



Technologia OptiScale™ dla routingu internetowego

Routing internetowy na dużą skalę dla urządzeń SLX instalowanych na brzegu sieci

W skrócie

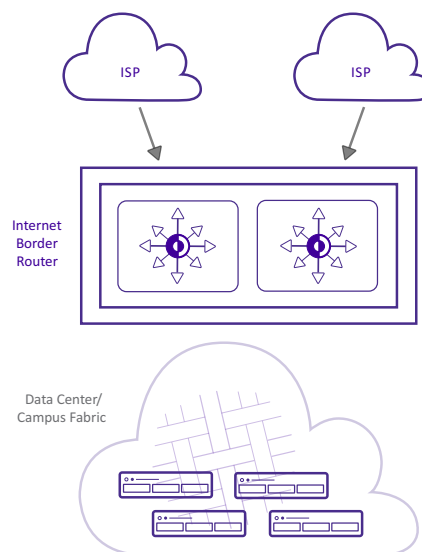
Elementy sieci w centrach danych są zazwyczaj zoptymalizowane pod kątem szybkiego przełączania pakietów. Ale istnieje wiele przypadków zastosowania routingu IP w centrach danych lub sieciach kampusowych, w których wymagana jest duża liczba tras IP. Przykładem może być wewnętrzna segmentacja, DCI (Data Center Interconnect) oraz routing internetowy.

Odpowiedzią firmy Extreme Networks jest technologia routingu internetowego OptiScale™ wspierająca routing w centrach danych. Technologia OptiScale to efektywne rozwiązanie do routingu brzegowego o dużej skali (może obsłużyć do 1.5 mln tras IP) i przez to zapewnia długotrwałą ochronę inwestycji. Podstawowym zastosowaniem tej technologii jest obszar brzegowy sieci korporacyjnych.

Wprowadzenie

Łączność internetowa korporacyjnych centrów danych stale rośnie na znaczeniu, głównie z uwagi na zwiększone wykorzystanie technologii cloud. Co więcej, rosną również rozmiary tablic routingu ze względu na coraz większe pofragmentowanie rozgłoszeń w przestrzeni adresowej IPv4 (z uwagi na wyczerpanie puli dostępnych adresów) oraz wzrost liczby rozgłaszanych tras IPv6.

Obszar brzegowy lub graniczny sieci korporacyjnej agreguje połączenia klientów, usługowe i internetowe, wchodzące lub wychodzące z centrum danych lub sieci kampusowej. Rysunek nr 1 pokazuje routery brzegowe łączące się z dostawcami usług internetowych ISP. Sieci, które one łączą mogą być sieciami IP lub Ethernet fabric, środowiskami agregującymi lub rdzeniowymi centrów danych lub sieciami kampusowymi.



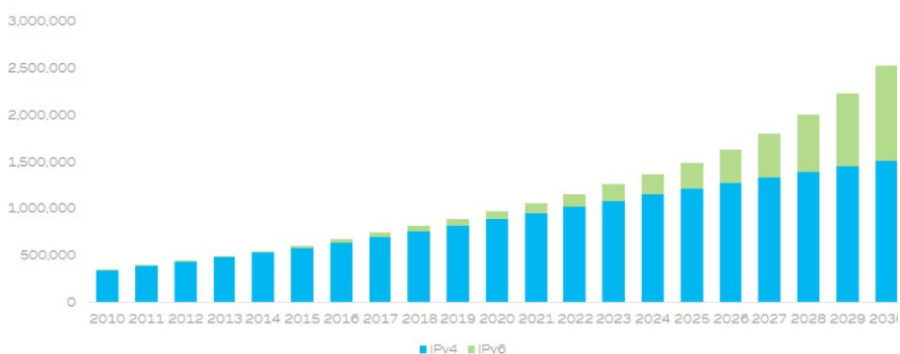
Rysunek 1: Komunikacja z dostawcami usług internetowych ISP z poziomu centrum danych lub sieci kampusowej

Wysoka dostępność i redundancja mają tutaj istotne znaczenie, ponieważ utrata ruchu w tej lokalizacji może negatywnie wpłynąć na działanie istotnych usług dla użytkowników. Korporacyjne centra danych nakładają dodatkowe wymagania na routing brzegowy. Przykładowo, muszą być obsługiwane wszystkie główne protokoły IPv4 i IPv6, a także sieci VPN MPLS i multicast. Jeżeli brana jest pod uwagę obsługa wielu użytkowników (multitenancy), wówczas ważne jest wsparcie dla logicznych domen poprzez konstrukcje takie jak VPLS lub BGP-EVPN z VXLAN.

Technologia Extreme Networks OptiScale obsługuje routing internetowy o dużej skali w sieciach komputerowych centrach danych. Technologia optymalizuje sprzęt i oprogramowanie, aby tworzyć tabele routingu internetowego w routerach SLX 9540 i SLX 9850, zapewniając obsługę do 1.5 mln tras. Opracowując tą technologię firma Extreme uwzględniła możliwość zwiększenia rozmiarów tabeli routingu internetowego i zapewniła odpowiednią pojemność bazy FIB (Formwarding Information Base) dla obsługi tego wzrostu.

Prognozy rozwoju sieci Internet

W ciągu ostatnich kilku lat obserwowany jest ok. 10% (rok do roku) wzrost rozmiaru tabel routingu internetowego dla IPv4 i ok. 30% dla IPv6. Na rysunku nr 2 przedstawiono prognozę wzrostu tras IPv4 i IPv6 do roku 2030, która wskazuje na konieczność stosowania baz FIB o większej pojemności.



Rysunek 2: Historyczny i przyszły wzrost liczby tras internetowych (źródło: <https://www.potaroo.net/>)

Do roku 2019 liczba tras wymaganych do obsługi ruchu internetowego najprawdopodobniej przekroczy 1mln, a współmiernie wzrosną również rozmiary FIB. Natomiast do roku 2023 lub 2024 przewiduje się, że liczba wpisów w tabeli routingu przekroczy 1.5 mln tras.

Prognozy rozwoju sieci Internet

Technologia OptiScale optymalizuje wielkość tablic w oparciu o analizę tabel routingu i trendy ich rozwoju. Przykładowo, na rysunku nr 3 przedstawione są najczęściej stosowane prefiksy w tabelach routingu IPv4 i IPv6.

Prefiks IPv4	Liczba tras internetowych	Udział %
/8	15	0
/9	13	0
/10	35	0.01
/11	106	0.02
/12	285	0.04
/13	549	0.08
/14	1,058	0.16
/15	1,851	0.28
/16	13,273	2.02
/17	7,647	1.16
/18	13,409	2.04
/19	24,635	3.74
/20	37,670	5.72
/21	42,772	6.5
/22	80,727	12.26
/23	65,054	9.88
/24	369,424	56.1
/32	9	0
658, 531		

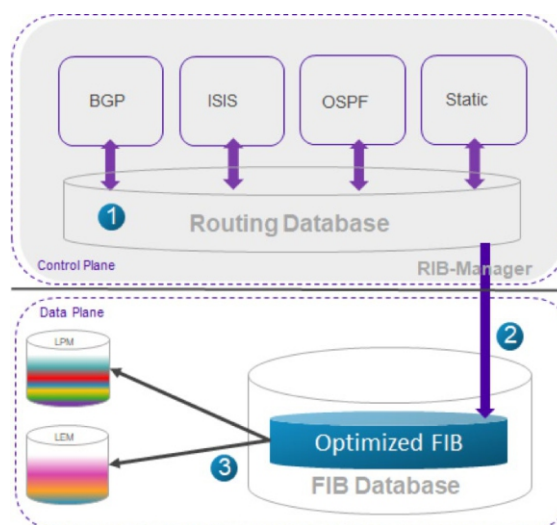
Prefiks IPv6	Liczba tras internetowych	Udział %
/20	9	0.02
/21	3	0.01
/22	4	0.01
/23	4	0.01
/24	19	0.04
/25	6	0.01
/26	16	0.04
/27	17	0.04
/28	78	0.18
/29	1857	4.39
/30	149	0.35
/31	110	0.26
/32	9306	22.01
/33	491	1.16
/34	359	0.85
/35	350	0.84
/36	1454	3.44
/37	201	0.48
/38	673	1.59
/39	149	0.35
/40	2283	5.40
/41	207	0.49
/42	351	0.83
/43	119	0.28
/44	2277	5.38
/45	183	0.43
/46	1451	3.43
/47	285	0.67
/48	19271	45.57
/49	19	0.04
/50	6	0.01
/52	4	0.01
/56	81	0.19
/64	480	1.15
42,285		100,00

Rysunek 3: Tabele routingu w podziale na prefiksy dla IPv4 (z lewej) i IPv6 (z prawej)

Okazuje się, że większość tras IPv4 poddanych sumaryzacji korzysta z maski /24 (56%), w przypadku masek /23 i /22 jest to odpowiednio 10% i 12%. Podobna sytuacja ma miejsce w przypadku prefiksów IPv6, gdzie na /48 przypada 46% sumaryzowanych tras a na /32 22%. Tym samym, wszelkie zabiegi związane z optymalizacją przetwarzania tabeli routingu powinny brać te dane pod uwagę i przechowywać tabele w mniej obciążonych urządzeniach.

Technologia OptiScale usprawnia profil skalowania tras korzystając z dwóch podstawowych technik: softwareowej optymalizacji FIB oraz sprzętowej optymalizacji tabel. Celem jest zagwarantowanie osiągalności wszystkich sieci znajdujących się w tabeli routingu internetowego, przy jednoczesnej eliminacji wpisów tabeli, które nie są potrzebne w przypadku routerów brzegowych.

W przypadku FIB wybrana tabela tras jest specjalnie obsługiwana w dwóch pod-tabelach: jednej dla LPM (Longest Prefix Match) i drugiej dla LEM (Longest Exact Match). Technologia OptiScale określa sposób zapisywania tras z wybranymi zakresami prefiksów i wykorzystaniem sekcji LEM bazy FIB, co zwalnia miejsce dla bardziej złożonych i zasobochłonnych LPM. Na rysunku 4 zilustrowano ten proces.



Rysunek 4: Optymalizacja tras z wykorzystaniem technologii OptiScale for Internet Routing

Optymalizacje mają następujący przebieg:

1. Optymalizacja FIB (Kroki 1 i 2)

- Trasy protokołu są pobierane do RIB, który z kolei pobiera trasy do FIB.
- Przeprowadzana jest wstępna optymalizacja tras, w efekcie której powstaje zoptymalizowana tabela FIB – poprzez eliminację nadmiarowych wpisów o następnym skoku.

2. Optymalizacja sprzętowa (Krok 3)

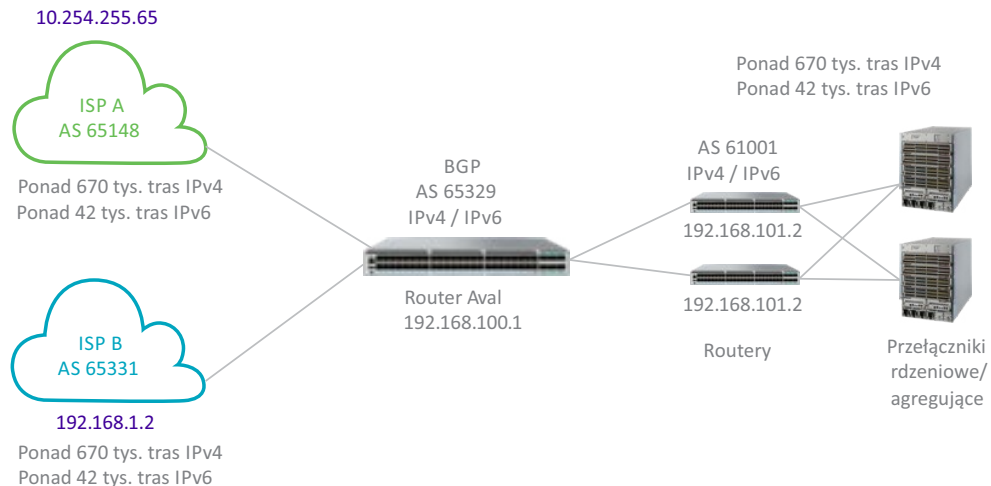
- Technologia OptiScale następnie dzieli wybrane trasy z wykorzystaniem wspólnych prefiksów (szczegóły na rysunku 4) i przechowuje je w oddzielnej tabeli.
- Mniej zasobochłonny LEM obsługuje wspólne prefiksy, natomiast LPM zachowuje miejsce na obsługę większej liczby wpisów tras dodawanych do tabeli.

Efekty

Klienci korzystają z technologii OptiScale for Internet Routing w wielu różnych scenariuszach. Opisane poniżej dwa przykłady cechują się mniejszym o ponad 50% wykorzystaniem sprzętu i oprogramowania przy obsłudze około 750 tys. tras.

Scenariusz zastosowania u klienta, nr 1

W tym scenariuszu zastosowania (rysunek 5), router brzegowy „Aval” otrzymuje pełne tabele routingu internetowego poprzez EBGp od dwóch różnych dostawców usług i rozgłasza te trasy do routerów w centrach danych w wielu lokalizacjach.



Rysunek 5: Powiązania w przypadku dwóch dostawców usług ISP

Można to traktować jako przykład routera brzegowego połączanego z dużą siecią korporacyjną. Ponieważ mamy do czynienia z dwoma dostawcami usług (A i B), router „Aval” jest w pełni powiązany z każdym z nich (dotyczy do tras IPv4 i IPv6).

Na rysunku nr 6 pokazano trasy IPv4 oraz powiązane tabele tras IPv4. W przypadku tabeli tras IPv4 istnieje ok. 745 tys. tras.

```
Aval# show ip bgp summary
BGP4 Summary
Router ID: 192.168.100.1 Local AS Number: 65329
Confederation Identifier: not configured
Confederation Peers:
Maximum Number of IP ECMP Paths Supported for Load Sharing: 1
Number of Neighbors Configured: 4, IP: 4
Number of Routes Installed: 1349931, Uses 276735855 bytes
Number of Routes Advertising to All Neighbors: 2235390 (745130 entries), Uses 56629880 bytes
Number of Attribute Entries Installed: 198593, Uses 32172066 bytes
Neighbor Address AS# State Time Rt:Accepted Filtered Sent ToSend
10.254.255.65 65148 ESTAB 1h33m47s 672333 0 677598 0
192.168.1.2 65331 ESTAB 0h16m46s 677598 0 67532 0
192.168.101.2 61001 ESTAB 2h35m13s 0 0 745130 0
192.168.102.2 61001 ESTAB 2h35m12s 0 0 745130 0

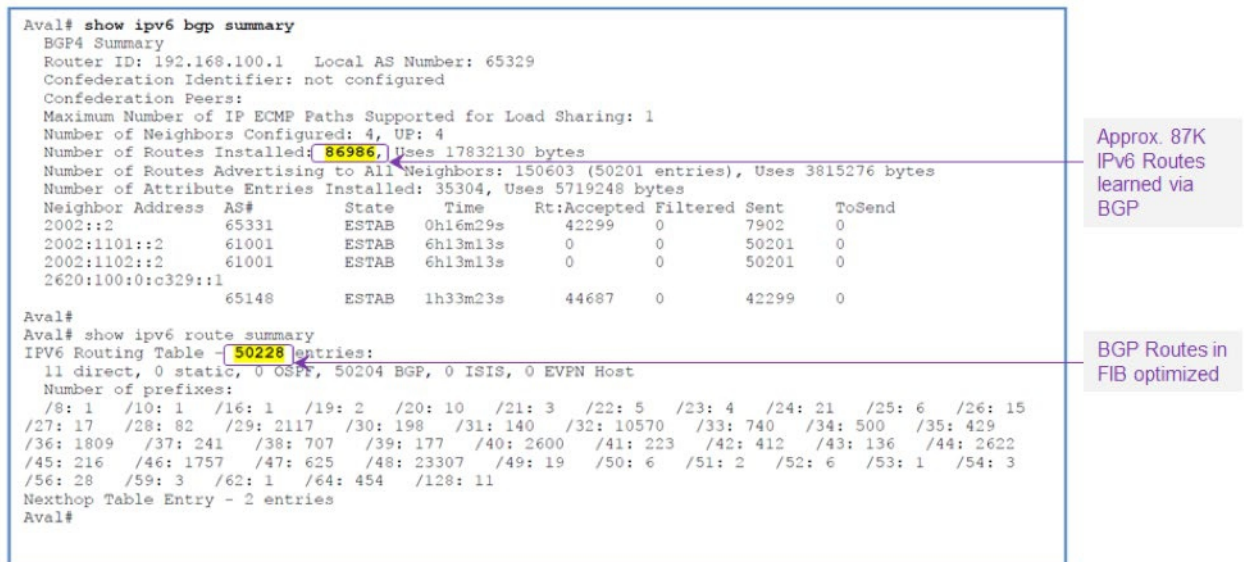
Aval# show ip route summary
IP Routing Table 745161 entries:
38 connected, 0 static, 1 OSPF, 745084 BGP, 0 ISIS, 0 unnumbered, 0 EVPN Host
Number of prefixes:
/0: 1 /8: 15 /9: 13 /10: 37 /11: 108 /12: 299 /13: 571 /14: 1121 /15: 1953 /16:
13744 /17: 8301 /18: 14510 /19: 26860 /20: 42059 /21: 49192 /22: 90273 /23: 76186
/24: 417389 /25: 927 /26: 661 /27: 693 /28: 76 /29: 43 /30: 50 /31: 12 /32: 67
NextHop Table Entry - 3 entries
Aval#
```

Approx. 1.35M
IPv4 Routes
learned via
BGP

BGP Routes in
FIB optimized

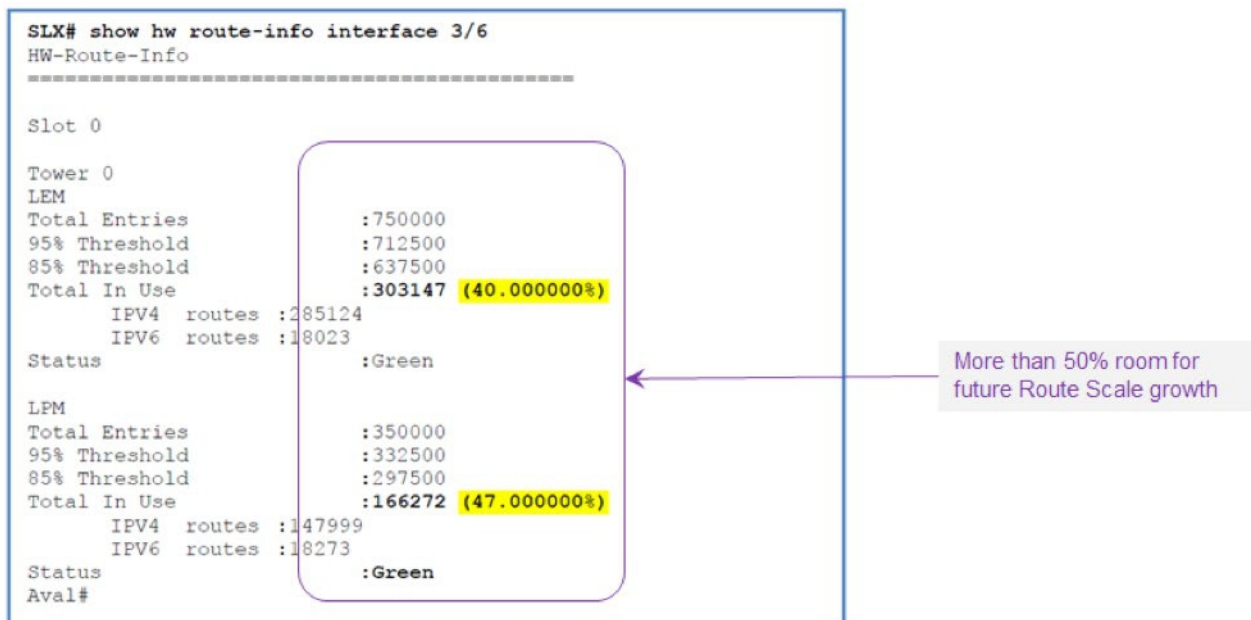
Rysunek 6: Szczegóły tras IPv4

Na rysunku nr 7 pokazano całkowitą liczbę tras IPv6 (od wszystkich peerów) oraz powiązane tabele tras IPv6. W przypadku tabeli tras IPv6 istnieje ok. 50 tys. tras.



Rysunek 7: Szczegóły tras IPv6

Wyniki wykorzystania zasobów dla routera „Aval” (rysunek nr 8) pokazują wykorzystanie LEM i LPM z aktywnym OptiScale for Internet Routing.

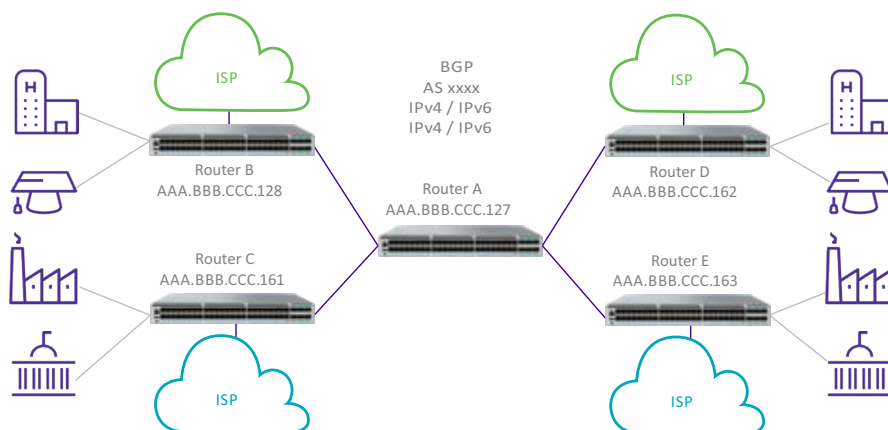


Rysunek 8: Wykorzystanie zasobów na routerze „Aval”

W przypadku tego routera (SLX 9850), nadal dostępnych jest ok. 53% zasobów LPM i ok. 60% zasobów LEM na potrzeby obsługi zwiększających się tabel routingu, a tym samym stanowi to ochronę inwestycji na kilka najbliższych lat.

Scenariusz zastosowania u klienta, nr 2

W tym przykładzie (rysunek nr 9) połączonych jest ze sobą 5 routerów brzegowych, które wspólnie tworzą rdzeń sieci BGP. Podłączeni dostawcy usług (ISP) obsługują powiązania wykorzystujące IPv4 i IPv6.



Rysunek 9: Pięć routerów brzegowych połączonych do 4 dostawców usług (ISP)

Wszystkie routery ISP są połączone poprzez EBGP, każdy dostarcza pełną tabelę tras internetowych do sieci rdzeniowej. Klient otrzymuje również trasy z obsługiwanych przedsiębiorstw. Jednocześnie, każdy z ISP wysyła także pełne tabele routingu internetowego IPv6 do sieci klienta.

Router A w rdzeniu sieci klienta komunikuje się z pozostałymi routerami poprzez powiązania IBGP. W ten sposób przechowuje do 4 tabeli tras IPv4 w bazie danych BGP, co przekłada się łącznie na 2.81 mln tras IPv4 BGP (rysunek nr 10).

Core RouterA# show ip bgp sum

```
BGP4 Summary
Router ID: AAA.BBB.CCC.127  Local AS Number: XXXX
Confederation Identifier: not configured
Confederation Peers:
Maximum Number of IP ECMP Paths Supported for Load Sharing: 1
Number of Neighbors Configured: 4, UP: 4
Number of Routes Installed: 2816454, Uses 577373070 bytes
Number of Routes Advertising to All Neighbors: 0 (0 entries)
Number of Attribute Entries Installed: 443420, Uses 71834040 bytes
Neighbor Address AS#      State Time  Rt:Accepted Filtered Sent ToSend
AAA.BBB.CCC.128 XXXX   ESTAB 7d1h58m 704112 0 0 0
AAA.BBB.CCC.161 XXXX   ESTAB 7d1h58m 704112 0 0 0
AAA.BBB.CCC.163 XXXX   ESTAB 7d1h58m 704112 0 0 0
AAA.BBB.CCC.165 XXXX   ESTAB 7d1h58m 704118 0 0 0
```

2.81M IPv4 Routes
learned via BGP

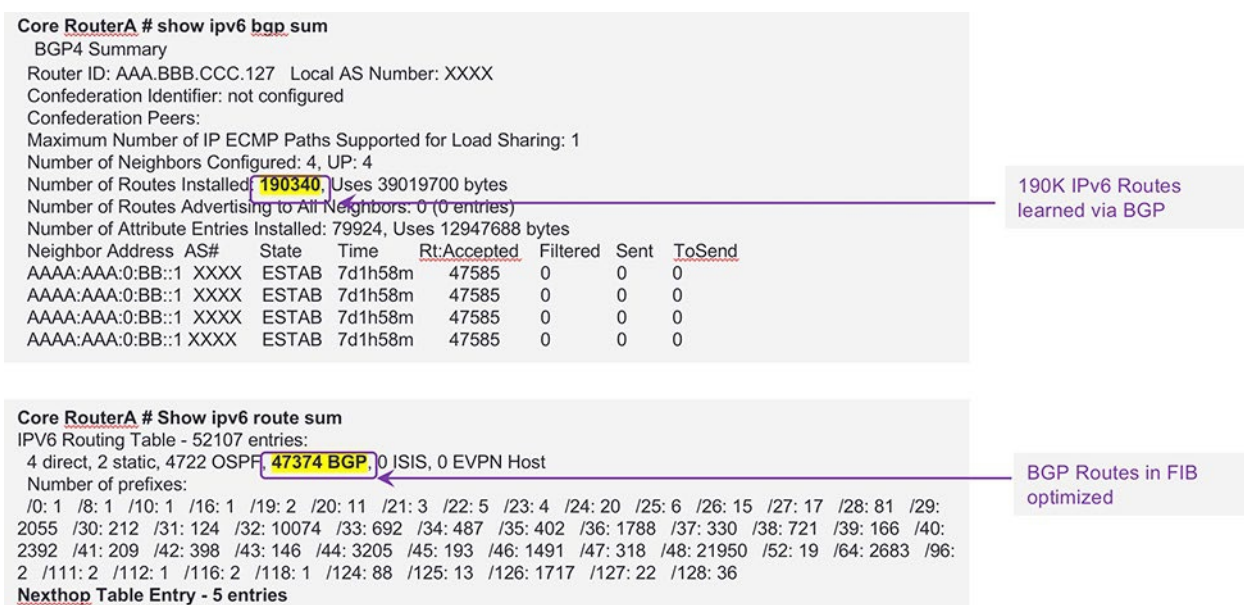
Core RouterA # show ip route sum

```
IP Routing Table - 710877 entries:
4 connected, 2 static, 7132 OSPF, 703736 BGP, 0 ISIS, 0 unnumbered, 0 EVPN Host
Number of prefixes:
/0: 1 /8: 14 /9: 11 /10: 36 /11: 98 /12: 287 /13: 561 /14: 1098 /15: 1903 /16: 13383 /17:
7859 /18: 13772 /19: 25254 /20: 38512 /21: 45472 /22: 86719 /23: 71689 /24: 395278 /25:
104 /26: 551 /27: 839 /28: 617 /29: 1042 /30: 3445 /31: 1 /32: 2331
NextHop Table Entry - 5 entries
```

BGP Routes in FIB optimized

Rysunek 10: Monitorowanie parametrów wyjściowych dla IPv4

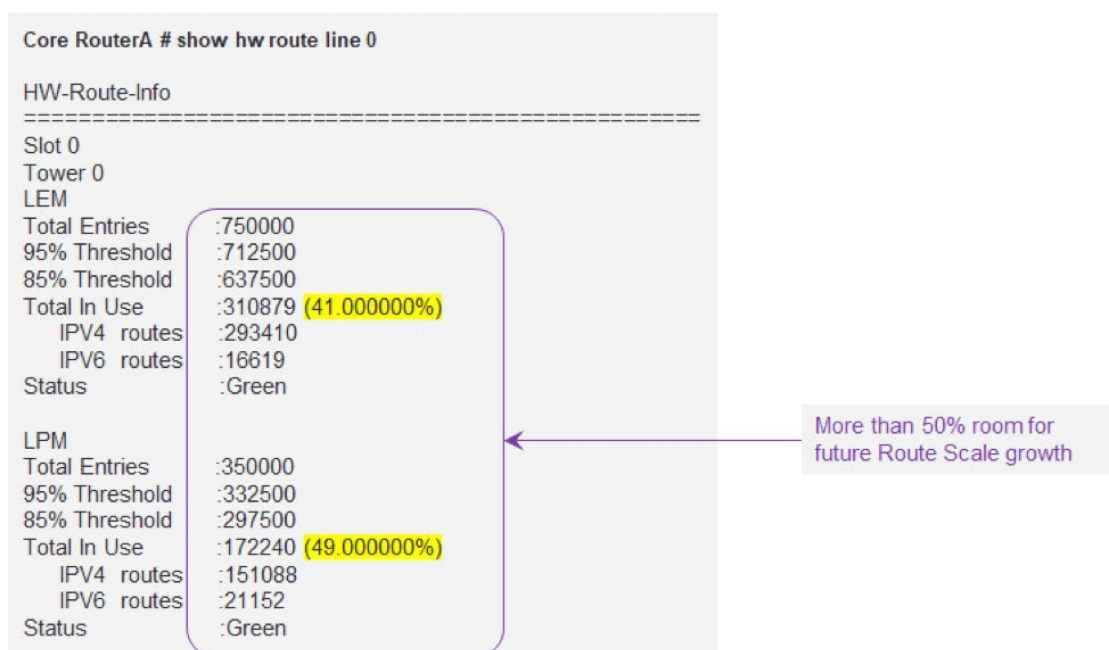
Istnieją także 4 tabele tras IPv6, które przekładają się łącznie na 190 tys. tras IPv6 BGP (rysunek nr 11).



Rysunek 11: Monitorowanie parametrów wyjściowych dla IPv6

W rezultacie Router A przechowuje jednocześnie pełne tabele tras internetowych dla IPv4 i IPv6 – łącznie ok. 704 tys. tras IPv4 i 47 tys. tras IPv6.

Na rysunku nr 12 przedstawiono wykorzystanie zasobów na Routerze A (SLX 9540). Zaprezentowano stopień wykorzystania LEM i LPM po optymalizacjach. Wyraźnie widać, że ponad 50% zasobów LPM i ok. 51% zasobów LEM jest nadal dostępnych na potrzeby wzrostu tablicy tras w przyszłości. Stanowi to kolejne potwierdzenie ochrony inwestycji na kilka najbliższych lat.



Rysunek 12: Poziom całkowitego wykorzystanie zasobów w przypadku Scenariusza nr 2.

