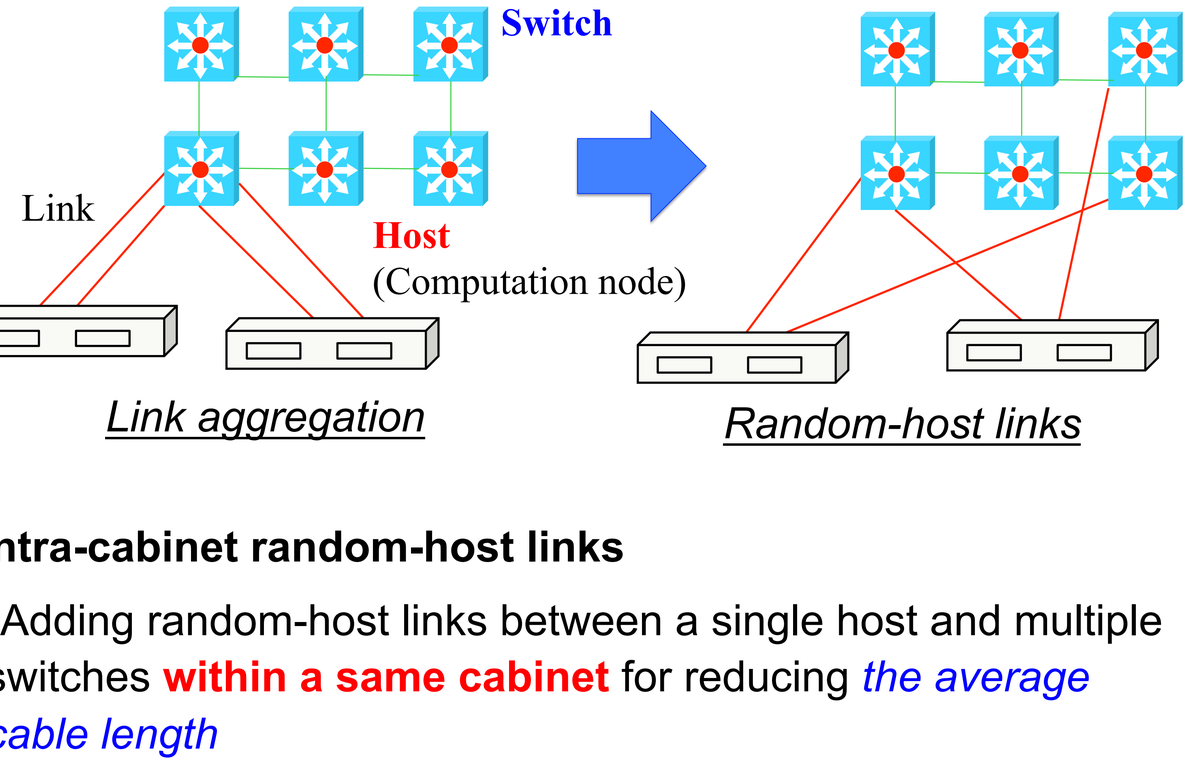
•Adding multiple host-links is typically used for **link aggregation** in a view of increasing bandwidth.

•Achieving lower switch hops between hosts by exploiting *small-world effects*.



Intra-cabinet random-host links

•Adding random-host links between a single host and multiple switches **within a same cabinet** for reducing *the average cable length*



Evaluation

Evaluation environment

Max / Average switch hops between hosts

Number of SWs	1,024
Number of hosts	8,192
Cabinet size	144 nodes/cab.

Average cable length

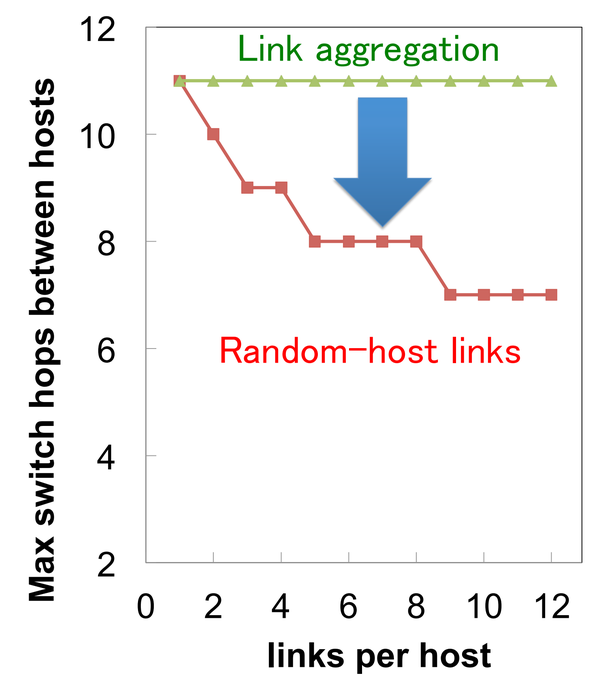
Number of SWs	1,024
Number of hosts	8,192
Cabinet size	144 nodes/cab.
Area occupied by one cabinet	0.6 x 2.1[m]
Cabling overhead	2[m]

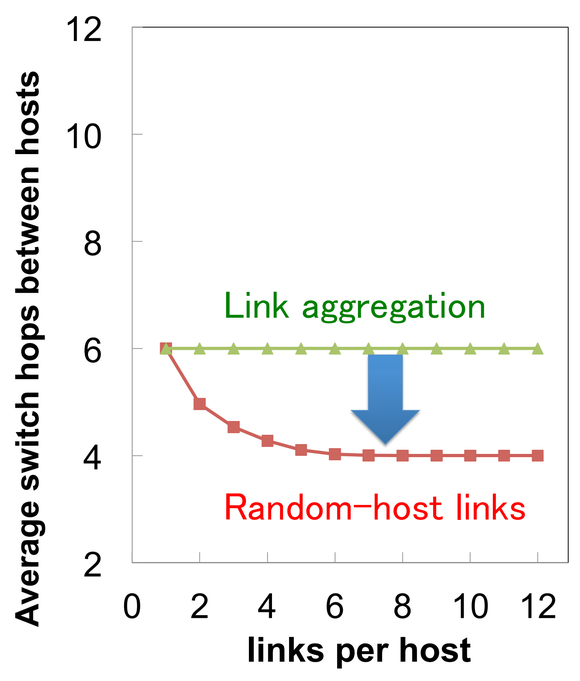
Latency

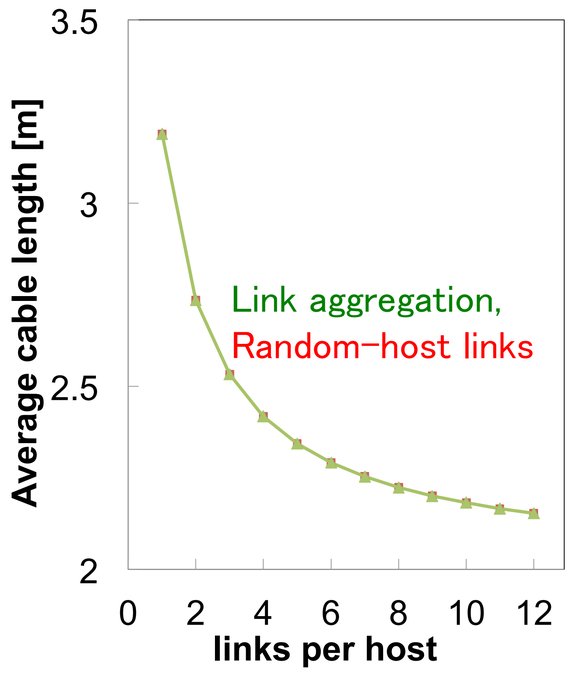
Number of SWs	256
Number of hosts	256
Number of links	4
Traffic pattern	Uniform
Packet length	33-flit
Flit length	256-bit
Number of VCs	4
Switching delay	> 100ns
Wire delay	5ns/m
Cabinet size	32nodes/cab.
Routing	Duato[2]
Escape paths	up*/down*

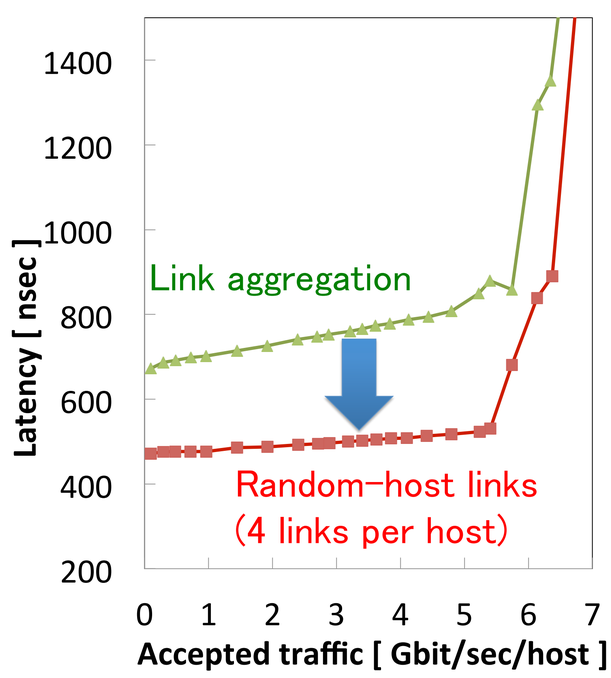
Evaluation result

•Compared to using link aggregation, using random-host links
- •largely reduces switch hops between hosts.
- •achieves almost the same average cable length.
- •reduces the network latencies by up to 40%.









References

[1] M. Koibuchi and et al. “A Case for Random Shortcut Topologies for HPC Interconnects”, Proc. of the International Symposium on Computer Architecture (ISCA) 177–188, 2012.

[2] F. Silla and J. Duato “High-Performance Routing in Networks of Workstations with Irregular Topology”, IEEE Trans. on Parallel Distrib. Syst. 699–719, 2000