

Network Architecture and Security for Big Data

Ms. Ashwini Sagar¹

M.tech Student Department of CS&E,
BMS Institute of Technology & Management
Yelahanka Bangalore, India

Mrs. Chethana. C²

Assistant Professor, Department of CS&E,
BMS Institute of Technology & Management
Yelahanka Bangalore, India

Abstract - A collection of large amount of data is called big data and transmitting over a network. It should find a better decision making in a critical areas like health care, economic productivity. Here we explained about how Google is detected flu disease in a short period. The large amount of information is collected from different devices, and are stored and worked in powerful data centers. As many users have a strong demand on unimpeded network infrastructure and effective delivery towards a destination so using multiple path in a network we can reduce bottleneck in a network. The network architecture contains access network, internet backbone, data center. In this we concentrated on unique challenges in constructing such a network infrastructure for a big data. Our study covers everything about in this network highway. Here we explained about how the data can secure. Using this AES algorithm we can protect our data explained in detail.

Keywords: *Big data, Networking, Cloud Computing, Network Based Application, Internet and Web Application*

I. INTRODUCTION

Our world producing the data is too much, moves too fast than ever before. The world is altering so rapidly that we really need to see beyond today's analytic capabilities and chart where enterprise grade architectures are headed in the future of big data. big data size of it is a constantly moving direct as of 2012 ranging from a terabytes to many pet bytes of data. Till 2020, over 40 or more the data is sharing range from a zetta bytes of data would be made, duplicated, ingested. The number of data pouring out into our daily lives, from any device, anywhere, and any time, we are going the era of big data in doubtlessly. Big data makes for large amount of value. Big data contains a number of techniques and technologies that wants new form of integrating to bring out large hidden values from large data sets that are various, building complex, and of a massive scale. big data examining engineering, insights can be find out to make better decision making for critical development areas like health care, economic productivity, and natural disaster prediction. For example Flu Trends service is developed by Google to find regional flu outbreaks in real time. Specifically the Google Flu Trends initially used a group of 50 to 300 search queries of common words of data each week from 2003 to 2008. Then to compute the correlation coefficient between each keyword search

history data and the sickness history data received from the Centers for Disease Control and Prevention (CDC) using the linear model method in United States. Then the keywords are selected based on nearest correlation coefficients, and searched keywords were combining to outbreaks the next flu in the United States. Before a week earlier than CDC, Google Flu Trends is found flu outbreaks, which can cut down the loss made by the flu and even, save lives. Google Flu Trends updated in 2014-2015. Good news for health care for at least two reasons. First, the health care community will be getting more exact information about flu incidence. Second, it providing opportunity to think about the challenges laid by the use of big data. It's possible to better predictions by gathering the information from the CDC and Google.

In general all the large information is stored in data warehouses. The different types of data are storing in a warehouses and proceeds in to a number of interconnected server nodes. We are using map reduce to produce better data. Big data is not a new word, has been around for decades handling tons of transitional data the years. In today the big data is using more in social networking like Face book, tweeter, Gmail etc. Generation, gathering, aggregation, processing, and application delivery are multiple stages of the big data. In data centers the data get from distributed devices and processing occurs in data centers. In Fig explains about Access network, Internet backbone, data center network. How the network plays a critical role among end users and data destination. In a network there is a demand to build an uninterrupted infrastructure to collect distributed data and rapidly generated data, and send them to data center for processing. Big data is presently defined by five dimensions: Bulk, Kind, speed, verification and value.

In this, we are analyzing the unique challenges about the network and big data and for securing our data we are using AES algorithm. Our study covers everything about this network highway and security.

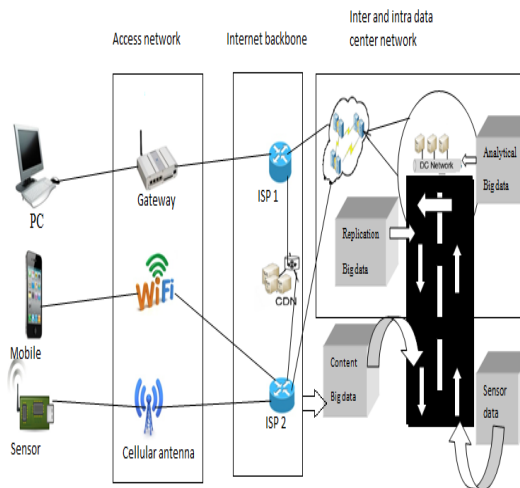


Figure: 1 Three layered network architecture interacting each other from the view of big data application

II. NETWORK ARCHITECTURE

Facebook has linked up successfully with efficient online advertisement, earning taxes and at the time satisfying its users. Other social networks like Twitter, G mail, and Instagram. Government funded number of projects have been launched recently to build up big data analyzing systems, involving critical areas from health care to climate change. The use of acceptances of big data is good and permits ratios in terms of price, productivity and invention. The above figure shows architecture of a network interacting each other from the view of big data application illustrated in fig.1. At first the user information's receive from a different devices and locations, which are collected by wired and wireless networks [1]. The all information's are combined and delivered to data centers through internet. In data centers the big data are examined then processed back to interested users.

Obviously network plays vital role in bridging over the different stages. Users demand to create a speed and authenticate interconnected network for the big data to move freely on this digital highway. The network highway contains not only just processing of data, but rather the whole section for the development of big data, from access networks to the internet backbone, and inter and intra data center networks.

Networks:-

- Access network
- Internet backbone
- Data center network

Access networks are directly connected to end devices. Other side the all received data from end devices is transmitted into the network system. Finally the data center analyzes and processed big data are sent back to the users. Big data applications such as movie quality video pouring require conformed performance over a long duration to assure the quality of user experience, which has become a critical challenge for wireless networking.

The internet backbone is an intermediate layer that connects data center networks and access networks. To make good user experience, the internet backbone require to forward massive distributed data to data centers with high output, and deliver working in a slow worked data to users from data centers with low latency. Big data applications like photo, video, file sharing allow users to upload multimedia contents to data centers and share them with their friends in real time. High performance end to end connections are needed for uploading, and effective content distribution networks (CDN) are needed for downloading. The request for more composite information is getting higher every year. In real time information pouring is becoming a challenge that must allocate more important video components with higher gain channel resources. Using this channel improves the quality of video over wireless.

In a data center, big data are processed analyzed using map reduce. A salable, ultra-fast and blocking free network is thus needed to interconnect the service to users. Service provides from an internet to users.

2.1 ACCESS NETWORK

Access network gives access to end devices such as mobile, laptops, sensors etc. through internet. An access network is the part of telecommunication network which connects subscriber to their immediate service provider. When a many users are access a specific resources then that network becomes a bottleneck. It makes a system to slow due to limited resources. This is a major challenge in a network. While sending a data congestion occurs when bandwidth is insufficient and network data traffic exceeds capacity. In a congested network, response time slows with reduced network throughput. So we are using multiple paths to deliver a data from source to destination. MIMO OFDM [2] channels, is proposed for video delivery in an optimized scheme. The video coding and wireless channel separates into independent components and allocates more important video resources with higher gain channel resources. This extends to significantly improved quality for video over wireless.

2.2 INTERNET BACKBONE

The principal of data routes between large, strategically interconnected computer networks and core routers on the internet is called an internet backbone. For example, the performance of resource constrained mobile devices, data from distributed mobile devices are sending

to the cloud for working in mobile cloud computing services [3].

2.2.1 BIG DATA APPLICATION IN INTERNET BACKBONE

END TO END TRANSMISSION With the developing capacity of access links, network bottlenecks are found in access networks to the core links in the internet backbone. To better the throughput of end to end with the developing capacity of access links, network bottlenecks are found in access networks to the core links in the internet backbone. To better the throughput of end to end data transmission, path should be searched, which applies multiple paths at the same time to avoid individual bottlenecks. The m path [4] uses a large set of distributed proxies to construct detour paths between end users. Using AIMD algorithm technique we are avoiding bottleneck.

- **CONTENT DELIVERY NETWORK**

A system of distributed systems of networks deployed in multiple data centers across the internet is a content delivery network (CDN). The main goal of a CDN is to serve content to end users with a high availability and high performance. To achieve high throughput in two ways: to avoid network bottlenecks optimize the path selection and increasing the number of peering points. Yu et al [5] introduce a model using both internet network typologies and synthetic, they show that changing the number of peering points to better throughput the most, while optimal path selection has only determined contribution. Liu et al [6] further find that video delivery optimized for low latency. High quality video delivery requires conformed performance over a long duration. This leads to an adaptive design with global knowledge of network distribution of end users. Jiang et al [7] advice's that the CDN structure can be extended to the edges of networks, broadband gateways.

2.3 DATA CENTER NETWORKS

Data center network is a large group of networked computer servers typically used by organizations for the remote storage, processing, or distribution of large amounts of data. Big data gathered from end devices are stored and processed in data centers.

2.3.1 DYNAMIC FLOW SCHEDULING

In a dynamic flow scheduling there are multiple equal cost paths between any pair of servers in a typical multiple rooted tree topology of a data center network. Hedera [8] is designed for better utilize paths to dynamically forward flows along these paths. From switches it collects flow information's, and instructs switches to reroute traffic accordingly. Hedera is able to maximize the overall network utilization with only for small data flows. In switches through monitoring the flow information, when congestion cause at certain location it schedules data flows only after

they cause. So avoiding congestion, data flows should be detected before they found. In flow comb [9], covers which data flows effectively by monitoring map reduce applications on servers through software agents.

2.3.2 INTER DATA CENTER LINKS

Social networks are big data applications usually exploits several distributed data centers for duplication and low latency service provision. Those data centers are interconnected by high capacity links rented from internet service providers (ISPs). Data duplication and synchronization like operations need high bandwidth transformation between data centers. Thus it is difficult to better utilization or reduce the cost for inter data center links.

Laoutaris et al [10] finds a form of user demand on inter data center bandwidth, which is in low bandwidth usage in off peak hours. So they propose netstitcher to usage the remaining bandwidth in off peak hours for backups and data migrations are non-real time applications. The netstitcher uses store and forwards algorithm to send big information among data centers. The data are partitioning into pieces and sending to their destination using a multiple paths, which contains a series of intermediate data centers. The data pieces should travel when and where decides scheduling module based on available bandwidth. Jet way can minimize the cost of inter data center links.

III. BIG DATA BASED NETWORKING APPLICATIONS

Big data applications are classified into two categories, online applications and mobile wireless network applications with respect to structure of networking.

- **ONLINE APPLICATIONS**

Netflix is closely related to big data applications in our daily lives. Using this application it provides a video on demand services to users. To support the combination of huge traffic and unpredictable demand bursts, Netflix has developed a universal distribution of video system using amazons cloud. Video streaming service is totally depends on users demand, Netflix front end services are running on Linux based tomcat java server and NGINX web servers.

Netflix first buys master copies of digital films from movie studios using the Amazon EC2 cloud machines, converts them to over 50 different versions with different video resolutions and audio quality. Many converted copies from master copies are stored in Amazon S3. In total, Netflix has over 1 petabyte of data stored on Amazon. As shown in below figure 2.

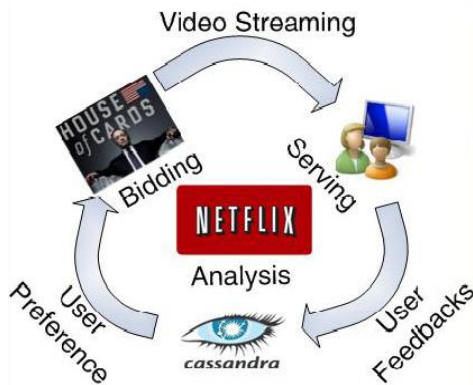


Figure 2: Netflix [1]

• MOBILE WIRELESS NETWORKS

Mobile wireless big data applications are entering in our lives. As an example, Amazon provides better service to users compare with early days. Amazon users can connect using wireless sensor devices and mobile phones. Amazon products are books, clothes, home appliances, gifts etc. The users can install Amazon apps in their smart phones, which collects information from the sensors through wireless connections. The Amazon apps enable social interactions among users. They can share their experience after using their products as reviews. Amazon becomes a big data platform that collects, stores, and processes data generated from more than 18 million users.

IV. IMPLEMENTATION

SECURITY

In network intrusions generates data traffic for network attack. This is a main challenge for securing a data. The different types of personal data such as health care record, credit card details are collected by big data applications these information send over networks, in between attackers also present to attack our information so we are keeping our data securely using Advanced Encryption Standard (AES). It helps to protect our large amount of data using encryption and decryption methods.

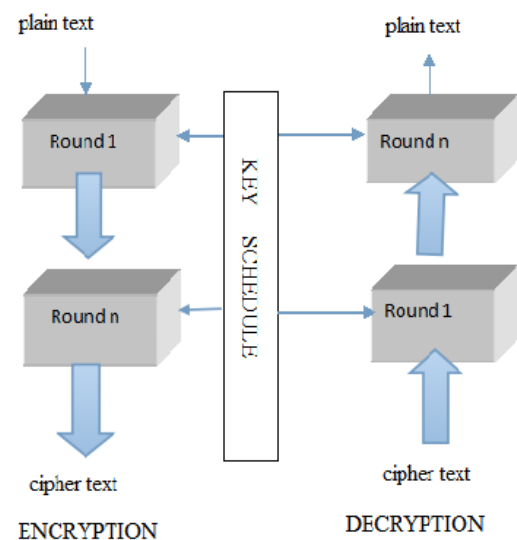


Figure 3: Structure of AES

Encryption is the process of converting an original message into a form that is unreadable to unauthorized peoples.

Decryption is the process of converting the cipher text into a message that conveys readily understood meaning.

Advanced Encryption System Algorithm steps

- AES is a block cipher with a block length of 128 bits. AES allows for three different key lengths: 128,192, or 256 bits.
- The number of rounds shown in figure 3 for the encryption key is 128 bit long.
- Before any round based processing for encryption can begin the input state array is XORed with the first four words of the key schedule. The same thing happens during decryption concept that now we XOR the cipher text state array with the last four words of the key schedule
- For encryption, it consists of
 - 1) Substitute bytes
 - 2) shift rows
 - 3) mixed columns
 - 4) add round keys.
 The last step consists of XORing the output of the previous three steps with four words from the key schedule.

- For decryption, it consists of

Inverse shift rows, 2) inverse substitute bytes, 3) add round key, 4) inverse mix columns. The third step consists of XORing the output of the output of the previous two steps with four words from the key schedule. Using this AES algorithm we can secure our large amount of data on network.

V. CONCLUSION

So we have brushed up about the network architecture, and big data application services. In a networking system we identified major challenges applications of a big data. We discussed here about AES algorithm using this we can secure our large amount of information when a data is transforming from source to destination. Big data is a emerging area it is attracted by both academic and industry.

REFERENCES

- [1] Yi liu jin "building a network highway for big data", IEEE network, 2014
- [2] X. L. Liu et al., "ParCast:Soft video delivery in MIMO-OFDM WLANs," proc. ACM Mobicom, 2010
- [3] F . Liu et al., "Gearing Resource-Poor Mobile Devices With Powerful Clouds:Architecture, Challenges and applications", IEEE Wireless Commun., Special Issue on Mobile Cloud Computing, Vol.20,2013
- [4] Y. Xu et al., "mPath: High Bandwidth Data Transfers with Massively-Multi-path Source Routing,"IEEE Trans. Parallel and distributed systems, 2013
- [5] M.Yu et al., "Tradeoffs in CDN Designs for Throughput Oriented Traffic," 2012
- [6] X. Liu et al., "A Case for a Coordinated Internet Video Control Plane",2012
- [7] W . Jiang et al., "Orchestating Massively Distributed CDNs", 2012
- [8] M. Al-Fares et al., "Hedera: Dynamic Flow Scheduling for Data Center Networks," 2010
- [9] Das et al., "Transparent and Flexible Network Management for Big Data Processing in the Cloud", 2013
- [10] N. Laoutaris et al., "Inter-Datacenter Bulk Transfer with NetStitcher," 2011