En-Route Web Caching in Trustworthy Structured Peer-To-Peer Networks

R.Jayamala
Dept. of Computer Science
Anna University of Technology
Tiruchirappalli, Tamil Nadu, India

Pon.L.T.Thai
Dept. of Computer Science
Anna University of Technology
Tiruchirappalli, Tamil Nadu, India

ABSTRACT
In the Peer-to-Peer networks Web caching plays an important role in reducing the network delays, latency of the user response and the server load. Still web caching is of great challenge as the maintenance of object copies at multiple nodes is the issue of trustworthiness, consistency and optimal resource placement. The en-route web caching strategy along with the reputation aggregation method can produce valuable solution to this issue. In this paper every request for web objects are traced at every node in the request path. These traces at the peers are used to predict the data popularity of each web objects. The popular web contents are requested and cached in different trustworthy nodes with effective resource placement strategy in order to improve the hit rate of the network. Trustworthiness calculation is the reputation aggregation of the Response provider node obtained at every node.

General Terms
Peer to Peer Networks, Web Caching, Resource Management, Network Latency.

Keywords
Distributed Hash Table, En-Route Web Caching, Trustworthiness, Resource Replacement.

1. INTRODUCTION
P2P network (Peer to Peer network) is the collection of nodes which are the members of internet, connected via the overlay network to form a closed group network. P2P network helps the users to find the web contents they want and to share what they have. But sharing the files and caching files in anonymous peers makes the system vulnerable to attack. The basic requirements of every peer in the network are, Every node needs to perform different roles such as, a server when a request arrives to the node, a client when it needs files from other nodes and a router when it is the intermediate node or when it knows the path of the requested web object. Peers need to provide simple and secure file sharing among nodes. The information about the nodes in the network and their keys are hidden from the users to reduce complexity. Redundancy of files at various nodes is suggested for high availability of data. It need to be noted that the cache memory of every node is limited to some Mega Bytes.

Routing of packets in the P2P network is widely classified into Table Driven, On demand and Location based routing. The combination of them is also used to increase the efficiency of the protocol. The best example of such Table driven and Location based routing protocol combination is Distributed Hash Table (DHT) based Protocol mainly used in P2P network.The four popular DHT based protocols are Content Addressable Network (CAN), Chord, Pastry and Tapestry. Every node in the DHT based P2P network maintains a hash table to store the information about all the neighbor nodes and their key, value pair.

The Trustworthiness of a peer can be measured using two methods. They are Subjective Trustworthiness and Objective Trustworthiness. Subjective Trustworthiness is measured using the direct communication of requesting node (Nr) with the Response provider node (Ns). Whereas Objective Trustworthiness is measured using the communication with the receiving node as well as the neighbor nodes (Nn). Making decision based on the information gathered from neighbor node and their communication history is more trustable than that of the trust information from single node.

2. RELATED WORK
About the efficient routing in the P2P networks, web caching and trustworthiness many research works has been carried out till now. In [1] the research of Hong Shen et al., the coordinated en-route web caching for multi server systems is discussed. They suggested deploying multiple servers in a system as an alternative solution for multiple copies at different nodes. This is because of the fault tolerance and high reliability of multi-server network. They also have calculated the cache gain in terms of caching benefit, which is defined as, the cost reduction at a particular node for requests to a given object to all servers.

In [2] Soichi Sawamura et al., have compared three types of Objective Trustworthiness algorithms, OT1, OT2 and OT3. OT1 floods the request to obtain the trustworthiness of the node from acquaintances. OT2 sends request only to the acquaintance and their neighbor nodes to calculate the trust level. Whereas the OT3 algorithm gets trust details only from its trusted acquaintances and consider the average as the trustworthiness of the node. They have concluded in their evaluation that the OT1 algorithm which floods message in the network is waste of resources, OT2 algorithm better suits for less trustworthy nodes and OT3 is good in case of trustworthy acquaintance. In [3] Weixiong Rao et al., have characterized some special nodes in the structured P2P as acceleration nodes in order to accelerate the search. Acceleration nodes and their neighbors form a group and direct long links exit between two or more acceleration nodes, known as PRing. PCache - popularity based web caching method caches the popular web contents by using Zipf-like distribution to decide the number of replicas and replacement.

3. OPTIMAL RESOURCE PLACEMENT

The P2P network is not stable due to the arrival and departure of nodes in the network. Different nodes join and leave the network, which questions the availability of data. This is addressed in the project by tracing the responses pass through every node in the network. For a particular time interval the traces are scanned for new popular web objects. This process is iterated to cache the popular web objects at the nodes on the response route. Another problem is that, the nodes in the network communicate with anonymous nodes to form the DHT. Malicious nodes can make use of it and can access the data that are not available to it. In order to provide optimal resource placement with trustworthiness, this paper proposes to store the trust level information of neighbor nodes in the DHT. This can be exchanged with other neighbors on demand.

3.1 Web Object Properties

From the study on the traces captured by Boston University for their research purpose it is clear that, a weak correlation exists between the resource request and the resource size. Another conclusion is that the rate of modification in the web object does not depend on its access frequency. For example multimedia files are often accessed by the users regardless of its size because of the growth of internet access technologies. While coming to sharing, every web object has some parameters associated with it, like object permission, object owner, access rights, cache version, etc. The object permissions possible are read, hold and manipulate. The read permits to access the file, hold allows caching of the web object and manipulate permits to modify the content.

3.2 Protocol Architecture

In the TCP/IP protocol suite, apart from the well known functions of layers some special functions are performed in this protocol. At transport layer trust level of the response provider is calculated and stored in the DHT. Routing of trust information and web caching is performed at Network layer. The physical layer is the wired and wireless connection established between two nodes in the overlay network. Routing and key value pair matching is performed in the network layer, where request for all the web objects are traced and popular web objects are cached. File and message transmission is performed in the Internet layer. This project works using the connection oriented, reliable Transmission Control Protocol (TCP) for file transfers and connection less User Datagram Protocol (UDP) for the control message transfers. The control messages are used to request file in the network and to request the trustworthiness of a particular node. In order to reduce the traffic the packet size of the control messages are small and they are not flooded in the network. The control messages are sent to the neighbor nodes to get the trust level of requesting node, based on their history of data transfer in the past. They are mentioned in the Figure 1.

The trust computation of neighbor nodes in the network is done in the Transport layer. The nodes in the P2P networks store all the neighbor details in its DHT. Whenever a response path is created, the neighbor node calculates the trust level of response provider. This information is stored in the DHT. It is a matter of little more space to store the trust level information for future use. Every response message through every node is traced and the web object details are stored in a data set as <Object ID, Object permission, Object owner, Cache Version>. The function of the application layer is to get the users input from console or to abstract the important information from end users. The user inputs are request to the web objects that are stored somewhere else in the network.

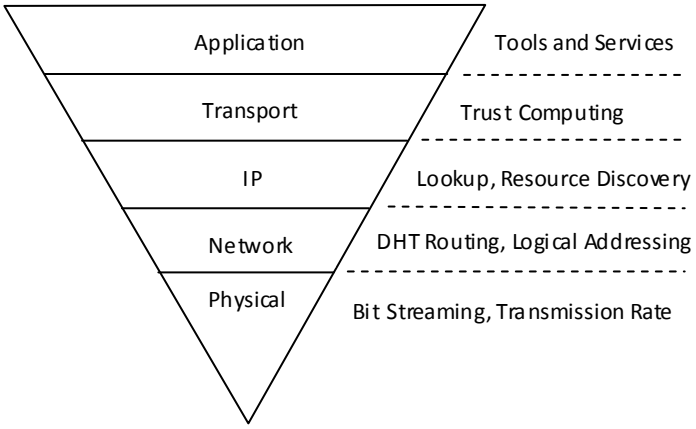

Fig 1: TCP/IP Protocol suite for Enhanced DHT based P2P network.

3.3 Optimal Resource Placement

In order to make efficient searching every node stores the list of web object present in the neighbor nodes and their cache details. So when a request arrives to a node, it is not flooded to its neighbors. The request is flooded only to the accelerator nodes. The nodes in the DHT based P2P networks stores all the node details, whose node IDs are numerically nearest to the current node, i.e. the neighbor nodes according to the numerical key value. The protocol need to consider the network locality. That is the node that is nearest to the current node in the DHT is far away in scalar proximity. This is because of the overlay network laid above the existing internetwork. So the node IDs are assigned by considering the distance exist between the nodes in the underlying network.

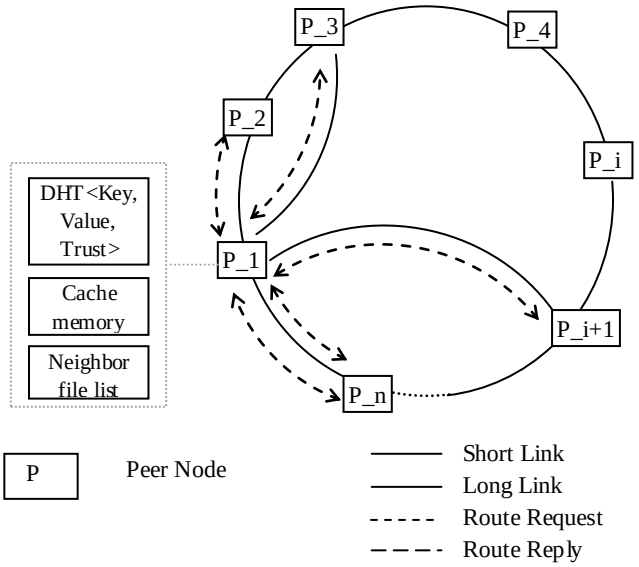

Fig 2: System architecture of Peer to Peer network with En-route web caching and trustworthiness calculation.

The web objects are of two types, static and dynamic. Static web objects do not vary with time. The other type is dynamic, such dynamic web pages. These dynamic web pages and web object of large size like multimedia are filtered out as Non cacheable web objects. Non- cacheable contents are time varying web objects and large size multimedia which require more memory to cache the content. The cache replacement policy suggested is en-route web caching. The popular web objects are stored in the intermediate node in the search path, so that the requests in future are not forwarded to the origin node of the web object. Simultaneously the tracing is done at node and new popular web objects are calculated.

The web objects in P2P network are cached and deleted based on the popularity of the data. So as soon as a cache is deleted, all the neighbor nodes having the cache location as current node need to be informed about the deletion of cache. Content management is done by routing of file list among the neighbor nodes for every time interval. The data set of the neighbor file list is <node ID, object, cache version>. If the cache version is null, the neighbor node is the owner of the content. The cache version is used to know the time of cache. If the cache is very old, request can be further routed for the fresh content. From the response the cache can be refreshed and the cache version is increased. The Figure 2 shows a standard P2P network with nodes connected in a ring structure. The dashed arcs show the duple connection established between the nodes. Every node has a successor and predecessor. Long links are established to accelerate the search. From the perspective of the peer node P_1 the predecessor is P_n, and its successor is P_2. Long links are established to the nodes P_3 and P_i+1. The node can communicate with these nodes directly.

3.4 Objective Trustworthiness

Trustworthiness is calculating the trust level of the neighbours from the transaction with them in the past and present. Every Distributed Hash Table (DHT) based P2P network suffers from man-in-middle attack, which mainly affects the Trustworthiness of the network. This attack is the entry of a malicious node into the network which cannot be differentiated from other nodes. Malicious node also looks like other nodes in the network and it gives wrong shortest path information in order to make file transfer through it. The insecure data transfer leads to misuse of data. So it is mandatory to choose the path through secure nodes for file transfer rather than malicious shortest path.

3.4.1 Computing Objective Trustworthiness

Trust level of a node depends on the time delay involved in the transaction in the past and the link strength. The request for the web object is created and flooded to the neighbor nodes and the nodes in the acceleration nodes. The time to live of the request message is set to remove the packets after sometime. The neighbor nodes search for the presence of web content in its neighbors by checking the neighbor file list. In this way a greedy search is initiated. If one of a neighbor has the content, then it is known by neighbor nodes file list. Then trust level request is sent to all its neighbors. From the responses the global score is calculated by the average of local trust level stored by the node and the local scores of its neighbors. If the accelerator node sends the response path, the request is sent to the nodes whose node ID is numerically nearest to the response provider. If the global score is acceptable to be secure then transaction link is established among them.

Trustworthiness can be given by a simple four-way handshaking method. Let us consider, the requesting node Nd wants to get web object from the Response provider node Ns. So Nd sends the request message for the required web object to the node Ns. The request message is forwarded to the node Ns through different nodes. All the addresses of the nodes along the path are attached in the request packet. From the request messages received through different paths the best path is chosen as the response path. The response message reaches the node Nd and file transfer starts.

Sometimes a malicious node Nm near the Response provider node Ns may rush a counterfeit shortest path to node Nd with its address as next node. If file transmission starts in this path the unauthorized malicious node will get access to data. In order to avoid that, before starting file transmission through the response path, the node Ns again sends request for trust level of next node to all the neighbor nodes. If any of the node is malicious and not present in the neighbors hash table it is informed and another route is chosen.

3.4.2 Example

Let us see, how this project avoids the major problem of P2P networks, man-in- middle attack by simple trust level calculation at every node. In the Figure 3 the node D tries to get the web object resides in the response provider node S. So the route request message is flooded to the neighbors through short links and to the accelerator nodes through long links. The malicious node M in the middle rush the reply packet before S replies with its path. So in order to maintain trustworthiness the node D, sends trust request to the next node in the response path mentioned by M that is E. But the node M is not present in the DHT of E. So node E sends the unsecure path error report to the requesting node D. The path with malicious node is rejected and file transfer starts using the next possible secure path,

S ----> C ----> B ----> A ----> D.

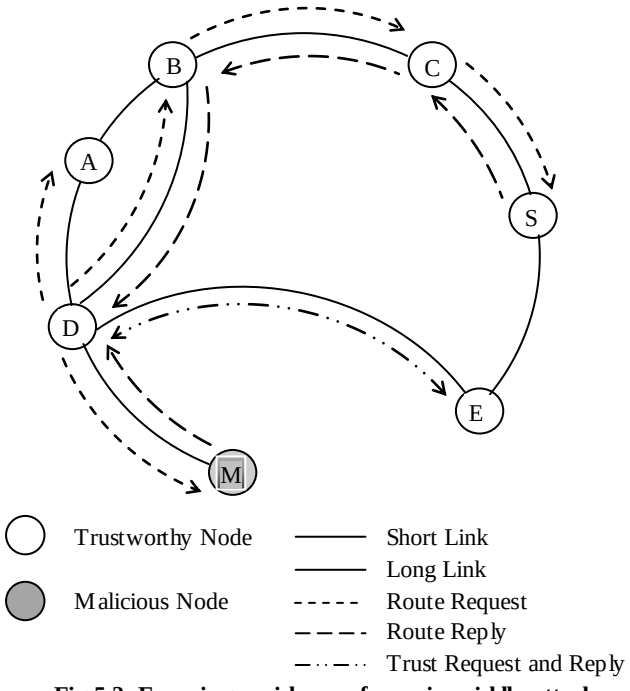

Fig 5.2: Ensuring avoidance of man-in-middle attack

4. ANALYSIS RESULTS

The emphasis of simulation is to compare the relative performance of the optimal file placement scheme with en-route caching schemes. Instead of simulating over a network containing a large number of web files, we simulate for a simple message and over a sub-network which is the tree rooted at the peer permanently maintaining the file. This simplified model serves as a valuable first step toward the simulation over the original networks with many files, provides us with a useful tool for studying and comparing the different behaviors of different file placement policies, which is important for understanding the placement algorithms at a deeper level. From the analysis done in the Pastry protocol, coded in Java programming language by routing varying number of message through different nodes in a single Java Virtual Machine (JVM), we obtained the time required for message creation, hop search and message delivery with respect to the increase in number of nodes. The simulation is performed by creating nodes varying from small number of nodes to large number of nodes using random node IDs from the Node factory. These nodes are then used to create and route message through the network. The messages are delivered among the nodes with minimum hop search. The resulting time delay in the simulation is given in Table 1.

Table 1. Time Delay with increase in the number of messages flooded in the network

Number of Messages / Number of Nodes	Time Taken (in Milliseconds)			
	5	**10**	**15**	**20**
5	18	22	24	27
10	36	41	46	51
15	59	65	73	80
20	74	84	95	102
25	90	109	124	126
30	107	123	134	149

The analysis result shows the time delay increases gradually with respect to the increase in the number of messages and nodes. This change is irrespective of the exponential increase in the number of process running from a single node in the P2P network because of the ring structure and the accelerator links used by the protocols.

Requests for the web object in consideration are generated by nodes and transmitted through many nodes in the path. Grouping of nodes into a subset of the network, and accelerator nodes as web object locator of the group are performed in these protocols for efficient hop search. Instead the number of request messages transferred over the network need to be reduced. Figure 4 shows the increase in time delay with flooding of request messages in the network. Figure 5 gives the performance analysis of the En-Route web caching project performance with the existing Pastry protocol performance.

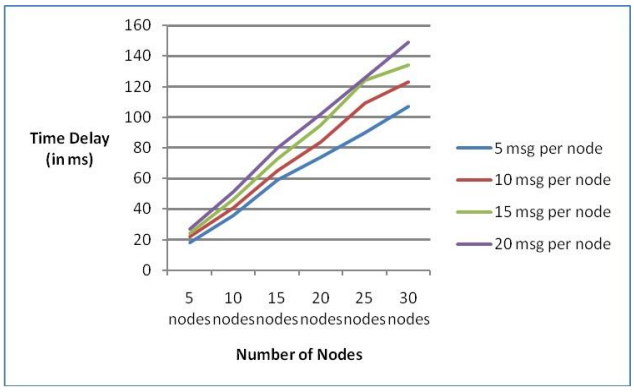

Fig 4: Gradual increase in time delay with respect to the increase in number of messages and nodes

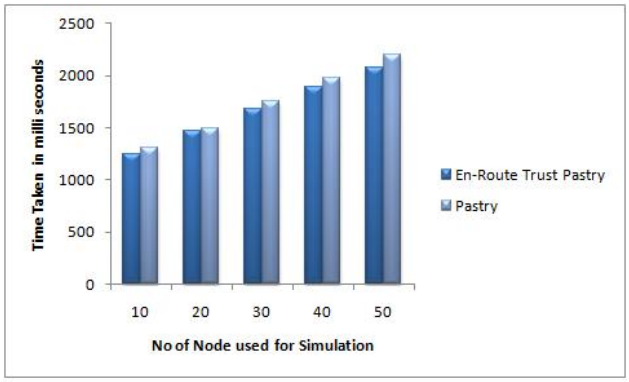

Fig 5: Performance Analysis with existing Pastry protocol and En-Route Trust Pastry

5. CONCLUSION

In existing P2P networks, the separation of concern is maintained between trustworthiness and routing technologies. This increases the number of request messages routed in the network, thus increase the network traffic. The proposed system presents a low-complexity solution which computes the optimal solution for file placement. It is shown both analytically and experimentally that the optimal solution given is significantly better than the sub-optimal solutions given by other schemes. This paper proposes to combine trust information with request and response message or file routing. The major modifications to the existing system are optimal number of web object replicas at different nodes and trust information exchange. The combined approach is showed to increase the hit ratio of web search, especially in DHT based routing networks.

6. FUTURE WORK

The future work of this project involves two aspects. First is performance tuning of the project. The time delay and the memory management associated with the project are not affordable when it comes to real time browsing. So the performance tuning needs to be done in future. The second one is relevant to analysis performed on the project. The analysis is done by creating so many nodes virtually on a single machine and based only on time. But it need to extend to memory requirement, traffic in the network, network latency, etc.

7. ACKNOWLEDGMENTS

We take this opportunity to express my hearty thanks to Dr. R. KRISHNAMOORTHI, Dean, Anna University of Technology, Tiruchirappalli, for providing us the essential backing and support. Our special thanks go to Mrs. R. S. PONMAGAL, Head of the Department, Computer Science and Engineering, who provided much encouragement and constant guidance. She also made the Computer Lab a wonderful workplace and home for the past two years by indulging our ever expanding bookshelf space and computer equipment needs. We also extend our thanks to the colleagues who have contributed by providing us various guidance to the completion of the project.

8. REFERENCES

[1] Hong Shen, and Shihong Xu, "Coordinated En-Route Web Caching in Multiserver Networks," IEEE Trans. Computers, Vol. 58, No. 5, May 2009.

[2] Soichi Sawamura, Ailixier Aikebaier, Tomoya Enokido, Valbona Barolli and Makoto Takizawa, "Confidence and Trustworthiness in Peer-to-Peer Overlay Networks", 13th International Conference on Network-Based Information Systems, Vol. 9, No. 6, pp. 24, Nov. 2005.

[3] Weixiong Rao, Lei Chen, Ada Wai-Chee Fu, and Guoren Wang, "Optimal Resource Placement in Structured Peer-to-Peer Networks," IEEE Trans. Parallel and Distributed Systems, Vol. 21, No. 7, pp. 1011 - 1026, July 2010.

[4] Antony Rowstron & Peter Druschel, "Storage management and caching in PAST, a large-scale, persistent peer-to-peer storage utility", In Proc. IFIP/ACM Middleware, Heidelberg, Germany, Nov. 2001.

[5] Peter Druschel and Antony Rowstron, "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems," Proc. of the 18th IFIP/ACM International Conference on Distributed Systems Platforms, November 2001.

[6] Lee Breslau, Pei Cao, Li Fan, Graham Phillips, Scott Shenker, "Web Caching and Zipf-like Distributions: Evidence and Implications", Proc. of IEEE INFOCOM, Vol. No 1, pp. 126 - 134, Mar. 1999.

[7] Runfang Zhou & Kai Hwang, "PowerTrust: A Robust and Scalable Reputation System for Trusted Peer-to-Peer Computing", IEEE Transactions on Parallel and Distributed Systems, Vol. 18, No. 4, pp. 460, Jun. 2006.

[8] Yoshio Nakajima, Alireza Goudarzi Nemati, Tomoya Enokidoy, and Makoto Takizawa, "Trustworthiness and Confidence of Peers in Peer-to-Peer (P2P) Network", 22nd International Conference on Advanced Information Networking and Applications, Apr. 2008.

[9] Shanshan Song, Kai Hwang, Runfang Zhou, and Yu-Kwong Kwok, "Trusted P2P Transactions with Fuzzy Reputation Aggregation", IEEE Internet Computing Magazine Special Issue on Security for P2P and Ad Hoc Networks, Vol. 9, No. 6, Mar. 2005.

[10] X. Tang and S. T. Chanson, "Coordinated en-route web caching", IEEE Trans., Vol. 51, No. 6, Jun. 2002.

[11] Masanori Yasutomi, Yo Mashimo, and Hiroshi Shigeno, "GRAT:Group Reputation Aggregation Trust for Unstructured Peer-to-Peer Networks," IEEE 30th International Conference on Distributed Computing Systems Workshops, pp. 126, Jun.2010.

[12] P. Yalagandula and M. Dahlin, "A Scalable Distributed Information Management System," ACM SIGCOMM Computer Comm. Rev., Vol. 34, pp. 379 - 390, 2004.

[13] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, "A scalable Content-Addressable network," Proc. of SIGCOMM, Aug. 2001.

[14] Stephanos Androustellis, Theotokis and Diomidis Spinellis, "A Survey of Peer-to-Peer Content Distribution Technologies," Proc. of the ACM Computing Surveys, Vol. 36, No. 4, pp. 335 – 371, Dec. 2004.