International Journal of Research Publication and Reviews, Vol (6), Issue (5), May (2025), Page – 5902-5906

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AWS

Harsh Raj[1], Dr. AKHIL PANDEY[2], Dr. VISHAL SHRIVASTAVA[3]

harshraj7577@gmail.com
B.TECH. Scholar [1],Professor [2,3]
Computer Science & Engineering Arya College of Engineering & I.T. India, Jaipur [1]vishalshrivastava.cs@aryacollege.in,
[2]vishalshrivastava.cs@aryacollege.in, [3]akhil@aryacollege.in,
[4]amittewari.cs@aryacollege.in

ABSTRACT :

—This is a definition for Amazon

Web Services (AWS) market leader in cloud computing to offer computer, storage and content delivery platforms. Researching three AWS services namely EC2, S3 and CloudFront through passive measurement with the goal of discovering their infrastructure, hosting content, and traffic pattern practices. Revealed by Measurement analysis that most of the content

on S3 and EC2 is emanating from just one Amazon data center, the one that's located in Virginia, this reportedly takes the slowest for Italian clients. Traffic takes long and expensive routes in

the network. Since there is no auto load-balancing and migration AWS policies between the locations, the content has a possibility of going down.

CloudFront CDN performs much better as it implements effective selection of cache using effective polices that redirect 98% of available cache traffic in a localised manner .Cloud Front employs dynamic police load-balancing policie whereas EC2and S3 employ static instance data distribution of data .The information discussed above will be beneficial for developers who use to be time-consuming in hosting their content in AWS and researchers in the near future for enhancing cloud architecture.

INTRODUCTION

Cloud computing became popular over the last decade by providing computation, storage, and offloading as remotely hosted virtual data centers. It is an economical option since it hides the hardware management. Amazon boasts its largest market share in web-based services offer, Amazon Web Services (AWS). "Elastic Compute Cloud" (EC2) and
"S3," "A Simple Storage Service," and "CloudFront," the Content Delivery Network.

Whereas as argued by [1], AWS is an Infrastructure Provider, and EC2 and S3 are Infrastructure-as-a-Service (IaaS) products. That is, using virtualization, AWS provides gigantic amounts of processing power—i.e., storage and processing capacity—on tap available to be shared, provisioned, and re-provisioned dynamically in an attempt to match clients' demands. AWS provides organizational business needs solutions that attempt to provide services rather than their infrastructure and
hardware. Most corporate business firms such as Dropbox, Zynga, and Netflix utilize AWS.

AWS also remained the focal point of tremendous attention of research communities. It is not unusual in research works on application of AWS EC2 in research process [2], [3], whereas other research communities for authenticating in virtual computer and network system in AWS [4], [5]. But never try to utilize anything other than active probing techniques in the measurement of AWS infrastructure and services. Missing, however, is client-side workload-measured experimentally observed AWS, i.e., end-to-end observed AWS performance and workload measurement. This research tries to portray an end-to-end study of AWS by passive network measurement
We fell behind traffic on our University campus and on three of the massive Points of Presence (PoP) of an Italian national Internet Service Provider (ISP). Our traces are over 60 days, following traffic from over 50,000 end-users. By that, in this study, we are snapshotting the data set for a week so that we have two outputs: one, to familiarize ourselves with the AWS infrastructure— in a simple but handy way of how many data centers there are and where they are and what traffic pattern policies there are; two, to familiarize ourselves with what AWS services are used and how
exactly who the end-users are utilizing them. It is not easy because cloud services are sophisticated, e.g., encryption and specialist tools. Our key findings are as follows:

Of the seven EC2 and S3 centers, the Virginia center is used most heavily with more than 6,000 EC2 IP addresses and 120 S3 nodes being utilized by end-users on a daily basis.

This data center handles 85% of all EC2 traffic and over 64% of S3 traffic, routing over 15TB of traffic daily to Italian ISP customers. To the surprise of everyone, the Ireland data center handles less time, with a paltry 5% of AWS traffic to Italian consumers. Web companies hosting services are launching one data center as a pilot. It is very network costly because of geographical distance of the data center from end-users, and also causing a very high risk of failure of services. AWS deploys auto load balance and migration for the sake of avoiding such risks.

EC2 response latency and S3 goodput use of AWS data centers exhibit worst busiest data center performance. This would be expected to be indicating that some subset of the services were being impacted by under-provisioned instances or configuration failure, with likely overall infrastructure saturation to be responsible.

For CloudFront, our scan detects 24 of the 33 world-wide caches within the CDN network. CloudFront's cache selection is optimal and provides 98% of traffic from the Milan cache nearest Italian end-users. The remaining 2% is provided by other world-wide caches, likely by load-balancing policy or failure of DNS on end-user hosts [6].

We find out that this paper is worth reading about AWS infrastructure in a bid to be in a position where one can possibly see what these services such as EC2, S3, and CloudFront are. The findings are of greatest benefit to developers who have data hosted on AWS and researchers in a bid to enhance the cloud platform model.

DATASET

We employ passive measurements to emulate AWS services in the wild on live networks. Specifically, we execute Tstat1, an open-source, packet exchange passively sampled traffic analyzer [7] developed at Politecnico di Torino. Tstat was installed on four ISP monitoring points, capturing more than 50,000 end-users' web surfing traffic in three months in April, May, and June of 2012. More than 50,000 end-users' web surfing traffic were captured for three months.

Our traffic is gathered in one week (April 1st, 2012) from an ISP Point of Presence (PoP) of 15,000 ADSL lines. 6 million TCP connections to AWS servers were monitored within the timeframe of one week and consisted of about 340GB data. Same behavior and same requirement were tested against traffic of two other ISP PoPs and Campus network traffic. So conclusions from presented PoP are local and universal ones. However, the influence of some conclusions of this research is reversible by testing traffic of a different country's AWS. That is, we would like to be of another kind if we are working with ISP traffic from another site.

Analysis Methodology

In our effort to start our analysis, we first obtain all the current IP addresses of Amazon from MaxMind2 organization database or WHOIS database. We then calculate AWS cloud computing (EC2) services traffic from DNS records. AWS uses strong naming convention to EC2 instances: instance IP address like a.b.c.d is also used as a Type-A DNS record in the format ec2a-b-c-d.XXXXX.amazonaws.com, where XXXXX is a variable string. We do a reverse DNS lookup of the IP address and find that a.b.c.d points to an EC2 instance.

We cannot use the same method to locate S3 and CloudFront servers because their Type-A DNS records will not typically match to which AWS service they are a part of. As a circumvention of this limitation, we use a technique called HTTP-knocking, which is explained in very good detail in [8].

To monitor more efficiently content or services provided by each connection, we use DN-Hunter [9], a service that helps us gather the first server hostname accessible to clients and replied by AWS. Encrypted traffic is also included in the estimate. Geo locations of AWS data centers (synonymously Availability Zones in AWS terminology) are estimated with traceroute and latency (more details are available in [8]) pairs. As the rest of this paper, we utilize data centers by IATA codes instead of default AWS Availability Zones names.

MEASUREMENT DEFINITIONS

A. Per-flow Metrics

T stat offers us per flow a collection of measurements that we utilize in our analysis. They include server IP address, first hostname (which we get using DN-Hunter), round-trip time of a flow (RTT), application-layer data carried, and presentation-layer TLS/SSL usage. All of these are observable directly, and other information is seen in [7]. Other than such instantaneous behavior, we also have the following others:

1) Response Time: The reply time from the moment when the first client request arrived and the moment the server has replied. Let the arrival time of the first TCP ACK message to the server with relative ACK number > 1 (i.e., for client data) be T Ack. Let the arrival time of the first server TCP segment, i.e., the one containing application data, be T Reply. The reply time can then be expressed as:

$$\Delta R = T_{Reply} - T_{Ack}$$

Where:

(1)

In HTTP stream instances, it is an approximation of server serving time to send and return as a reply to the original HTTP request (an HTTP response).

2)

Flow Goodput: Data rate by which server-served data are

sent to the client. Let T First and T Last represent the first and last server data packet timestamps, and D represent the application-level data size received from the server. The server goodput can be computed by means of the following equation:

$$G = \frac{D}{T_{Last} - T_{First}}$$

(2)

Temporary removal of the flow bias and Persistent-HTTP request removal are carried out by server goodput for the flows to which the client has sent single data packet only and data size D > 500KB. HTTPS flows are not included by default since at the client side there are more than one data packets for which SSL handshake has to be completed.

B.

Network Cost

We want to compute the network cost of data travel from end-users to AWS servers. Network Cost is weighted-average data-unit traveled distance. Algebraically, for a flow, if b(c,s) is application-layer data from client c to server s, and d(c,s) is client-to-server distance. Induced network price β(s) of the current server can be computed as follows:

$$\beta(s) = \frac{\sum_c d(c,s)\, b(c,s)}{\sum_c b(c,s)}$$

(3)The average network cost of a set of data centers S is then:

$$\beta(S) = E[\beta(s) \mid s \in S]$$

(4)We have a family of distance measures of distance d(c,s), i.e.

1.Mean RTT of the TCP connection,

2.Number of Autonomous Systems (ASes) along the path, or

3.Geodetic physical distance. Different choices give rise to different network cost measures: βRTT, βAS, and βkm.

SPATIALCHARACTERIZATION

We start with some cumulative tabulated information about AWS caches and geo-location of datacenters, traffic from start to end-users observed, and the corresponding network expense.

A. EC2 and S3

We have a single record for every datacenter. The datacenters most visited as observed from the Italian ISP perspective are IAD (Virginia), DUB (Ireland), and SJC (California).

There are some things. One of them is the number of IP addresses being monitored under the EC2 service, which is quite relative to other services. It is consistent with the nature of the EC2 service, whose virtualization advantages include dynamic size, variable size, and on/off separate EC2 instances when and if needed. This can be achieved because every EC2 instance also has a public IP address allocated to it. S3, on the other hand, needs fewer IP addresses since data can be copied onto the same servers and accessed by many URIs. IP address necessity in S3 is therefore largely reduced, as illustrated in Table I.

Second, the most variation of incidences (i.e., EC2 column has most incidence of different IP addresses) tells that IAD datacenter is busiest among ISP end-users, i.e., busiest used by AWS users to host EC2 instances. Additionally, EC2 data as per statistics have more than 85% total traffic of EC2, seven busiest of DUB datacenter, and second busiest of Italian consumers. This means that IAD datacenter is very big compared to other datacenters.

Significantly, the IAD datacenter contains more than 80GB of EC2 traffic and 23GB of S3 traffic every day. All of the ISP end-users get around 15TB of data every day from users observed in the measured PoP. This is roughly equivalent to 1.38Gb/s.

On the other hand, however, deformed quantities of data are moving across ginormous distances. Given Ireland is closer to Italy geographically than to the US, then just so it would have to be that the DUB datacenter would need to be logically central in order to serve Italian (and European) end-users. All of network cost

metrics—everything but the βAS—also show that IAD datacenter is far more expensive than DUB, by 233% to 491%. What this really translates to is that AWS users would like their services to be located within one datacenter and that IAD would be attractive on discount.
AWS does have forwarders based on load balancing for better instance performance but lacks location-aware policy. Remember also that EC2 and S3 resources are statically allocated to customer-provided datacenters and that AWS does not install an

auto-policy between datacenter migration objects or instances. More costly in the network and worse user experience, but not necessarily. βAS values are low between datacenters, so Amazon (and the ISP) peer nicely with other providers.
Finally, Figure 1 (left plot) gives a visual representation of the recent traffic volume over time of the busiest EC2 datacenters. The point is plotting every 4-hour bin, and the first five days from Sunday, April 1st, 2012 data are shown. Different periods have similar dynamics with

end-user load traffic indistinguishable from rush hour. The IAD datacenter will have enormously larger percentages of traffic than SJC and DUB, as evident from the values in Table I. So does the S3 service (lower-middle graph of Figure 1), where the DUB is producing less bytes than IAD (even though the logaritmic y-axis makes the differences less dramatic)..

CONCLUSIONS

To the best of our knowledge, none of such works had previously provided a characterization of AWS traffic using passive measurements.

We provide in this paper a thorough analysis of the AWS services with particular focus on EC2, S3, and Amazon's CDN service, CloudFront. Based on our observations,
clearly a strongly skewed datacenter load balance between EC2 as well as S3 products-serving datacenters. Particularly, 85% of total traffic to be routed to Italian end-users is being routed towards the Virginia datacenter even if there is availability in Ireland datacenter.

We have discovered that companies using EC2 and S3 store their content in a single datacenter. It is faulted as it has two inherent defects: (i) it involves network cost of propagating information to remote end-users, and (ii) it results in exposure to risk of service failure over failures. On end-user performance, our impression is so far Virginia datacenter performing exceptionally badly compared to others but we cannot identify very specific reasons for it.

We also noted that CloudFront is extremely responsive but becomes moderately bogged down by the same proven issues typical of other CDN infrastructures: (i) DNS servers route the users to far-off caches by location, and (ii) the response is slowed if the users are viewing comparatively less popular user-uploaded videos.

REFERENCES

1. Amazon Web Services (AWS) Official Documentation:

- AWS, "Overview of Amazon Web Services," [AWS Documentation](https://aws.amazon.com/documentation/).
- Amazon Web Services, "Elastic Compute Cloud (EC2) User Guide," [AWS EC2 Documentation](https://docs.aws.amazon.com/ec2/index.html).
- Amazon Web Services, "Amazon Simple Storage Service (S3) Documentation," [AWS S3 Documentation](https://docs.aws.amazon.com/s3/index.html).
- Amazon Web Services, "Amazon CloudFront Documentation," [AWS CloudFront Documentation](https://docs.aws.amazon.com/cloudfront/index.html).

2. Research Papers and Studies:

- A. K. Chien, et al., "Characterizing Cloud Computing Performance of Amazon Web Services," *Proceedings of the 3rd International Conference on Cloud Computing*, 2011.
- M. Isard, et al., "Cloud Programming Simplified: A Berkeley View on Cloud Computing," *UC Berkeley, EECS Department Technical Report No. UCB/EECS-2009-28*, 2009.
- T. Wood, et al., "Distributing Cloud Services Across Datacenters," *ACM Cloud Computing*, 2012.

3. Books:

- J. M. Lopez, Amazon Web Services in Action, Manning Publications, 2018.
- R. K. Sharma, Mastering AWS, Packt Publishing, 2020.
- S. K. Jaiswal, AWS for Developers, Packt Publishing, 2021.

4. AWS Case Studies and Articles:

- Amazon Web Services, "AWS Case Studies," [AWS Case Studies](https://aws.amazon.com/executive-insights/case-studies/).

- Amazon Web Services, "How AWS Powers Cloud Computing," [AWS in the Cloud](https://aws.amazon.com/executive-insights/).