Cloud Performance Benchmark Series

Network Performance: Amazon EC2

Sumit Sanghrajka
Nilesh Mahajan
Radu Sion

1. Overview

This document details a network performance benchmark of the Amazon EC2 cloud. TCP/UDP bandwidths and latencies within and across different deployment regions of Amazon are included.

2. Setup

We explored whether different instances are connected differently and concluded that the connection backplane is the same, or at least bottlenecks are clearly lying elsewhere. The experimental setup was thus based on m1.small EC2 instances. As observed network performance varies naturally with usage peaks and time of day, experiments were repeated at four different times of day.

Bandwidth, delay jitter and datagram loss were measured using the Iperf benchmarking framework. Unlike other network benchmarking tools, Iperf consumes very few system resources, thus resulting in more precise measurements. Round-trip latency was defined as ½ of the time taken for a single packet to be echoed back, and measured using custom designed code. This was also performed for both TCP and UDP.

The following peer pairs were considered in the experimental setup:
- *Within* the US East, US West, Europe, and Asia Pacific regions
- Between local clients (Stony Brook, NY) and all regions.
- Between US East and US West, US East and Europe, US East and Asia Pacific, US West and Europe, US West and Asia Pacific, and Europe and Asia Pacific regions

All benchmarks were (repeatedly) repeated at four different times of the day – Morning (around 6am), Afternoon (around 12noon), Evening (around 6pm), and Night (around 12am) Eastern Time mid-week. Averages are presented for each time.
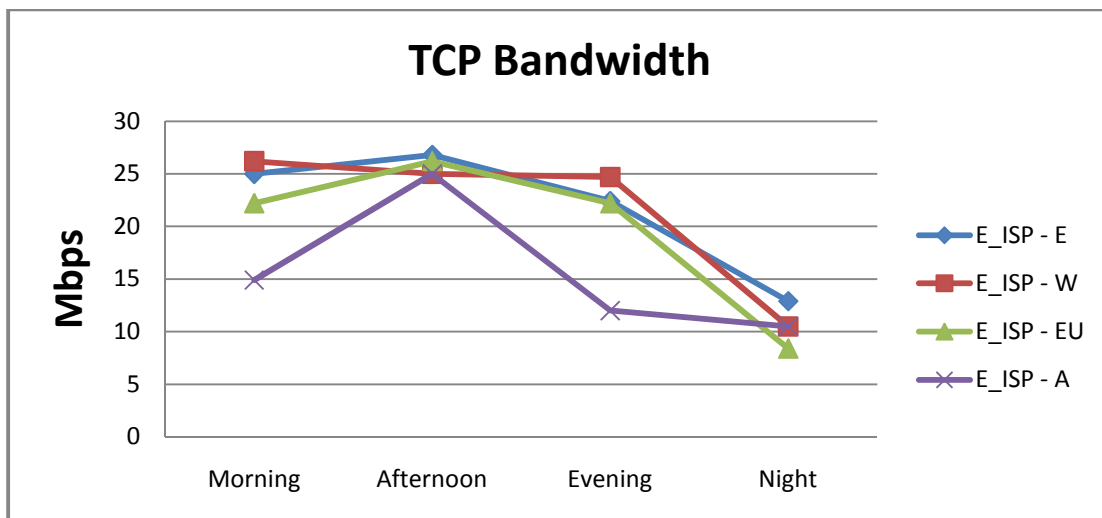
3. Results

3.1. Bandwidth

3.1.1. TCP Bandwidth. Intra-region.
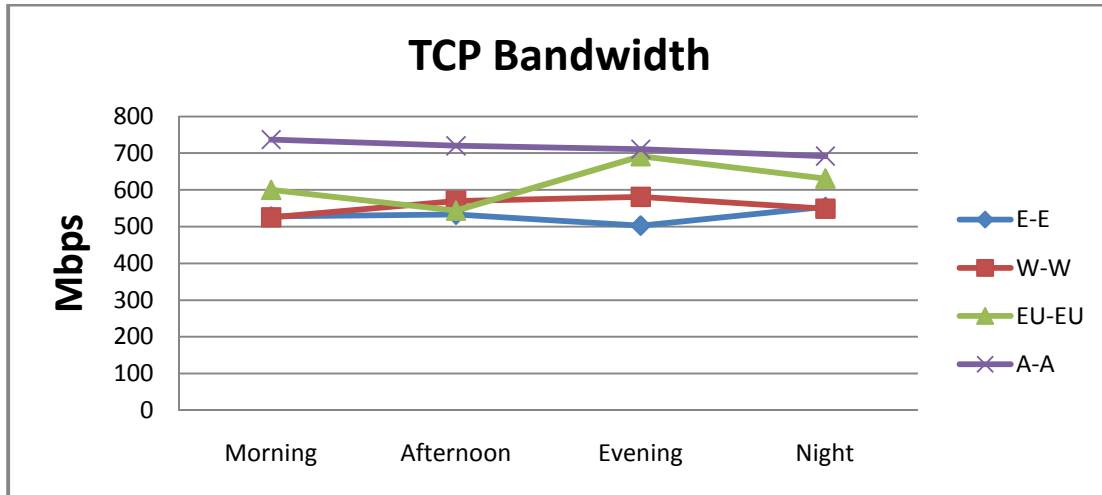
Observation: The intra region bandwidth of Asia Pacific region appears slightly higher than the other regions at all the different times. This is likely due to the fact that this is the latest region in the line-up and as such it has less traffic.

3.1.2. TCP Bandwidth. Stony Brook, NY to/from different regions.

Observation: Bandwidth is lowest during night time for all the regions, most likely confirming the urban myth that multimedia is consumed mainly at night-time (at least in New York, our main vantage point). We performed a brief set of validation experiments

with clients from a different provider in New York (Optimum Inc.) and similar behavior has been observed, thus ruling out a Stony Brook campus – specific bias.

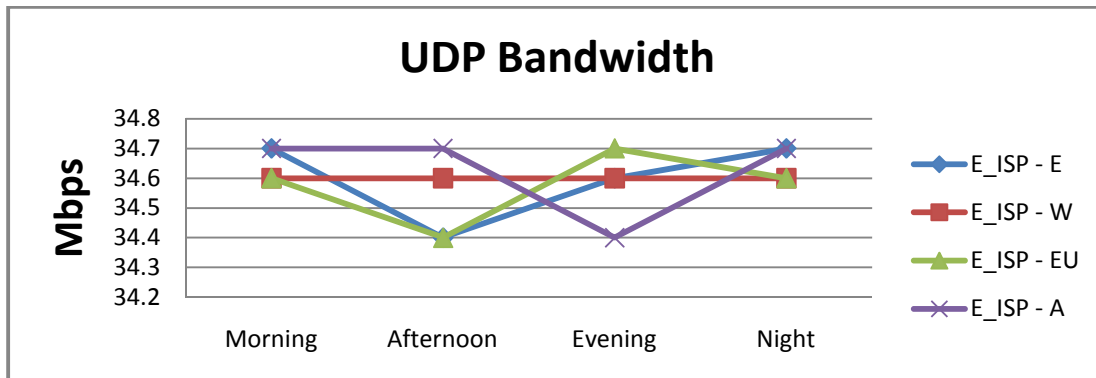### 3.1.3. TCP Bandwidth. Inter-region.
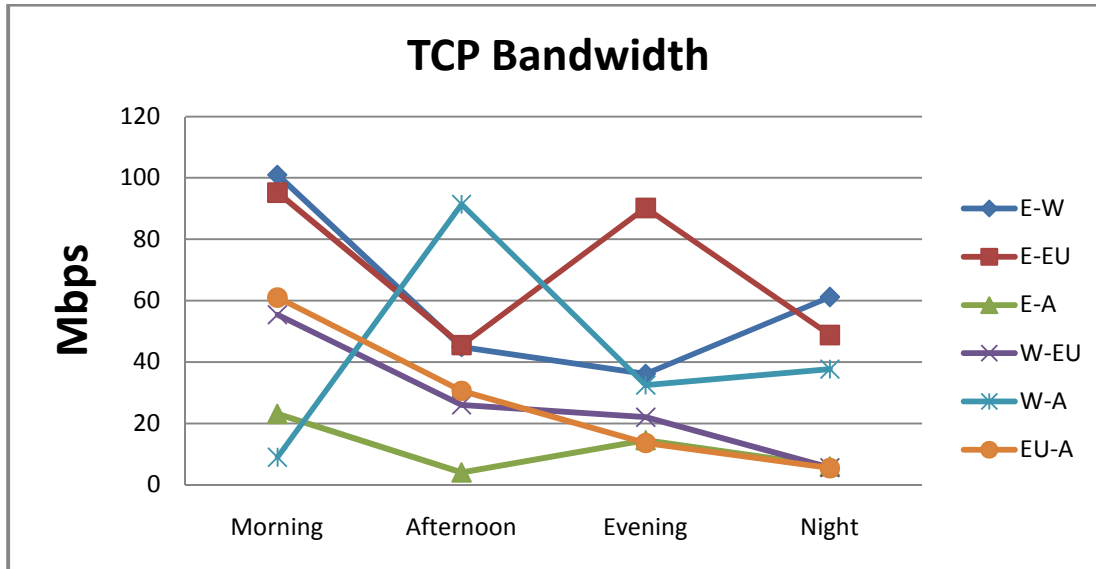
Observation: Not surprisingly, the best connected pair is US E-W. On the other hand, somewhat unexpectedly, the Eastern US region is strongly connected with Europe early in the morning. This may illustrate a combination of lunchtime in Europe with closed business / stock market in the US.

### 3.1.4. UDP Bandwidth. Intra-region.
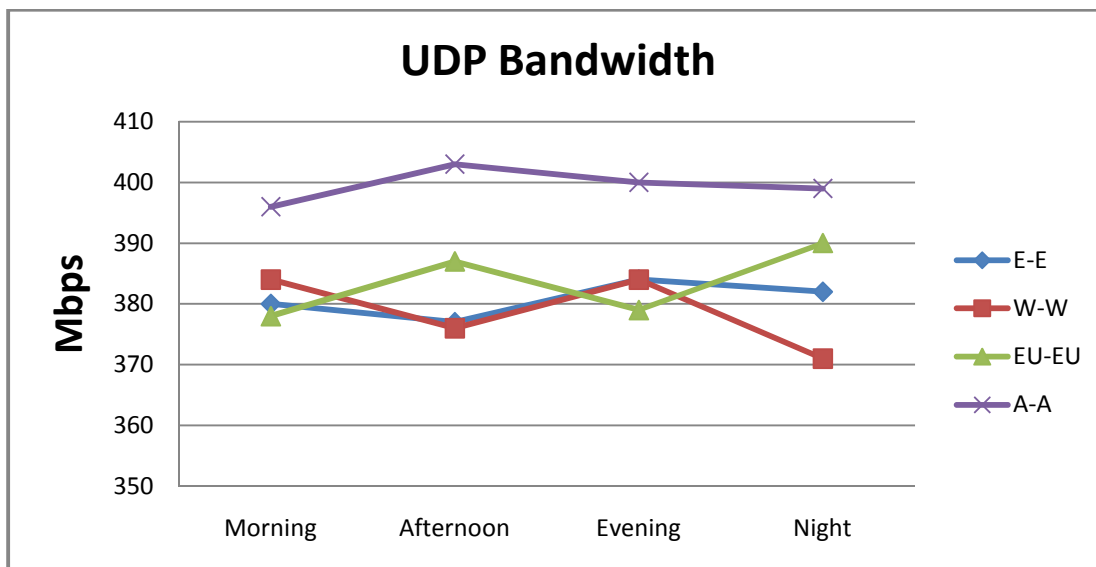
Observation: The intra-region bandwidth of the Asia Pacific region seems slightly higher than the other regions at all the different times. Reason may be as above for TCP.

3.1.5. UDP Bandwidth. Stony Brook, NY to/from different regions.

Observation: Surprisingly, the numbers are almost identical for all regions. Varying drop rates combined with ISP imposed rate-caps may be to blame for the discrepancy between this behavior and the TCP observations.
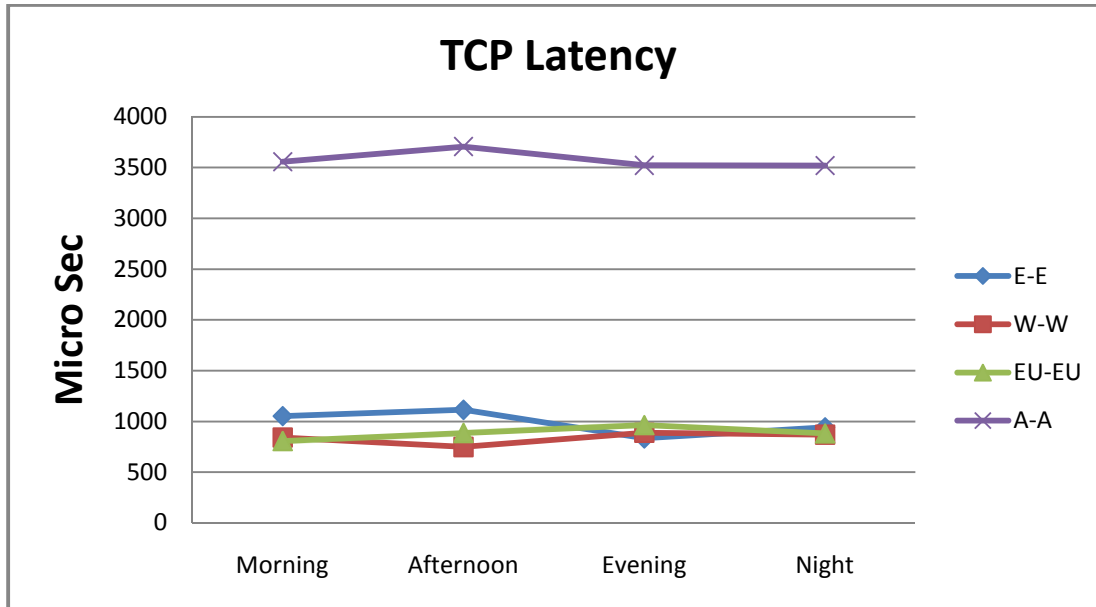
3.1.6. UDP Bandwidth. Inter-region.
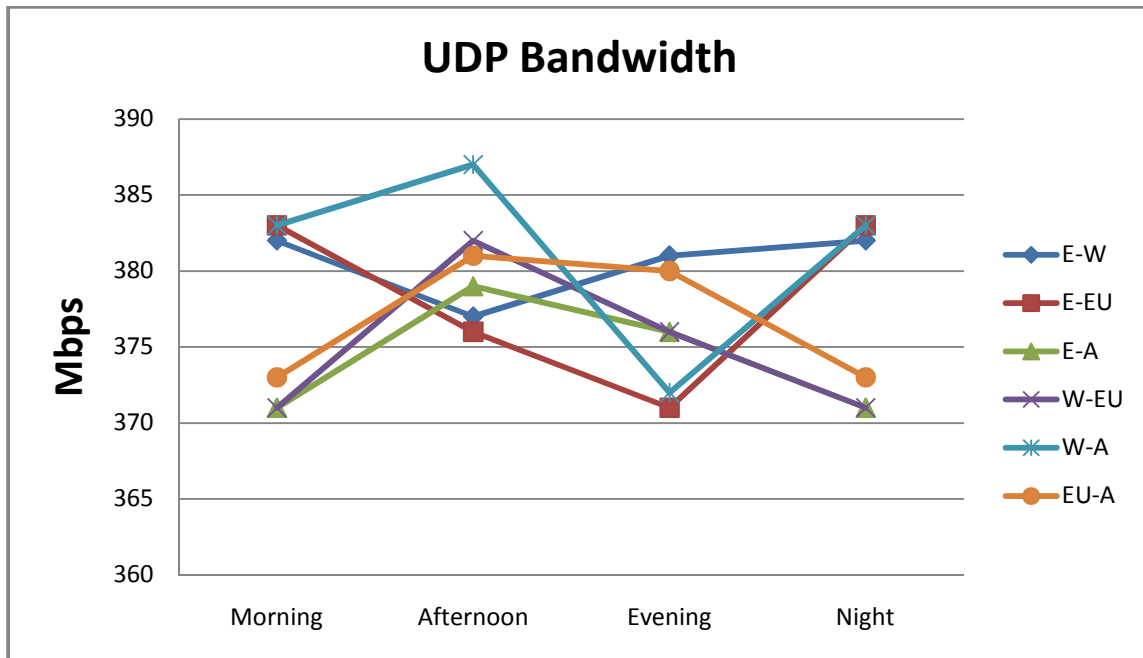
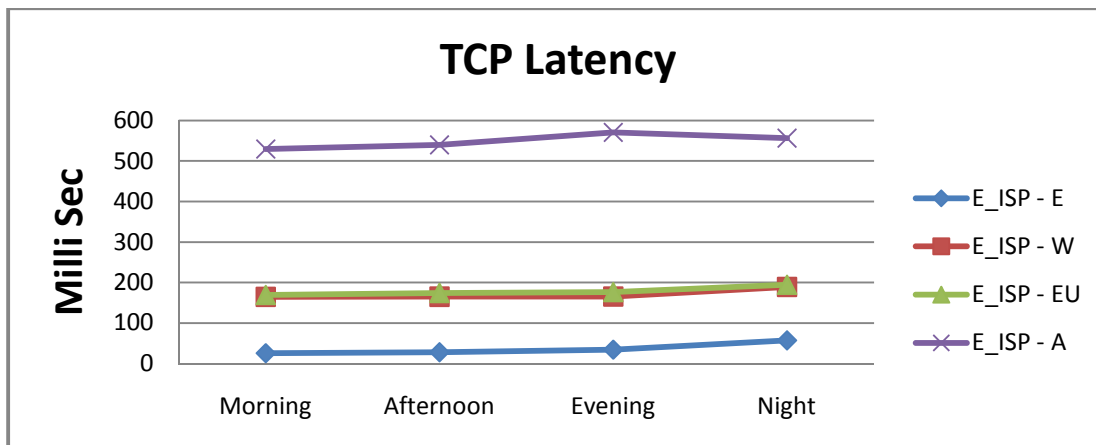**Observation:** As expected, intra-region clusters are strongly connected with 350Mbps+ trunks.

3.2. Latency

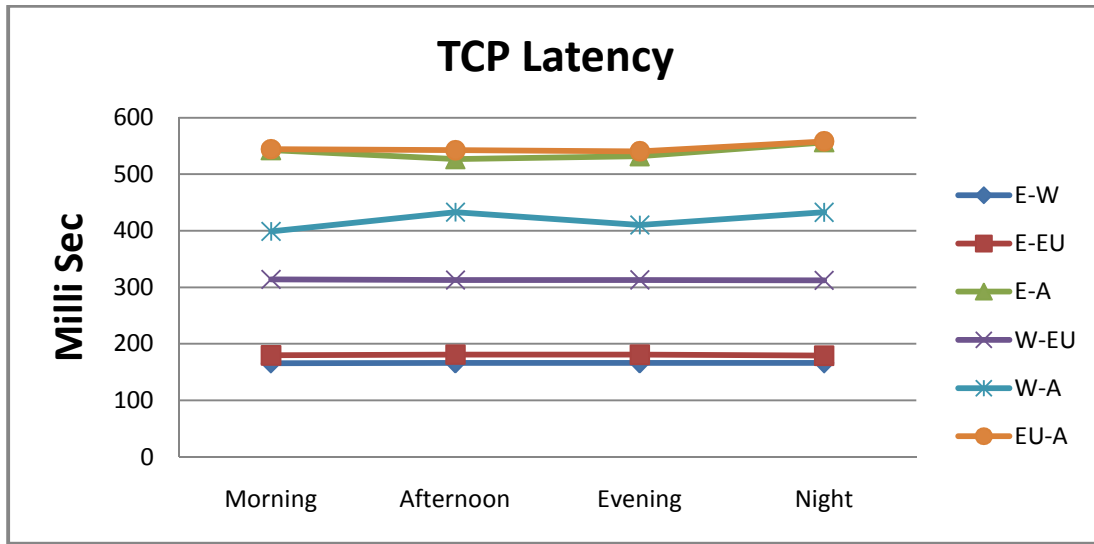### 3.2.1. TCP Latency. Intra-region.

Observation: Latency within Asia Pacific region is higher than all the other regions.

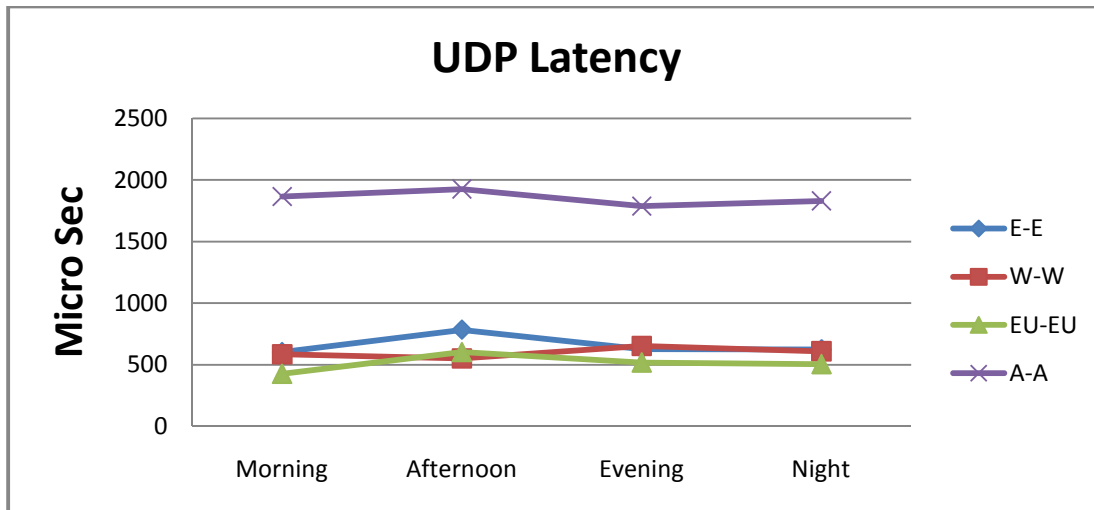### 3.2.2. TCP Latency. Stony Brook, NY to/from different regions.

Observation: As expected, latency between Stony Brook (US East Coast) and the Asia Pacific region is highest (reflecting the physical distance). Distance plays a role also in the latencies observed from the East Coast Vantage point to either the West Coast or Europe which are almost identical.
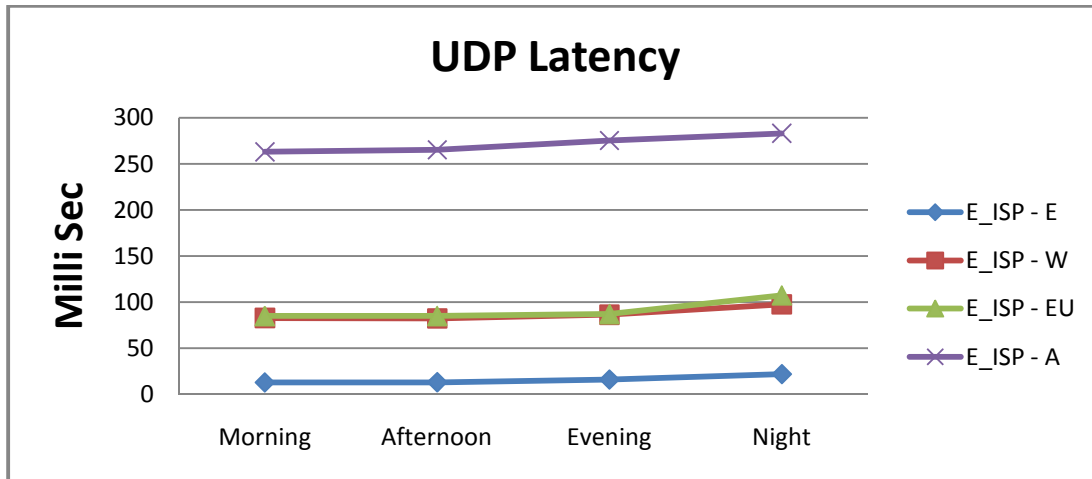
3.2.3. TCP Latency. Inter-region.

TCP Latency

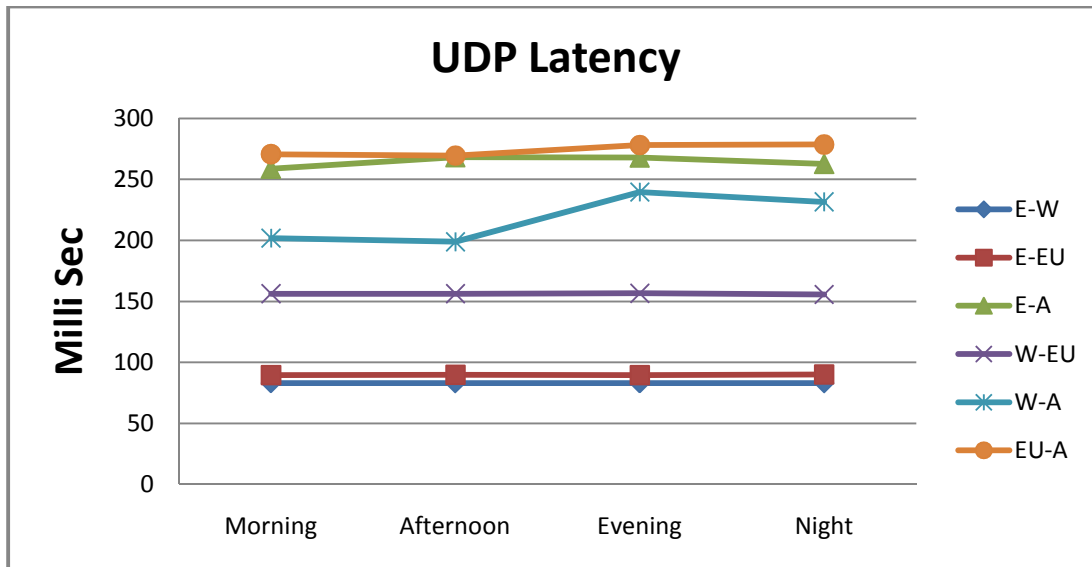### 3.2.4. UDP Latency. Intra-region.

UDP Latency

Observation: Latency within Asia Pacific region is higher than any other region.

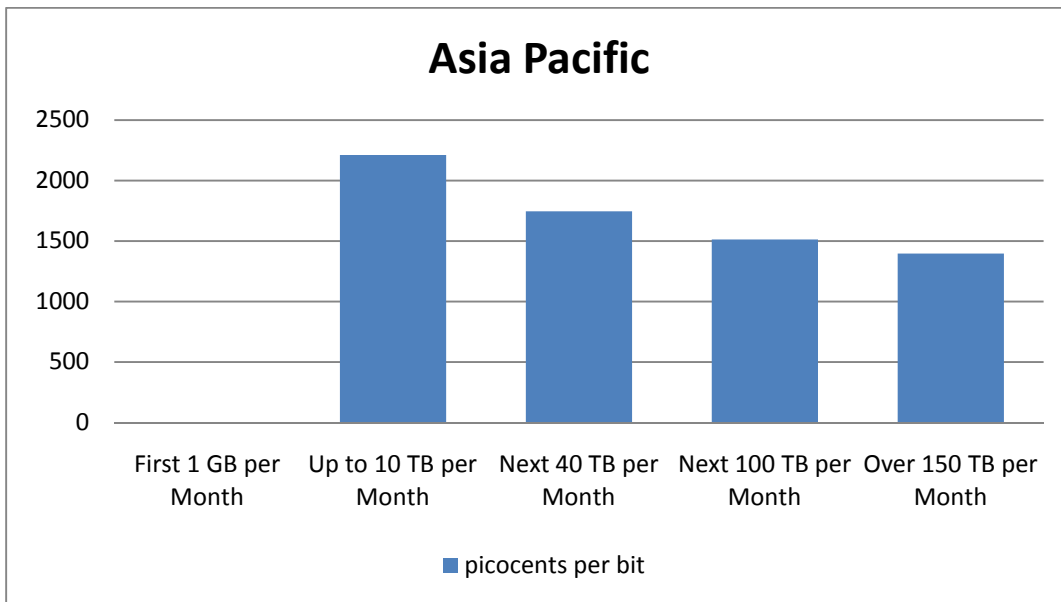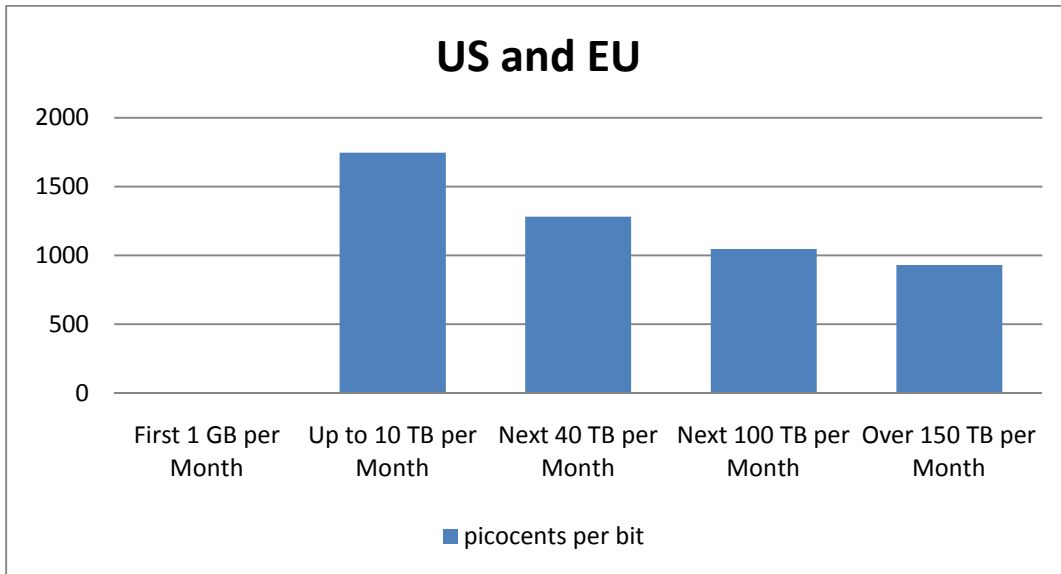**3.2.5. UDP Latency. Stony Brook, NY to/from different regions.**

Observation: Similar behavior as in the TCP case.

3.2.6. UDP Latency. Inter-region.

Observation: Similar behavior as in the TCP case.

3.3. Data Transfer Costs

US and EU

legend: picocents per bit

Categories: First 1 GB per Month, Up to 10 TB per Month, Next 40 TB per Month, Next 100 TB per Month, Over 150 TB per Month

Asia Pacific

legend: picocents per bit

Categories: First 1 GB per Month, Up to 10 TB per Month, Next 40 TB per Month, Next 100 TB per Month, Over 150 TB per Month

Interestingly, according to our earlier estimates, Amazon is paying a bottom line of about 500-600 picocents/bit within the US. These numbers suggest a *markup of at least 30% on all network traffic*.