Privacy Preserving of Data Transmission and Efficient Routing Process of Hierarchal Clustering in Named Data Networking

Ellapu Jyothi[1], P. Mohana Roopa[2]
Final M.Sc Student[1], Lecturer[2]
[1, 2] M. Sc Computer Science, Chaitanya Women's PG College, Old Gajuwaka, Visakhapatnam Andhra Pradesh

Abstract:

In Named Data Networking one of the unique features is caching of content in the routers along with shortest path. In-network caching allow users to obtain the contents from nearer intermediate routers, reducing the need for content fetching from the servers often located deep in the network. In the network caching of content is taken by intermediate routers and providing content can be loaded into server. Before performing content forwarding we can generate peer-to-peer shortest route for reduce traffic and also overcome chance of congestion. Finally the response time and the transmission overhead for fetching of contents are reduced. In this paper we are proposed three concepts i.e. authentication of clients or users, generating hierarchical clustering of shortest path, data encryption and decryption. In the named data networking each user will be identify of server by using hash based signature generation algorithm. After completion of authentication process the server will generate hierarchical clustering of nodes in the network. The completion of clustering process the server will generate shortest routing from source node to destination node. Before performing shortest routing the server will find out cluster head by using cluster head election process. After completion of cluster head the server will find out shortest route from source node to destination node by using Nodes Signal Strength Shortest Path Algorithm. After finding shortest route the server will send path to both nodes of source node and destination node. The source node will enter transferred message and encrypt that message by using hybrid bit transpose bioinformatics technique. After completion of encryption process the source node will send to destination node. The destination node will retrieve cipher format data and perform the decryption process of hybrid bit transpose bioinformatics technique. By implementing those concepts we are reduce response time and transmission overhead of content.

Keywords: *Data Mining, Cryptography, Privacy, Hierarchical Clustering, Named Data Network*

.

I. INTRODUCTION

Named Data Networking (NDN) architecture geared toward addressing the Internet's most pressing problems in security, management complexity, and sustainability, and meeting requirements of current and emerging applications. While TCP/IP was a unique and ground-breaking architecture, it solved a problem of the telephony world: enabling point-to-point conversations between two communication endpoints. The world has changed dramatically in the last 30 years, driven by the spectacular success of this architecture. Growth in e-commerce, digital media, social networking, smartphone applications, and the Internet of Things (IoT) has resulted in the Internet primarily being used as an information distribution network. Operators, equipment designers, application developers, users, and policymakers have all struggled with the complexity (and kludges) required to accommodate the inherent misalignment between the TCP/IP architecture and its primary use today. The conversational nature of IP is embodied in its datagram format: IP datagrams identify communication endpoints, i.e., the IP destination and source addresses. NDN generalizes today's Internet architecture by removing this restriction: names in an NDN datagram are hierarchically structured but otherwise arbitrary identifiers. The name in an NDN datagram can identify any content object, including a communication endpoint, a chunk of data in a movie or book, a command to turn on a light bulb, etc. This conceptually simple change - rooted in the recognition that data, rather than virtual channels that contain data is the foundation of modern communication - facilitates profound improvements in security, scalability of content distribution, support for mobility, and ease of application development. NDN research challenge is to realize this data-centric vision using an architectural framework capable of solving real problems, particularly in application areas poorly served by today's TCP/IP architecture.

In the initial NDN project, our focus was to validate the basic ideas by prototyping a range of applications, and to develop the architecture by designing various network mechanisms. Early application prototypes included streaming video,

serveries chatroom, peer-to-peer file sharing, multiplayer games, web browser support, and vehicular networking. Developed network mechanisms include routing protocols, forwarding strategies, fast and scalable name lookups, mitigation, and privacy preservation via name obfuscation. We developed several approaches to evaluation of research ideas and artifacts: a global overlay tasted a simulator, and a substantial open source software base including core forwarding and routing implementations (completed during NP phase), supporting libraries, and applications. Our application-driven focus proved successful. In the follow-on NDN project we chose three network environments that required critical thinking about key components of the architecture: mobile health, building automation and management, and multimedia conferencing. Interest from physicists and climate scientists in applying NDN to some of their persistent data management problems motivated a fourth network environment. This phase of the project enabled further evolution of the architecture itself - including packet formats, protocol extensions, trust management models, and routing and forwarding algorithms - in tandem with development and refinement of applications that demonstrated its utility. In parallel, interest from the academic, commercial, and government research communities grew rapidly, so much that in we augmented our annual project retreats with international community meetings to support expanding collaborations. We also launched the NDN Consortium as a forum for broader discussion of the research and its implications, including responding to industry priorities and concerns. We made tremendous progress in the last five years, but unexpected collaborations have revealed the importance of demonstrating NDN capabilities in IoT and big data environments, and highlighted needs for accessible software platform support and emulation capabilities to facilitate R&D on both the NDN architecture and applications that leverage it.

In-network caching is expected to improve the performance of content transfers in Information-Centric Networks (ICN), by allowing any network element to become a temporary content server. As a result, requests can be fulfilled by any element along the path if the content is cached locally. This Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. However, bears the challenge of how to manage and co-ordinate such distributed storage infrastructure in a scalable, efficient and cost-effective way. According to, there are three main challenges with regard to the setup, co-ordination and management of a ubiquitous in-network caching system. Firstly, there is a cache placement challenge, which addresses the issue of which nodes within a domain are upgraded to in-network caches. This constitutes mainly a network planning problem, as the decision on which nodes to upgrade should take into account the domain's network topology, traffic characteristics and position within the Internet hierarchy . In this study, we assume that a domain has already deployed caches internally and focus on the two remaining challenges: the content placement challenge, which addresses how to distribute contents across in-network caches of a domain, and the request to-cache routing challenge, which determines how content requests are resolved to the corresponding cache nodes.

II. RELATED WORK

The researches about catching have been widely carried out in the past [2]. While catching in CCN has its own content-oriented features. First of all, caching is a native property of routers in CCN. In CCN, request catching and content catching should be handled at a same network layer. That makes content retrieval and replacement be considered at line speed [3]. More clearly, a router should check if its local content store has the requested content before it sends a request to next hop. Secondly, the placements of the ubiquitous caches are arbitrary, but not hierarchical, which make caching in CCN different from Web-catching and Content Delivery Network (CDN). At last, as content chunks in CCN are identified by the unique names, different applications could use same cache space in a router at the same time. This is the most basic feature of catching in CCN, which make it different from web, CDN and P2P. In-networking catching is the distinctive feature of CCN infrastructure and plays an important role in terms of system performance. In-network caching mechanism can avoid wasting network bandwidth due to the repeated delivery of popular content. Additionally, it can reduce response time for content by placing the content closer to users. The challenges surrounding in-networking caching involves cache placement, cache replacement and network cache model, etc. However, Kutscher, et al [4] define three key issues which influence the performance of in-networking caching system, i.e., cache placement, content-placement, request-cache routing. Cache placement mainly focus on deciding which nodes are supposed to upgrade for in networking caching in a domain, which are mainly related with the whole network planning, such as, the network topology, traffic and positions [5], [6]. As for content-placement, it is an issue about the distribution policy of contents across in-networking caches in a domain. However, request-

cache routing solves the problem of actions took for a content request corresponding to node caches. Above all, in this paper, we focus on the content placement issue and the request-cache routing issue. To manage distributed locations of cache storage, we use a hierarchical clustering approach to manage the in-network caching. Clustering approach is common in network design to manage distributed entities in a network, such as those in Mobile Ad hoc Networks (MANETs) [5] [6], Wireless Sensor Networks (WSN) [7] [8]. As clustering approach permits scaling of a potentially large network into several smaller autonomous groups as well as scoping of operational functions and message exchanges within a cluster, this approach generally offers high scalability and efficiency. To take the advantage of clustering approach for management of distributed caching, we present a Hierarchical Cluster based Caching (HCC) solution. Our design has a two-layer hierarchical clustering architecture. The routers in Core Layer are not used for caching so that they can focus on content routing. In Edge Layer, routers are designed to cache contents for prompt user responses. Furthermore, we introduce importance rating where nodes of higher (resp. lower) importance rating in Edge Cluster cache more (resp. less) popular contents. Cluster head has a responsibility to collect and allocate the information of node importance based on betweenness centrality, content popularity and probability matrix in its cluster. In additional, all the important nodes execute their respective caching decision whether a content should be cached, and if so, where. Our proposed HCC solution is implemented in ndnSIM [10] considering two different network topologies. We show that HCC out forms other strategies in several aspects.

(Size 10 & Normal)An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III.PROPOSED SYSTEM

In the proposed system we are implementing an efficient hierarchical cluster based named data network. Before building named data network we are preform the authentication of each node in the network. The authentication of nodes can be done by server and the server will send authentication status to all other nodes in the network. After completion of authentication process the server will perform the hierarchical clustering process. In the hierarchical clustering process the server find out each cluster head randomly. After choosing cluster head the server will group nearest nodes into group based on number of clusters in the network. In the grouping one of group member will work as gateway and the gateway group member will connect to other gate way group member. Suppose to transfer content to other group the gateway will send that information to other group gateway member. By performing this operation we can reduce traffic in the network. After that the source node will transfer content to destination node through shortest path. The generation of shortest path can be done by using shortest path routing protocol. Before transferring content from source node to destination node the source node will encrypt content and send to destination node. The transferring cipher format data can be traverse through shortest path and reach destination node. The destination node will retrieve cipher format data and perform the decryption process. After completion of decryption process the destination node will get plain format data. So that by implementing those concepts we can improve efficiency for transferring content from source node to destination and also provide more security of transferring content. The implementation procedure of authentication of group members in the network is as follows.

A. Authentication of Group Members:

In this module each member in the network will be authenticated by the server. The implementation process of authentication of group members is as follows.

1. The server will generate p be a large prime such that computing discrete logarithmic modulo p is difficult
.
2. The server will also generate g < p be randomly chosen generator of the multiplicative group of integers of modulo p.

3. After generating p and g values the server will send those two values to all members in the network.

4. Each member will retrieve p, g values and choose secret key x_i with $1<x_i<p-1$.

5. Each member will calculate public key y_i by using following formula.

$$y_i = g^{x_i} \bmod p$$

6. To generate signature each member will randomly choose k_i with $1<k_i<p-1$ and gcd $(k_i, p-1) = 1$.

7. After that each member will calculate r and s values by using following equation.

$$r_i = g^{k_i} \bmod p$$
$$s_i = (H(id) - x_i r_i)\, k^{-1}_i \bmod (p-1)$$

8. Then the pair (r_i, s_i) is the digital signature of each member and sends those pairs to server.

9. The server will retrieve each member pairs and verify the each member in the network.

$$g^{H(m)} \bmod p = y_i^{r_i}\, r_i^{si} \bmod p$$

The server verify the all the signatures if the condition is satisfied send authentication status to all members. If the condition is not satisfied reject the member. After completion of authentication process the server will perform the hierarchical clustering of nodes in the network. The implementation process of hierarchical clustering of nodes is as follows.

B. Hierarchical Clustering Structure:

In this module the server will perform hierarchical clustering nodes in the network. Before performing clustering of nodes the server will choose cluster heads by calculating nearest distance from cluster heads to other nodes. Take each cluster head and perform clustering process is as follows.

1. Take each cluster head and remaining nodes we can calculate mse of each cluster head to all nodes. The calculation of mean square error is as follows.

$$MSE=1/n \sum_{i=1}^{n} (x_i - c)$$

Where n is number of nodes in network, x_i is the remaining nodes and c is cluster heads.

2. The algorithm goes next iteration and finds mse compare with previous iteration fitness. Find out better mse from the both iteration and repeat this process will stop if the best mse value seems not to be changed in next generation.

3. Take those centroids are output of the process and again calculating a pairwise distance between all clusters.

4. After calculating pairwise distance of all clusters the algorithm picks all pair of clusters distance smaller than overall average of cluster distance.

5. Once we are calculating pair of distance the algorithm again re cluster using same process.

6. By performing this process we can get small sub set of data and algorithm easily finds partitions which are more distant to each other.

After completion of clustering of nodes the server will choose the gateway nodes of each cluster by using nearest distance between remaining two cluster gateways. By using the gateways we can connect with all other cluster heads and also transfer content from one cluster to other cluster nodes. After that the source node will enter transferred content and perform encryption process.

C. Hybrid Bit Transpose Bioinformatics Technique:

In this module the source node will encrypt the transferred content and convert into cipher format. The implementation process encryption is as follows.

1. Read input data and convert into binary format.

2. Take the binary format data and convert binary format data into DNA characteristics like A, C, G and T. Here we are take binary value 00 is A, 01 is G, 10 is C and 11 is T.

3. Take the p and g values as key and perform the xor operation it will get ascii value.

4. Take that ascii value convert binary format and those binary format data also write into DNA characteristics format.

5. Apply xor operation of step 2 and step 4.

6. After completion xor operation take those values convert into binary format and perform complementary operation.

7. Take the output of step6 and convert into DNA sequence.

8. Take the DNA sequence from step7 and convert into ascii format.

9. Take the ascci format data from the step 8 and convert into binary format.

10. Take the binary format data and write into 32 * 32 matrix format.

11. After that we can transpose that matrix and get those binary values. Take those binary values and convert into ascii format and that ascii format data is cipher format data.

After completion of encryption process the source node will transferred cipher format data to destination node. Before reaching content to destination node the server will find out destination node contains which cluster. After that the cluster head will generate path to destination node by using shortest path routing protocol. The implementation process of shortest path routing protocol is as follows.

D. Nodes Signal Strength Shortest Path Algorithm:

The implementation Average Nodes Strength Shortest Path algorithm finding shortest path. Before finding shortest path we are initialize each node signal strength and stored into database. After that we can also specify the out signal strength node in network. The implementation process of Average node strength shortest path algorithm is as follows.

1. Get the all nodes of signal strength (S_i).

2. Finding distance source node to other nodes by using following formula

```
int max=0;

int min=S[i];

if (max<min)

  {
  Max=min;
}
```

3. After finding distance of each node we can arrange the path from source node to destination node.

4. So that the data send through path and reached the destination node.

After reaching data to destination node will decrypt the cipher format data and get original plain format content. The implementation process decryption of Hybrid Bit Transpose Bioinformatics Technique is as follows.

E. Decryption Process:

In the module the destination node will perform the decryption process for getting original content.

1. Take the ascii format cipher data as input and convert into binary format.

2. Take the binary format data and generate 32 * 32 matrix format. After generating matrix we can transpose binary format matrix data.

3. Take the binary format data and convert into ascii format.

4. Take the ascii format data of stpe3 and convert into values of DNA sequence.

5. Convert DNA sequence from step4 into binary sequence.

6. Take the binary sequence data from stpe5 and perform the complementary operation

7. The destination node will take p and g values will perform the xor operation.
8. Take those xor data and convert into binary format. Take the binary format data and convert into DNA sequence format.

9. Apply xor operation between rule step 6 and step 8.
10. Convert those xor data into binary format and Take each eight bit from binary data convert into character it will get original plain format data.

So that by implementing those concepts we can get better performance in the authentication and also get efficient Hierarchical clustering of nodes in the network. Another process we are implementing a simple way of routing process for finding shortest path using nodes signal strength shortest path algorithm. After that we are perform the secure data transfer from source node to destination node.

IV. CONCLUSIONS

In this paper we are propose a named data network for maintaining caching strategy. By maintaining caching strategy we are implement hierarchical cluster based caching. Before performing hierarchical clustering the server will identify all nodes are authenticated nodes or not. By performing authentication process we are implementing hash based signature algorithm. After completion of authentication process the server will performing clustering all nodes in named data network. In the clustering process we are using hierarchical clustering process. By performing clustering process we are generating number of cluster group nodes in the network. Take the any node as source node and send data to destination node. Before transferring data to destination node the server will generate shortest route based on clusters. After completion of routing process the source will encrypt transferring message and send to destination node. The destination node will retrieve cipher format data and decrypt it. In this paper we are implement one of cryptography technique for performing encryption and decryption process. The cryptographic technique hybrid bit transposes bioinformatics technic for data encryption and decryption. By implementing those concepts we can improve performance of data transferring and also improve security of transferred data.

REFERENCES

[1] Jacobson V, Smetters D K, Thornton J D, et al. Networking named content[C]//Proceedings of the 5th international conference on Emerging networking experiments and technologies. ACM, 2009: 1-12.

[2] Breslau L, Cao P, Fan L. Web caching and zipf-like distributions: evidence and implications. In: INFOCOM' 99. Proceedings of the IEEE 18th annual joint conference of the IEEE computer and communications societies, vol.1.IEEE; 1999.p.126–34.

[3] Rossi D, Rossini G, Caching performance of content centric networks under multi path routing (and more). Technical report, Telecom ParisTech; 2011.

[4] D.Kutscher and et al. Icn research challenges. IRTF, draft-kutscher-icnrg-challenges-00, Februady 2013.

[5] P.Krishnan, D. Raz, and Y. Shavitt. The cache location problem. IEEE/ACM Trans. Netw., 8(5), 2000.

[6] V.Pacifici and G. Dan. Content-peering dynamics of autonomous caches in a content-centric network. In IEEE INFOCOM, 2013.

[7] J.A. Torkestani and M. R. Meybodi, "A ability-based cluster formation algorithm for wireless mobile ad-hoc networks," *Cluster Computing*, vol. 14, no. 4, pp. 311–324, 2011.

[8] J.Y. Yu and P. H. J. Chong, "A survey of clustering schemes for mobile ad hoc networks." IEEE Communications Surveys and Tutorials, vol. 7, no. 1-4, pp. 32–48, 2005.

[9] H.J.Choe, P. Ghosh, and S. K. Das, "QoS-aware data reporting control in cluster-based wireless sensor networks," Computer Communications, vol. 33, no. 11, pp. 1244–1254, 2010.

[10] G.Chen, C. Li, M. Ye, and J. Wu, "An unequal cluster based routing protocol in wireless sensor networks," Wireless Networks, vol. 15, no. 2, pp. 193–207, 2009.

BIOGRAPHIES

Ellapu Jyothi is student in M.Sc (Computer Science) in Chaitanya Women's pg college, Old Gajuwaka, Visakhapatnam. She has received his Degree B. Sc (M.P.Cs) from Mahathi Degree College, NAD Junction, Visakhapatnam. Her interesting areas are data mining, network security and cloud computing

P. Mohana Roopa is Head of Department of M.Sc Computer Science in Chaitanya Women's PG College. Old Gajuwaka, Visakhapatnam, Andhra Pradesh. She Receive her M. Sc Computer Science from Andhra university.Her research areas include Network Security and Computer Networks