CSE224 HW4

Yawen Zhao A53280596

Provisioning virtual machines

Leader

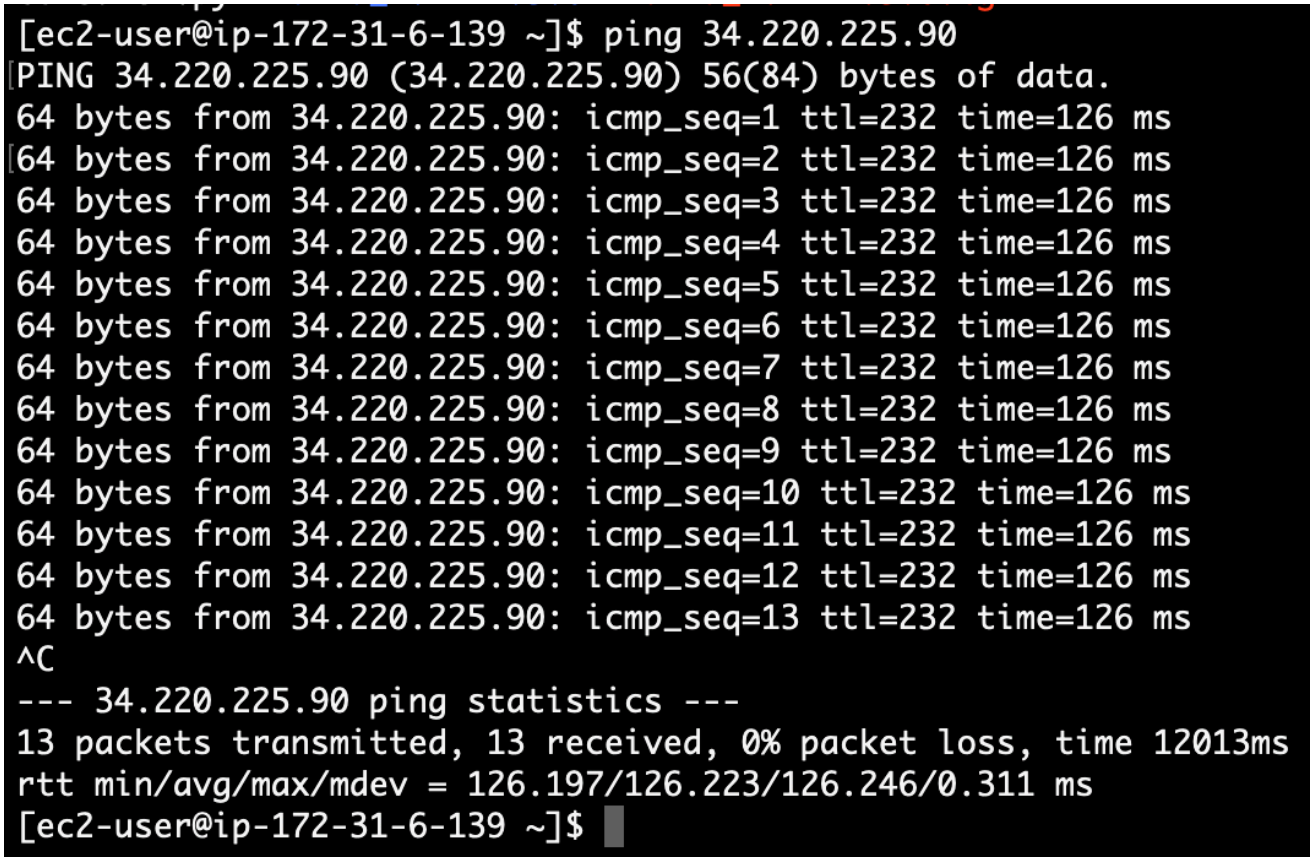
Instance ID	i-0cfa02e3426be62fa	Public DNS (IPv4)	ec2-34-220-225-90.us-west-2.compute.amazonaws.com
Instance state	running	IPv4 Public IP	34.220.225.90
Instance type	m5.large	IPv6 IPs	-
Elastic IPs		Private DNS	ip-172-31-17-73.us-west-2.compute.internal
Availability zone	us-west-2b	Private IPs	172.31.17.73
Security groups	cse224-fa19. view inbound rules. view outbound rules	Secondary private IPs	
Scheduled events	No scheduled events	VPC ID	vpc-122cbb75

Follower

Instance ID	i-0ce45a53da7e01db6	Public DNS (IPv4)	ec2-52-78-243-169.ap-northeast-2.compute.amazonaws.com
Instance state	running	IPv4 Public IP	52.78.243.169
Instance type	m5.large	IPv6 IPs	-
Elastic IPs		Private DNS	ip-172-31-6-139.ap-northeast-2.compute.internal
Availability zone	ap-northeast-2a	Private IPs	172.31.6.139
Security groups	cse224-fa19. view inbound rules. view outbound	Secondary private IPs	

Above are the two instance I created. One in Oregon. One in Seoul.

Performing basic benchmarking

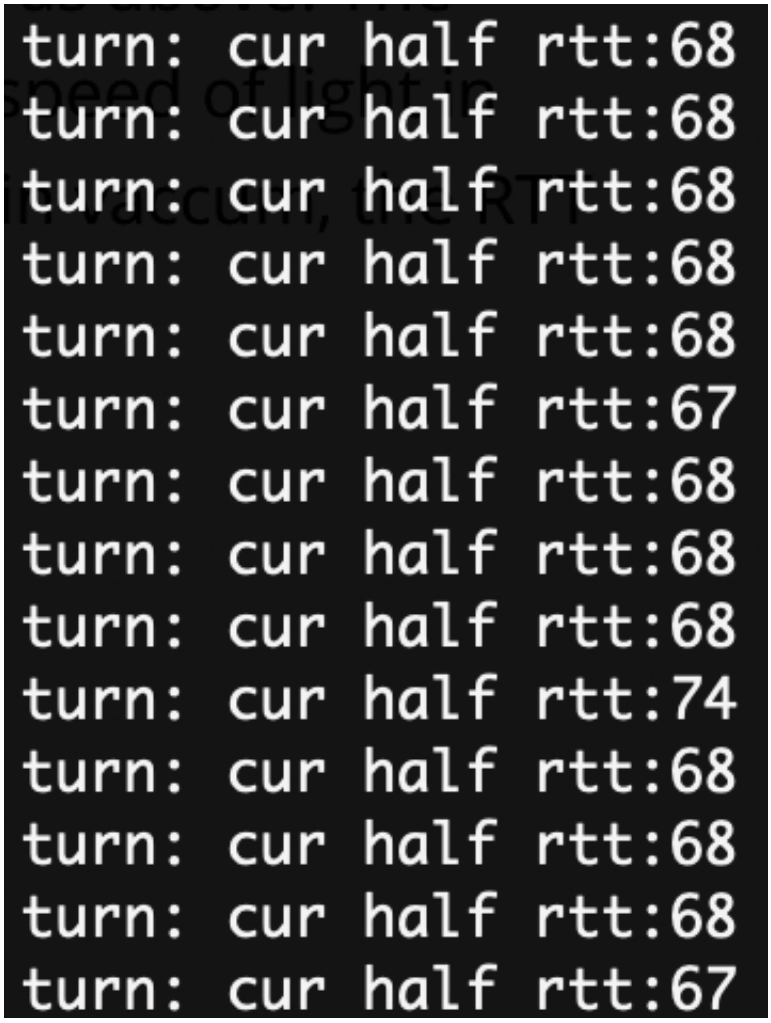


The RTT I measure is averagely 126.223ms, which is shown as above.

The distance between Seoul and Oregon is 8697km. Given the speed of light in fiber-optic cable which is equal to 2/3 of the speed of light in vaccum, the RTT can be computed as

$$8697\,km \times 2/(3.0 \times 10^5\,km/s \times 2/3) = 86.97\,ms$$

Benchmarking Kafka



The mean half RTT calculted by Kafka is 68.3 ms. So the RTT is approximately 136.6ms.

The result is a little bit larger than the result meatured by ping method, which may cause by the inconsistent of the clock of the two machine.

Below are the code of producer and consumer:

producer.py

```
from kafka import KafkaConsumer
import time

consumer = KafkaConsumer('testt', bootstrap_servers="52.12.165.225:9092")
sum = 0
i = 0
for msg in consumer:
    i += 1
    cur = int(round(time.time() * 1000)) - int(msg.value)
    sum += cur
    print("%d turn: cur half rtt:%d"%(i,cur))
```

consumer.py

```python
from kafka import KafkaConsumer
import time

consumer = KafkaConsumer('testt', bootstrap_servers="52.12.165.225:9092")
sum = 0
i = 0
for msg in consumer:
    i += 1
    cur = int(round(time.time() * 1000)) - int(msg.value)
    sum += cur
    print("%d turn: cur half rtt:%d"%(i,cur))
```