Huawei NE5000E Core Router
Throughput, Resiliency and Power Consumption Tests Report

Introduction
Every four years on average, an upgrade to the Ethernet interfaces in networks arrive. We have seen the initial deployment of 10GigabitEthernet (10GbE) in 2002 with the technology becoming common place in 2006. 100GigabitEthernet (100GbE) then started appearing in various vendor announcements in 2009 in a single interface per slot configuration. From here the first capacity growth will be in the single slot configuration - 4 or more ports of 100GigabitEthernet in a single slot.

When Huawei first engaged EANTC for this test we were excited to get the opportunity to apply our rigorous testing methodology to mature line cards that are clearly aimed at alleviating core-networks congestion as well as helping service providers increase their core capacity without requiring a fork-lift upgrade. This increase is gained by offering more capacity per slot and by extension per router.

In theory, and hopefully for Huawei also in practice, service providers could replace their existing single port 100GigabitEthernet line cards with their 4x 100GbE cards and with this increase the capacity of their core network four-fold.

Our task was clear. We took our extensive experience with service provider Proof of Concept testing and applied it to these line cards. We asked ourselves “if we were a service provider looking to upgrade our core capacity, what aspects will we investigate to help us make a purchasing decision?” The answers to this question are the tests you see herein.

Tested Devices and Test Equipment
Huawei provided their mature Flexible Card Line Processing Unit (LPUF-400) carrier cards for the test. The interfaces were installed in Huawei’s flagship core router - NetEngine 5000E (NE5000E).

The NE5000E router has 16 line card slots, each able to host two sub-slots. Each sub-slot of LPUF-400 holds one flexible card (CP400) with different number of 10GbE, 40GbE, and 100GbE interfaces.

Test Results Highlights

<table>
<thead>
<tr>
<th>Throughput and Scale</th>
<th>Test Results Highlights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>➔ 400 Gbit/s per slot capacity for all frame sizes tested</td>
</tr>
<tr>
<td></td>
<td>➔ 6.4 Tbit/s per chassis capacity for all frame sizes tested</td>
</tr>
<tr>
<td></td>
<td>➔ 512,000 RSVP-TE tunnels distributed into 8 RSVP-TE processes</td>
</tr>
<tr>
<td></td>
<td>➔ High-scale BGP: 200 BGP peers, each with the same 500K prefixes, supported by 5 distributed BGP processes</td>
</tr>
<tr>
<td></td>
<td>➔ Virtual Router (VR) and UCMP tested</td>
</tr>
<tr>
<td>Resiliency</td>
<td>➔ Multi Service Non-Stop Routing:</td>
</tr>
<tr>
<td></td>
<td>➔ Main Processing Unit (MPU) failover doesn’t affect control plane and traffic</td>
</tr>
<tr>
<td></td>
<td>➔ IS-IS average convergence time 278 ms IPv4/IPv6 dual stack for 20,000 IS-IS routes</td>
</tr>
<tr>
<td></td>
<td>➔ eBGP average convergence time for IPv4/IPv6 dual stack 4.67 s with 500,000 routes for IPv4 only 2.87 s with 400,000 BGP routes</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>➔ 8,106 Watt weighted average power consumption</td>
</tr>
<tr>
<td></td>
<td>➔ 1.27 Watt/Gbps weighted power consumption rate</td>
</tr>
</tbody>
</table>

In our test, CP400 of 2x 100GbE, 6x 40GbE, and 20x 10GbE were used.

The breadth of the testing required several devices to be used for the project:

- NE5000E router with eight 4x 100GbE, one 12x 40GbE, and seven 40x 10GbE line cards. In total the chassis was fully loaded with 16 line cards each able to handle 400Gbit/s of traffic.
• NE5000E router with one 4x 100GbE, and one 40x 10GbE line card.
• NE40E router with one 2x 10GbE and two 8x GbE interface cards.
• Two Main Processing Units (MPU) for each router. MPU is the control plane for NE5000E and NE40E routers.
• Four Switching Fabric Unit (SFU) for each NE5000E router.

The routers where running Huawei Versatile Routing Platform VRP (R) software version 8.60 in all test cases.

In real deployment scenarios multiple routers will be positioned in the core of the network. To simulate such a topology, we used the second NE5000E router and the NE40E router to setup test scenarios involving control plane protocols such as Intermediate System to Intermediate System Protocol (IS-IS) and Border Gateway Protocol (BGP) and to implement services like Multi Protocol Label Switching (MPLS) Virtual Private LAN Service (VPLS), BGP/MPLS L3VPN, and multicast.

For this test Spirent brought the tester, together with a sizeable support team, to EANTC’s lab. Spirent supplied their brand new tester platform, the SPT-N11U. We used 4 x Spirent SPT-N11U chassis running TestCenter 4.33 firmware for all tests.

Performance

The Times of India technical news (http://articles.timesofindia.indiatimes.com/2013-07-09/internet/40468638_1_growth-rate-population-10-fold-growth) recently reported that the Asia-Pacific (APAC) region will generate 43.4 EByte (1 EByte equal 1,048,576 TByte) IP traffic per month by 2017. The amount of IP traffic in the Middle East and Africa (MEA) will continue to grow from 2012 - 2017 (5-fold growth, 38 percent compound annual growth rate). The number of Internet users is expected to grow to 3.6 billion by 2017 and global network users will generate 3 trillion Internet video minutes per month.

Replacing core router line cards with new cards that promise four-fold capacity, has obvious benefits to service providers, but also a potential hazard. In our initial assessment of these line cards we were looking to discover potential risks for service providers before they use these cards to deliver services to their customers. Such risks could be decreased performance under certain loads or configurations and interface errors.

Single Slot Throughput Performance

To ensure that these revenue-supporting line cards perform as advertised, we ran an extensive testing and analysis campaign under a variety of conditions that mirror real-world traffic mixes.

The NE5000E router can be positioned in different places in the network connecting high-speed Ethernet links, multiple types of NE5000E line cards providing the same maximum throughput (400 Gbit/s). It gives the service providers the flexibility to interconnect with transport providers and internet peers using options of interfaces like 10GbE and 100GbE.

We verified under which conditions line-rate forwarding performance could be reached using two line card configurations: 4x 100GbE and 40x 10GbE. We used standard methodology as defined in the Internet Engineering Task Force (IETF) Request for Comments (RFC) 2544 for IPv4 throughput measurements.

The smaller the packet received by the router, the more packets need to be switched to achieve wire-rate performance; while on the other hand, larger
packets will require less packets per second (PPS) throughput to achieve wire-rate.

As a slight deviation from the frame sizes defined in RFC 2544, Huawei asked us to start with 128-Byte as the smallest packet size which better reflect real network scenarios. At 400 Gbit/s (bidirectional) each line card was still processing 675.7 million packets per second per line card for 128 Byte frames. We also extended the frame sizes mentioned in the RFC by adding 9,000 Byte frames and additionally used IMIX frames to reflect the behavior of the IP traffic in reality. We used a customized IMIX mixing frame sizes of 64, 128, 570, 373, 1400, 1518, and 9000 Byte according to our experience testing multiple vendors and service providers’ networks.

The table below shows the results of the forwarding performance tests:

<table>
<thead>
<tr>
<th>Test Focus</th>
<th>Test Parameters</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two 4x 100GbE (All across the backplane)</td>
<td>Frame Sizes 128-9000 Bytes and IMIX IPv4 and IPv6 traffic mix</td>
<td>Full-Line rate for all packet sizes.</td>
</tr>
<tr>
<td>Two 40x 10GbE (All across the backplane)</td>
<td>Frame Sizes 128-9000 Bytes and IMIX IPv4 only&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Full-Line rate for all packet sizes.</td>
</tr>
<tr>
<td>Full-Chassis</td>
<td>Frame Sizes 128-9000 Bytes and IMIX IPv4 only&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Full-Line rate for all packet sizes. 6.4 Tbit/s</td>
</tr>
</tbody>
</table>

<sup>a</sup> Main focus was IPv4 forwarding performance. IPv6 was only tested in the first scenario.

**Throughput Performance on a Fully Loaded Chassis**

After verifying that a line card support 400Gbit/s per slot, we turned our attention to the complete chassis. The router is expected to support multiservice core with high-performance and a flexible backbone architecture that can carry a wide range of services with high scale bandwidth. Can it perform just as well as it did in the single slot test when running all interfaces at the same speed?

The NE5000E can be loaded with 16 carrier cards, each supporting 400 Gbit/s of performance. In theory this means that the full router can support 6.4Tbit/s! Such capacity allows transporting 32 Ultra HDTV films per second (assuming each film size is 25 Gbyte).

In order to show that the NE5000E can reach the maximum throughput performance mixing variety of CP400 cards together, we used two 6x 40GbE in addition to the 10GbE and 100GbE based cards. We connected the router to the Spirent tester with 32x 100GbE, 10x 40GbE, and 280x 10GbE interfaces.

We verified that the NE5000E router can deliver 6.4 Tbit/s forwarding performance between all slots crossing from ingress to another egress slot through the backplane using all frame sizes (128 to 9,000 Byte and IMIX).

We then ran the test for a duration of 12 hours. Our intention was to try and get a closer look at the performance over an operational duration longer than the typical RFC 2544 test and much closer to the way that the router will be used in the real world. Since repeating all test runs with each packet size for this duration was impractical, we picked one packet size - 256 Byte. This size was not chosen at random. It is smaller than the average of packet used in the Internet nowadays. We performed the 12 hours test in a snake setup and did not observe any packet loss for the whole duration of the test.

**Figure 2: 400Gbit/s Per Line Card Test Results**

**Power Efficiency**

Service providers are sensitive to the amount of power consumed by their network devices. The more energy efficient a device is, the more a service provider can save in energy costs, which leads to reduction in operating costs (OPEX) and in some countries, to healthy government subsidies.
We measured the power consumption of Huawei’s NE5000E router based on test methodology defined by the Alliance for Telecommunications Industry Solutions (ATIS) ATIS-0600015.03.2009 standard. As the methodology defines, we performed three measurements at different traffic loads: 100% and 30% of capacity as well as during idle state. Each measurement was performed for 15-minutes duration.

The NE5000E supports two power modes. Besides the basic (default) power mode, the NE5000E also supports the so called “deep mode”. In “deep mode” the router will turn off unused components. This is done by dynamic energy consumption adjustments according to constant monitoring of traffic. Huawei explained that the power consumption is reduced when traffic is not at full rate. This is achieved by adjusting both frequency and number of active CPU cores in the forwarding chips. Huawei stated that they designed the forwarding chips with multi-cores to be used as a resource pool. In addition, Huawei built check points in the chips to monitor the load of the traffic. Based on the check point states the power management unit controls the number of cores and frequency to suit the loading rate of traffic.

We of course insisted on testing both power consumption modes and afterwards comparing them to see if there are tangible savings when using deep mode. We also made sure that saving on power did not translate to reduction in performance.

The results are Energy Consumption Rating (ECR) values that describe the amount of energy consumed by the device to move one Gigabit of line-level data per second. This value is taken at maximum throughput. We also measured another value, called Energy Consumption Rating Weighted (ECRW). ECRW is a weighted ECR ratio which takes the power consumption under a realistic load scenario into account. Running in basic mode, we measured an ECR rating of 1.38 watts per Gbit/s and 1.27 watts ECRW. The average power consumption per 100% load was 8,848 watts.

While the router was running in deep mode, we did not observe any packet loss or an increase of latency. Huawei recommend configuring deep mode when the traffic is not high. If there is no traffic on the line card, 5 to 7 percent power consumption can be saved of the line card at typical traffic model.

**Core Scale**

Once we were convinced that the line cards’ packet forwarding performance is in line with the demands put on core routers, we turned our attention to the next crucial elements - control plan scalability. To forward any packet, a core router needs to know where the packets are headed - which is where routing protocols as well as path establishment protocols come into play.

**FIB Scalability**

BGP is the undisputed king of inter-domain routing protocols. Since BGP routes are growing year by year, route table capacity is a very important metric for a network operator when designing and operating networks. Huawei confirmed that the maximum Forwarding Information Base (FIB) capacity for the NE5000E is 4,000,000 routes. This value can be reached for IPv4 only. In case the FIB would be loaded with IPv6 routes only, the maximum capacity would be 2,000,000. To better reflect the reality, we decided to use a mix of 80% IPv4 and 20% IPv6 routes. With this distribution the maximum number of total routes was expected to be 3.6 million. We emulated one BGP router with two BGP sessions between the tester and the NE5000E, one IPv4 and another IPv6 session. We advertised 3.2 million IPv4 routes and 400,000 IPv6 routes to the NE5000E, in total 3.6 million routes. We observed the resources and the forwarding behavior of the router when advertising all these routes. To verify that all routes are installed into the FIB, we sent test...
traffic toward all the advertised routes and measured packet loss.

After advertising the routes, the NE5000E populated all IPv4 and IPv6 routes into its FIB successfully. The maximum CPU utilization was less than 20% while learning the routes and less than 5% after the routes were installed into the FIB.

To verify that all routes were learned and installed into the FIB, we sent traffic for each of the routes and measured packet loss. We sent the traffic for a duration of two minutes so that BGP updates had to occur during the time that traffic was sent. Not a single packet was lost during the whole test duration.

To verify reliability and stability, we also tested what happens when the router is exceeding the maximum number of routes in the FIB by advertising additional 400,000 IPv4 and 200,000 IPv6 routes.

We identified that the NE5000E did not show an alarm message to indicate the overload. Huawei provided a software patch to implement this feature. We verified a correct alarm indication if the maximum number of routes have been exceeded with the new version.

**Distributed Processing for BGP and RSVP**

Now that we knew the scale of the router’s FIB we turned our attention to other control plane features supported by Huawei. One such feature that is meant to increase the number of maintained sessions and the processing power is a distributed processing function for both BGP and RSVP protocols.

Distributed RSVP-TE multi-processing supports more RSVP PATH and RESV refresh messages’ pressure. Processing of PATH and RESV messages is done per distributed RSVP-TE process independently, while in a non-distributed RSVP-TE structure, only one RSVP process handles all RSVP-TE tunnels with all PATH and RESV refresh messages. For service providers, these distributed control plane features are architectural improvements in terms of performance and scale, but not a functional feature. There is no change with regards to the protocol itself.

We verified 8 independent RSVP-TE processes, each handling 64,000 tunnels resulting in a total number of 512,000 RSVP-TE tunnels. We sent traffic and recorded less than 51 μs latency for traffic forwarded between all tunnels. None of the packet got lost.

In addition to distributed RSVP-TE, we tested distributed BGP processes at a high scale of routes. For service providers the implications of our findings are significant. Using 3 distributed BGP processes we were able to maintain 100 BGP peers with 50 million BGP routes, each peers advertised the same 500,000 prefixes. When we asked Huawei to increase the number of BGP processes to 5 we were able to maintain 200 BGP peers advertising 500,000 of the same prefixes. This resulted in a total number of 100 million BGP routes, from which 500,000 were stored in the FIB based on the best path selection process.

To reach this number, Huawei increased the memory of the MPU from 8 GByte to 16 GByte for this test.

As a conclusion, the NE5000E router was able to establish the neighborhood to 200 peers, each handling the same 500,000 BGP prefixes using 5 BGP processes.

**Resiliency Test**

Being so that core routers are the aggregation of all customer traffic, a failure could mean direct hit to many different revenue generating services. These days core routers carry voice over IP, mobile, residential and business traffic. Applications for interactive multimedia services such as Voice over IP (VoIP) and Video can be very sensitive to traffic loss, such as occurs when a link or router in the network fails. The cost of network failures in service provider networks is therefore significant and directly effect the customers’ satisfaction.
In order to test that the Huawei NE5000E, armed with its 400 Gbit/s-capacity line cards meets service providers’ service continuity requirements we simulated an array of device and link failures scenarios and measured the impact on running user traffic.

For extra-realism effect, we created a configuration, both on the tester and the solution under test, with a number of services and routes. We performed all the resiliency tests while these services where enabled:

- 128,920 MAC addresses in 1,172 MPLS VPLS instances (VSI)
- 489,896 routes in 1,172 BGP/MPLS L3VPNs
- Internet service with 780,000 routes within three L3VPNs
- IPv4 and IPv6 Multicast using PIM with 2 listeners
- 10,000 OSPF (Open Shortest Path First) routes
- 10,000 IP over MPLS services using LDP-based LSPs and 10,000 IS-IS routes
- One 6PE session with 2,000 IPv6 routes

We used IIMIX frame size for the test traffic sending 576 million frames per second in total.

**Hardware Redundancy: Control Plane and Switch Fabric Failure**

As the complexity of core routers increases on the path to higher throughput performance, so do the possibilities for hardware issues. At the heart of the NE5000E there are two main components: The Main Processing Unit (MPU) and the Switching Fabric Unit (SFU). The former manages, monitors, and maintains the boards, fan modules, and power modules of the entire router while the later is responsible to switch the traffic between the line cards. As per Huawei’s recommendation to service providers, the NE5000E router in the test was equipped with two MPU and four SFU cards. In both instances we emulated failure of one of the active units (or one of the units when all were active) by pulling the card out of the router while sending test traffic. We used the missing packets to calculate out of service time per flow. We observed that removing and inserting the MPU was hitless and did not affect the test traffic. After removing one SFU, we recorded an out of service time between 3.8 and 4.7 milliseconds. To calculate the time we have chosen the flow with the longest out of service time. The results were consistent, we repeated each failure scenario three times. The following figure shows the distribution of the flows interrupted by the SFU failover.

![Figure 5: SFU Out of Service Time Distribution](image)

**Routing Protocols Convergence**

When a network event causes topology changes, routers send routing update messages through the network. That causes the best path selection algorithms to recalculate optimal routes. We expect that the faster a router converges its routing table, the more available the network will be. Routers that converge slowly may cause temporary routing loops or temporary network unavailability.

We calculated the convergence time by measuring the timestamp of first lost packet on old path and the timestamp of last arrival packet on new path.

**IS-IS Convergence**

We measured IS-IS protocol convergence using link loss (Loss of Signal) as the trigger event to reflect the situation where a router directly connected to its adjacent experiences link issues. We used three Huawei routers for this test (2xNE5000E and 1xNE40E) and advertised 20,000 of IPv4 and 20,000 of IPv6 routes to the network reflecting a realistic IS-IS setup.

As a first step we tested IPv4 and IPv6 separately. For IPv4 traffic we measured an out of service time between 140 and 143 ms, for IPv6 between 100 and 106 ms.

Finally we tested a mix of IPv4 and IPv6 traffic. The measured convergence times ranged from 242 to 319 ms (average 278 ms).
BGP Convergence with Loss of Signal

BGP typically carry a lot more routes than internal routing protocols such as IS-IS. This applies especially for Internet gateways. Therefore we also measured the time the network takes to converge when BGP routes are added and removed from the network following a link failure.

We tested two scenarios, one dual stack scenario with a mix of IPv4 and IPv6 routes and one scenario with IPv4 routes only.

For the dual stack scenario we advertised 500,000 routes in each direction (1 million routes in total). we measured an out of service time between 3.3 and 5.7 s (average 4.67 s). For the IPv4 only scenario we advertised 400,000 routes in each direction (800,000 in total) and measured an out of service time between 1.8 and 4.1 s (average 2.87 s).

Service provider customers use packet loss as a key metric of their network service quality so minimizing the possibility for packet loss is obviously desired. It is important that networking devices remain operational and that failures are imperceptible to the user.

Functional Tests

Huawei explained that the NE5000E router is targeted for Internet backbone networks, IP dedicated backbone networks, or IP metropolitan core networks covering large cities and regions. It can also function as an egress router to the Internet in data centers. It is intended for network operators, enterprises, and public institutions. Therefore, the NE5000E need to provide a wide range of functions to fulfill different customers’ needs.

Unequal Cost Multipath (UCMP)

Today the requirement to handle large aggregations of traffic can be handled by a number of techniques which are collectively called multipath. Service providers use multipath links to connect to their customers, peers, or internally in the network. Multipath helps service providers to:

- Increase the total amount of link bandwidth and enjoy a higher bandwidth without replacing devices.
- Increase the reliability of the links. When one link is faulty, traffic can be switched to other links.

In Equal Cost Multiple Path (ECMP) mode, traffic is evenly load balanced among links to a destination regarding the differences of link bandwidths. When the link bandwidths differ greatly, traffic congestion occur on low-speed links, and the bandwidths of high-speed links are wasted. To fully utilize bandwidths of different links, traffic must be load balanced according to bandwidth ratio of these links. UCMP feature benefits the service provider and give them the flexibility to balance the traffic among different interfaces’ bandwidth which in turn allows them for better links’ and cost utilization.

Unequal Cost Multipath (UCMP) allows traffic to be distributed according to the bandwidth ratio of multiple unequal-cost paths that point to the same destination.

The Huawei’s UCMP implementation is based on interface bandwidth. When advertising routes, the Forwarding Information Base (FIB) module checks whether UCMP is enabled on outbound interfaces and records bandwidth information of the UCMP-enabled interfaces in the FIB. The Line Processing Card (LPU) calculates the distribution ratio of traffic according to the bandwidth ratio of interfaces that are involved in load balancing. Therefore, Huawei’s UCMP solution is independent to particular routing-protocols, it applies to all UCMP-enabled interfaces.

We tested UCMP functionality of the NE5000E router by advertising two different sets of IP routes for the same IP prefixes via BGP protocol through two links connected to the NE5000E router, one link connected to 100GbE and the other connected to 40GbE interfaces.

The figure below shows the setup of the test:

Figure 6: NE5000E UCMP Test Setup
The NE5000E showed expected UCMP functionality. We sent IPv4 test traffic for 60,000 advertised IPv4 routes at 4.6 Gbit/s, and observed that 5/7 of the IPv4 traffic (3.3 Gbit/s) was received on the 100GbE link from the NE5000E, and 2/7 of the IPv4 traffic (1.3 Gbit/s) on the 40GbE link. We calculated a weight ratio as 5:2 for IPv4 test traffic on NE5000E at UCMP enabled links, equal to ratio of the speed (100:40) of that links as expected.

**Virtual Router**

An important feature for a new generation of core routers is virtualization. By means of virtual router (VR), service providers can divide a large physical router into multiple small separate VRs, optimizing physical resource allocation and reducing capital and operational cost.

We verified that two separated and independent Autonomous Systems (ASes) co-existed on one physical NE5000E router, each AS implemented in separate virtual routers.

The figure below shows the setup of the test:

![Figure 7: NE5000E Virtual Router Test Setup](image)

During the test of two virtual router instances we concluded that:

- VRs can be configured and managed independently
- The interfaces on the DUT are split between different VRs, each AS in one VR
- Each AS is using dual stack IPv4/IPv6
- Both ASes could be accessed at the same time and configuration changes could be made at the same time without one AS affecting the other.

**Summary**

The capabilities enabled by the Huawei NE5000E router running Versatile Routing Platform (VRP) software were impressive. The NE5000E with its three different 400 Gbit/s line cards showed a high forwarding performance on both line card and full chassis load. EANTC sees Huawei pushing the boundaries concurrently for Ethernet density and services’ scale, as well as power efficiency.

The fact that the NE5000E router is used for multiple positions in the network, the port density such as 10GbE, 40GbE and 100GbE interfaces supported in the same NE5000E router has a big advantage to the service providers. It gives them the ability to use the router for different purposes in the network.

During this test campaign, EANTC verified the forwarding performance of 4x 100GbE, 10x 40GbE and 40x 10GbE line cards as well as the performance of the fully loaded chassis. We moved then to measure the route scale and route convergence time. Testing the power consumption using different traffic loads also took a place in this test. We finally tested the next-generation core routers’ features like distributed BGP, distributed RSVP, virtual router and unequal cost multipath load balancing.

**About EANTC**

The European Advanced Networking Test Center (EANTC) offers vendor-neutral network test services for manufacturers, service providers and enterprise customers. Primary business areas include interoperability, conformance and performance testing for IP, MPLS, Mobile Backhaul, VoIP, Carrier Ethernet, Triple Play, and IP applications.

EANTC AG
Salzufer 14, 10587 Berlin, Germany
info@eantc.com, http://www.eantc.com/