

Technical Validation

Maximizing Network Performance with Amazon Elastic Compute Cloud (Amazon EC2)

By Jack Poller, Senior Analyst

November 2019

This ESG Technical Validation was commissioned by Amazon Web Services and is distributed under license from ESG.



Contents

- Introduction 3
 - Background 3
 - Amazon Web Services Global Cloud 4
 - Amazon EC2 C5n Instances Featuring 100 Gbps of Network Bandwidth 4
- ESG Technical Validation..... 5
 - Network Throughput 5
 - What the Numbers Mean 7
 - Network Latency 8
 - What the Numbers Mean 9
 - Packet Rates..... 10
 - What the Numbers Mean 11
- The Bigger Truth..... 12

ESG Technical Validations

The goal of ESG Technical Validations is to educate IT professionals about information technology solutions for companies of all types and sizes. ESG Technical Validations are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objectives are to explore some of the more valuable features and functions of IT solutions, show how they can be used to solve real customer problems, and identify any areas needing improvement. The ESG Validation Team’s expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments.

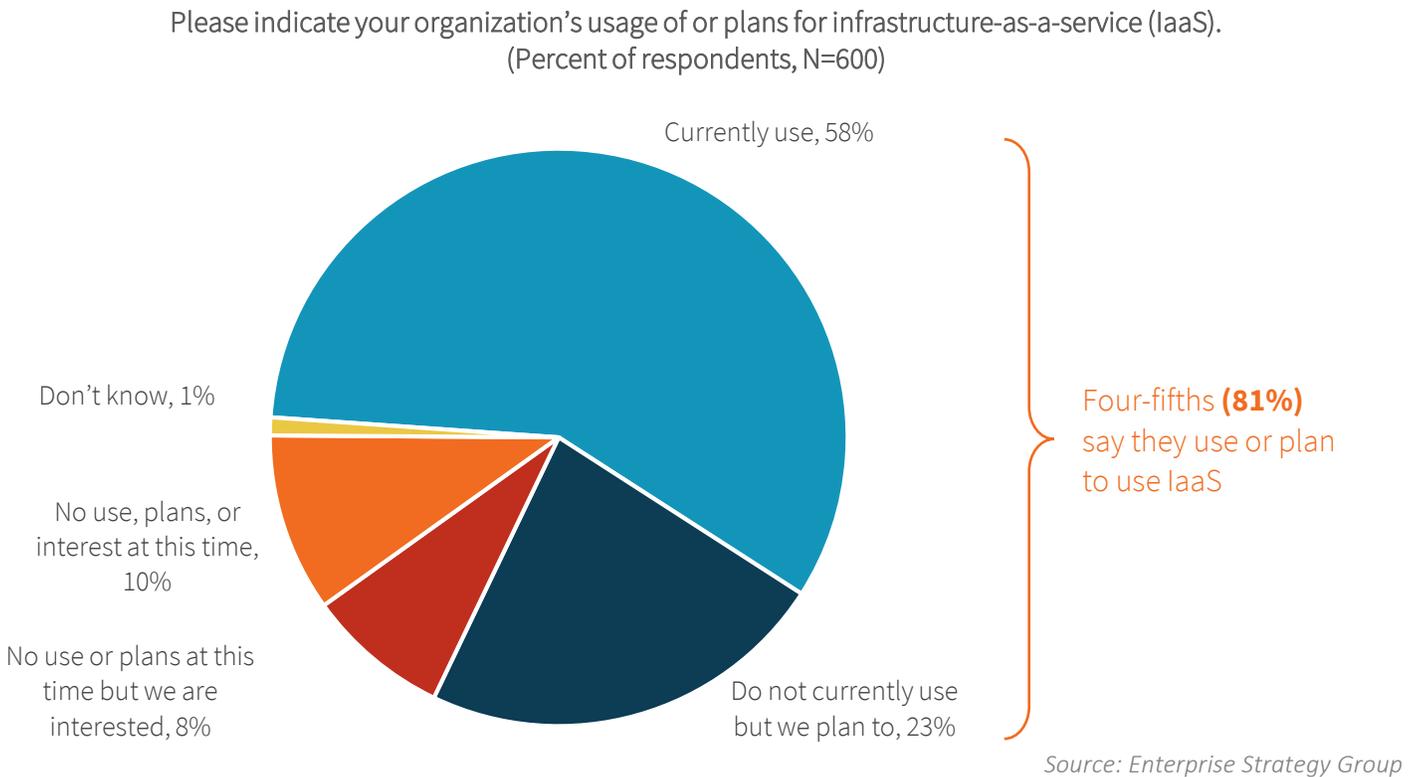
Introduction

This report documents ESG’s network performance testing of Amazon Web Services (AWS) cloud infrastructure using real-world customer-available compute server instances over six months. Performance of Amazon Elastic Compute Cloud (Amazon EC2) instances are compared against two cloud infrastructure compute server solutions.

Background

IT is growing more demanding every year, driven by increasing amounts of data, applications, devices, and users, as well as digital transformation initiatives. Organizations are turning to cloud infrastructure to meet these demands. Indeed, according to ESG research, four-fifths (81%) of organizations say that they use or plan to use infrastructure-as-a-service (IaaS).¹

Figure 1. Usage of or Plans for IaaS



Data drives business processes, making speed of access a business priority. Organizations are executing transactions and analyzing trends in real time to make better business decisions. New insights come from aggregating and analyzing volumes of data in data lakes, as well as the application of new techniques such as machine learning and deep learning. Generating valuable business insights depends on fast access to data.

Organizations need an IaaS stack that provides computational power for data analysis with network performance for access to large volumes of data. High network throughput and low latency enable organizations to leverage even the largest and most complex AI models and data lakes.

¹ Source: ESG Research Report, [2019 Public Cloud Computing Trends](#), April 2019.

Amazon Web Services Global Cloud

Amazon Web Services' global cloud infrastructure was designed to provide flexible, reliable, scalable, and secure cloud computing solutions with high quality network performance. Amazon Web Services incorporated redundancy and reliability into core components and locations. Amazon Elastic Compute Cloud (Amazon EC2) is a key component of Amazon Web Services' solution, providing resizable compute capacity.



Using a simple web service interface or an API, organizations can obtain and configure compute capacity on demand.

Organizations using Amazon Web Services' IaaS can benefit from:

- **Performance**—Amazon Web Services designed the infrastructure to provide high-throughput, low-latency networking.
- **Availability**—Availability zones provide physical redundancy, resiliency, and uninterrupted access.
- **Security**—The core infrastructure was designed to meet the more stringent large enterprise, military, and government security requirements, and is continuously monitored to help ensure the confidentiality, integrity, and availability of data. Amazon Virtual Private Cloud (VPC) provides additional security, isolating resources and providing access using an IPsec-based VPN.
- **Reliability**—Amazon Web Services builds data centers in multiple geographic regions, and each region is isolated from others. Redundancy of components, networks, and availability zones enables Amazon Web Services to provide a 99.99% uptime commitment.
- **Scalability and elasticity**—Customers can scale from one to many thousands of simultaneous server instances, and capacity can be automatically or manually increased or decreased within minutes.
- **Complete control**—Root access and complete control of the virtual server enable administrators to treat Amazon EC2 instances the same as any other compute server. API access enables automation and orchestration of system startup and shutdown.
- **Flexibility**—Multiple configurations of memory, CPU, instance storage, boot partition size, network capacity, and operating systems enable organizations to optimize server configuration to requirements. Amazon Web Services offers bare metal servers and access to ARM CPUs in addition to Intel and AMD x86 CPUs.
- **Integration**—Amazon EC2 is integrated with most Amazon Web Services including Amazon Simple Storage Service (Amazon S3), Amazon Relational Database Service (Amazon RDS), and Amazon Virtual Private Cloud (Amazon VPC), providing a complete IaaS solution for computing, query processing, and cloud storage.

Amazon EC2 C5n Instances Featuring 100 Gbps of Network Bandwidth

Amazon Web Services designed the C5n instance type to meet demanding data transfer needs, reducing data ingestion wait times and speeding up delivery of results. C5n instances can utilize up to 100 Gbps network bandwidth depending on the instance size.

The C5n instances use 3.0GHz Intel Xeon Skylake CPUs with support for the Intel Advanced Vector Extensions 512 (AVX-512) instruction set. The instances leverage the Nitro system, a collection of AWS-built hardware and software components enabling high performance, high availability, and high security.

ESG Technical Validation

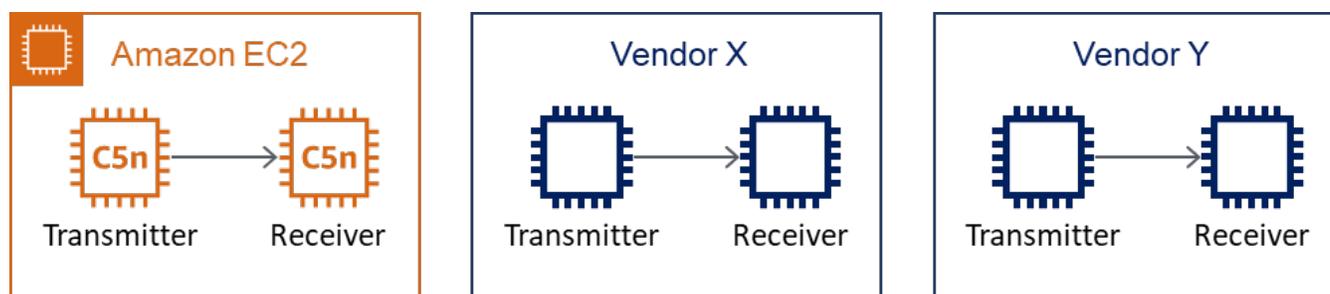
ESG tested and compared the network performance of Amazon Web Services with two comparable cloud IaaS solutions. Like Amazon Web Services, Vendor X and Vendor Y provide a broad offering of virtual servers running in multi-region global data centers. Each solution provides VMs with persistent storage, and configurations can be optimized for memory, CPU, and storage requirements.

ESG tested the largest general-purpose compute server configuration for each vendor. We used the Amazon EC2 C5n.18xlarge instances, consisting of 72 Intel Xeon vCPUs, 144 GiB memory, and no local storage. The general-purpose compute servers from Vendor X and Vendor Y included 64 to 96 Intel Xeon vCPUs and 256 to 360 GiB memory.

Testing used open source and Linux tools to measure network throughput, latency, and packet rates. Testing did not use local storage and did not consume either the available memory or CPU. Thus, differences in the number of vCPUs, the amount of memory, or the amount or type of local storage did not affect the results.

Testing used a simple environment consisting of two server instances in each IaaS provider environment, with one server acting as a data transmitter and the other server acting as a data receiver, as shown in Figure 2.

Figure 2. Test Environment



Source: Enterprise Strategy Group

To ensure the results replicated real customer experiences, testing was conducted at random times every two to four days over a six-month period from May through October 2019. Testing was automated using each IaaS provider's API, and servers were programmatically instantiated. The Amazon EC2 instances were tested using both Amazon Linux AMI release 2018.03 and Ubuntu 16.04.6 LTS. Both Vendor X and Vendor Y were tested using Ubuntu 16.04.6 LTS.

Network Throughput

ESG measured network throughput using [iperf2](#), an open source software tool designed to help users measure and tune network performance. Iperf2 has both client and server functionality and can create data streams to measure throughput between the two ends of a connection in either or both directions. Iperf2 parameters are shown in Table 1. All other iperf2 and operating system parameters were set to the default values.

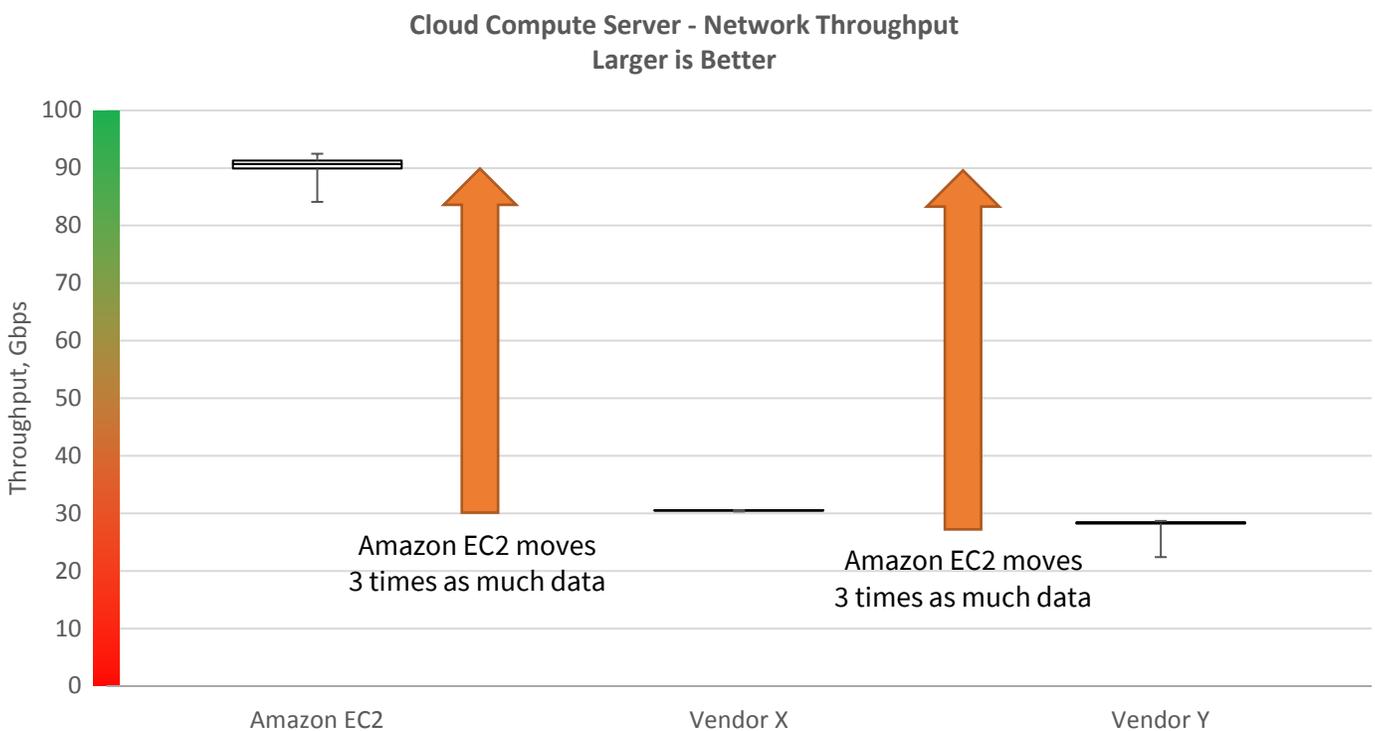
Table 1. Network Throughput: Iperf2 Configuration

Parameter	Value
Protocol	TCP
Direction	One-way
Duration	60 seconds
Buffer size	1K
Parallel data streams	32

Source: Enterprise Strategy Group

Throughput measurements are shown in a box-and-whisker plot (boxplot), a standardized method for displaying the distribution of data based on a five-number summary. The minimum and maximum values are shown with the whiskers. The median, or middle value, of the data is shown by the line between the two boxes. The first quartile (25th percentile) is the middle number between the smallest number and the median and is represented by the lower edge of the lower box. The third quartile is the middle value between the median and the highest value and is represented by the upper edge of the upper box. The size and location of the boxes indicates whether the data is symmetrical, skewed, or tightly grouped.

Throughput measurements are shown in Figure 3, and are summarized in Table 2. Amazon EC2 C5n instances achieved a median throughput of 90 Gbps, moving three times as much data as Vendor X (30 Gbps) or Vendor Y (28 Gbps).

Figure 3. Network Throughput


Source: Enterprise Strategy Group

Table 2. Network Throughput

	Amazon EC2	Vendor X	Vendor Y
Minimum (Gbps)	84.10	29.70	24.71
1st Quartile (Gbps)	89.00	30.31	28.26
Median (Gbps)	90.11	30.34	28.32
3rd Quartile (Gbps)	90.84	30.36	28.46
Maximum (Gbps)	92.03	30.40	28.67

Source: Enterprise Strategy Group

What the Numbers Mean

- Amazon EC2 achieved a median throughput of 90 Gbps, moving three times more data than Vendor X (30 Gbps) or Vendor Y (28 Gbps).
- Amazon EC2 throughput was consistent, ranging from 84.1 Gbps (6.7% below the median) to 92.0 Gbps (2.1% above the median). With 50% of the measurements grouped tightly around the median (1.2% below to 0.8% above the median), Amazon EC2's consistent throughput leads to consistent and predictable performance for demanding applications such as HPC data analysis and AI model development.
- Vendor X also maintained consistent throughput from 2.1% below the median to 0.2% above the median.
- Vendor Y's throughput was much more variable, ranging from 12.7% below the median to 1.2% above the median. This inconsistent and variable performance was mostly measured as reduced throughput (below the median), indicating that organizations using Vendor Y may not achieve the performance they expect.



Why This Matters

IaaS makes it possible for organizations to leverage the scale and economics of the solution providers—but if that scale hinders data throughput, analyses may be delayed, and business insights and decision making may suffer. Thus, cloud compute servers need to meet and exceed application compute power and network throughput requirements.

ESG validated that over six months of testing, Amazon EC2 C5n instances achieved a median of 90 Gbps network throughput, moving three times as much data as comparable solutions. Amazon EC2 network throughput was consistent, ranging from 84 to 92 Gbps (6.7% below the median to 2.1% above the median). Large and consistent throughput enables organizations to deploy HPC, AI, and data analysis applications in the cloud, leveraging cloud scale and economics to achieve faster insights.

Network Latency

ESG measured network latency using qperf, an open source software tool available in standard Linux distributions. Qperf was designed to help users measure and tune network performance. Qperf has both client and server functionality and can create data streams to measure latency between the two ends of a connection in either or both directions. Qperf parameters are shown in Table 3. All other qperf and operating system parameters were set to the default values.

Table 3. Network Latency: Qperf Configuration

Parameter	Value
Protocol	TCP
Duration	2 seconds
Buffer size	1K
Parallel data streams	32

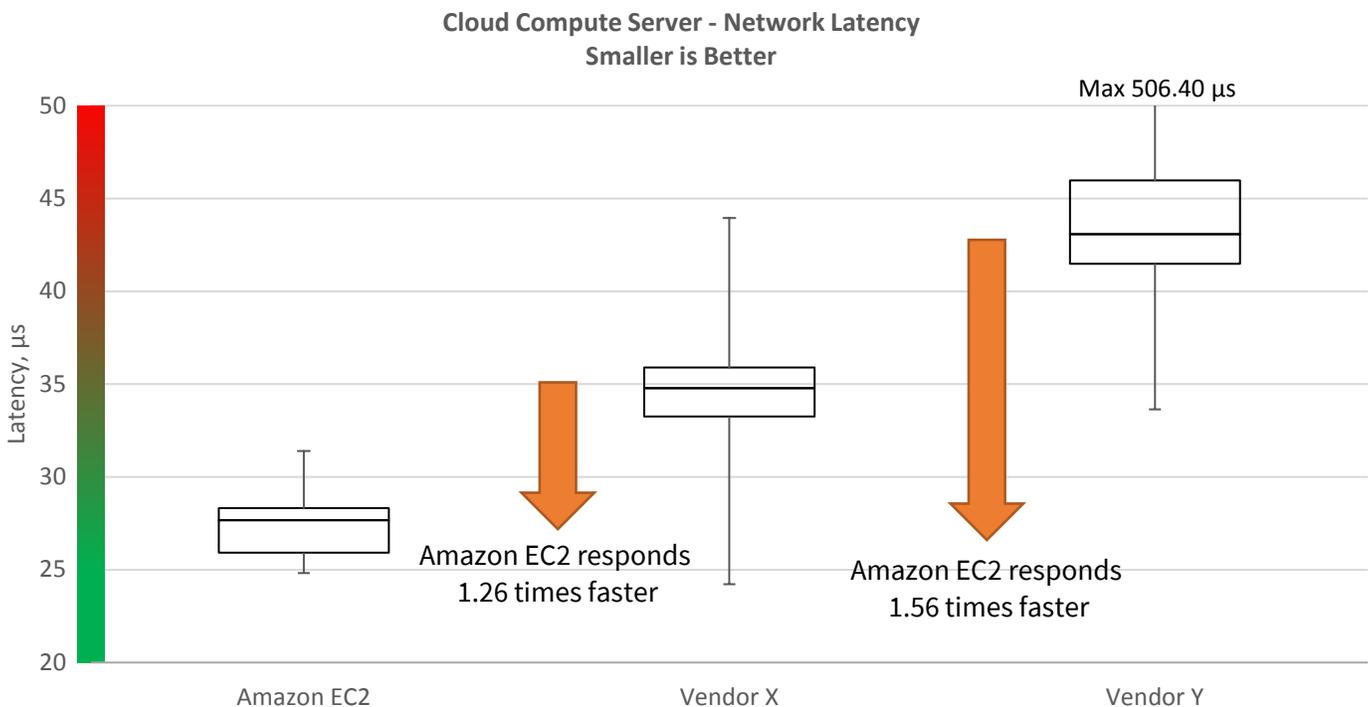
Source: Enterprise Strategy Group

As with network throughput, latency measurements are shown in a box-and-whisker plot (boxplot).

Latency measurements are shown in Figure 4 and are summarized in Table 4. Amazon EC2 instances achieved a median latency of 27.66 μ s, 26% faster than Vendor X (median latency 34.78 μ s) and 56% faster than Vendor Y (median latency 43.08 μ s).

For Vendor Y, ESG recorded multiple measurements of latency over 50 μ s and measured a maximum latency of 506.40 μ s. Displaying the maximum recorded value on the boxplot would make the graph unreadable. Therefore, the boxplot displays data for latency between 20 and 50 μ s.

Figure 4. Network Latency



Source: Enterprise Strategy Group

Table 4. Network Latency

	Amazon EC2	Vendor X	Vendor Y
Minimum (μ s)	24.82	24.22	33.64
1st Quartile (μ s)	25.93	33.26	41.49
Median (μ s)	27.66	34.78	43.08
3rd Quartile (μ s)	28.32	35.89	45.98
Maximum (μ s)	29.66	39.22	506.40

Source: Enterprise Strategy Group

What the Numbers Mean

- Amazon EC2 instances achieved a median latency of 27.66 μ s, 26% faster than Vendor X (median latency 34.78 μ s) and 56% faster than Vendor Y (median latency 43.08 μ s). Amazon EC2's low latency means that applications requesting data don't have to wait as long for data and can consistently respond to user requests faster.
- Amazon EC2 latency was consistent, ranging from 24.82 μ s (10.3% below the median) to 29.66 μ s (7.2% above the median). With 50% of the measurements grouped tightly around the median (6.3% below to 2.4% above the median), Amazon EC2's consistent latency leads to consistent and predictable performance for demanding applications such as HPC data analysis and AI model development.
- Vendor X's latency was much more variable, ranging from 30.4% below the median to 12.8% above the median. With inconsistent and variable latency, organizations using Vendor X may see a variable application performance and underutilize compute resources waiting for network packets.
- Vendor Y's latency was much more variable, ranging from 21.9% below the median to an astounding 1,076% above the median. With inconsistent, variable, and exceptionally long latency, organizations using Vendor Y will see a wild variance in application performance and a measurable impact in user experience.

Why This Matters

Modern performance-oriented workloads require high throughput, low latencies, and flexible scalability to keep up with today's demands for data. Response time is a crucial component of performance, affecting user experience, productivity, decision making, and the time to execute business-critical jobs.

ESG validated that over six months of testing, Amazon EC2 C5n instances achieved a median network latency of 27.66 μ s, 26% to 56% faster than comparable solutions. Amazon EC2 network latency was consistent, ranging from 24.82 μ s (10.3% below the median) to 29.66 μ s (7.2% above the median). Fast and consistent latency enables applications to keep data pipelines full and CPUs working rather than idling waiting for data. These results validate that Amazon EC2 C5n instances are well suited for HPC, AI, data analytics, and other applications that require fast response to data requests.

Packet Rates

ESG measured packet rates using iperf2 with the parameters shown in Table 5. All other iperf2 and operating system parameters were set to the default values.

Table 5. Iperf2 Packet Rate Testing Configuration

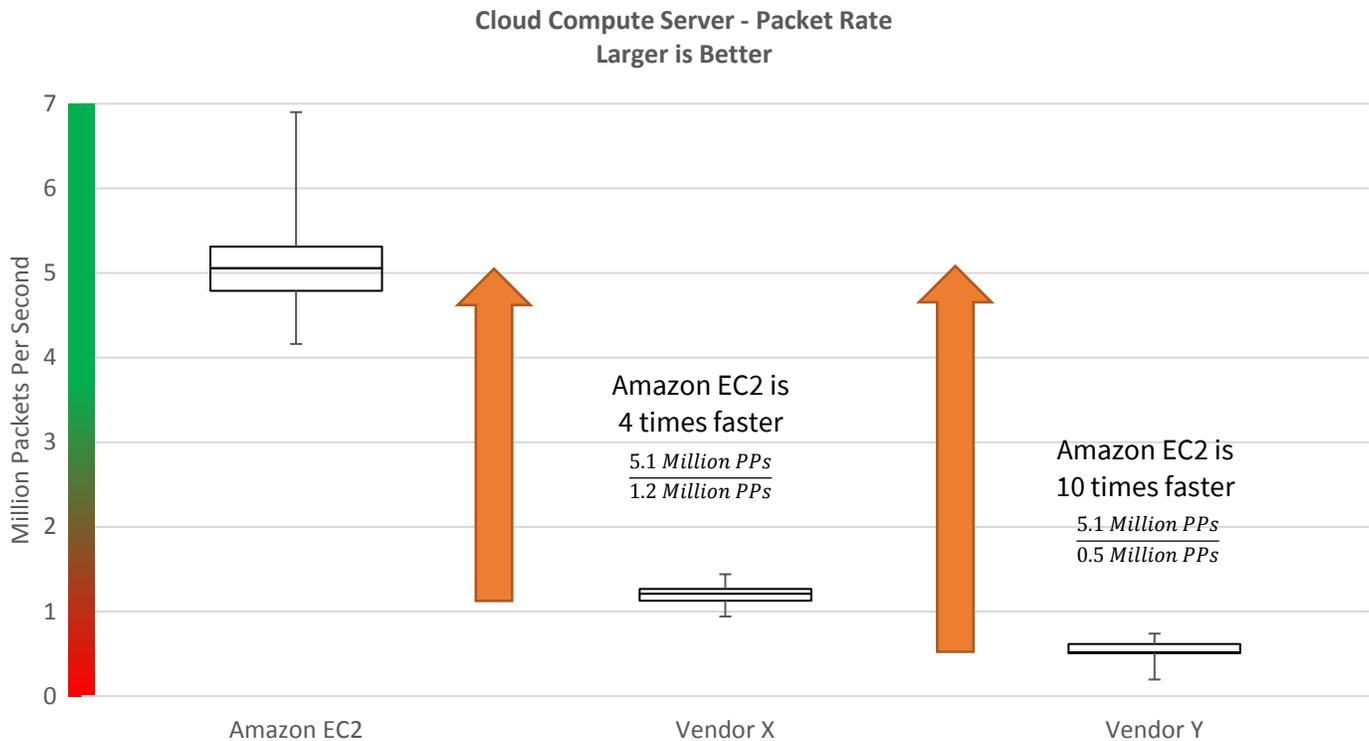
Parameter	Value
Protocol	TCP
Duration	60 seconds
Buffer size	96 bytes
TCP maximum segment size (MSS)	96 bytes
Parallel data streams	64

Source: Enterprise Strategy Group

As with network throughput and latency, packet rate measurements are shown in a box-and-whisker plot (boxplot).

Packet rate measurements are shown in Figure 5, and are summarized in Table 6. Amazon EC2 instances achieved a median of 5.05 million packets per second, more than four times as fast as Vendor X (median 1.2 million packets/sec) and almost ten times as fast as Vendor Y (median 0.52 million packets/sec).

Figure 5. Packet Rates



Source: Enterprise Strategy Group

Table 6. Packet Rates

	Amazon EC2	Vendor X	Vendor Y
Minimum (packets/sec)	4,161,392	940,575	196,063
1st Quartile (packets/sec)	4,789,554	1,127,122	507,513
Median (packets/sec)	5,055,019	1,209,995	516,831
3rd Quartile (packets/sec)	5,312,579	1,267,615	616,769
Maximum (packets/sec)	6,900,719	1,440,055	738,358

Source: Enterprise Strategy Group

What the Numbers Mean

- Amazon EC2 instances achieved a median of 5.05 million packets per second, more than four times as fast as Vendor X (median 1.2 million packets/sec) and almost ten times as fast as Vendor Y (median 0.52 million packets/sec). Amazon EC2's higher sustained packet rate performance helps ensure that applications driving high network throughput with small packets such as analytics and real-time communication can achieve peak performance.
- Amazon EC2's small packet transfer rate was consistent, ranging from 4.16 million packets/sec (18% below the median) to 6.90 million packets/sec (27% above the median). With 50% of the measurements grouped tightly around the median (5% below to 5% above the median), Amazon EC2's consistent packet transfer rate means consistent and predictable performance for network appliances such as firewalls and routers, and video and audio conferencing applications.
- The consistency of Vendor X's packet transfer rate was similar to Amazon EC2, ranging from 22% below the median to 19% above the median. However, with slower packet transfer rates, organizations using Vendor X may not be able to maximize performance of their packet rate bound applications.
- Vendor Y's packet transfer rate was much more variable, ranging from 0.20 million packets/sec, 62% below the median, to 0.74 million packets/sec, 43% above the median. With inconsistent, variable, and low packet transfer rates, organizations using Vendor Y will achieve lower performance of their packet rate constrained applications.

Why This Matters

Analytics, real-time communication, and network applications rely on consistent high packet transfer rates to achieve peak application performance.

ESG validated that over six months of testing, Amazon EC2 C5n instances achieved a median of over 5 million 96-byte packets per second, four to ten times faster than comparable solutions. Amazon EC2 packet rates ranged from 4.2 million to 6.9 million packets/sec, and packet rates were consistent with 50% of the measurements ranging from 5% below to 5% above the median. Applications can depend on Amazon EC2 C5n instances for fast and consistent data transfers.

The Bigger Truth

Public cloud compute servers deployed to support tier-1 data-intensive applications need to provide consistent network performance, sustaining high throughput and providing low-latency data access. Predictable performance is critical to maximize productivity throughout the organization.

Amazon EC2 is designed to provide scalable, elastic, controllable, flexible, reliable, and secure cloud-based compute resources. Amazon Web Services developed the C5n instance to meet the demanding needs of HPC, AI, data analytics, and other data-intensive applications. Leveraging the Nitro system, an Amazon Web Services-designed hardware and software platform, enables C5n instances to utilize up to 100 Gbps network bandwidth. AWS provides C5n instances with 3.0GHz Intel Xeon Skylake CPUs and one-third more memory than C5 instances to process data in volume.

ESG validated that over six months, Amazon EC2 C5n instances:

- Sustained 90 Gbps median large data block throughput (ranging from 84-92 Gbps), transferring three times more data than two alternative IaaS solutions.
- Provided consistent and predictable throughput, with 50% of the tests sustaining 89.0-90.1 Gbps.
- Provided 27.66 μ s median latency (ranging from 24.81-29.66 μ s), responding 26-56% faster than two alternative IaaS solutions.
- Provided consistent and predictable latency, with 50% of the tests responding between 25.93-28.32 μ s.
- Sustained 5 million median small packets/sec, four to ten times the transfer rate of alternative solutions.
- Provided consistent and predictable packet transfer rates, with 50% of the tests sustaining 4.8-5.3 million packets/sec.

The testing used in this report attempted to replicate customer experiences, using customer-available instances and industry-standard benchmarking tools deployed in a controlled environment. However, IaaS vendors practice continuous improvement, and new compute server instance types may provide different results. Due to the many variables in each IaaS solution, performance planning and testing with your own applications and environment are recommended. Readers are well advised to always explore the details behind any vendor testing to understand the relevance to your environment.

Competition drives enterprises today to ever-higher levels of performance for analytics, HPC, AI, and transaction-oriented workloads. But IT must always make a tradeoff between performance and cost, even with technology innovations like cloud-based compute servers. If your organization needs consistent and predictable high-performance low-latency data transfers for business-critical data-intensive workloads, Amazon EC2 is worth a hard look.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of The Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.



Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

© 2019 by The Enterprise Strategy Group, Inc. All Rights Reserved.

