Privacy Preserving Algorithm based on Big Bang-Big Crunch and Simulated Annealing method for Adaptive Routing in TOR Network

Dr. S. Shakila
[1]*Government Arts College, Tiruchirappalli - 22*
TamilNadu
(E-mail: shakilamuthusamy@gmail.com)

Abstract— Digital empowerment is a multi-phased process where every citizen is empowered through digital literacy and universal access to digital resources with the help of networking, communication and cooperation opportunities. There is a dearth of information privacy and security and thus anonymity in communications has become mandate for the people due to government policies. Onion routing is a solution where packets are rerouted through intermediate routers in an encrypted format. TOR is the onion routing network that provides anonymity during information access. This paper proposes a secure node selection method based on trust levels demanded by the user to improve anonymity. Clustering is applied to improve the efficiency of convergence. A modified form of Simulated Annealing with Big Bang-Big Crunch method is proposed for node selection in TOR. The proposed algorithm shortens the running time for node selection and improves the efficiency of TOR. Applying clustering improves the efficiency of convergence to obtain better solution with proper level of exploration and exploitation.

Keywords— *Onion Routing, TOR, Big Bang-Big Crunch, Simulated Annealing, Principle of Dominance*

I. INTRODUCTION

The digital world today, provides every civilian a bright prospect that transforms their lives with the help of technological innovations. 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[1]. 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 [2]. 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 [3][4].

The effectiveness of TOR network 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 browsing [2]. 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 [5]. 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 [6].

TOR is a free software, that was presented as a component of an anonymous routing project named; The Onion Router [4]. 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. 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, TOR exit node block, Bad apple attack, Sniper and heartbleed bug [7]. 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 is to be used whenever possible [8]. A comparison showing pros and cons of TOR networks is presented in [9]. 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 [10][11]. 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 [12]. 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 optimum circuit using Big Bang-Big Crunch and Simulated Annealing algorithms , Section IV presents the results and discusses them and Section V concludes the study.

II. REVIEW OF RELATED WORKS

The TOR network using self-reported bandwidth is replaced with an opportunistic bandwidth measurement mechanism which can accurately predict 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 [13]. This improved flexibility should make TOR palatable to a wider range of users, and thus increase anonymity for everyone due to a larger community [14]. 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 about such strategies is that they enable the evaluation of the features of current network, thereby potentially providing an easily deployable solution for improving the performance[15]. 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 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 and Syverson [16] presented new attack strategies to detect the location of hidden servers using only one TOR node. The authors proposed changes in route selection and relay selection to increase anonymity. The average duration of 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 [17] introduce a flexible path selection design in which applications select a trade-off between performance and anonymity based on the user's specifc requirements. Furthermore the notion of trust is employed in [18][19][20][21]. In addition [22] proposed a method for improving performance of TOR by changing its transport layer design. There are also a number of proposals with the goal of optimizing the path while considering the latency between relays.

A trust based routing methodology for onion networks that guards specifically against interference attacks has been presented in [23]. 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 provides a trust degree based methodology, that helps defeat the inference attacks. A similar trust based approach that uses trust graphs is proposed in [23].

This attack has been thwarted by [23] using restricted user knowledge. It uses three unique properties for performing routing namely; group trust is maintained, that verifies the trust levels assigned by users, an adaptive trust propagation system is maintained, which derives the global trust from the trust graphs and it works on a completely decentralized environment. 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 [24].

In the existing path selection algorithms trust over TOR nodes is not properly distributed by prioritizing few high-bandwidth nodes This leads to single points of failure to break the anonymity level guaranteed to achieve. Based on this insight [25] proposed an alternative path selection algorithm, coined DISTRIBUTOR, that maximally distributes the trust without decreasing the overall performance of the TOR network, thereby achieving significantly better anonymity while preserving TOR's throughput and latency.

III. PRIVACY PRESERVING ALGORITHM BASED ON BIG BANG-BIG CRUNCH AND SIMULATED ANNEALING METHOD FOR ADAPTIVE ROUTING IN TOR NETWORK

In this section first the relevant theoretical aspects of Big Bang-Big Crunch and Simulated Annealing is discussed followed by the proposed hybrid algorithm to find a secure and adaptive route in TOR network.

A. *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.

B. *Big Bang-Big Crunch Theory*

Big Bang-Big Crunch (BB-BC) is an optimization algorithm developed by Erol and Eksin [26]. This method involves low computational cost and a high convergence speed. It is a two stage method. In Big Bang stage, for the given optimization problem candidate solutions are generated randomly and they are spread over the search space. In Big Crunch stage the random points are shrunk to a single point using a convergence operator that has many inputs but one output named as centre of mass denoted as X_c which is calculated from the equation

$$Xc = \frac{\sum_{i=1}^{sp} \frac{1}{f(Xi)} Xi}{\sum_{i=1}^{sp} \frac{1}{f(Xi)}} \qquad (1)$$

Where Xi is the ith candidate in an D-dimensional search space, f(Xi) is a fitness function value of this point, sp is the population size in Big Bang. The populations produced by the Big Bang phase will be gradually reduced in Big Crunch phase in order to reduce computational time and have quick convergence. The best candidate solution in the population represents centre of mass which attracts other neighbor solutions.

C. *Simulated Annealing*

It is a probabilistic method proposed by [27] for finding the global minimum of a cost function. SA is becoming popular as it plays a vital role in the field of industrial chemistry, computer science and metallurgy. It successfully solves many complex optimization problems. SA is analogous to the process of physical annealing in which on heating a crystalline solid and allowing it to cool very slowly turns into crystalline lattice structure. The position of particles specify the state of the crystalline solid. Thus the transition to new state is accomplished by small movements of randomly chosen particles. The change in energy ΔE has a profound influence on acceptance or rejection of a state and is calculated using the Metropolis acceptance condition [28]. If $\Delta E < 0$ the new state is accepted and if $\Delta E > 0$ the new state is accepted with the probability

$$E = 1/(1 + e^{-\frac{\Delta E}{t}}) > \text{r} \qquad (2)$$

where t is the temperature or control parameter and r is a random number 0 ≤ r ≤1. The value of p is decreased as the algorithm progresses. Simulated annealing is one of the most popular meta heuristic algorithm to solve numerous combinatorial optimization problems. In metallurgy when a material is heated and cooled in a controlled manner it causes the atoms of the material to move from high energy to low energy increasing the size of crystals and reduce their defects.

The relationship between physical annealing and simulated annealing is, the states correspond to the feasible solution, energy is related to the cost of objective function and change of state defines the neighbouring solution. The annealing schedule is an important aspect in SA that provides an algorithmic approach for exploitation of connection(nodes). For successful annealing it is important to use a good annealing schedule which can be determined. The initial temperature T0 is set at a high value and it is decided how to decrease the temperature gradually as a function of time f(t). We start from an initial situation with 'energy level' f(0),and a small perturbation in the state of the system is brought moving the system to a new state with energy level f(1) . If f(1) < f(0), then the new state is accepted. If f(1) > f(0) then the change is accepted with a certain probability to escape from local minima.

D. System Architecture

Privacy Preserving Algorithm based on Simulated Annealing and Big Bang-Big Crunch algorithm for dynamic, distributed and Adaptive routing in TOR Networks presents a routing mechanism that is both robust and secure. Trust level is used as a base for node selection which is based on the network characteristics such as delay, jitter, throughput and vulnerability of the nodes. Simulated Annealing is used to find the optimal nodes for transmission.

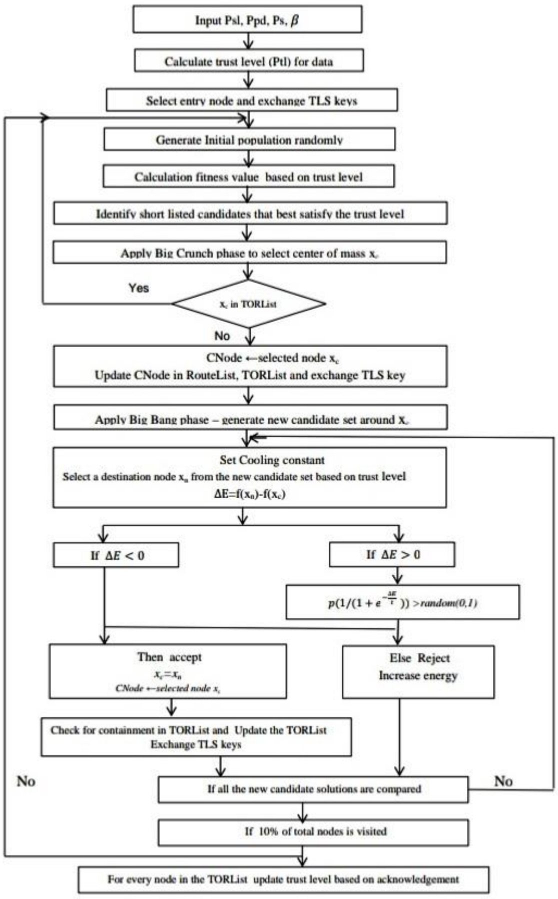
In TOR network nodes are of three categories entry nodes, exit nodes and relay nodes. The entry and exit nodes play a pivotal role in the TOR network by enhancing the security level while entering and leaving the TOR network which are the most vulnerable phases of the packet transmission. The process begins by using Big Bang –Big Crunch algorithm to regenerate new candidate solution based on required trust level reducing the size of population to form clusters and eliminating the worse solutions. Most stable and high performance nodes are selected as the entry and exit nodes. The best solutions of the previous generations are stored in the TORList which will be exploited as reference solutions in the next phase. At each iteration, the TORList is updated by replacing the worst solution cost in the current center of mass. If there are too many constraints involved, in order to enhance security and at the same time to improve throughput, Principle of dominance is used in node selection.

1) Principle of dominance:

The principle of dominance states that if one strategy of a player dominates over the other strategy in all conditions then the later strategy can be ignored. A strategy dominates over the other only if it is preferable over other in all conditions.

The principle of dominance in investing states that Among investments with the same rate of return, the one with the least risk is most desirable. In addition, given a group of investments with the same level of risk, the one with the highest return is most desirable.

Fig. 1 shows the workflow of the proposed system.



Input Psl, Ppd, Ps, β
Calculate trust level (Ptl) for data
Select entry node and exchange TLS keys
Generate Initial population randomly
Calculation fitness value based on trust level
Identify short listed candidates that best satisfy the trust level
Apply Big Crunch phase to select center of mass x_c
Yes
x_c in TORList
No
CNode ←selected node x_c
Update CNode in RouteList, TORList and exchange TLS key
Apply Big Bang phase – generate new candidate set around x_c
Set Cooling constant
Select a destination node x_n from the new candidate set based on trust level
ΔE=f(x_n)-f(x_c)
If ΔE < 0
If ΔE > 0
p(1/(1 + e^{-ΔE/T})) > random(0,1)
Then accept
x_c=x_n
CNode ←selected node x_c
Else Reject
Increase energy
Check for containment in TORList and Update the TORList
Exchange TLS keys
No
If all the new candidate solutions are compared
No
If 10% of total nodes is visited
For every node in the TORList update trust level based on acknowledgement

Fig. 1 Workflow of the proposed system

2) TOR circuit:

A packet transmission is initiated by selection of the entry nodes. Nodes with efficient performance, highly secure and reliable nodes are designated as entry nodes. Selection of entry nodes follows TLS key exchange between source node and entry nodes. After the entry node selection, TLS key exchange takes place between the source node and the entry node, which marks the entry of the packet into the TOR network and the entry node is set as the current node CNode. To avoid exploitation a Queue data structure TORList is maintained, which manages the nodes that have been recently visited. If the list is full, the next node is added by automatically popping out the first element.

The TORList is actually a modified Tabu List [29][30][31]. The TORList allows maintain the exploration and exploitation ability of the system, by constraining the system algorithm from selecting a node that has already been visited. Search is restricted to 10% of total nodes in the TOR

Network and size of TORList is set accordingly and it can be varied based on the exploration and exploitation ability.

3) Trust Requirement Calculation and TOR List Initialization:

QoS requirements in TOR network is application based which is concerned with the end-to-end quality of realtime applications such as voice and video. At the application layer QoS is concerned with protection level, integrity, speed, delay , packet loss and throughput based on human perception. The above characteristics are represented as Trust values which play a vital role in selection of TOR router nodes based on the requirement of the user. Every packet transmitted via TOR network is embedded with security level(Psl), Speed(Ps) and permissible delay(Ppd). These characteristics are scaled using pair wise comparison and form the basis for identifying the required trust levels(Ptl). The trust levels are encoded as a table in the routers. In this paper Trust values are assigned values from 1 to 5 with 1 indicating very low trust level, 2 indicating low trust level, 3 indicating nominal trust level, 4 indicating high trust level and 5 indicating very high trust level. To avoid exploitation as the network ages, trust levels are encoded to automatically increase /decrease constrained to a maximum of 5 and a minimum of 1.

4) Big Bang- Big Crunch Approach to support large population and clustering

In the Big Bang phase, some candidate solutions (nodes) to the optimization problem are randomly generated and distributed in the search space. In the Big Crunch phase the center of mass of candidate solutions is calculated. To explore best solutions in a large space the population produced by the Big Bang phase will be shrunk in Big Crunch phase to reduce computational cost and facilitate quick convergence. .After the Big Crunch phase, the Big Bang phase is repeated again and the new solutions(cluster of nodes) are generated using the Euclidean distance method

$$D(Oi,Oj) = \sqrt{\sum_{p=1}^{d} (Oi^{p} - Oj^{p})^{2}} \qquad \textbf{(3)}$$

Successively the Big Bang phase(*cluster*) and Big Crunch phase(contraction) are carried out repeatedly until a termination criteria has been met. The pseudo code is as follows

```
a) Big Bang phase (solutions construction):
    Step 1: Generate initial population
b) Big Crunch phase (Local Search move):
    Repeat
    Step2: Generate some neighbors for all solutions in the population and replace the parent with its best offspring for each solution in the population
    Step 3: Find the centre of mass
    Step 4: Apply local search to the centre of mass
    Step 5: Update the TORList with the best found solution
    Step 6: Eliminating some poor quality solutions
    Step 7: Return to Step 1 to generate new candidate solution centered around centre of mass if stopping criterion is not met
```

5) Exploration and Exploitation using Simulated Annealing:

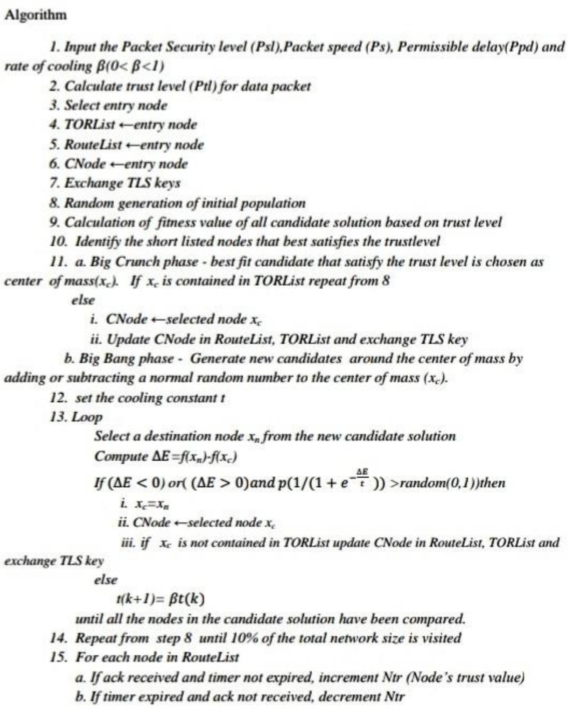
Simulated Annealing is a generic heuristic technique, applied to a combinatorial optimization problem by starting with a random solution in a solution space. Simulated Annealing(SA) is an iterative approach that generates global optimal solution whose convergence speed depends on neighborhood solution. This implies that to solve a problem transformation of solution is vital and in order to improve the convergence speed, the traditional simulated annealing method is combined with clustering using Big Bang theory. Big Bang algorithm is executed in parallel with SA. During each iteration it searches continuously for a better solution using the cooling schedule and acceptance probability. Every time when it finds a new solution it either accepts or rejects based on the fulfillment of objectives adopted by the problem. Acceptance probability is calculated using the formula $P = e^{-\Delta E/T}$, where ΔE is the change of energy which decides the move from the existing state to new state and T refers to the current temperature which is a cooling constant.

The value for the cooling constant is made to decrease periodically. Initially the cooling schedule is kept at a high value allowing the algorithm to explore new nodes and subsequently it narrows down to restrict and refine the searching process in further iterations [32]. A high temperature indicates that the route has been preferred mostly, hence we try to explore new nodes for convergence to global solution and avoid local minima. So, higher the cooling constant value, the larger is the solution space. As, the cooling constant keeps decreasing, the solution space shrinks gradually. If T0 is the initial temperature, a cooling schedule is

$$T(k+1)= \beta T(k) \qquad (4)$$

β is the cooling rate which controls the exploitation level and depends upon the application requirement. If higher trust level is required then β assumes a very low value.

To accelerate the convergence of simulated annealing to the global optima a cybernetic optimization by simulated annealing as a method of parallel processing method is proposed by[33].To deal with clustered data simulated annealing is combined with K-means clustering by performing clustering in the first stage and applying simulated annealing in the next stage[34] . Here Big Bang-Big Crunch is used for clustering which is executed in parallel with Simulated annealing .



```
Algorithm
    1. Input the Packet Security level (Psl),Packet speed (Ps), Permissible delay(Ppd) and rate of cooling $\beta(0< \beta<1)$
    2. Calculate trust level (Ptl) for data packet
    3. Select entry node
    4. TORList ←entry node
    5. RouteList ←entry node
    6. CNode ←entry node
    7. Exchange TLS keys
    8. Random generation of initial population
    9. Calculation of fitness value of all candidate solution based on trust level
    10. Identify the short listed nodes that best satisfies the trustlevel
    11. a. Big Crunch phase - best fit candidate that satisfy the trust level is chosen as center of mass($x_c$). If $x_c$ is contained in TORList repeat from 8
          else
            i. CNode ←selected node $x_c$
            ii. Update CNode in RouteList, TORList and exchange TLS key
        b. Big Bang phase - Generate new candidates around the center of mass by adding or subtracting a normal random number to the center of mass ($x_c$).
    12. set the cooling constant t
    13. Loop
            Select a destination node $x_n$ from the new candidate solution
            Compute $\Delta E=f(x_n)-f(x_c)$
            If $(\Delta E < 0)$ or( $(\Delta E > 0)$and $p(1/(1+e^{-\frac{\Delta E}{t}}))$ >random(0,1))then
                i. $x_c=x_n$
                ii. CNode ←selected node $x_c$
                iii. if $x_c$ is not contained in TORList update CNode in RouteList, TORList and exchange TLS key
            else
                $t(k+1)= \beta t(k)$
        until all the nodes in the candidate solution have been compared.
    14. Repeat from step 8 until 10% of the total network size is visited
    15. For each node in RouteList
        a. If ack received and timer not expired, increment Ntr (Node's trust value)
        b. If timer expired and ack not received, decrement Ntr
```

Simulated Annealing randomly selects best next neighbour around center of mass from the new candidate solution. Best neighbouring node is selected by computing the difference in objective function values. The best node is checked for containment in the TORList. If it is absent it is designated as the next node else the process is repeated for the remaining nodes. The recently selected node is the best node and is moved to the first position in the TORList. The selected node is set as the new center of mass and TLS key exchange takes place between the source and the selected node.

This process is repeated until 10% of the network nodes have been reached. In this study, in order to provide high security, 10% of the total nodes in the network is considered as the node access limit and results are recorded.

6) *Acknowledgement based Trust update:*

After every transmission, a timer is set for receiving the acknowledgement. If the timer expires and the acknowledgement has not yet been received, then the trust value of the node is decremented. If the acknowledgement arrives before the timer expiry, the trust value is incremented. The increment and decrement levels are minor, so as to maintain the system from incorporating a huge bias in a single transaction. Further, boundary values are set for the trust levels (1-5) and the trust values are not allowed to reduce beyond these boundaries.

IV RESULTS AND DISCUSSION

Experiments were conducted on a network containing 500 nodes and 50 transmissions were conducted, each with varying number of packets. The initial temperature facTOR is set as 10000, the acceptance probability is set as 0.8, the optimal value of β is set as 0.9 or it can take any value from 0.6 to 0.99 based on required speed , number of generations is 200 and 10% of the total nodes were considered for the routing process irrespective of the transmissions. Required Packet Security level (Psl) , required speed (P_s), permissible delay(Ppd) and are provided by the application and these values contribute to the trust level, which determines the final path to be taken during transmission. The proposed algorithm is implemented in Java and executed for range of nodes from 10 to 50. The test inputs were tested on normal Simulated annealing and improved Simulated annealing methods. The inference is represented graphically.

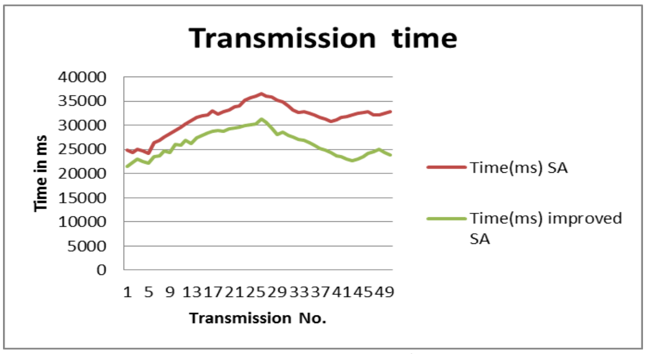


Fig. 2 Time Taken for transmissions

Figure 2 shows the time taken by normal SA method and improved SA for selection of nodes during each of the transmissions. It can be observed that the time taken by the normal SA method ranges from 25000ms to 36000ms and improved SA takes the time range between 24000 and 31000ms. There is a variation of 4 to 5 sec. Thus the improved SA proves to be fast in selecting nodes.

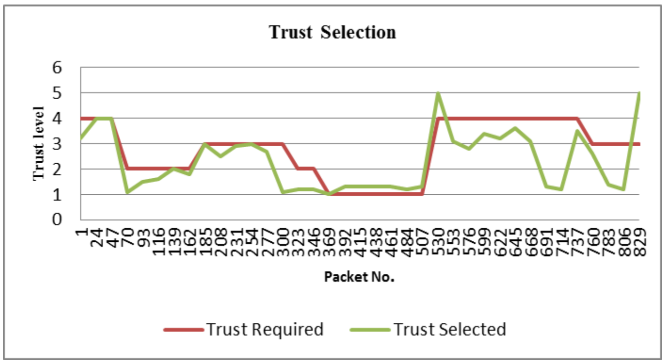


Fig. 3 Trust level selection

Figure 3 shows the trust requirement for each packet being transmitted and the trust provided for each transmission. It can be observed that most of the graph lines overlap, which shows that the algorithm functions in an effective manner in redirecting the traffic to the nodes with appropriate trust levels.

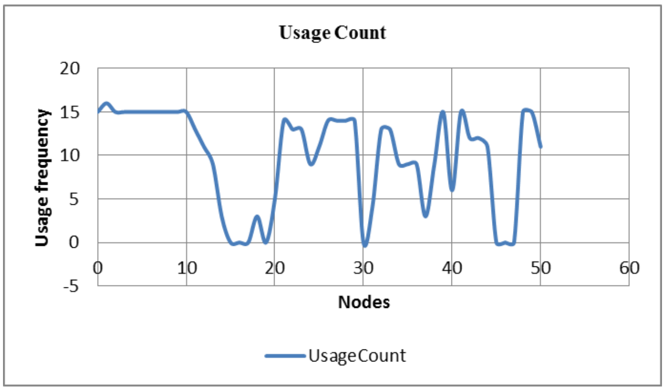


Fig. 4 Usage level

Figure 4 shows the level of usage of nodes. The graph, maps the nodes and their usage levels. Spikes are seen in entry and exit nodes, while the other nodes show uniform usage. This infers that the distribution of traffic is uniform among all the other routing nodes.

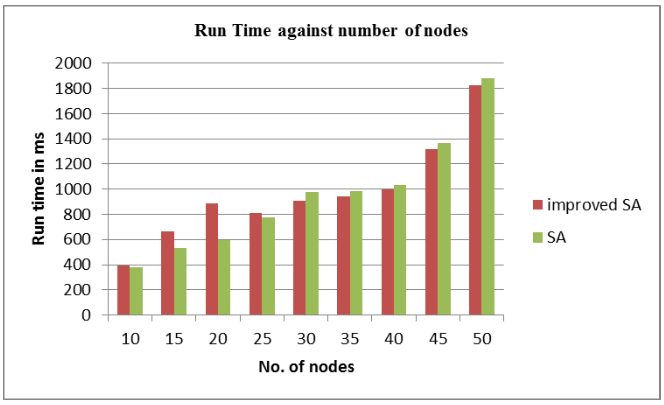


Fig. 5 Runtime of the improved SA for range of nodes

Figure 5 shows the bar chart representing the time complexity with increasing number of nodes. When 20 nodes are selected for transmission the improved SA yields better result when compared to normal SA. For fifty nodes and above both the algorithms take a lot of time to complete (about 18000ms or more). In Simulated Annealing increase in number of nodes increases the runtime. However it can generate optimal solution for large size population with the help of Big Bang method. The runtime may also be affected by varying the set of parameters used.

V. CONCLUSION

This paper presents a route selection algorithm for a TOR network based on Big Bang-Big Crunch and Simulated Annealing meta heuristic method. It works as an improved form of normal Simulated annealing method using Big Bang method for clustering. It works as an enhancement to reduce the time taken for node selection. Further, the speed required for transmission is provided by the application during the packet initiation itself, hence the node selection is based totally on the parameters specified, which makes the transmission effective and fast. The probabilistic behavior of the algorithm establishes a secure and unpredictable route for every transmission. The algorithm may be improved further by adding fuzziness to speed and trust values to add more accuracy.

REFERENCES

[1] S. Shakila, Gopinath Ganapathy, 'Privacy for interactive web browsing: A study on Anonymous communication protocols," International Journal of Advanced Research in Computer Science and Management Studies, vol 2, Issue 5, May 2014, pp. 270-280

[2] S. Shakila, Gopinath Ganapathy,"A Hybrid privacy preserving algorithm based on ants and reinforcement learning for distributed and adaptive routing in TOR network," Australian Journal of basic and applied science, June 2015, pp. 306-312.

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

[4] Li, Bingdong, Erdin, Esra, Güneş, Hadi, M. Bebis, George, Shipley and Todd, 2011. An Analysis of Anonymity Usage". Springer Berlin Heidelberg.

[5] Dingledine, Roger, Nick Mathewson, and Paul Syverson. TOR: The second-generation onion router. Naval Research Lab Washington DC, 2004.

[6] Owen, Michael. "Fun with onion routing." Network Security 2007.4 (2007): 8-12.

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

[8] https://www.ibm.com/developerworks/community

[9] Liška, Tomáš, T. Sochor and H. Sochorová, 2010. Comparison between normal and TOR-anonymized web client traffic. Procedia-Social and Behavioral Sciences, 9: 542-546.

[10] Milajerdi, S., Momeni and M. Kharrazi, 2015. A Composite-Metric Based Path Selection Technique for the TOR Anonymity Network. Journal of Systems and Software.

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

[12] Gopinath Ganapathy, S. Shakila ,"A survey on anonymity based solutions for privacy issues in web mining", International Journal of Computational Intelligence and Information Security, January 2014 Vol. 5, No. 1 ISSN: 1837-7823

[13] Robin Snader and Nikita Borisov , "Improved security and performance in the TOR network through tunable path selection", IEEE Transactions on dependable and secure computing, 2012

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

[15] Fallon Chen, Joseph Pasquelle , "Towards improving path selection in TOR", https://cseweb.ucsd.edu/~pasquale/Research/Papers/globecom10c.pdf

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

[17] Micah Sherr, Matt Blaze, and Boon Thau Loo, \Scalable linkbased relay selection for anonymous routing," in Proceedings of Privacy Enhancing Technologies, 9th International Symposium, PETS 2009, Seattle, WA, USA, August 5-7, 2009

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

[19] A. Johnson and P. Syverson, \More anonymous onion routing through trust," in Computer Security Foundations Symposium,2009. CSF'09. 22nd IEEE. IEEE, 2009, pp. 3-12.

[20] V. Sassone, S. Hamadou, and M. Yang, \Trust in anonymity networks," CONCUR 2010-Concurrency Theory, pp. 48-70, 2010.

[21] Anupam Das, Nikita Borisov, Prateek Mittal, and Matthew Caesar, \Re3: Relay reliability reputation for anonymity systems," in Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security (ASIACCS 2014), June 2014.

[22] Mashael AlSabah and Ian Goldberg, \PCTCP: Per-Circuit TCP-over-IPsec Transport for Anonymous Communication Overlay Networks," in Proceedings of the 20th ACM conference on Computer and Communications Security (CCS 2013), November 2013

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

[24] K. Kiran, B. Vignesh , PD Shenoy, "Client requirement based path selection algorithm for Tor network", 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) IEEE 2017]

[25] Michael Backes, Aniket Kate, Sebastian Meiser,"Monitoring the Anonymity of Tor's path selection", In Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS '14), ACM, 2014.October 2014

[26] K. Erol Osman, Ibrahim Eksin, "New optimization method : Big Bang-Big Crunch", Elsevier, Advances in Engineering Software, Vol. 37, No. 2, pp. 106–111, Febraury 2006.

[27] S. Kirkpatrick, C. J. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," Science, vol. 220, pp. 671– 680, 1983.)

[28] Metropolis, N, et al. Equation of State Calculations by Fast Computing Machines. Florida State University. [Online] 1953. www.csit.fsu.edu/~beerli/mcmc/metropolis-et-al-1953.pdf)

[29] Salhi and Said, 2002. Defining tabu list size and aspiration criterion within tabu search methods. Computers & Operations Research, 29(1): 67-86.

[30] Glover, F., 1989. Tabu Search - Part 1. ORSA Journal on Computing 1 (2): 190–206. doi:10.1287/ijoc.1.3.190. [31]

[31] Glover, F., 1990. Tabu Search - Part 2. ORSA Journal on Computing, 2(1): 4–32. doi:10.1287/ijoc.2.1.4

[32] Hajek, Bruce. "Cooling schedules for optimal annealing." Mathematics of operations research 13.2 (1988):311-329

[33] Fleischer, M.A. and Jacobson, S.H. (1996) Cybernetic optimization by simulated annealing: An implementation of parallel processing using probabilistic feedback control, pp. 249–264.

[34] Wayan Firdaus Mahmudy,"Improved simulated annealing for optimization of vehicle routing problem with time windows (VRPTW), Kursor Journal, Department of Computer Science, University of Brawijaya (UB) Vol. 7, No. 3, October 2014 ISSN 0216 – 0544

Author Profile:

Dr. S. Shakila is the Assistant Professor in the Department of Computer Science, Government Arts College, Tiruchirappalli, TamilNadu, India. She obtained her Bachelors degree and Masters degree from Bharathidasan University, Tiruchirappalli, India in 1988 and 1990 respectively. She also did her Master's Degree in Engineering from Anna University, Chennai, India. Received Master of Philosophy in Computer Science from Bharathidasan University, Tiruchirappalli, India. Successfully completed her Ph.D programme in Bharathidasan University, India in the area of Information Security and privacy under the supervision of **Dr. Gopinath Ganapathy,** Registrar, Bharathidasan University. Her Research Interest includes Information Security, Distributed Computing, Grid Computing and Embedded Systems.