**Networking performance of EC2 instances**

**TL;DR**

The network bandwidth available to an EC2 instance depends on factors such as the destination of the traffic, the number of vCPUs, and the instance size. To maximize bandwidth and reduce latency, users can utilize different techniques such as cluster placement group, Multipath TCP, and enhanced networking with the Elastic Network Adapter (ENA). Network performance metrics can be used to monitor instances when traffic exceeds network allowances, while CloudWatch metrics can be used to monitor network bandwidth and packets sent and received.

**Full text:**

Amazon EC2 instance bandwidth specifications apply to both inbound and outbound traffic. The available network bandwidth for an instance depends on several factors such as destination of the traffic, number of vCPUs, and instance size. Instances with 16 vCPUs or fewer have a baseline bandwidth and can burst beyond their limit for a limited time using network I/O credit mechanism. CloudWatch metrics can be used to monitor instance network bandwidth and packets sent and received.

To increase single-flow bandwidth and reduce latency, users can use a cluster placement group, set up multiple paths with Multipath TCP, or enable enhanced networking with the Elastic Network Adapter (ENA). Network performance metrics provided by the Elastic Network Adapter driver can be used to monitor instances when traffic exceeds the network allowances defined by Amazon EC2 at the instance level. However, it is possible for CloudWatch instance metrics to not reflect microsecond spikes in demand for network resources.

A cluster placement group is a logical grouping of instances in a single Availability Zone that provides higher per-flow throughput limits and lower network latency for applications that require high network throughput or low network latency, or both. It is recommended to use the same instance type for all instances in the placement group and to launch all instances in a single launch request to minimize the chances of getting an insufficient capacity error. Additionally, instances in the placement group must be launched in the same Availability Zone and enhanced networking is recommended for the lowest latency and highest packet-per-second network performance.

Enhanced networking is a feature that uses single root I/O virtualization (SR-IOV) to provide high-performance networking capabilities on supported instance types. SR-IOV provides higher I/O performance and lower CPU utilization compared to traditional virtualized network interfaces. This feature provides higher bandwidth, higher packet per second (PPS) performance, and consistently lower inter-instance latencies without any additional cost. The Elastic Network Adapter (ENA) supports network speeds of up to 100 Gbps for supported instance types.