Computing a path to more profits

The benefits of a centralized Path Computation Element using Bell Labs Self-Tuned Adaptive Routing

Technology White Paper

According to recent Nokia Bell Labs study results, network operators can optimize network utilization and support more revenue-generating traffic with a centralized Path Computation Element (PCE) based on the Bell Labs Self-Tuned Adaptive Routing (STAR) algorithm. Using STAR, network operators can improve ROI in their IP/Multiprotocol Label Switching (MPLS) networks by accommodating more service requests with the installed network capacity and can improve service quality by reducing link-congestion situations.

IP/MPLS networks use link-state protocols such as Open Shortest Path First (OSPF) and Intermediate System to Intermediate System (IS-IS) to route Label Switched Paths (LSPs). Although network operators can apply various link metrics and path constraints to influence route selection, congestion points and deadlock situations can easily occur on the most popular routes while leaving bandwidth stranded on the less-used routes.

A software-defined networking (SDN) approach with a centralized PCE addresses these issues by helping network routing protocols to select more efficient paths. The Bell Labs STAR algorithm is designed to balance link utilization and network utilization, helping network operators to achieve their objectives of optimized network utilization and increased revenues.

Contents

Introduction

With virtually all services now converged on IP, effective management and control of IP traffic is challenging. Network routing protocols such as Open Shortest Path First (OSPF) and Intermediate System to Intermediate System (IS-IS) operate autonomously in a distributed control plane. Network operators need effective tools to help ensure that network assets yield the highest ROI. Moreover, the emerging cloud era and new traffic dynamics require elastic networks that can adjust to rapid demand fluctuations without failing.

In this context, software-defined networking (SDN) introduces a powerful, centralized control plane. The SDN control plane is more open and accessible by IT and operations support system (OSS) applications than the traditional Simple Network Management Protocol (SNMP) management information bases (MIBs) and command-line interface (CLI) device scripts used for configuring a network-embedded control plane.

Centralized PCE

A Path Computation Element (PCE), defined in RFC 4655,[1] applies SDN concepts to assist network-embedded routing protocols in selecting paths that more efficiently utilize the available network capacity, carry more revenue-generating traffic, and improve overall service performance. The PCE concept separates path computation and path signaling to give operators more control over their networks, enables the application of network operator policies, and addresses concerns regarding multivendor integration.

A centralized PCE can use sophisticated path-computation algorithms to optimize IP/Multiprotocol Label Switching (MPLS) networks and link-state protocols such as OSPF and IS-IS. These network routing protocols use Dijkstra's algorithm[2] to select the shortest open path based on applicable path constraints.

Network routing protocols are designed for speed—fast route-convergence times—and scalability, which is critical because memory and processing resources are limited. However, links on popular routes can easily become congested and can cause network hotspots while other links remain underutilized. Control of routing outcomes by manually changing link metrics is difficult and risky because these changes may have far-reaching consequences.

With no constraints on available memory, processing resources or route-convergence times, a centralized PCE can consider multiple possible routes to find the optimal path. A centralized PCE can also employ path-computation algorithms that use dynamic metrics based on link utilization without risking potential route instabilities in the network control plane itself.

---

[1] IETF, RFC 4655: "A Path Computation Element (PCE)-Based Architecture," August 2006. https://tools.ietf.org/html/rfc4655

[2] Dijkstra's algorithm finds the shortest path between a given source node and all other nodes in a graph.

Benefits of a centralized PCE with Bell Labs STAR algorithm

A recent Nokia Bell Labs study reported that the combination of a centralized PCE with the Bell Labs Self-Tuned Adaptive Routing (STAR) algorithm can yield significant benefits for network operators:

- Depending on topology size, the use of STAR can yield 12 to 24 percent more revenue-generating traffic by using deployed capacity more efficiently and accepting more services than OSPF routing algorithms.
- STAR balances network traffic over available links, reducing the chance of network hotspots and creating sufficient headroom to absorb traffic peaks and demand growth without regular traffic engineering.
- Network operators can apply STAR selectively to a subset of paths or portions of the network while retaining the overall benefits of balanced network traffic, improved service performance and higher yields.
- Network operators can apply STAR to existing IP/MPLS networks and routing protocols in multivendor deployments. Operators can also use STAR in a complementary role with offline planning and traffic engineering tools.

By deploying centralized path computation based on the Bell Labs STAR algorithm, network operators can increase profitability and optimize their network investments while reducing the chance of network hotspots as a result of link congestion.

STAR path-computation algorithm

The STAR algorithm addresses the optimization of network utilization—a key concern for network operators, with a direct relation to ROI and overall profitability. Network capacity is a finite resource that is allocated on behalf of service requests made by customers. From a mathematical perspective, this is a multidimensional bin-packing problem, with each service request representing an item of a certain volume and value. As the network fills up, it becomes progressively harder to accommodate more service requests.

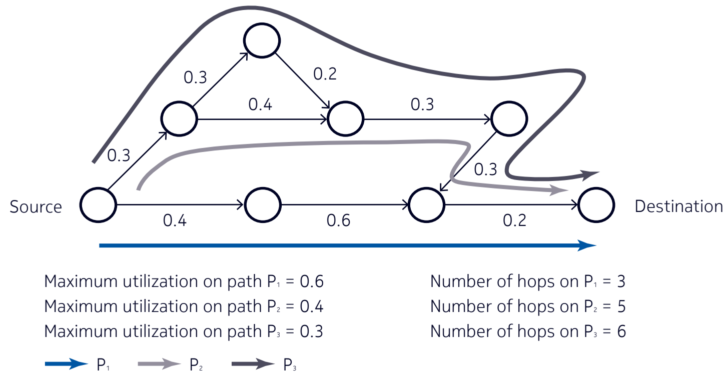
Optimized network utilization

For optimized network-capacity utilization, a routing algorithm must meet two objectives:

- Efficiency: Consume the least total network bandwidth possible to optimize resource usage.
- Balance: Avoid overloading any links to avoid congestion and deadlock situations.

The objectives of efficiency and balance may be contradictory, requiring trade-offs. Figure 1 shows an example of a trade-off for a given topology and link utilization.

Figure 1. Balancing link utilization and congestion avoidance



There are three different paths to consider in accommodating a path request between the source and destination nodes: P_1, P_2 and P_3. The options are:

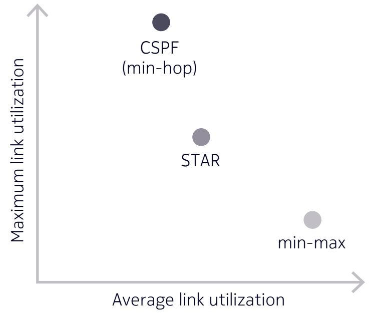
- Select the longer six-hop path P_3 to avoid the link with utilization of 60 percent on the much shorter three-hop path P_1
- Select the most cost-effective path for this request or risk the lock-out of links for future requests.

A Constrained Shortest Path First (CSPF) or min-hop routing protocol in OSPF and IS-IS networks selects path P_1 with the fewest hops unless administrative link metrics or explicit route constraints dictate otherwise. CSPF achieves efficiency, but at the expense of balance.

An alternative routing protocol such as min-max would ensure that the maximum current link utilization is as small as possible and pick path P_3. Min-max achieves balance at the expense of much longer routes, which also negatively impacts network utilization and ROI.

The Bell Labs STAR algorithm is designed to find the optimal balance between link utilization and network utilization, as shown in Figure 2. The algorithm avoids congested links while ensuring that the path does not consume too much network bandwidth.

Figure 2. Bell Labs STAR algorithm

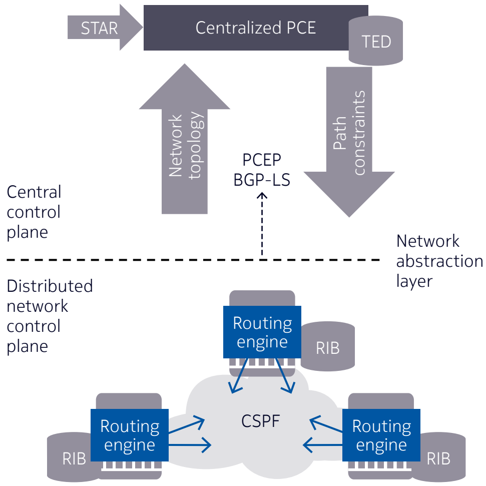


Interaction between centralized PCE and routing protocols

STAR performs a series of alternative path calculations for each path request, including min-hop and min-max, and compares link- and network-utilization results. The optimal path is subsequently conveyed to the network control plane as an EXPLICIT_ROUTE object (ERO)[3,4,5] using the PCE Communication Protocol (PCEP), defined in IETF RFC 5440.[6]

Routing protocols such as OSPF and IS-IS can implement the computed source routes using their existing routing functionality. A variety of Path Computation Clients (PCCs),[7] such as network management systems, customer self-service portals, and traffic engineering applications, can use the PCEP to make path-computation requests on behalf of CSPF routing engines in the network control plane, as shown in Figure 3.

Figure 3. Interaction between centralized PCE and distributed routing protocols



Network layers and path-computation algorithms

The PCE can be considered as part of a central control plane that communicates with the distributed network control plane through a network abstraction layer. The network abstraction layer used by a PCE consists of various open and standards-based protocols to acquire the topology and link states of the underlying routing domains in multivendor environments.

---

[3] IETF, RFC 3209: “Extensions to RSVP for LSP Tunnels,” December 2001. http://www.rfc-base.org/txt/rfc-3209.txt

[4] IETF, RFC 3473: “Codepoint Registry for the Flags Field in the Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Session Attribute Object,” April 2007. https://tools.ietf.org/rfc/rfc4859.txt

[5] IETF, RFC 3477: “Signalling Unnumbered Links in Resource Reservation Protocol - Traffic Engineering (RSVP-TE),” January 2003. https://tools.ietf.org/html/rfc3477

[6] IETF, RFC 5440: “Path Computation Element (PCE) Communication Protocol (PCEP),” March 2009. https://tools.ietf.org/html/rfc5440

[7] IETF, RFC 5862: “Path Computation Clients (PCC) - Path Computation Element (PCE) Requirements for Point-to-Multipoint MPLS-TE,” June 2010. https://tools.ietf.org/html/rfc5862

Network topology, link-state information, and deployed and allocated capacity are tracked in a Traffic Engineering Database (TED).[8] Multiple PCEs may be involved to address different network layers—for example, an IP PCE for the routing layer and a transport PCE for the optical layer—or network domains, such as different autonomous systems.

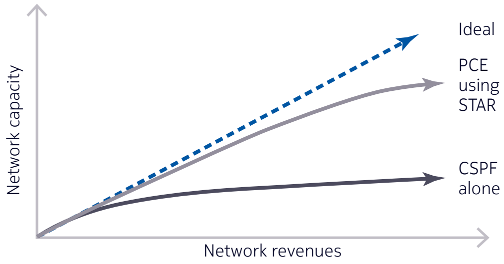
While the architecture, role and interface protocols of the PCE are standardized by the IETF, the actual path-computation algorithms are open because different network layers require different optimization algorithms. An algorithm such as STAR, which balances network traffic over all available links, optimizes resource efficiency and minimizes latency in the IP/MPLS layer. However, the optical layer requires different algorithms to minimize the number of optical-electrical conversions and the use of wavelengths while latency is determined predominantly by the path length. PCE algorithms are also open to vendor differentiation.

A centralized IP PCE can apply sophisticated path-computation algorithms to complement and optimize distributed, network-embedded CSPF routing algorithms, which are constrained by compute resources and the need for fast convergence times. A centralized PCE capability also plays an important role in the IETF Segment Routing standards.[9] Segment Routing improves the scalability of the MPLS control plane by reducing state information and signaling in the network but must rely on a centralized PCE to maintain state information such as allocated link capacity.

Study objectives and approach

The objective of the Bell Labs financial network modeling study was to determine the extent to which a PCE with the STAR algorithm can improve network utilization and convert these efficiency gains into incremental revenue gains compared to CSPF. In the theoretical ideal case, deployed network capacity would be fully utilized for revenue-generating traffic, as shown in Figure 4.

Figure 4. Comparison of CSPF and STAR path optimization



---

[8] IETF, RFC 6825: "Traffic Engineering Database Management Information Base in Support of MPLS-TE/GMPLS," January 2013. https://tools.ietf.org/html/rfc6825

[9] IETF, Segment Routing web site. http://www.segment-routing.net/home/ietf

The study evaluated link utilization and revenue generation for two different topologies: a small six-node core topology and a larger 59-node network topology. For each topology, a traffic matrix was defined to represent the forecast traffic demand for all service requests. All links were dimensioned to accommodate this demand.

Identification of revenue-generation potential

Based on the given traffic matrix, random path requests were generated between arbitrary endpoints, with a variable capacity up to the total forecast demand. Revenue was calculated for each successfully routed service request, based on the capacity requirement of the path and the distance in terms of the minimal number of network hops required by the routed path.

For example, a 2 Gb/s path between two endpoints with a three-hop distance would generate a maximum of six revenue units, regardless of the actual path taken. Paths that cannot be routed generate no revenue. A less effective routing protocol loses potential revenue opportunities because it requires more than the minimal number of hops to accommodate a path request, consuming additional network resources that cannot be compensated by incremental revenue. Moreover, the required additional hops decrease the remaining available link capacity for future path requests and may result in lost revenue opportunities when these path requests must be rejected.

To illustrate these points, consider a scenario with an enterprise ordering a 10-Gb/s IP virtual leased line (VLL) service to connect its headquarters in San Francisco with a venue in Las Vegas for a live demonstration of its latest cloud entertainment suite. The network operator has a direct link between Las Vegas and San Francisco but this link has insufficient free capacity available and cannot be upgraded on short notice. If the network operator decides to route the service over Los Angeles it cannot charge a higher service fee to the enterprise even though it will in fact incur a larger cost to provide the service over this longer route. Moreover, because the rerouted service traffic consumes 10 Gb/s of bandwidth on the links between San Francisco, Los Angeles and Las Vegas, the network operator may potentially lose revenue opportunities on these links in future.

Assessment of network utilization

The study included the effect of network churn to assess the impact of capacity fragmentation on the results. Much like memory space on a disk drive, network capacity can become fragmented when capacity is allocated and released as paths are created, changed in size and deleted. To simulate network churn, each path request was assigned a random lifetime in the network.

Randomly generated path requests were processed by both the default CSPF routing algorithm—with link weight set as inversely proportional to link capacity—and by STAR in a series of five runs. For each run, the study counted the total number of accepted path requests and bandwidth as well as the total amount of associated revenue. To assess the relative network balance, the study measured the relative utilization of all links after the network had fulfilled approximately 50 percent of path requests. In practice, it is impossible to achieve the ideal of 100-percent network utilization for several reasons:

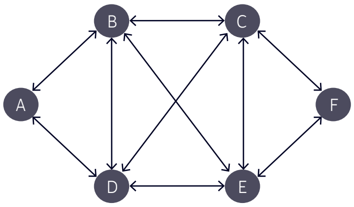
- Forecast demand is an estimate that is likely to deviate from actual demand.
- Service planning and network provisioning take time and must be proactive to stay ahead of demand.
- The order in which path requests arrive is random, so it is not possible to try out different packing orders.
- One-hundred percent utilization implies the existence of congested links and large latency, which degrade service performance.
- Traffic patterns are bursty, and networks need some headroom to accommodate busy-hour peaks.
- Some extra overflow capacity is always needed to handle network maintenance or failure events.

Study results

Small six-node topology

The study analyzed traffic in a small six-node core topology, as shown in Figure 5. All traffic was assumed to be symmetric, with the same Quality of Service requirements. Figure 5 also shows the hop count of each aggregate demand flow for the purpose of revenue calculation. For example, A-to-B traffic is forecast to be 200 Gb/s, and the distance is one hop. For A-to-F traffic, the forecast is 100 Gb/s, and the minimum hop count is three. The minimum hop count is used as a surrogate for the distance between the nodes. The revenue generated by carrying a connection is a product of the request size and the distance between the source and destination for the connection.

Figure 5. Six-node topology and traffic matrix

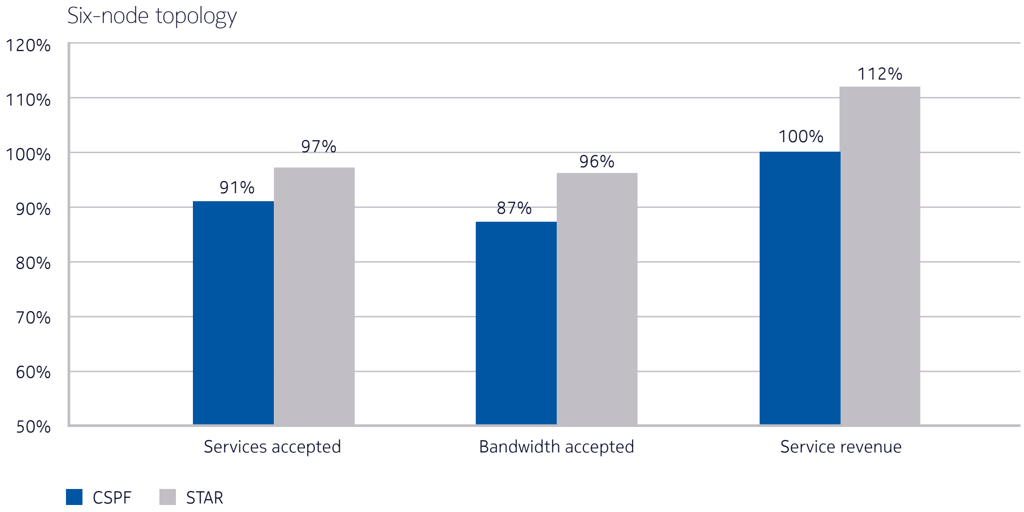


Bandwidth/ Number of hops	B	C	D	E	F
A	200/1	300/2	400/1	200/2	100/3
B		200/1	200/1	100/1	100/2
C			300/1	200/1	200/1
D				200/1	200/2
E					100/1

A-to-F path requests would generate three times more revenue per Gb/s of capacity than A-to-B path requests because they require a distance of three hops instead of only one hop. However, if the actual path selected by the routing protocol was more than the minimal number of hops required, no additional revenue would be generated for the extra hops taken, and there would be an obvious cost to the network operator.

Path requests with capacities between 1 Gb/s and 10 Gb/s were generated with random network lifetimes to simulate service churn. Researchers routed all path requests using both the default CSPF min-hop routing algorithm and STAR. They then compared the number of path requests that each algorithm was able to accommodate. If insufficient capacity existed between the source and destination, the request was rejected. Figure 6 shows the path-request and revenue-generation results for the six-node topology.

Figure 6. Six-node topology: path-request and revenue-generation results

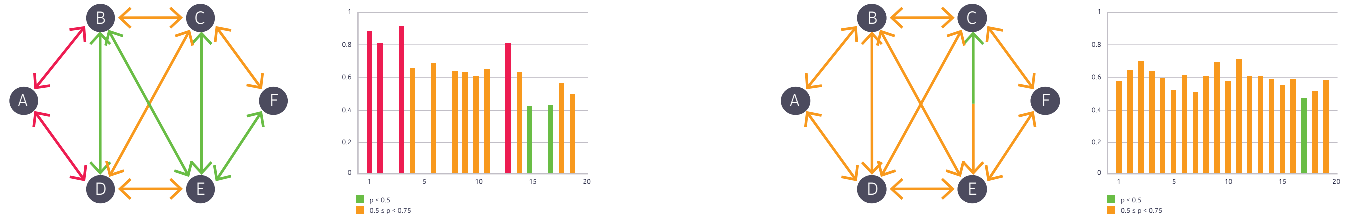


The six-node topology yielded the following results:

- STAR successfully completed more path requests (97 percent compared to 91 percent by CSPF).
- STAR accommodated 12 percent more revenue generating traffic than the default CSPF routing.

The study also compared link utilization after the network had fulfilled approximately 50 percent of the total service requests. Figure 7 shows the increased efficiency of STAR compared to CSPF.

Figure 7. Six-node topology: link-utilization results



CSPF link load distribution after 50% of requests

- Unbalanced link load distribution
- Less efficient packnig andmore rejected demands

STAR link load distribution after 50% of requests

- Balanced link load distribution
- Fewer rejected requests than CSPF

CSPF link-utilization results

CSPF had already utilized over 75 percent of the total link capacity on the A-to-B and A-to-C links in both directions. Four links—in green—had less than 50-percent utilization (spaces in the bar chart represent zero link utilization).

If more path requests are processed, the already congested A-to-B and A-to-D links would become fully loaded, resulting in future rejected path requests. Existing service traffic on these links would start to experience increased latency because of increased buffering. Network operators would therefore need to trigger actions to rebalance network traffic, causing network churn, or add more capacity on congested links.

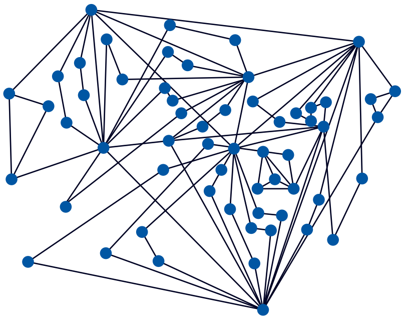
STAR link-utilization results

The STAR algorithm resulted in a more even link distribution, with no links loaded more than 75 percent and only one link with a load below 50 percent. Much more headroom was left on all links to accept additional service requests. With all network traffic fairly equally distributed over links, network operators would not need to rebalance traffic or upgrade link capacity. On average, STAR had to reject only 3 percent of the total number of path requests compared to CSPF, which rejected 9 percent of the requests.

Large 59-node topology

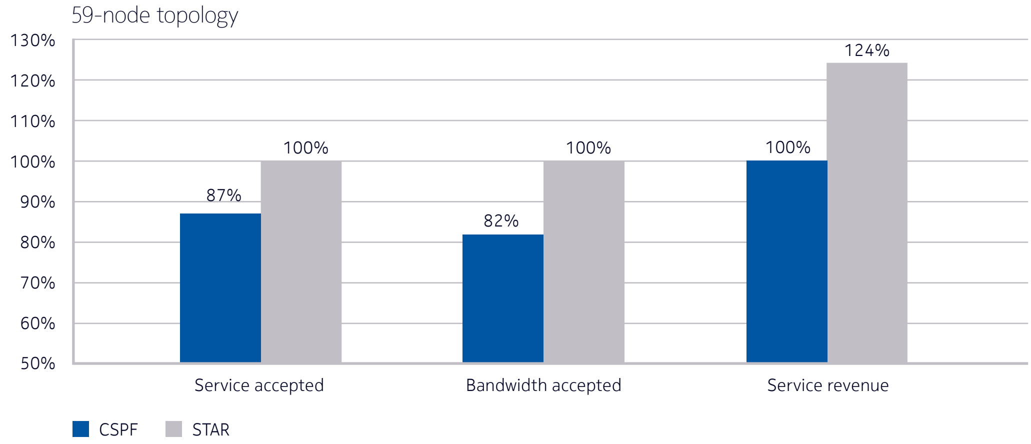
The study examined how the efficiency of the CSPF and STAR algorithms depends on the size and complexity of the routing topology. A large topology, shown in Figure 8, represents a wider network scope that could encompass a national network or region.

Figure 8. Large topology



The study defined and evaluated a traffic matrix for both CSPF and STAR in a 59-node topology. Random path requests were generated, with and without service churn. Figure 9 shows the path-request and revenue-generation results for the 59-node topology.

Figure 9. 59-node topology: path-request and revenue-generation results



The 59-node topology yielded the following results:

- STAR accepted nearly all service requests while CSPF rejected 13 percent of the requests.
- STAR was able to accommodate 24 percent more revenue-generating traffic than the CSPF algorithm.

The study also compared link utilization after the network had fulfilled approximately 50 percent of the total service requests. Figure 10 shows the increased efficiency of STAR compared to CSPF.

Figure 10. 59-node topology: link-utilization results

CSPF link-utilization results

The left chart shows the imbalance that can result when operators deploy default CSPF routing. Multiple links—in red—show 75-percent to 100-percent utilization while several other links show less than 25-percent utilization.

STAR link-utilization results

With the STAR path-computation algorithm, no link was loaded above 75 percent or below 25 percent. STAR path computation results in more balanced link utilization compared to CSPF. Balanced link utilization enables more efficient packing of bandwidth, fewer rejected path requests, and increased network profitability.

Applications of the STAR algorithm

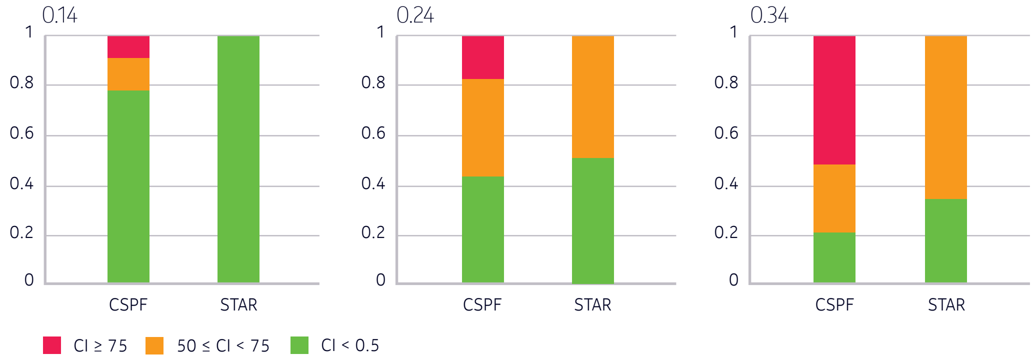
Network operators can leverage the STAR algorithm to optimize routing in their IP/MPLS networks. The principle use of STAR is to optimize network utilization or “run the network hotter.” The more balanced link utilization achieved by STAR also results in better end-to-end performance for connections.

Improved connection-level performance

Congestion index (CI) values are useful in comparing the connection-level performance of different routing algorithms. The CI for a connection is the highest link utilization among all the links on the path that a connection uses. Higher CI values indicate that the packets belonging to a connection will likely face delays in the network. Figure 11 shows the CI distribution achieved by STAR and CSPF for different network loads.

- The red bar represents connections with CI above 75.
- The orange bar represents connections with CI between 50 and 75.
- The green bar represents connections with CI below 50.

Figure 11. Performance of CSPF and STAR for different network load factors

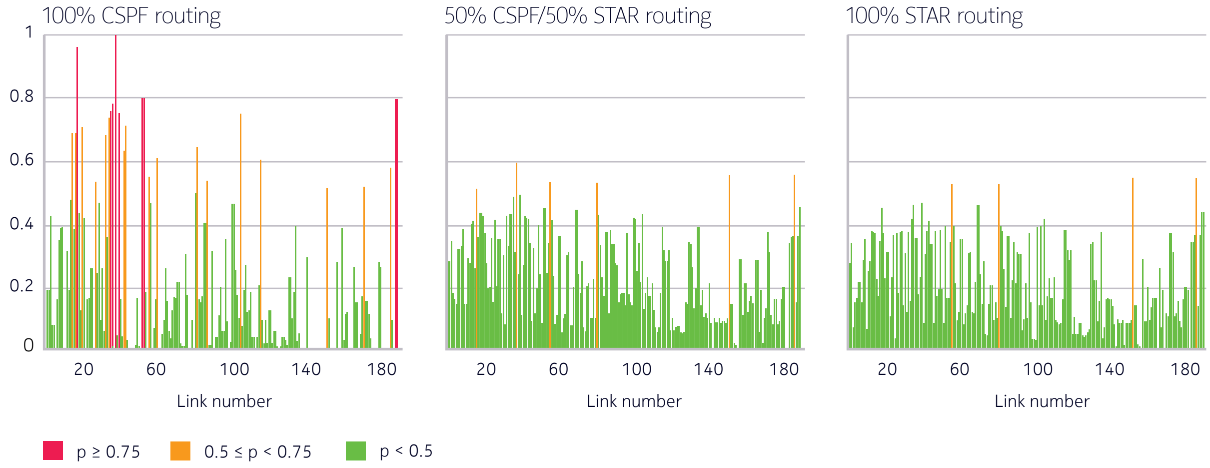


For all three load factors, the CSPF bars show many connections with CI above 75 while STAR keeps the CI for all connections below 75. It is not necessary for STAR to route all path requests to achieve a more balanced network.

Figure 12 shows link utilization for a network topology with:

- 100-percent use of CSPF-based route computation
- CSPF and STAR routing, each used for 50 percent of the paths
- 100-percent use of STAR to compute routes.

Figure 12. Link utilization with different amounts of STAR path computation



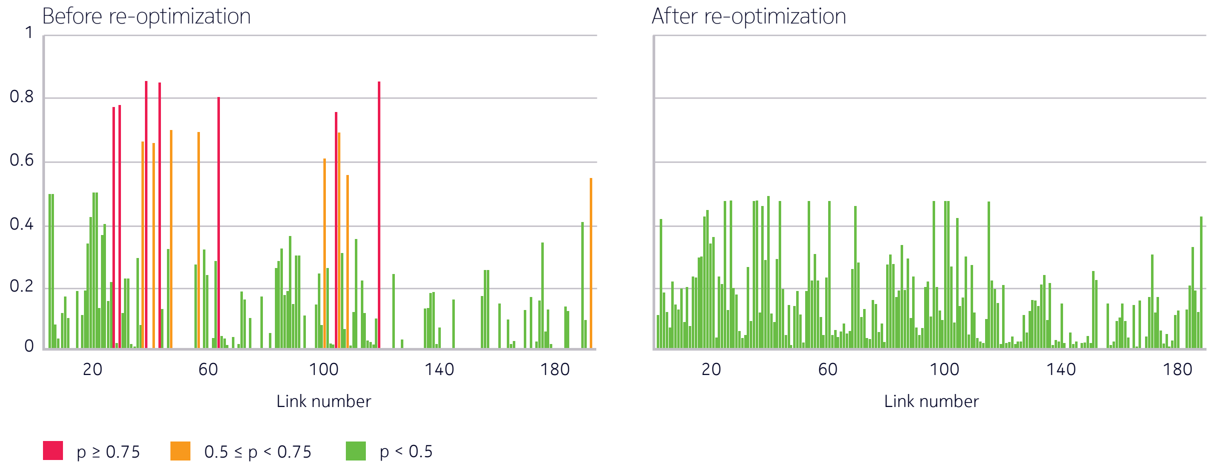
Results show that the network is already far more balanced when STAR is used for just half the total paths. Network operators can therefore apply STAR-based path computation to new path requests at any time, and the network automatically balances itself following normal service-churn events.

Online traffic engineering

STAR can also be used as an online traffic-engineering tool to alleviate congestion from heavily utilized links. To minimize service churn, operators can apply network rebalancing on the most heavily loaded links. It is also possible to restrict the total number of paths that may be rerouted. The STAR algorithm automatically selects the set of connections to be rerouted for maximum impact and suggests new routing paths.

Figure 13 shows the results of a network re-optimization of a CSPF routed network with STAR.

Figure 13. Network re-optimization with STAR



A centralized PCE with STAR complements offline planning and traffic-engineering tools by adding an online capability to help implement and enforce network-usage policies, performance objectives and operational planning parameters.

Elastic bandwidth services

Network operators can introduce STAR to enable new elastic service capabilities, such as Bandwidth on Demand (BoD) and bandwidth calendaring. BoD enables users to create temporary paths or dynamically scale up the capacity of existing paths when needed. Bandwidth calendaring enables the scheduling of paths and associated capacity to meet regular demand fluctuations, such as day/night usage cycles, data-center backups, and end-of-month reporting. These dynamic bandwidth services introduce a certain amount of network churn, with opportunities for STAR to rebalance network traffic.

Conclusion

While the management and control of IP traffic is challenging, the combination of a centralized PCE and the Bell Labs STAR algorithm can yield significant benefits for network operators. Suited to a range of applications, a centralized PCE and STAR can yield more revenue-generating opportunities by optimizing network utilization, avoiding congestion, and balancing traffic over the available links.

Acronyms

BGP-LS	Border Gateway Protocol Link State
CLI	command-line interface
CSPF	Constrained Shortest Path First
ERO	EXPLICIT_ROUTE object
GMPLS	Generalized MPLS
IETF	Internet Engineering Task Force
IP	Internet Protocol
IS-IS	Intermediate System to Intermediate System
IT	information technology
MIB	management information base
MPLS	Multiprotocol Label Switching
OSPF	Open Shortest Path First
OSS	operations support system
PCE	Path Computation Element
PCEP	PCE Communication Protocol
RFC	Request for Comments
RIB	routing information base
ROI	return on investment
SDN	software-defined networking
SNMP	Simple Network Management Protocol
STAR	Self-Tuned Adaptive Routing
TED	Traffic Engineering Database

References

1. IETF. RFC 3209: “Extensions to RSVP for LSP Tunnels.” December 2001. http://www.rfc-base.org/txt/rfc-3209.txt
2. IETF. RFC 3473: “Codepoint Registry for the Flags Field in the Resource Reservation Protocol - Traffic Engineering (RSVP-TE) Session Attribute Object.” April 2007. https://tools.ietf.org/rfc/rfc4859.txt
3. IETF. RFC 3477: “Signalling Unnumbered Links in Resource Reservation Protocol - Traffic Engineering (RSVP-TE).” January 2003. https://tools.ietf.org/html/rfc3477

4. IETF. RFC 4655: “A Path Computation Element (PCE)-Based Architecture.” August 2006. https://tools.ietf.org/html/rfc4655
5. IETF. RFC 5440: “Path Computation Element (PCE) Communication Protocol (PCEP).” March 2009. https://tools.ietf.org/html/rfc5440
6. IETF. RFC 5862: “Path Computation Clients (PCC) - Path Computation Element (PCE) Requirements for Point-to-Multipoint MPLS-TE.” June 2010. https://tools.ietf.org/html/rfc5862
7. IETF. RFC 6825: “Traffic Engineering Database Management Information Base in Support of MPLS-TE/GMPLS.” January 2013. https://tools.ietf.org/html/rfc6825
8. IETF. Segment Routing web site. http://www.segment-routing.net/home/ietf

Nokia Oyj
Karaportti 3
FI-02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Product code: PR1606020443EN (June)