IJASRM International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

Optimized Circuit Construction for Privacy Enhanced Onion Routing Based on Genetic Algorithm

Dr. S. Shakila

Department of Computer Science, Government Arts College
Tiruchirappalli, TamilNadu

Abstract

Anonymity during transmission is achieved using the process of Onion Routing by rerouting a packet through various intermediate routers in an encrypted format. Security seems to be an indispensable aspect for onion routing network. Security in onion routing network is determined by the strength of the circuit. Though there are a number of protocols to enhance the security, providing security at required level demanded by an application is a matter of serious concern. One of the other major discrepancies that is unheeded here is the slow nature of transmission when an onion routing network is used. This paper provides a mechanism that determine the level of application traffic and choose the circuit levels accordingly. Fuzzy logic is used to distinguish the nature of application as whether throughput sensitive or delay sensitive. Further Genetic algorithm is being used in this paper to construct an optimal path dynamically in a reasonable computational time. The QOS requirements like delay, latency, jitter, relay uptime are considered to find an optimal solution. This work seems to be efficient and secure in the process of data transmission.

Keywords *- AHP, Anonymity, Circuit Construction, Fuzzy set, Genetic Algorithms, Latency, Onion Routing, QoS, Roulette wheel selection Tor*

1. Introduction

With the tremendous growth of Internet users and rapid advancement in the technology over the years paved way for the people to depend on networked distributed systems to carry out their daily activities. However, people show great concern over their personal details being tracked by the third parties. It can be forthrightly said that, privacy is a major issue on the web because the IP address of the user, domain name, organization, referred website, information requested etc. are being advertised by the browser (S.Shakila, et al, 2014). Preserving privacy has become one of the major requirements of the current internet age. Anonymous Communication Systems (ACS) provide collaboration between online users in a secure fashion. Need for privacy is increasing due to the increased attacks by exploiting the vulnerabilities in the system to identify the user's credentials (S.Shakila, et al, 2015) . Onion routing is one of the techniques that provides privacy by means of encrypting the data and by shuttling the packet through several routers in the Tor network (Panchenko et al., 2009).

The effectiveness of Tor networks depends on the route selection strategy which is very complex because the selected route should be the fastest route and unpredictable. It is infeasible to achieve both these functionalities because speed is a trade off for the need of security. Hence a Tor network always remains a slow and secure routing structure. A lot of research has been carried out to measure anonymity provided by Tor and to guard against potential attacks. Due to the wide use of Internet based applications, Hyper Text Transfer Protocol(HTTP) traffic comprises an overwhelming majority of the connections and it is unclear whether Tor can facilitate interactive web (S.Shakila, et al., 2014). Any packet that is passed through a Tor network has always been found to reach the destination taking at least 3x or 4x times of the transmission time taken by a normal transmission.

The mode of operation of an onion routing network begins by establishing a secure initial connection to an entry node. The next phase starts with the exchange of the TLS key (Dingledine et al., 2004). Similar node selection and key exchange takes place in the Tor network until the exit node is reached. The source is aware of the number of nodes in the route, hence it encrypts the packet accordingly. Every node, on receiving the packet decrypts or strips off the encryption layer using the exchanged key and passes it to the next node in the route (Owen et al., 2007).

Tor is a free software, that was presented as a component of an anonymous routing project named; The Onion Router currently, Tor contains six thousand relays worldwide for transmitting traffic to incorporate anonymity to the information being transmitted. It has been funded by the National Security Agency (NSA) and is considered as the most popular anonymous internet communication system. Onion Routing is implemented by encrypting the packet in the application layer of the protocol stack (Li et al., 2011). The encryption includes the source and destination IP addresses, hence user anonymity is maintained throughout the communication process. Traffic passed via Tor network is not 100% fool proof and it is also prone to a few attacks. The sender has no information about number and identities of compromised nodes. The route selection therefore does not rely on knowledge about which nodes are compromised. Thus, some compromised nodes may be on the rerouting path. The adversary has full knowledge of the path selection algorithm. In particular, the adversary knows the path length distribution.

The entry and exit nodes are the most vulnerable points on a Tor network, as the packets in those nodes contain certain crucial details about the sender or the receiver. Other attacks that a Tor network is prone to are eavesdropping, traffic analysis (Zhou, Peng et al., 2013), Tor exit node block, Bad apple attack, Sniper and heartbleed bug. Anonymity should not be confused with security. Hacking into the Tor network is difficult but hacking into Tor browser is a whole new story. The need for HTTPS is still relevant in Tor and is advisable that it be used whenever possible. A comparison showing pros and cons of Tor networks is presented in (Liska et al., 2010). It shows that the major problem of a Tor network is the implicit delay incorporated into it, which need to be addressed.

Efforts to improve the performance of the Tor network can often decrease the anonymity, and vice versa. To address this problem, an optimised path construction which can be tuned to the requirements of the user is suggested. As Tor network is designed using complete graph topology each router will have a chance to interact with other routers and their performance can be observed empirically. Also over-utilized routers will show decreased performance, hence exploration and exploitation of routers need to be considered in selection. A metric based path selection technique is presented in (Milajerdi, et al., 2015)(Backes et al., 2012)This method employs a combination of metrics in the path selection process. The metrics used are bandwidth, uptime of relays, node conditions, jitter and delays between the relays. Tor is presented as a solution for the existing privacy concerns in web mining in (Gopinath et al., 2014). It presents the security issues and loop holes arising in web mining and explicates how Tor can help overcome these issues.

The remainder of this paper is structured as follows; Section II presents the related works, Section III explains the approach of obtaining an optimized circuit using Genetic Algorithm, Section IV presents the results and discusses them and Section V concludes the study.

2. Related Works

2.1 Overview of Tor

Tor is an Internet networking protocol used to anonymize the data relayed across it. The Tor network uses several thousands of volunteers comprising of computer servers spread throughout the world. Tor is the second generation Onion routing protocol where data is bundled into an encrypted packet in layers when it enters the Tor network. Each layer of the packet contains addressing information about sender and receiver which is learnt by stripping the layers. The encrypted data packet in the form of onion is then routed through many of these servers, called relays, on the way to its final destination.

The Tor network using self-reported bandwidth is replaced with an opportunistic bandwidth measurement mechanism which can accurately predicts the performance of the routers and is significantly less susceptible to low-resource attack. Experiments with Tunable Tor show that users can achieve great improvements in performance without sacrificing much anonymity, or significantly increase anonymity protection without any loss in performance (Robin Snader et al., 2012) .This improved flexibility should make Tor palatable to a wider range of users, and thus increase anonymity for everyone due to a larger community (Dingledine et al., 2007). Tor routers are registered with a directory service. Each router advertises its IP address, public key, policies about what traffic it will accept, and a bandwidth value that is determined by monitoring the peak bandwidth achieved by the router over a period of time.

To improve the performance of Tor network the impact of different Tor path selection strategies on bandwidth, circuit failure rates, and the number of attempts required to build a circuit need to be studied. The benefit of studying such strategies is that they enable the evaluation of features of the current network as well as potentially providing an easily deployable solution for improving performance. Choosing appropriate metric to measure performance and reliability of Tor network is of paramount importance. Reasonable amount of throughput is needed to transfer large files and hence measuring the throughput is an integral part of how Tor builds faster circuits through its network, which

shows that it is a useful metric when considering the performance of Tor. As only a small percentage of Tor relays have very high bandwidths, path selection approaches which only consider relay bandwidths would tend to select such nodes, creating more deterministic paths and also congesting these few nodes by placing heavy loads on them. On the other hand, only selecting relays based on their geographical location, will put heavier load on some relays near specific high traffic locations. Hence, it would be beneficial to select paths while considering three parameters, up-time for reliability, bandwidth for performance, and geographical locations for latency.

(Overlier et al., 2006) presented new attack strategies to detect the location of hidden servers using only one Tor node. They proposed changes in route selection and relay selection to increase anonymity. The average duration of the attack varied from minutes to a few hours. The various attacks they considered included the timing signature analysis attack, service location attack, predecessor attack and distance attack. Their proposed solution included introducing middleman nodes to connect to rendezvous points, introducing dummy traffic, extending hidden server path to rendezvous points and using guard entry nodes. A path selection method in which overloaded nodes are avoided is proposed by (Micah Sherr et al., 2009) and (Ian Goldberg et al., 2009). They introduced a flexible path selection design in which applications select a trade-off between performance and anonymity based on the user's specifc requirements.

A trust based routing methodology for onion networks that guards specifically against interference attacks has been presented in (Zhou et al., 2013). The problems in conventional routing methods have always been the fact that if the intruders have prior knowledge about the trust degrees present in the system, then anonymity becomes compromised. Hence the paper (Zhou, Peng, Xiapu et al., 2013) provides a trust degree based methodology, that helps defeat the interference attacks. A similar trust based approach that uses trust graphs is proposed in (Zhou, Peng et al., 2013). Due to the usage of relays at various geographical locations Tor circuit connection gets further prolonged and it is essential to analyse the factors influencing the performance of Tor network. The present works in Tor are focussed to improve the functionalities contributing to low latency anonymous browsing. Based on the derived bandwidth, relays are segregated to generate path for clients browsing to suit their requirement .

2.2 Fuzzy Set

Lotfi A. Zadeh developed the concept of ‘fuzzy sets’ in the year 1965. His study provided the basis for the Fuzzy sets theory (Zadeh et a.l., 1965). This theory helps understand and program numerous concepts used in human reasoning which are vague and imprecise e.g. tall, old. Fuzzy refers to managing uncertainty in complex systems. It is considered as an approximation theory that can represent uncertain data. The three major components of a Fuzzy logic include, Fuzzy sets, linguistic variables and possibility distributions. Fuzzy sets are characterised by its membership function. The membership values refer to the degree to which an object belongs to a fuzzy set. This value can range from 0 to 1. It is referred to as the membership value. The membership functions maps the elements of a universe of discourse to their corresponding membership values. In Fuzzy logic, a statement can be both true or false and also can be neither true nor false. Fuzzy logic is a non monotonic logic. Law of excluded middle does not hold true in fuzzy logic.

2.3 Evolutionary algorithms (EA)

Evolutionary algorithms are a class of meta heuristic techniques inspired by nature. The usage of these algorithms here is justified due to the intrinsic randomness associated with them. The EAs are characterised as population based, fitness oriented and variation driven. Genetic Algorithm (GA) is a special kind of stochastic search algorithm that depicts the biological evolution as the problem solving technique. GA works on the search space called population (Ian Goldberg et al., 1989). A working model for the Travelling Salesman Problem (TSP) was proposed, which provided efficient optimization. Genetic Algorithm is used in this proposed work due to its random nature, efficiency in giving optimal solution for complex problems. Genetic Algorithm (GA) is a guided random search technique to solve large-scale optimization problems and combinatorial problems. It works based on the principle of evolution and uses payoff function to guide the random search (T.N. Bhui et al., 1996)(Davis et al., 1991).

2.4 Genetic Algorithm : Basic steps

Genetic algorithms are class of evolutionary algorithm inspired by nature. Genetic algorithms(GA) are stochastic, population-based search and optimization algorithms iterative in nature. GA work with individuals also called chromosomes contributing to a possible solution to a given problem. Chromosomes require additional encoding/decoding steps to be defined in the algorithm.

Figure 1 depicts the basic steps involved in genetic algorithm process:

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

- **Representation** : The genetic algorithm work with individuals called chromosomes encoded as bit strings of fixed length which require encoding and decoding steps. In some applications encoding and decoding may be skipped that single solution is called the individual.

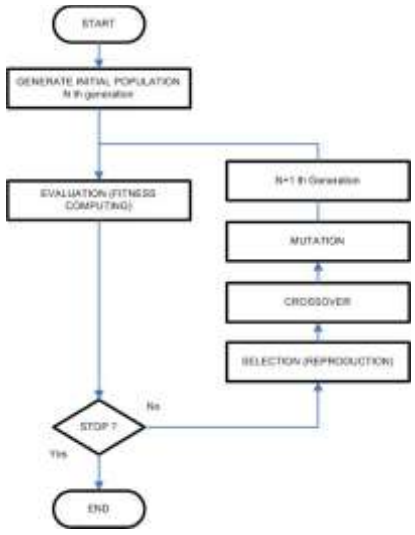

Fig. 1 Genetic Algorithm Process

- **Population and initialization** : Initial set of chromosomes called initial population is created. The size of initial population defines the optimality of the solution as local or global optimum to be found.
- **Evaluation and selection for reproduction :** Selection operator is used to identify the chromosomes used for reproduction to survive in the next generation.
- **Crossover :** Gene exchange, is done between two or more individuals and this variation operaTor is called crossover or recombination.
- **Mutation :** A mutation operator is used with intention to prevent getting stuck in the local optimum and increase a probability to find the global optimum. In the mutation operator, a new offspring is created from the single solution by changing some characteristics within it.
- **Termination :** Termination define states in which the evolutionary search terminates and the best individuals are presented. They usually vary accordingly to the type of applications.

3. Genetic Algorithm Based Enhanced Privacy Preserving Circuit Construction For Tor Network

In Tor network the three basic components are the entry nodes, exit nodes and other nodes. Entry nodes and Exit nodes are vulnerable points which contain crucial information. Hence the entry nodes are high performance, highly stable and secure nodes which determine the level of encryption that cannot be compromised easily. The exit nodes are the ones which strip off the final layer of encryption, hence they are also made reliable and secure because the last layer contains the identity of the original receiver which may break the receivers anonymity if compromised. The remaining nodes are intermediary nodes which strip off the encryption layers and forward the packet to the next node. Packets sent through a Tor network requires different levels of encryption based on the nature of application and their importance. The encryption level for a packet is determined by the application transmitting the packet (Gopinath et al., 2014). Higher level of encryptions are provided to throughput sensitive applications, while encryption levels are reduced as the transmission type comes down to delay sensitive applications.

In this work, a dynamic circuit construction algorithm is proposed considering the QoS requirements. Besides bandwidth it uses various QoS parameters like relay uptime, latency and delay to select a router dynamically. To choose less congested onion routers, round trip time(RTT) of the circuit is calculated during circuit construction phase. Tor network relies on speed and faster transmissions. Tor decentralises path selection in clients and they can choose the paths independently based on their requirements. Thus like ordinary routing, clients make decisions locally based on their view of the network. This may leave the network resources in a suboptimal state. Tor uses centralised servers to store the directory information to keep track of relays which can be downloaded by the clients and stored in their files. As Tor network imposes additional overhead of increased packet transmission and encryption a mechanism which can reduce transmission is evitable but problem arises when security is lowered as a trade off for time.

The objective of new path selection algorithm for multiple-path Tor circuit is to achieve the three key factors.

(i) Performance : The bandwidth provided to the Tor users by the new path selection algorithm must be equal or better than the bandwidth provided to the current path selection algorithm, at average.
(ii) Security : The probability to get compromised nodes must be lower than the probability with the current path selection algorithm.
(iii) Stability : If a node crashes in the circuit, the other handle the stream and the connection on the transport layer is not closed. (Rochet et al., 2011)

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

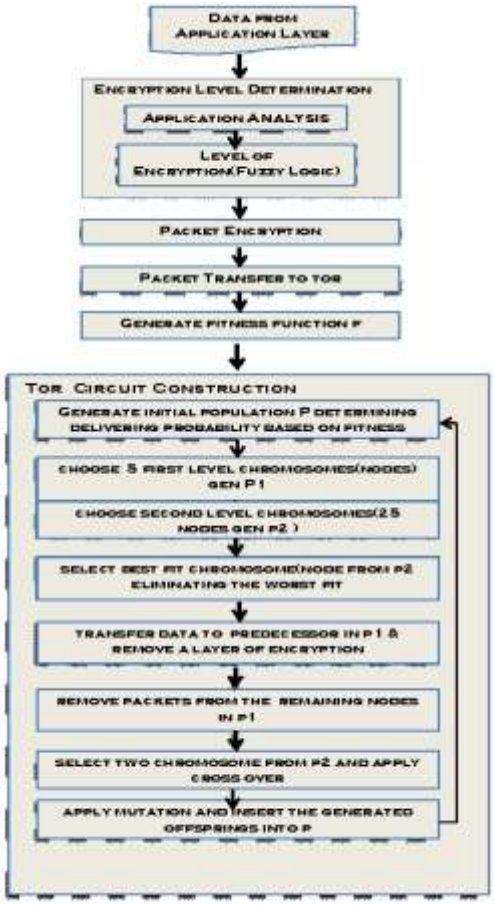

Fig. 2 System Architecture

Hence considering the QoS parameters related to transmission, while discovering a route provides an effective solution for determining the best route, and by incorporating the information about network traffic, the route that has been selected can be considered much more reliable. In this paper genetic algorithm is used to determine best route for QoS based Tor network at minimal cost by using the resources effectively and efficiently. This paper concentrates on the encryption levels based on the type of data transmitted from the application layer (Pingley, Aniket, et al., 2012). Some applications like browsing, video, and audio transfer, streaming etc. involve bulk amount of data to be transferred which demands a large bandwidth. While certain applications require faster transmission and can tolerate packet loss. The fuzzy based representation is used to identify the type of application to determine the levels of encryption.

In Fuzzy systems the membership values are indicated by a value in the range [0,1] with 0 for absolute falsity and 1 for absolute truth. Fuzzy sets are assumed to indicate probability, but even though they are somewhat similar, the membership grades are not probabilities. Probabilities on a finite universal set must add to 1 while there is no such requirement for membership grades. For our application, we have chosen the triangular membership function represented as,

$$\mu F(x,a,b,c) = \begin{cases} 0 \ if \ x < a \ (x-a)/(b-a) \\ (c-x)/(c-b) \ if \ b \le x \le c \\ 0 \ if \ c < x \end{cases}$$

It takes the form represented in the below figure 2

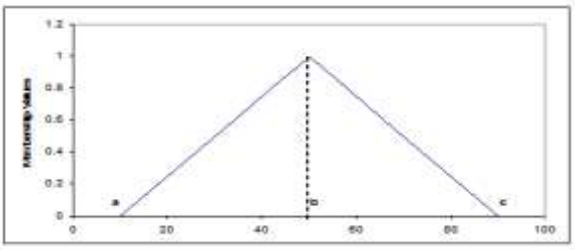

Fig. 3 Triangular Function

Based on the membership value the application data is categorised as throughput sensitive or delay sensitive and appropriate level of encryption is chosen. The proposed work uses Privacy Homomorphism (PH) as the encryption scheme (Jonker et al., 2013). Homomorphic cryptosystem is a cryptosystem with the additional property that there exists an efficient algorithm to compute an encryption of the sum or the product, of two messages given the public key and the encryptions of the messages but not the messages themselves.

The Privacy Homomorphism scheme allows users to encrypt a cipher text using algebraic operations and decrypt them using the reverse process. The advantage of using a PH scheme is that they are malleable. Let + denote an operation on a plain text and × be an operation on the cipher text, then *encr(m1)×encr(m2)=encr(m1+m2)*
This property of the PH scheme provides the best advantage for the proposed scheme with multiple encryptions.

Thus the Tor Client gets the application data fuzzify it to know the level of encryption to be applied on the base data. The base data packet is encrypted along with the destination address, in order to maintain the user's privacy. The process of encryption is carried out repeatedly depending on the *encrl* value. The packet is then forwarded to an entry node in the Tor network.

The process begins by categorizing the most stable and high performance nodes as the entry and exit nodes. The metrics used for node selection are delay, bandwidth, jitter, loss and node availability. Tor networks do not usually maintain the latency values, but the value of RTT is actively determined in the

IJASRM International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

Tor network using the method described in (Panchenko et al., 2009). As the data packet reaches the exit node, it is encapsulated by a single layer containing the identity of destination and data making the exit nodes more vulnerable. Thus usually exit routers in Tor deny connections to local host. To measure RTT a packet is sent to the network with local host as destination and an error is triggered in the exit node. Bandwidth in a router can be passively estimated by counting the user traffic that is transferred through these links. In the proposed work delay is calculated using the equation

$$D = \sum_{i=1}^{S-T} Di + \sum_{i-1,j=1}^{S-T} Li,j \quad \text{Eq. (1)}$$

where D_i is the end-to-end delay and $L_{i,j}$ is the link from node i to j. S is the start node and T is the Terminal node.

$$\text{Li,j} = \begin{cases} 1 \; if \; there \; exist \; a \; connectivity \; from \; i \\ 0 \quad otherwise \end{cases} \quad \text{Eq. (2)}$$

Eq. (2)

Node availability indicates the load on the current node based on the usage of resources. This helps in selecting less congested nodes. If a node is already in another connection it is omitted. Loss rate is the ratio of packets lost during the transmission from source to destination. Thus, the nodes are initially segregated based on delay, jitter, loss ratio and latency which act as the input to GA approach and bandwidth availability at the links for developing fitness functions to find an optimal path.

3.1 Tor Circuit Construction

The aim of using Genetic algorithm for circuit construction is, high diversity and high flexibility that prevents significant changes on computational load despite the change in network size(Chang Wook et al., 2002) Based on the type of application (throughput sensitive or delay sensitive) the QoS parameters are set.

Algorithm:

1. *Get the data from the application layer.*
2. *Get input parameters like population size, crossover probability, mutation probability, and number of generation for iterations*
3. *Analyse the application to distinguish whether throughput sensitive or delay sensitive using the QoS requirements to be met.*
4. *Fuzzify to determine the encryption level based on the type of application.*
5. *Encrypt the packet.*
6. *Generate fitness function f exclusively for throughput sensitive and delay sensitive applications respectively.*
7. *For each packet encountered perform the following*

i. Decrypt the packet to get $Packet_{new}$.

ii. Randomly choose usable set of routers(Population P) and apply appropriate fitness function to determine the delivering probability.

iii. Choose five first level chromosomes (routers)Using appropriate fitness function to form first generation population P.

iv. Send $Packet_{new}$ to all selected chromosomes(routers).

v. Wait for the acknowledgement or retransmit if necessary.

vi. Select 5 second level Chromosomes(nodes) from each node of the first level Chromosomes (nodes) to form second generation population P2.

vii. From the available 25 second level chromosomes(nodes) select a chromosome (node) with the highest delivering probability.

viii. Retain the best fit chromosome(node) from P2 and eliminate the worst fits .

ix. Find the predecessor node of bestfit. if it is in the Exploitation list(EL) repeat steps vii-viii for the next best fit else add the node to the EL.

x. Intimate remaining nodes connected to paths(chromosomes)in P1 to drop the packets.

xi. select two parent chromosomes(node) .

xii. perform one point crossover on the selected parents.

xiii. Apply Mutation on each offspring with a probability of 0.01.

xiv. Insert the offspring into the population P.

xv. Repeat the sequence of steps from 7 until the exit node is reached.

8 . If the exit node is reached, strip off the final layer of encryption and transmit packet to destination.

3.1.1 Representation Of Genes

Since selecting the secure and best route between the source and destination is our issue, representation of chromosomes is very important as its genes define the fitness of the route nodes. Munemoto's algorithm(Munemoto et al., 1998) uses variable length chromosomes which is feasible both for wired and wireless routes. (Inagaki et al., 1999) proposed an algorithm with fixed length chromosomes. The chromosomes in the algorithm are strings of real values which represent corresponding weights for QoS parameters. Each parameter is provided with weights ranging from -10 to 10. All the routers in the network work with the same parameter weights, hence this is a onetime process. The method of Analytic Hierarchy Processing (AHP) is used to perform weight assignments. This can be performed using pairwise comparison or using direct weight assignments(Saaty et al., 2008). In this paper the length of chromosomes is 5(bandwidth, uptime, delay, jitter, loss ratio) and fixed.

IJASRM International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

3.1.2 Population Initialization

Genetic algorithm starts with generating an initial population by random selection of the individuals named chromosomes that each encodes the solution of the problem. The chromosomes are evaluated based on the fitness function. Larger the fitness of the individual, better is the performance, more can satisfy the QoS parameters, and individuals will have higher probability to survive. Individuals with lower fitness is discarded. To start the evolution process it is required to initialize the population with popsize. Popsize specifies the number of chromosomes in a generation. If popsize is small there will be limited possibilities to perform crossover restricting the exploration level of search space. On the other hand a large population will slow down the performance of algorithm increasing the latency.

3.1.3 Designing fitness functions

The fitness function of each route in the Tor network depends on parameters like available bandwidth, end-to-end delay, Uptime of router node, Jitter and Loss ratio in such a way that it satisfies set of QoS requirements expected by the application. Each parameter is assigned with a weight. Weights differ for throughput and delay sensitive applications because a throughput sensitive traffic is tolerant towards delay in data delivery, while a delay sensitive application cannot tolerate delays. Further, throughput sensitive applications require data delivery without any packet loss, while delay sensitive applications can tolerate packet loss. These weighted parameters are combined into a single function, which is known as fitness function. In general a fitness function f may be of the form

$$f=w1B+w2D+w3UT+w4J+w5LR$$

where B-Bandwidth availability, D- End-to End Delay, UT- Uptime, J – Jitter, LR – Loss ratio and w1,w2….w5 are weights corresponding to the above parameters.

3.1.4 Selection

Selection operator plays a crucial role in the diversity of population. The selection operator aims to improve the quality of population yielding high quality chromosomes(reproduction) operator is intended to. Roulette wheel selection method is used which selects chromosomes based on their fitness values relative to the fitness of the other chromosomes in the population. The probability of kth chromosome is calculated as

$$pk = \frac{fk}{\sum_{i=1}^{n} fi} \quad \text{Eq. (3)}$$

3.1.5 Crossover and mutation:

Crossover also called recombination, is a genetic operator which generates new offspring by combining the genetic information of two parents. Crossover may be single point crossover, used to combine the genetic information of two parents to generate new offspring. Crossover are of various types single point, N-point, Uniform, Arithmetic, order, cycle etc. Mutation operation changes the genes of the selected chromosomes to maintain the genetic diversity of the population. Mutation contributes to the exploration of search space. Mutation can be done using bit flipping, insertion, inversion etc.

The choice of crossover rate and mutation rate greatly influence the behaviour and performance of GA. Higher crossover rate allows fast exploitation and smaller mutation rate controls exploration of individuals(off springs). The GA converges too fast when the mutation rate is too small. To get optimum solution a higher crossover rate and lower mutation rate is preferred. The proposed work uses single point crossover with crossover rate 0.7 (70%) and mutation rate 0.01(1%).

3.2 Implementation

The process of secure circuit construction is implemented using Java. The process is simulated with 40 nodes. The Implementation begins by initializing the population size, crossover probability and mutation probability. Data packet from the application layer is encrypted in various levels based on the type of application and transferred to the Tor network. Initial population P of chromosomes(nodes) are randomly created . Weights are directly assigned to the nodes in P based on each QoS parameter. Using weighted sum method, the importance value of each router is found. First generation population P1 is created by selecting five first level chromosomes from the fitted individuals in P. Packet is then forwarded to all the selected five first level chromosomes. For each first level node, 5 second level neighbour nodes are selected constituting second generation population P2 containing 25 nodes. From P2, a best node is chosen based on the fitness value. Predecessor of the best node is the first hop where the packet is decrypted and the remaining nodes in P1 discard the packets. Using roulette's wheel selection two chromosomes are randomly chosen and crossover and mutation is made according to the crossover probability and mutation probability. Using deterministic crowding new offsprings which compete with parents or close to parents are added to the population while the losers are discarded. The process continues for determining the next hop until the exit node is encountered ,

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

where the final layer of encryption is stripped off and packet is forwarded to the destination.

4. Results and Discussion

Analysis is carried out using 40 nodes. The algorithm is analysed using packets of size 50 KB and 100KB. Delay sensitivity and throughput sensitivity is measured for various levels of encryption. The success rate and failure rate of each node is measured to know the level of exploration and exploitation. The properties used for analysis in the current simulation are throughput, delay, latency and success rate of the router, failure rate of the router, network traffic, bandwidth, jitter and delay.

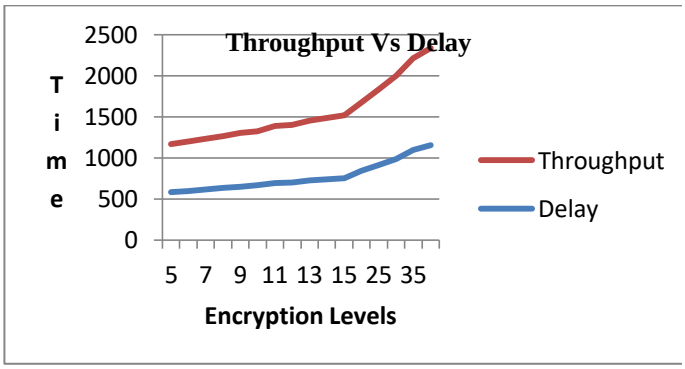

Fig. 4 Algorithm Analysis(50 KB)

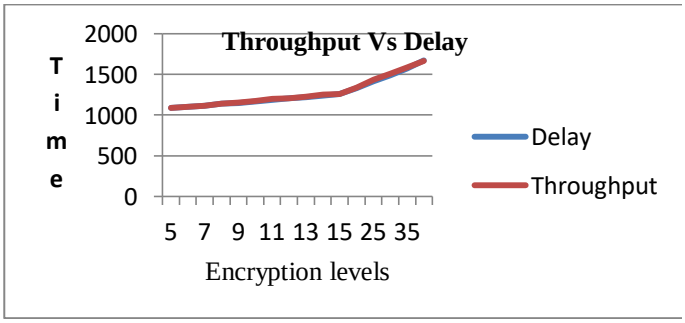

Fig. 5 Algorithm Analysis(100 KB)

Figures 4 and 5 shows the performance of the proposed work when files of sizes 50 KB and 100 KB are transmitted in the Tor network for throughput sensitive and delay sensitive traffic. Different encryption levels (5 to 40) were used and the time taken for transmission were analyzed. It can be inferred that there is a significant difference in transmission time taken to transfer throughput sensitive and delay sensitive file of size 50 KB. While to transfer file size of 100 KB in throughput sensitive and delay sensitive traffic, difference is not apparent. Hence the proposed algorithm shows better performance for both delay sensitive and throughput sensitive applications.

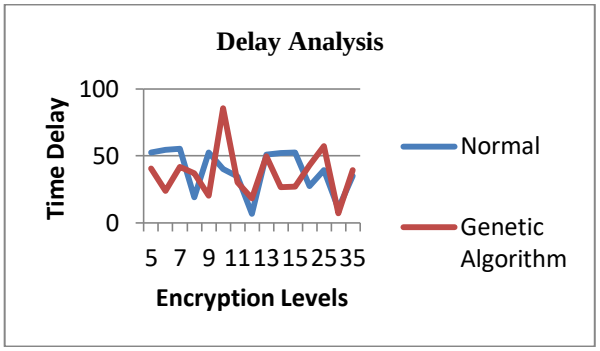

Fig. 6 Algorithm analysis(50 KB)

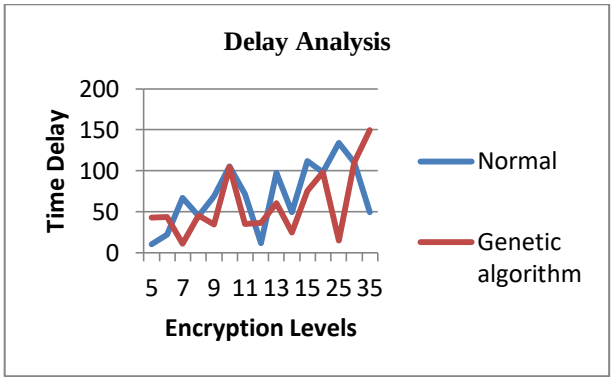

Fig. 7 Algorithm analysis(100 KB)

Figures 6 & 7 shows the comparative difference in travel time of transmitting 50 KB and 100 KB packets in delay sensitive traffic using Genetic algorithm and during Normal transmission. Encryption levels were set from 5 to 40 and time taken for the packet to travel from source to destination is measured. It is inferred that the proposed Genetic algorithm performs well on par with Normal transmission when packet of larger magnitude is transferred. Small variation is seen while transmitting 50 KB packet.

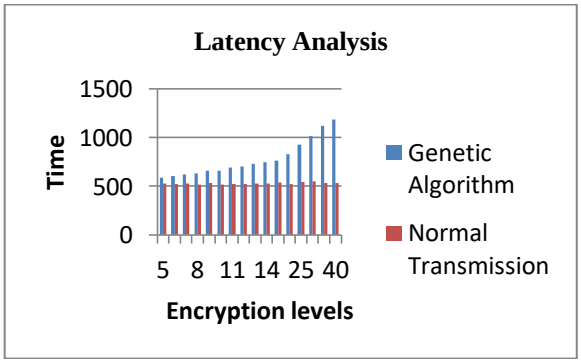

Fig. 8 Algorithm analysis(50 KB)

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 7, July 2018
www.ijasrm.com
ISSN 2455-6378

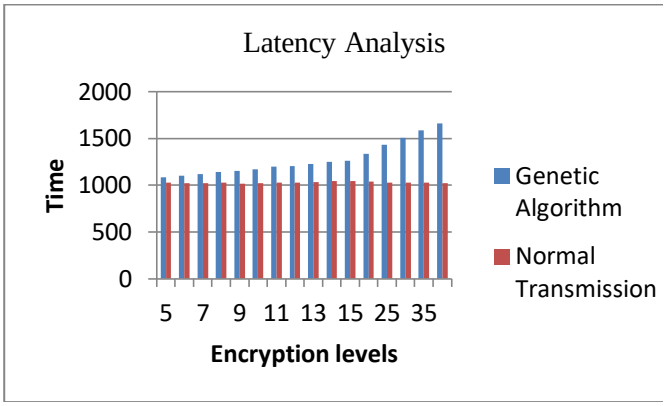

Fig. 9 Algorithm analysis(100 KB)

Figures 8 and 9 shows the latency observed during the transmission of packets of sizes 50KB and 100KB. During the initial phase while transmitting 50KB packet, latency observed was found to be high when compared to the normal transmission, but as the size of the data increases, the latency is found to be moderate and acceptable.

Every Tor node is made to maintain the success and failure rate of all its neighbours along with the bandwidth availability details. Network traffic is computed dynamically during transmission and delay is calculated using the difference between the packet sending and packet receiving time.

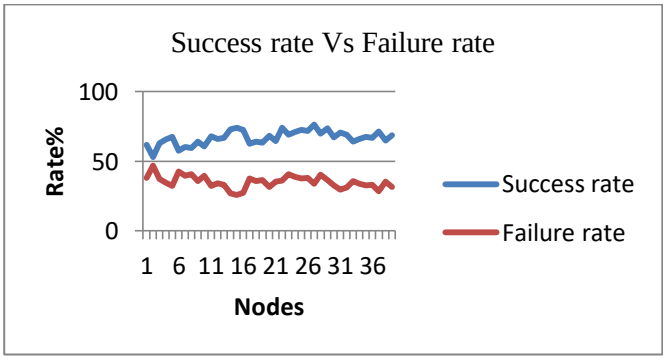

Fig. 10 Success rate and failure rate of nodes during transmission

The success and failure rates of the nodes in the network are depicted in Figure 10. As selection of a node needs evaluation using fitness function, only best fit nodes are considered and the worst fit nodes are discarded. Thus every attempt is recorded in the directory information of a router. It can be observed that the success rate graph is always above the failure graph, which proves that the deployed system performs efficiently in terms of selection, by maintaining low failure rates.

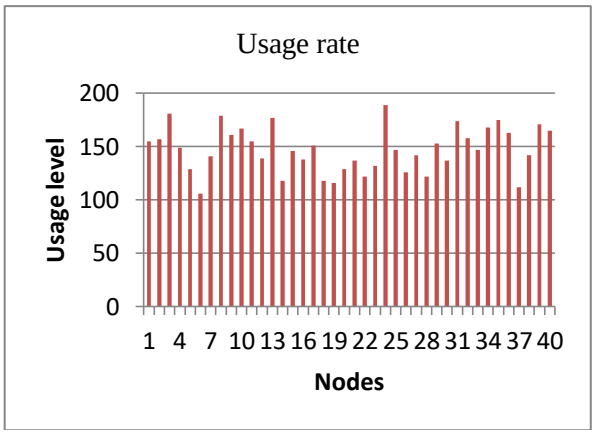

Fig. 11 Usage Rate

Due to the probabilistic nature and a demanding need to provide anonymity and preserve privacy it is essential to maintain the level of exploration and exploitation of nodes. Fig. 11 shows the number of times every node has been used for transmission. It is found that except a few set of nodes almost 85% of the nodes were used in a uniform manner. This is due to the fact that as the success rate of a node increases, the usage level of the node also increase.

The current proposal is found to be efficient in terms of lesser latency and high level security. In the usual process of route construction it is the responsibility of destination node to send acknowledgement to the intermediate nodes by sharing session key. The acknowledgement scheme is modified such that every intermediate node takes the responsibility of sending acknowledgement to its predecessor. Delivery becomes the responsibility of the new source and not the original source. Hence, the probability of a trace-back is almost impossible. Unlike the usual process of selecting a random node for the next level, this method selects five nodes for the next level based on fitness function thereby discarding the worst fit nodes. Randomness is added in node selection using Roulette wheel selection improving anonymity. Since this process is carried out till the packet reaches the exit node, tracing back is not possible, but reliable delivery is maintained.

The probabilistic nature of the routing mechanism has made the routes efficient, and also unpredictable. Since it is probabilistic and non deterministic, we have an advantage that any adversary who has access to the protocol could not exactly predict the next node.

5. Conclusion

The GA based circuit construction methodology proposed here is found to exhibit better security and efficient route construction that reduces latencies to a great extent. Further, the fuzzy based decision

making of the encryption level makes certain that the appropriate levels of encryptions are provided to the text. The process of selecting the nodes is based on the QoS parameters; hence reduction of latencies was observed to the maximum extent. As the selection process is carried out by every router node in the network, the complexity is kept to the minimum of O(n).

References

[1] Backes, Michael, I. Goldberg, A. Kate, and E. Mohammadi, 'Provably secure and practical onion routing", In Computer Security Foundations Symposium (CSF), IEEE 25th, pp: 369-385, 2012

[2] Chang Wook Ahn, and R. S. Ramakrishna, "A Genetic Algorithm for Shortest Path Routing Problem and the Sizing of Populations", IEEE Transactions on Evolutionary Computation, VOL. 6, December 2002.

[3] Davis, Handbook of Genetic Algorithms, Van Nostrand Reinhold, Chapter 6 & 7, 1991

[4] Dingledine, Roger, Nick Mathewson, and Paul Syverson, "Tor: The second-generation onion router", Naval Research Lab Washington DC, 2004

[5] Dingledine and N. Mathewson, "Anonymity loves company: Usability and the network effect," in Designing Security Systems That People Can Use. O'Reilly Media, 2007

[6] Goldberg E. David Genetic Algorithms in search, optimization and machine learning, Pearson Education, 1989

[7] Gopinath Ganapathy and S. Shakila, "Fuzzy Based Optimized Circuit Construction for Privacy Enhanced Onion Routing", European Journal of Scientific Research, Vol 123 Issue 2, ISSN 1450-216X/1450-202X, pp. 157-168, June 2014

[8] Ian Goldberg and Mikhail J. Atallah, Eds. vol. 5672 of Lecture Notes in Computer Science, pp. 73-93, Springer, 2009

[9] J. Inagaki, M. Haseyama, and H. Kitajima, "A genetic algorithm for determining multiple routes and its applications," in Proc. IEEE Int. Symp, Circuits and Systems, pp. 137–140, 1999.

[10] Jonker, Hugo, Sjouke Mauw, and Jun Pang. "Privacy and verifiability in voting systems: Methods, developments and trends." Computer Science Review 10 1-30, 2013

[11] L. Overlier, P. Syverson, Locating hidden servers, in Tor: Symposium on Security and Privacy, IEEE , pp. 15, 2006

[12] Li, Bingdong, Erdin, Esra, for anonymous routing," in Proceedings of Privacy Enhancing Technologies, 9th International Symposium, PETS 2009, Seattle, WA, USA, August 5-7, 2009

[16] Munemoto, Y. Takai, and Y. Sato, "A migration scheme for the genetic adaptive routing algorithm," in Proc. IEEE Int. Conf. Systems,Man, and Cybernetics, pp. 2774–2779, 1998.

[17] Owen and Michael, "Fun with onion routing Network Security", Elsevier Issue 4, pp: 8-12, April 2007.

[18] Panchenko, Andriy, and Johannes Renner, "Path selection metrics for performance improved onion routing Applications and the Internet" SAINT'09. Ninth Annual International Symposium on. IEEE, 2009.

[19] Pingley, Aniket, et al. "A context-aware scheme for privacy-preserving location-based services", Computer Networks, Vol 56, Issue 11, pp: 2551-2568, 2012

[20] Robin Snader and Nikita Borisov, Member, IEEE "Improving security and performance in the Tor network through tunable path selection", IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, Vol 8, 5, pp: 728-741, 2011

[21] Rochet, Universite Catholique deLouvain ICTEAM Crypto Group Louvain-la-Neuve, Belgium et all Moving Tor Circuits Towards Multiple-Path: Anonymity and Performance Considerations.

[22] Saaty, Thomas L, "Relative Measurement and its Generalization in Decision Making: Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors – The Analytic Hierarchy/Network Process". Review of the Royal Academy of Exact, Physical and Natural Sciences, Series A: Mathematics (RACSAM) 102 (2): pp: 251–318, June 2008.

[23] Shakila, Gopinath Ganapathy, "Privacy for Interactive Web Browsing: A study on Anonymous communication protocols", International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 5, pp: 270-280, 2014

[24] S. Shakila, Gopinath Ganapathy, "A Hybrid Privacy Preserving Algorithm based on Ants and Reinforcement Learning for Distributed and Adaptive routing in Tor Networks", Australian Journal of Basic and Applied Sciences, 9(20) , pp: 306-312, 2015

[25] T. N. Bui and B. R. Moon, "Genetic algorithm and graph partitioning," IEEE Transaction on Computers, Vol. 45(7), pp. 841-855, 1996.

[26] Zadeh, Lotfi A. "Fuzzy sets." Information and control 8.3 (1965): 338-353.

[27] Zhou, Peng, et al. "SGor: Trust graph based routing", Computer Networks 3522-3544, 2013.

[28] Zhou, Peng, Xiapu Luo, and Rocky KC Chang. "Inference attacks against trust-based onion routing: Trust degree to the rescue." Computers & Security 39.. 431-446, 2013.