# Node clone detection using a stable overlay network

**Balika J. Chelliah, M. S. Antony Vigil, M. S. Bennet Praba**
Department of Computer Science & Engineering, SRM Institute of Science and Technology, Chennai, India

**Article Info**

***Article history:***

ReceivedApr20, 2019
Revised Aug16, 2019
Accepted Aug29, 2019

***Keywords:***

Clone detection
Distributed detection
Overlay network

**ABSTRACT**

Wireless sensor networks consist of number of sensor nodes widely distributed in particular region to communicate and sharing the environmental information and also these data's are stored in central location for further data prediction. Such nodes are susceptible to cloning attack where the adversary captures a node, replicates with the same identity as that of the captured node and deploys the clone back into the network, causing severe harm to the network. Hence to thwart such attacks, a distributed detection protocol is used with initiator-observer-inspector roles assigned randomly for the nodes to witness the clone and thereby broadcast the evidence through a balanced overlay network. Use of such balanced network provides high security level and reduces the communication cost when compared to other overlay networks with a reasonably less storage consumption.

***Corresponding Author:***

Balika J. Chelliah,
Department of Computer Science and Engineering,
SRM Institute of Science and Technology,
Chennai, India.
Email: bjchelliah@outlook.com

## 1. INTRODUCTION

Wireless Sensor Networks (WSNs) is a network of spatially distributed sensors that monitor physical and environmental conditions and collectively pass the data to a centre location through the network [1, 2]. Such networks bring in a vast improvement in a wide spectrum of technologies ranging from health care to surveillance in military. Each wireless network can scale from tens to hundreds of nodes and seamlessly integrate with existing wired measurement and control systems. These networks are deployed in hostile environments where the opponents are equally intelligent to cause various types of attacks like routing attacks, denial of Service attacks, node subversion, node malfunction, node replication, node outage, passive information gathering etc. In a node replication attack, an adversary captures the identity of an existing node, replicates the node and places the replica back into the sensor network. This replicated node can severely disrupt the network's performance by mis-routing packets, dropping the packets leading to distorted sensor readings.

Wireless sensor networks can be static or dynamic. To implement a network service that is not originally available in the existing network, an overlay network is used. An overlay network is a virtual network of nodes and logical links built on top of physical network to effectively maintain and distribute among the nodes [3]. Internet is the basis for many overlaid networks that can be constructed in order to allow routing of messages to destinations that are not specified by an IP address. Distributed Hash Tables (DHTs) can be used to route messages to a node having a specific logical address, whose IP address is not known in advance. A DHT [4] is a decentralized key-based caching and checking system for key- value pairs. Any participating node can effectively retrieve a value for a given key. They can scale up to a plethora of nodes and are strongly unaffected by node joins, departures and failures. Several DHT mechanisms like Chord, CAN, Tapestry are rapidly becoming popular.

For any key k, each node either has a node ID that owns k or has a link to a node whose node ID is closer to k, in terms of the keyspace distance defined. At each step, forward the message to the neighbor whose ID is closest to k. When there is no such neighbor, then we must have arrived at the closest node, which is the owner of k. This style of routing is referred to as key-based routing [5] followed in distributed hash tables. It must be guaranteed that the maximum number of hops in any route is low, so that requests complete quickly.

Many distributed hash table mechanisms can be employed to catch the replicated nodes effectively. Such systems incur a significantly high communication cost in some scenarios. Chord lookup protocol is based on a clockwise lookup, hence causes a high delay when communicating with nodes in anti-clockwise direction. The average path length between two random node varies from to O(log n) to O(√n ) dependent on underlying sensor networks and the average chord hop length is log n. Hence the distributed hash table employed on a chord ring leads to high communication cost ranging from $\log^2 n$ to √n log n. Chord must update the routing information when a node joins or leaves the network; a join or leave requires $\log^2 n$ messages. Chord lacks a cache mechanism to preserve useful information for future session establishment.

To overcome the above drawbacks, a balanced overlay network becomes indispensable. A balanced network is self-adjusting to data skew and is height balanced. Because the height is balanced, the routing information is maintained both horizontally and vertically for effective search. We make use of a distributed binary balanced tree structure where each node has links to its parent, children, adjacent nodes and selected neighbor nodes at the same level maintained by a left and right routing table. This mechanism alleviates the delay incurred if the node is located anti-clockwise by maintaining both left and right routing tables. The average path length in such bigreedy search is $\log n / 2 - \sqrt{\log(n / 2\pi)} + 1$ which is way less than the traditional chord systems. Maintenance of a cache in every node can bring down the communication cost to as low as O(1), so that the next time a source knows the destination, it directly retrieves the destination's IP address. Node join and leave operations always take only log n steps as opposed to $\log^2 n$ in Chord.

## 2. RELATED WORK

### 2.1. Centralized techniques

The detection of node replication attack in static wireless sensor networks are categorized mainly into two types, centralized and distributed techniques. In centralized techniques, base station is considered to be a powerful center which is responsible for collating information and decision making. In this process of detection every node in the network sends its location claim (ID, Location Information) to a base station through its neighboring nodes. Upon receiving the entire location claims, the base station checks the node Identities along with their location and time, and if it finds two different locations with the same ID, it alarms about the clone node.

Random Key Predistribution [6, 7] is a centralized technique where the keys employed by this scheme should follow a certain pattern and those keys whose usage exceeds a defined threshold can be adjudged to have been cloned. SET [8] is yet another technique in which the network is randomly divided into mutually exclusive subsets. There is a designated subset leader in each of the subsets, and the subset members are one hop away from their subset leader. Each subset leader collates its members information and forwards it to the root of the subtree. On every root of the subtree a SET intersection operation is executed to detect the cloned nodes.

Albeit the existence of several centralized techniques, they do not go well with applications that may not use base stations. They may be exposed to a single point failure where the vandalization of the base station will result in the failure of the entire network. Also, the nodes near the base station get exhausted easily because the number of messages that pass through them will be drastically high.

### 2.2. Distributed techniques

In distributed techniques, there is no central authority and claimer-reporter-witness mechanism is provided in which the detection is performed by locally distributed node sending the location claim not to the base station but to a randomly selected node called witness node. Deterministic Multicast protocol [9] is a claimer-reporter-witness framework. The claimer node locally broadcasts its location claim to its neighbors, each neighbor serving as a reporter, and employing a function to map the claimer ID to a witness. Then the neighbor node forwards the claim to the witness, which can possibly receive two different location claims for the same node ID if the adversary has cloned a node. Randomized Multicast (RM) [10, 11] scatters the location claims to a randomly selected set of witness nodes. Line-Selected Multicast (LSM) [12], exploits the routing topology of the network to select witnesses for a node location and utilizes geometric probability to detect replicated nodes.

Randomized, Efficient, and Distributed protocol [13] called RED is executed at fixed intervals of time and occurs in two steps. In first step, a random value is shared between all the nodes through base station. In the second step, i.e detection phase, each node broadcasts its claim (ID and location) to its neighboring nodes. Every node in the path (from claiming node to the witness destination) forwards the message to its neighbor which is nearest to the destination. Hence, the replicated nodes will be detected in each detection phase.

In two other distributed protocols called Single Deterministic Cell (SDC) [14] and Parallel Multiple Probabilistic Cells (P-MPC) the entire sensor network is divided into grid cells to form a geographic grid. Such distributed protocols incur a high communication overhead, while having an implication that every node must be aware of all other nodes with an adequate geographic knowledge. Distributed Hash tables can also be employed over a chord overlay network as suggested by Zhijun Li *et al*., [15] where the communication cost would be as high as log n to n log n because of the chord ring structure though it provides high security and performance. Chord ring also has several shortcomings which we try to alleviate by a balanced overlay structure.

## 3. SYSTEM MODEL

The system model makes use of a decentralized key-based distributed dectection protocol that operates over a balanced tree overlay network.

### 3.1. Distributed detection protocol

A key based caching [16] and checking system is used which sends a message along with a key over the network to a destination that is entirely decided based on the key. The key is generated based on a random seed sent by a trusted initiator. Each node A is assigned a public and private key, where $K_a$ is the public key, which identifies the node uniquely, and could be the IP address or MAC address of the node, $K_a^{-1}$ is the private key assigned to a node by a trusted third party. Clone is detected by the occurrence of nodes with same identity, but located at a reasonably distant place at a designated time. As shown in Figure 1, the model primarily involves an initiator which is a trusted node that starts a round of detection by broadcasting an action message to all other nodes in the network. Every node on reception of the action message, stores the random seed that appeared in the message and becomes an observer to generate a claiming message for all of its neighbors. A claiming message is the message that provides the identity and the location of a neighbor node which is the examinee of the observer. The observer sends this claiming message over the overlay network attached with a key, which is the hash of the identity of the examinee concatenated with the random seed [17].

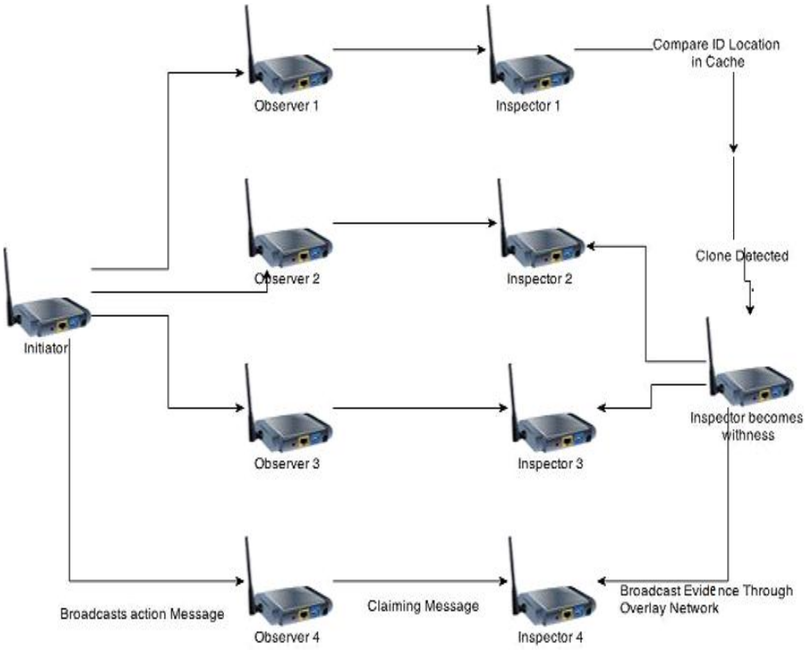


Figure 1. A round of detection

The destination to which this message must be sent is decided based on the key valued attached. Once the message reaches the destination, the destination node maintains a cache that stores the {Id, Location} of the examinee, where Id is the identity of the examinee and Location is the location of the examinee that is determined by some secure localization protocol. As the number of detection rounds increases, if the destination node, which is the inspector of the claiming message finds in its cache that for the same identity $Id_a$ there exist two different locations $Loc_a$ and $Loc_a$' then the inspector becomes a witness of the clone and broadcasts an evidence containing the claiming messages to the entire network. All the valid nodes verify this evidence message and stop communicating with the replicated node. Each node also maintains a revocation list of the compromised nodes' identities.

### 3.2. Balanced overlay network

A balanced overlay network [18-20] is constructed with the nodes in the network such that at each level L, nodes are numbered from 1 to $2^L$. Thus the level and the number determine the location of the node in the binary tree. Each node has links to its parent, children, adjacent nodes and selected neighbor nodes at the same level. Links to the selected neighbor nodes are preserved in a left and right routing table. Each of these two routing tables contains links to the nodes at the same level with numbers that are less or greater than the number of the source node by a power of 2.

Definition 1: A routing table is considered full if all of the valid links are not null.

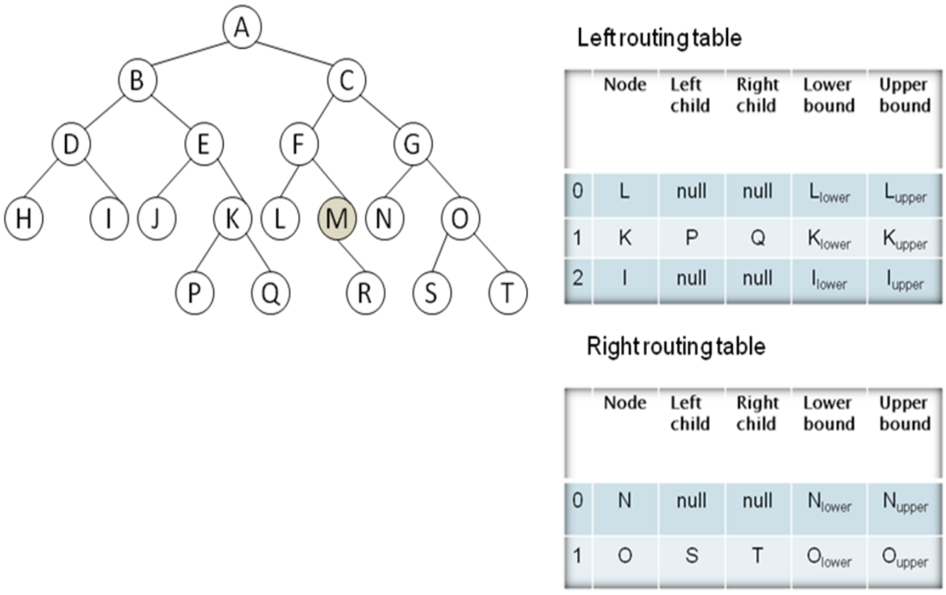
Definition 2: A tree is a balanced tree if every node in the tree that has a child, also has both its left and right routing tables full.

Definition 3: The traversal of the tree is inorder, where the left subtree is traversed recursively followed by the root and then the right subtree.

Given a node x, the node immediately prior to it in traversal is the left adjacent to it and the node immediately after x is right adjacent to it.

In Figure 2, for the node M, left routing table contains 3 entries which are at the $2^i$ positions from the node, where i>=0 till the no. of nodes available in the level. L is at $2^0$, node K is at $2^1$ position and node I is at $2^2$. The left and the right child of L are null and node's lower and upper bound values are recorded. The left and right child of K are not null, K has left child as P and right child as Q. The node I also doesn't have children and hence both left and right child values are null.

In the same way the right routing table is also constructed. In the right routing table, we have two entries, one for node Nwhich is at the $2^0$ position from M and the other for node O which is at $2^1$ position. The left and right child of N are both null, while the left and right child of O are S and T respectively. The parent node of M is F, left child is null, right child is R, left adjacent node is F and right adjacent nodeis R.



Left routing table

| | Node | Left child | Right child | Lower bound | Upper bound |
|---|---|---|---|---|---|
| 0 | L | null | null | $L_{lower}$ | $L_{upper}$ |
| 1 | K | P | Q | $K_{lower}$ | $K_{upper}$ |
| 2 | I | null | null | $I_{lower}$ | $I_{upper}$ |

Right routing table

| | Node | Left child | Right child | Lower bound | Upper bound |
|---|---|---|---|---|---|
| 0 | N | null | null | $N_{lower}$ | $N_{upper}$ |
| 1 | O | S | T | $O_{lower}$ | $O_{upper}$ |

Figure 2. A binary balanced tree

## 4. PROTOCOL DETAILS

### 4.1. Generation of action message and key

Initiator uses a broadcast scheme to release an action message as in (1) including a monotonously increasing nonce, a random round seed, and an action time.

$$M_{act} = \text{nonce, seed, time, } \{ \text{ nonce} \parallel \text{seed} \parallel \text{time } \} \text{ key}^{-1}_{initiator} \tag{1}$$

On receiving this action message, every observer constructs a claiming message for each of its neighbor and sends the message with a key as in (2) which is the hash of the concatenation of seed sent along with action message and the examinee ID. Examinee is the neighbor of the observer for which the claiming message is constructed and the observer becomes the examiner.

$$\text{Key} = \text{Hash ( seed} \parallel \text{Id}_{examinee}) \tag{2}$$

### 4.2. Generation of claiming message

Upon receiving the action message, each node updates the nonce and stores the seed. The node ‘a’ then operates as an observer that generates a claiming message as in (3) for each neighbor ‘b’ (examinee) and transmits the message through the overlay network with the message containing the identity of the observer $ID_a$, identity of the examinee $ID_b$ and location of the observer $L_a$ and the location of the examinee $L_b$ encrypted by the private key of the observer $key^{-1}_a$. $M_{a4b}$ is the claiming message of ‘a’ constructed for ‘b’.

$$M_{a4b} = ID_b, L_b, ID_a, L_a, \{ID_b \parallel L_b \parallel ID_a \parallel L_a \parallel \text{nonce }\}key^{-1}_a \tag{3}$$

### 4.3. Routing through the overlay network

This claim along with key which is Hash [21] (seed || $ID_b$) is sent through the balanced overlay network. For a key issued or received at node x, the node will first check its own range of values, i.e the upper and the lower bound values assigned to the node when the tree was constructed. If key is within the current range, the local index is searched for the value, and the search stops. Otherwise, x routes the key to the destination node.

Figure 3 indicates that if we route a message from H with a key 73 , H refers to its right routing table and routes the message to L which is the farthest node whose lower bound is less than or equal to 73. L then routes to M by the same logic and M to R which is the right child of M. R again routes the message to C which is the right adjacent node of R.

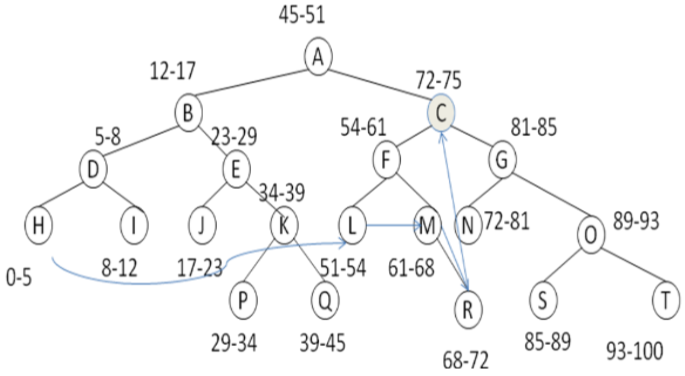


Figure 3. Routing through network

**<u>Algorithm: search(node n, message $M_{a4b}$, value v)</u>**

```
If ((v >= LowerBound(n)) and (v <= UpperBound(n)))
Inspect $M_{a4b}$
Else
If (v >UpperBound(n))
f=TheFarthestNodeHaving(v>=LowerBound(f))
If (there exists such an f)
Forward m to f
Else
```

```
If (RightChild(n)!=null)
Forward m to RightChild(n)
Else
Forward m to RightAdjacentNode(n)
End If
End If
Else
//A similar process is followed towards the left
End If
End If
```

### 4.4. Inspection of the claim

The destination which acts as an Inspector buffers and checks messages for detection. Inspector checks whether there exists two messages $M_{a4b}$ and $M'_{a4b}$ with same ID $ID_b$ but different locations $L_b$ and $L_b$'. If it detects a clone, inspector becomes witness and broadcasts the evidence which contains the two messages with same IDs, but different locations to notify the whole network. All uncompromising nodes verify the evidence message and stop communicating with the cloned nodes.

**Algorithm: Inspect($M_{a4b}$)**

```
Verify the signature of $M_{a4b}$
If $ID_b$is found in cache
If $ID_b$has two distinct locations
Broadcast evidence
Else
Store $ID_b$and $L_b$in cache
```

## 5. EXPERIMENTAL RESULTS

Communication cost refers to the average number of messages sent per node. The average path length between two random node varies from to O(log n) to O(√n) dependent on underlying sensor networks. Chord hop length provides an additional log n and hence overall communication cost ranges from $\log^2 n$ to √n log n. When compared with chord overlay network, our proposed balanced network for clone detection alleviates the delay incurred if the node is located anti-clockwise by maintaining both left and right routing tables. The average path length in a bidirectional search is log n / 2 [22] and the average path length in a bigreedy search is $\log n / 2 - \sqrt{\log(n / 2\pi)} + 1$ [12] which is significantly less than the communication cost using a chord overlay[23].

The representation of two results, first chart Figure 4 (Number of nodes Vs overlay hop length) indicates how balanced overlay network performs better than traditional chord overlay network in terms of hop length. Second graph Figure 5 (Number of nodes Vs Communication cost) is an indication of the impact of hop length on the communication cost. Maintenance of a cache in every node can bring down the communication cost to as low as O(1), so that the next time a source knows the destination, it directly retrieves the destination's IP address [24]. Nodejoin and leave operations always take only log n steps as opposed to $\log^2 n$ in Chord [25].

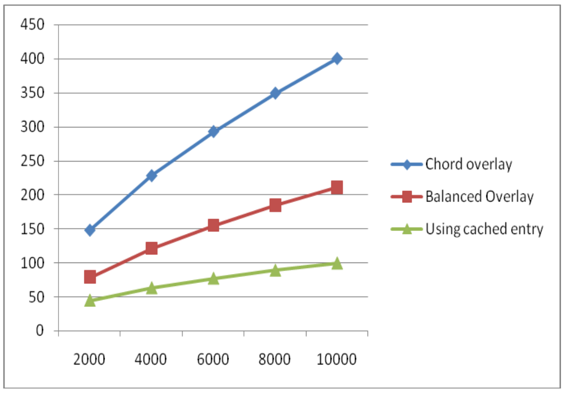


Figure 4. No. of nodes Vs Overlay hop length

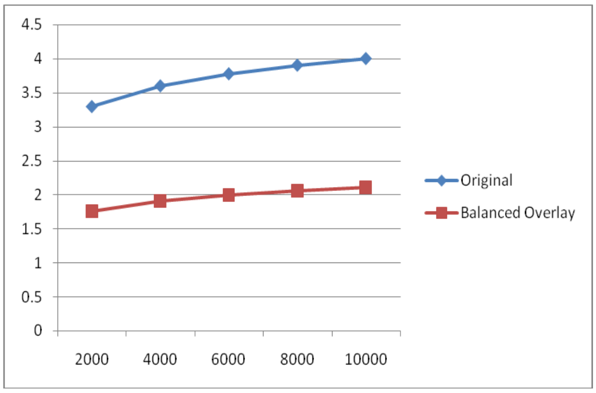


Figure 5. No. of nodes Vs Communication cost

## 6. CONCLUSION AND FUTURE WORK

Wireless sensor networks are susceptible to node replication attacks [26] and hence a decentralized key based caching system is used to identify the clones with the presumption that nodes that have same identity but are geographically distant are considered clones. It is inferred that the use of a balanced network for node clone detection reduces the delay and communication cost when compared to the other overlay networks. This protocol provides high security level with reduced communication cost and minimal storage consumption. As a future work, we can design the overlay based on the physical location and geographic co-ordinates [27] so that the overhead in communication because of the physical network being different from logical network can be solved.

## REFERENCES

[1] M. Udin Harun Al Rasyid, *et al.*, "Beacon-enabled IEEE 802.15.4 wireless sensor network performance," *2013 IEEE International Conference on Communication, Networks and Satellite,* 2013.

[2] M. Udin Harun Al Rasyid, *et al.*, "Performance of multi-hop networks using beacon and non-beacon scheduling in Wireless Sensor Network (WSN)."*International Electronics Symposium,* 2015.

[3] http://www.cs.virginia.edu/~cs757/slidespdf/757-09-overlay.pdf

[4] D. Kato, "Latency model of a distributed hash table with big routing table." *Proceedings. Fourth International Conference on Peer-to-Peer Computing, Proceedings, 2004.*

[5] John Risson, *et al*., "Topology Dissemination for Reliable One-Hop Distributed Hash Tables," *IEEE Transactions on Parallel and Distributed Systems, Volume: 20, Issue: 5,* May 2009.

[6] R. Brooks, *et al*., "On the detection of clones in sensor networks using random key predistribution," *IEEE Transactions on Systems, Man and Cybernetics C,* vol. 37, pp. 1246-1258, 2007.

[7] M.G. Sadi, etal, "GBR: Grid Based Random Key Predistribution for Wireless Sensor Network." *11th International Conference on Parallel and Distributed Systems*, 2005.

[8] H. Choi, *et al*., "SET: detecting node clones in sensor networks," *Proceedings of the 3rd International on Security and Privacy in Communication Networks (SecureComm '07)*, pp. 341-350, 2007.

[9] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," *Proceedings of the 9th ACM Conference on Computer and Communications Security, Washington, DC, USA*, pp. 41-47, 2002.

[10] Parno, *et al*., "Distributed detection of node replication attacks in sensor networks," *Proceedings of the IEEE Symposiumon Security and Privacy (IEEE S and P '05)*, pp. 49-63, 2005.

[11] N. Malhotra, *et al*., "LRRM: a randomized reliable multicast protocol for optimizing recovery latency and buffer utilization." *24th IEEE Symposium on Reliable Distributed Systems,* 2005.

[12] Zhiyang Guo, *et al*., "On-Line Multicast Scheduling with Bounded Congestion in Fat-Tree Data Center Networks." *IEEE Journal on Selected Areas in Communications,* 2013.

[13] M. Conti, *et al*., "A randomized, efficient, and distributed protocol for the detection of node replication attacks in wireless sensor networks," *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '07)*, pp. 80-89, 2007.

[14] B. Zhu, *et al*., "Localized multicast: efficient and distributed replica detection in large-scale sensor networks," *IEEE Transactions on Mobile Computing*, vol. 9, pp. 913-926, 2010.

[15] ZhijunL., *et al*., "On the node clone detection in wireless sensor networks."

[16] Preetha Theresa Joy, *et al*., "A key based cache replacement policy for cooperative caching in mobile ad hoc networks." *3rd IEEE International Advance Computing Conference (IACC),* 2013.

[17] Masahito Hayashi, *et al*., "More efficient privacy amplification with less random seeds."*IEEE International Symposium on Information Theory,* 2015.

[18] Qu Hua, *et al*., "A routing algorithm with Multiple Constrained Balanced Path for overlay network." *6th International Conference on Signal Processing and Communication Systems,* 2012.

[19] Abdullah Faruq Ibn Ibrahimy, *et al*., "Two Dimensional Array Based Overlay Network for Reducing Delay of Peer-to-Peer Live Video Streaming." *International Conference on Computer and Communication Engineering,* 2014.

[20] A. Anitha, *et al*., "A survey of P2P overlays in various networks." *International Conference on Signal Processing, Communication, Computing and Networking Technologies,* 2011.

[21] Mahima Singh, *et al*., "Choosing Best Hashing Strategies and Hash Functions."*IEEE International Advance Computing Conference,* 2009.

[22] X.Zheng and V.Oleshchuk, "Improvement of chord overlay for p2psip-based communication systems."

[23] P. G. G. S.Manku, "Optimal Routing in Chord."

[24] Ramanpreet Kaur, *et al*., "Analysis of different churn models in chord based overlay networks."*Recent Advances in Engineering and Computational Sciences,* 2014.

[25] H. V. Jagadish, *et al*., "BATON: A Balanced Tree Structure for Peer-to-Peer Networks."

[26] Moirangthem Marjit Singh, *et al*., "Preventing node replication attack in static Wireless Sensor Netwroks." *Proceedingsof 3rd International Conference on Reliability, Infocom Technologies and Optimization,* 2015.

[27] Maryam Bagheri, *et al*., "Tree Optimization In Overlay Multicast Based on Location and Bandwidth." *International Conference on Computer Technology and Development,* 2009.