International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:4, No:9, 2010

Enhanced-Delivery Overlay Multicasting Scheme by Optimizing Bandwidth and Latency Discrepancy Ratios

Omar F. Hamad, T. Marwala

***Abstract*—**With optimized bandwidth and latency discrepancy ratios, Node Gain Scores (NGSs) are determined and used as a basis for shaping the max-heap overlay. The NGSs - determined as the respective bandwidth-latency-products - govern the construction of max-heap-form overlays. Each NGS is earned as a synergy of discrepancy ratio of the bandwidth requested with respect to the estimated available bandwidth, and latency discrepancy ratio between the nodes and the source node. The tree leads to enhanced-delivery overlay multicasting – increasing packet delivery which could, otherwise, be hindered by induced packet loss occurring in other schemes not considering the synergy of these parameters on placing the nodes on the overlays. The NGS is a function of four main parameters – estimated available bandwidth, B_a; individual node's requested bandwidth, B_r; proposed node latency to its prospective parent (L_p); and suggested best latency as advised by source node (L_b). Bandwidth discrepancy ratio (BDR) and latency discrepancy ratio (LDR) carry weights of α and $(1{,}000 - \alpha)$, respectively, with arbitrary chosen α ranging between 0 and 1,000 to ensure that the NGS values, used as node IDs, maintain a good possibility of uniqueness and balance between the most critical factor between the BDR and the LDR. A max-heap-form tree is constructed with assumption that all nodes possess NGS less than the source node. To maintain a sense of load balance, children of each level's siblings are evenly distributed such that a node can not accept a second child, and so on, until all its siblings able to do so, have already acquired the same number of children. That is so logically done from left to right in a conceptual overlay tree. The records of the pair-wise approximate available bandwidths as measured by a pathChirp scheme at individual nodes are maintained. Evaluation measures as compared to other schemes – Bandwidth Aware multicaSt architecturE (BASE), Tree Building Control Protocol (TBCP), and Host Multicast Tree Protocol (HMTP) - have been conducted. This new scheme generally performs better in terms of trade-off between packet delivery ratio; link stress; control overhead; and end-to-end delays.

***Keywords*—**Overlay multicast; Available bandwidth; Max-heap form overlay; Induced packet loss; Bandwidth-latency product; Node Gain Score (NGS).

I. INTRODUCTION

WHEN the upper positioned nodes fall short in serving the lower nodes in an overlay scheme, due to mismatch of the bandwidths between the parents and the children/grand children, low packet delivery ratio, especially in a heavily loaded traffic, has been one of the critical issues requiring intensive scheme while dealing with such behaving systems. Dealing with heavy loads worst cases of traffic using the existing schemes is surely prone to a lot of packet loss. The case becomes more alarming when we think of the induced packet loss where the higher leveled nodes in an overlay spread the loss they have incurred to their respective lower leveled nodes. This study, hence, aims at addressing the need of devoting the available and simple measurable and estimatable parameters in creating a better overlay multicast tree which can utilize the same available resources optimally to prevent or at least to largely reduce the induced packet loss for the purpose of achieving higher delivery ratios.

The main contribution of this proposal is to play a great role in reducing a vital induced packet loss caused by the present schemes which do not consider these parameters on placing the nodes on the overlay tree. It is proposed that each node be positioned according to the NGS it earns from a function governed by the four main influencing parameters – the estimated available bandwidth, B_a; the individual node's requested bandwidth, B_r; the proposed node latency to its prospective parent, L_p; and the suggested best latency as advised by the source node, L_b. An optimized NGS function in which the NGSs will be used as the Node_IDs has been worked on in this study. The NGS of each node is pre-calculated as an integrated measure from a fraction of the bandwidth discrepancy ratio *(BDR)* and that of the latency discrepancy ratio *(LDR)* with the weights of α and $(1{,}000 - \alpha)$, respectively and with arbitrary chosen α ranging between 0 and 1,000 to make sure that the NGS values, used as node IDs, maintain a good possibility of uniqueness and a good balance between the most critical factor between the BDR and the LDR. The NGSs are expected to be unique nearest integers such that if two or more nodes possess the same NGS values, the NGS of the newest node is recursively decreased by a numeric one and so on. A max-heap-form tree is then constructed with an assumption that all the nodes possess, as it must practically be, NGSs less than the source node's NGS. This scheme tries to maintain a sense of load balance by evenly distributing the children of each level's siblings such that a node can not accept or can not be assigned the n^{th} child until all its siblings have been able to register $(n - 1)$ children if they are capable of doing so as dictated by their out-degree boundaries and their bandwidth capabilities. That assignment is so logically done from left to

† The authors are with the Faculty of Engineering and the Built Environment, University of Johannesburg, Auckland Park Kingsway Campus, Johannesburg 2006, REPUBLIC OF SOUTH AFRICA.

right in a conceptual overlay tree construction. The bandwidth estimation tool proposed by M. Jain et al in [8] and that proposed by Melander et al in [9] can be applied and enhanced, but, in this work, pathChirp [4] seems to be more efficient in providing estimation with minimal errors. The record of the estimated available bandwidths as measured by a pathChirp scheme [4] at individual nodes is maintained. The two parameters, B_r and L_p, are fed to the source node from individual nodes as per individual nodes requirements for better delivery of the content. The available bandwidth, *Ba*, and the source-based proposed best latency available, L_b, are the measured ones through probing. Comparing to other schemes, this ne proposal seems to generally perform better in terms of trade-off between packet delivery ratio, maximum link stress, control overhead, and end-to-end delay.

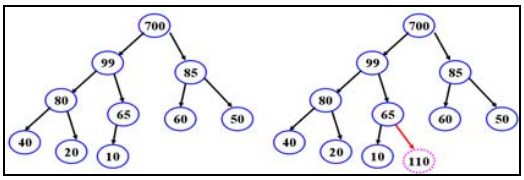


Fig. 1 Example of a max-heap form formation of nodes with a new node number 110 added

Consider the max-heap illustration given in Fig. 1 with numbers representing performance, say NGS, at each node. The right-hand side of Fig. 1 shows a simple example of the possible max-heap overlay while the one on the right-hands side shows the max-heap rule being broken because of the new node number 110. When an element is added to a heap, it should be initially placed as the rightmost leaf to maintain the completeness property, but by doing so the max-heap ordering property becomes broken!

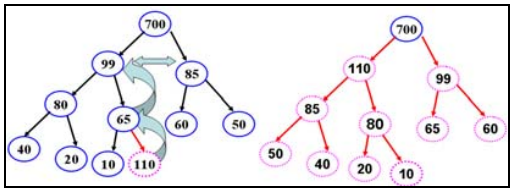


Fig. 2 Restoration of a max-heap form ordering properties by shifting the added node 110 upward and adjusting all other affected nodes

To restore the heap ordering property, the newly added element must be shifted upward, that is bubbled up, until it reaches its proper place by recursively swapping with parent. While doing so, the other nodes, too, need adjustment so that the max-heap ordering rules are maintained. Likewise, creating the min-heap overlay just follows the reverse procedure of the max-heap form. For the case of bandwidth-based, we prefer that the higher bandwidth-ed nodes are settled on the top and gradually decreasing when moving down and from left to right. Hence, max-heap becomes an ideal option among the forms. However, on the other hand, when costs are involved, it is better to have higher costs down and, therefore, the ideal overlay becomes a min-heap form.

Immediately after this introduction, this work addresses some related works in section 2. The concept of bandwidth-latency-product measures (BLPMs) is discussed in section 3 while section 4 is about the NGS-based max-heap overlay multicast scheme where a gist of the limitations of the fundamental bandwidth-only-based max-heap overlay scheme is given, and the details of the proposed NGS scheme described. The node gain score (NGS) function is also expressed in this section. In addition, the section is devoted to logical membership positioning in NGSs based overlay. Simulation setup and simulation results and inference are laid down in section 5. To conclude this paper, the conclusion, discussion, and future direction have been put in sections 6.

II. RELATED WORKS

In overlay multicast, as an alternative scheme to IP multicast where data packets transmission can relay between group members, the fear has been that it introduces larger link stress - a number of times that the same data traverses a particular physical link than IP multicast. This results into packet loss and, in turn, causing network and/or node congestions. Besides the fact that the traditional TBCP [11] is a simple mechanism to implement and run, the fact remains that its disadvantage of the parent node having the RTTs only from one layer - between itself and a newcomer, between its children and a newcomer, and between itself and its children - make this algorithm be not optimal. The parent has no RTTs from the next lower layers. The limitations of the TBCP and the optimal TBCP [10] includes the possibility that the link from the parent node to the child to have less available bandwidth than the one from the child to the grandchild. This is one of the sources of induced packet loss in multimedia applications.

A serious problem to ponder is that the packet loss incurred on one overlay's group member can affect all the descendants' packet reception. Therefore, it is very important to construct another type of overlay multicast tree (OMT) to minimize packet loss, especially the induced packet loss. Among the proposed scheme is the one that suggests to build a new overlay multicast tree called BASE (Bandwidth aware overlay multicast architecture) as described eloquently by Kim in [1]. The proposed BASE mechanism uses available bandwidth metric instead of hop counts or delays in order to construct and reconstruct overlay multicast trees. BASE reduces packet loss probability by locating group member with more available bandwidth at upper level so that packet loss seldom happens at upper position. To resolve both network and node congestions, especially at a varied traffic load, BASE designates proper dynamic number of children.

2.1 Bandwidth-Only Based Overlays

In [1], K. Kim has introduced a scheme named BASE (Bandwidth Aware overlay multicaSt architecturE) and explained the work as an effort to minimize link stress by connecting each link over overlay multicast tree with available bandwidth metric. In this proposal, Kim has suggested that

BASE can minimize group member's delivery failure influence by locating group members with more available bandwidth at higher level on overlay multicast tree. In doing so, Kim has observed that BASE can obtain higher packet delivery ratio because packet delivery failure will rarely happen at higher levels on overlay multicast tree. Kwan et al in [2] have addressed a major challenge in designing application layer multicast protocols by improving the joining and maintenance procedures of an overlay multicast tree. They have addressed the challenge by speeding up the formation of the tree and by enhancing the efficiency of the tree maintenance and they have taken both the bandwidth availability and the round-trip-time (RTT) into consideration when a newcomer selects its parent node, but they did not consider the relationship between the discrepancy of the latencies affordable with respect to the one to be offered. They have neither considered the discrepancy between the requested bandwidth and the availability of the bandwidth. These constraints are of critical importance especially when the overlay under discussion is constituted of the worst case multicast groups.

In [3], Zhu et al have presented a bandwidth-efficient scheme called CRBR that seamlessly integrates network access control and group key management. This scheme seems to incur much smaller communication overhead than two other well-known schemes when they are directly applied in overlay multicast, but again the efficiency is just based on the bandwidth, without taking into consideration the other critical parameters. Further more, Yang et al in [5] have proposed a scheme with key metrics judging the quality of a multicast tree being the normalized aggregate delay, *D*, the normalized aggregate cost, B and the weighted sum of delay and bandwidth consumption (WSDB). The individual discrepancies of bandwidths and latencies have not, though, kept into serious consideration. All the above schemes do not, or just consider, the available bandwidth as the core metric in positioning the overlay multicasting nodes in a tree. These can heuristically induce tremendous packet loss as the tree gets deeper with bandwidth demand of the nodes randomly distributed.

2.2 Fundamental Max-Min-Heap/Rate Overlays

The fundamental max-heap overlay schemes have been thought to partially solve the bandwidth demand challenge of individual nodes. In yet another recent research, M. Hosseini et al in [6] have based their work on three dynamic network metrics (available bandwidth, latency, and loss) and the two main components to adaptation process. They have devised a mechanism to detect poor performing current parents and/or the determination for parents' switch on hosts. Rather than considering comparative latencies and comparative bandwidths, they have just considered the absolute values which do not help much in reducing the induced packet loss than the proposal of bandwidth-latency product max-heap form construction described in this research work. The aspect of dynamic bandwidth estimation, to serve as an important basis for performance optimization of real-time distributed multimedia applications, has been stressed by Wang et al in [7]. They have developed a bandwidth estimation algorithm for the fast fluctuated internet and analyzed the relationship between the one way delay and the dispersion of packets train. Their work is based on the proposal of an available bandwidth estimation algorithm using the two features while eliminating administrative access to the intermediate routers along the network path. For robustness and efficiency, the top-down approach has been used to infer available bandwidth, but nothing about the BDR and/or LDR has been taken as a basis for the overlay tree construction. The idea falls short as it is required to analyze the effect of each parameter and combine them properly. The NGS values proposed in this work, based on bandwidth-latency product (and hence based on BDR and LDR), at a go, integrates all the four critical parameters to reduce the induced packet loss in an overlay.

III. BANDWIDTH-LATENCY-PRODUCT MEASURES

3.1 Available Bandwidth Estimation

The concept of self-induced congestion is the one that pathChirp mechanism [4] applies, and it relies on a simple heuristic that if the probing rate exceeds the available bandwidth over the path, then the probe packets become queued at some node/router resulting in an increased transfer time. If the probing rate is below the available bandwidth, the packets face no queuing delay. Therefore, based on this belief, the available bandwidth can then be estimated as the probing rate at the onset of congestion. The schemes are equally suited to single and multiple hop paths, since they rely only on whether the probe packets make it across the path with an unusual delay or not. There is a feature unique to pathChirp operation, as depicted in Fig. 3. It uses an exponentially spaced chirp probing train. The chirp probe train of N pulses are transmitted with an exponentially time period with geometric propagation ratio γ called a spread factor. Fig. 3 shows node-pairs $N_{S1} \rightarrow N_{R1}$, $N_{S1} \rightarrow N_{R2}$, …, $N_{S1} \rightarrow N_{RN}$ through $N_{SN} \rightarrow N_{R1}$, $N_{SN} \rightarrow N_{R2}$, …, $N_{SN} \rightarrow N_{RN}$ probing the chirps to determine the available bandwidths between them and reporting the resultant available bandwidths to the master-node (N_M). The master-node N_M can be a source node, SN, or a node designated for the purpose and with SN information exchange.

According to V. J. Ribeiro et al in [4] and following the ideology of the concept presentation as in Fig. 3, the pathChirp methodology is based in to estimation of the rate A[1, m] in such a way that for m > tight link, A[1,m] remains constant. Therefore, the available bandwidth of the path can be step by step estimated. The pathChirp makes an estimate, $E_k^{(m)}$, of the per-packet available bandwidth and takes a weighted average of the $E_k^{(m)}$ corresponding to each chirp *m* to obtain estimates *D(m)* of the per-chirp available bandwidth which, for N packets per chirp with Δ_k inter-spacing time between packets *k* and *k+1* is as in Eq. 1.

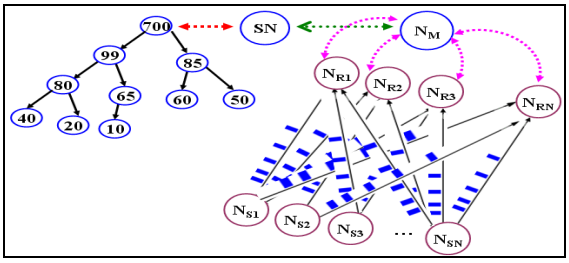


Fig. 3 Schematic diagram showing the bandwidth estimation in a pathChirp operation embedded in max-heap overlay creation

$$D^{(m)} = \frac{\sum_{k=1}^{N-1} E_k^{(m)} \Delta_k}{\sum_{k=1}^{N-1} \Delta_k} \quad (1)$$

Finally, pathChirp makes estimates of the available bandwidth, B_a, by averaging the estimates $D(m)$ obtained in the time interval, τ . This estimated available bandwidth is the one which is critically used in determining the NGSs which are used as the node IDs of the OMT.

3.2 Node Gain Score (NGS) Function

The basic aforementioned components, the Node Gain Score (NGS) has, in this paper been defined in terms of two critical components; the bandwidth discrepancy ratio (BDR) and the latency discrepancy ratio (LDR). The α scaled component associated with BDR, NGS_1, is the ratio of the difference between the requested bandwidth from the available bandwidth to the requested value of the bandwidth. NGS_1 is mathematically expressed as in Eq. 2.

$$NGS_1 = \frac{\alpha (B_a - B_r)}{B_r} \quad (2)$$

where α is the constant determining the weight in importance of the BDR factor as compared to the LDR parameters, B_a is the available bandwidth as estimated by the pathChirp mechanism for the path that is suggested to connect that particular node, and B_r implies the requested bandwidth that a node under consideration feels comfortable to be served on. In a similar fashion, NGS_2 is an NGS component describing the weighted importance of the said LDR. With a logical numerical figure '1' standing for any complete set value as per the proportion of the projected number of members, the scale (1- α), for example $(1{,}000 - \alpha)$, is the representation of the factor which an LDR contributes to the computation of the overall NGS. If the proposed latency which a node prefers is L_p and the best latency that the node under context can be offered is L_b, then NGS_2 can be expressed as in Eq. 3.

$$NGS_2 = \frac{(1-\alpha)(L_b - L_p)}{L_p} \quad (3)$$

Having described the two main components dictating the positioning of the nodes in a max-heap form overlay tree, we can, hence, simply combine the two NGS functions into a general NGS function as a sum of the two components and hence write it as in Eq. 4.

$$NGS = NGS_1 + NGS_2 \quad (4)$$

With Eqs. 2 and 3, we substitute the respective expressions and obtain the integrated NGS as in Eq. 5.

$$NGS = \frac{\alpha (B_a - B_r)}{B_r} + \frac{(1-\alpha)(L_b - L_p)}{L_p} \quad (5)$$

Therefore, the NGS value of each node can be collectively determined by the two main components; the BDR and the LDR, scaled by arbitraries α and $(1{,}000 - \alpha)$ as their factors and with a consideration of the possibility of generating as much unique number as required of NGS values which will stand as the unique node IDs on the max-heap form overlay tree.

3.3 Bandwidth-Latency-Product Formulation

Recalling the main combinatorial measure, (Eq. 4), of the overlay construction based on the max-heap form, the NGS, upon substitution, leads to a complete expression, (Eq. 5), for NGS including the constants and the variables. Upon a minor simplification and re-arrangement, the above NGS function can be expressed as in Eq. 6.

$$NGS = \frac{\alpha L_p B_a + B_r L_b - B_r L_p - \alpha B_r L_b}{B_r L_p} \quad (6)$$

Now, the simplified and re-arranged NGS can be visualized as the sums and differences of the products of the bandwidths and the latencies, that is the bandwidth-latency product measures (BLPMS). We can safely say that the NGS is a directly increasing function of $\alpha L_p B_a$ and $L_b B_r$. It is also a directly decreasing function of $L_p B_r$ and $\alpha L_b B_r$. The NGS is also a directly decreasing function of $L_b B_r$. The fact is strongly true as we always and practically deal with only the positive bandwidths and the positive latencies. Therefore, the NGS measure is, at the user point of view, nothing but the bandwidth-latency product parameters. This can be quite easily controllable and the trade-off can be undergone by choosing parameters of individual importance.

To test the applicability of the NGS function, we do simple and obvious mathematical tests which can comfortably be extended to general applicability. Since we are interested in positive numerical numbers to stand as node IDs, the main numerical focus in estimating the NGS function is to make sure that the combinations of the ranges of the variables will never result into a negative NGS values for the practical possible values in use of B_a, B_r, L_p, and L_b. That being into consideration, we start by estimating the first component of the NGS, that is NGS_1, at the minimum expected B_r of 20MB and setting the value "1" equal to 1,000. If the weight of importance for each of the BDR and the LDR is equally balanced, that is α = ("1" - α) = 500, and if we consider the said B_r, we can express NGS_1 as in Eq. 7.

$$NGS_{1Br=20} := B_a \rightarrow 25B_a - 500 \quad (7)$$

The above expression implies that by fixing B_r equal to 20MB, the NGS_1 component behaves as a linear equation of a variable B_a. The contribution of NGS_1 to the whole NGS for the given sample figures can be plotted with the expected available bandwidth ranging between 20MB and 100MB. Similarly, the contribution of NGS_2 to the whole NGS can be represented as in Eq. 8 with a varying L_b.

$$NGS_{2L_p=5} := L_b \rightarrow 100L_b - 500 \quad (8)$$

Both, the NGS_1 and the NGS_2, show positive monotonic incremental contribution towards the total NGS. We can avoid the possibility of introducing negative node IDs.

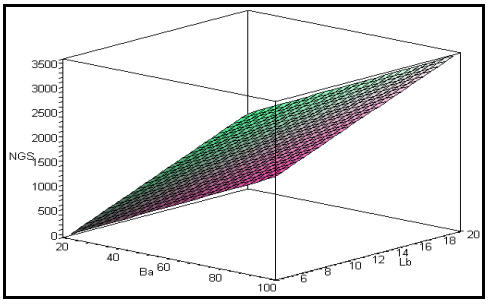


Fig. 4 Desired NGS's B_a - L_b matrix: The positive NGS values which indicates that, for practical ranges of B_a and L_b, we can have up to 3,500 unique whole numbers to represent individual nodes in an overlay

If we contradict the test for the NGS function by considering the undesirable, non-practical values of B_r and L_p and we involve nodes with the requested bandwidth just equal to the possible available bandwidth of 120MB and change the available bandwidth from a minimum of 20MB to that upper limit, with L_p set at 25ms which is beyond the expected convenient latency of 20ms and the value of L_b varied between 0 to 20ms - practically viable and acceptable, then expressions for NGS_1 and NGS_2 have the results as in Eqs. 9 and 10.

$$NGS_{1Br=120} := B_a \rightarrow \frac{25}{6}B_a - 500 \quad (9)$$

$$NGS_{2L_p=25} := L_b \rightarrow 20L_b - 500 \quad (10)$$

The respective components-combined NGS expressions for the desired and the undesired ranges of B_r and L_p are, respectively, given in Eqs. 11 and 12.

$$NGS_{Br=20,Lp=5} := (B_a, L_b) \rightarrow 25B_a - 1000 + 100L_b \quad (11)$$

$$NGS_{Br=120,Lp=25} := (B_a, L_b) \rightarrow \frac{25}{6}B_a - 1000 + 20L_b \quad (12)$$

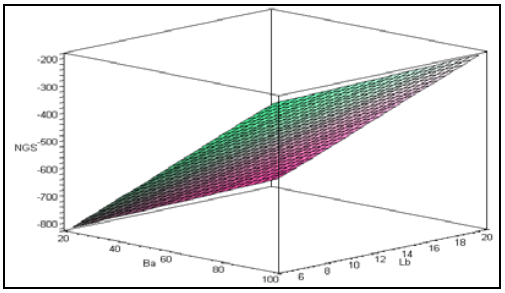


Fig. 5 Undesired NGS's B_a - L_b matrix: all negative, undesired whole numbers for similar ranges of B_a and L_b

The matrices representing Eqs. 11 and 12 are illustrated in Figs. 4 and 5. The former being the positive NGS values which indicates that, for practical ranges of B_a and L_b, that we can have up to 3,500 unique whole numbers to represent individual nodes in an overlay multicasting construction. On the contrary, Fig. 5 displays all negative, undesired whole numbers for undesired ranges of B_a and L_b. The 3-D matrices suggest that we can have a unique assignment of up to 3,500 nodes in our overlay tree without any ambiguity if we design based on the desired NGS's B_a-L_b matrix.

IV. NGS-BASED MAX-HEAP OVERLAY SCHEME

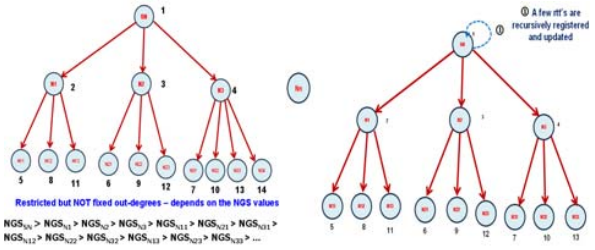
4.1 The Proposed NGS Scheme

The combination of the requested bandwidth from a newly joining member to its expected parent, B_r, the preferred latency from the newcomer to an expected parent, L_p, and the proposal of the best latency available from the newcomer to a proposed parent, L_b, is of equal importance for consideration while constructing and reconstructing an overlay multicast tree with an aim of reducing induced packet loss to the end users. Therefore, the novel proposed scheme in this work aims at building a max-heap form of an overlay tree where the major key of nodes' positioning is a value, NGS, determined by a function of integrated metrics of B_r, L_p, L_b, and B_a. The bandwidth-latency product oriented max-heap-form overlay construction scheme intends to reduce the incurred induced packet loss and hence improve the packet delivery ratio by imposing integrated metric values, Node Gain Scores (NGSs), acting as node IDs for the nodes positioning, in such away that the higher NGSed nodes logically take upper-to-down and alternating left-to-right positions. The reason for the upper-to-down placement is to make sure that the higher metric valued serve the lower ones so that the packet loss causing nodes do not recursively influence the lower children nodes. On the other side, the logical alternating left-to-right approach is for the sake of ensuring load balancing throughout the tree.

4.2 Logical Member Positioning in NGSs Overlay

4.2.1 Max-Heap Form Overlay Tree Construction

Based on the NGS values as calculated by the NGS function, the members are ranked from top to down and alternating from left to right in such away that the overlay tree constructed maintains a max-heap form. Therefore, the max-heap form of the members in overlay tree is mainly determined by the products of the bandwidths and latencies as illustrated by the NGS's general and simplified expression. In this proposal, we introduce restricted-but-not-fixed out-degree allocation of paths. The numbers of out-degrees will mainly depend on the values of NGS the nodes earn, and hence will depend on the BDR and LDR values which determine the amount and the extent in which those particular nodes can sustain induced packet loss.



(a) Members positioning (b) Intermediate step before N_n joins

Fig. 6 Logical member positioning in NGS based overlay: The positioning of the members numbered by considering their NGS values they have earned as a result of their BDRs and LDRs

The max-heap form suggested in this design can be compared with the heap form described in Fig. 6 (a) and the positioning of the members numbered by considering their NGS values they have earned as a result of their BDRs and LDRs. The members are alternated across the level in such away that each sibling at a given level evenly serves equal number of children before the others serve more, unless that particular sibling has no capability to hold more. In Fig. 6 (a), a source node, *SN*, logically takes position *"1"* because it has, or assumed to have, the highest NGS value, NGS_{SN}, while nodes N_1, N_2, and N_3 respectively having NGS values NGS_{N1}, NGS_{N2}, and NGS_{N3}, are respectively positioned at positions "2", "3", and "4". From there then, the newly coming nodes are spread evenly to be served as the children of N_1, N_2, and N_3 with positions "5" as a child of N_1, "6" falling under N_2, and N_3 carrying a new node positioned at " 7".

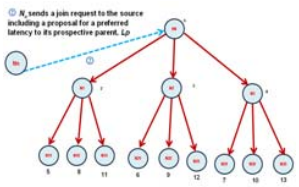
The same fashion is followed until when N_1 and N_2 can no longer hold any more children and hence the subsequent new node, N_{34}, is positioned at "14". Fig. 6 (b) shows the intermediate stage which a new node N_n, meets. There is an updated list of a few best rtt's registered at the *SN* which maintain the best L_b values that will be used to offer the new members wanting to join so that the later can use the values to compare with their L_p values. A source node (SN) starts with an rtt to itself registered as 0 and SN stores rtt's of only a few members with remaining out-degree (s) with a consideration that non-fixed out-degree restriction applies. For making it easier in design, we set very large values for the out-degrees, say a nodal out-degree =30! The membership procedure is detailed illustrated shortly in the coming subsection.

4.2.2 Overlay Membership Based on NGSs

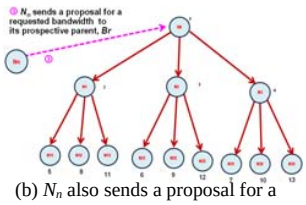
Fig. 7(a) through Fig. 7(h) depict the overlay membership based on NGS values of the member nodes. In Fig. 7(a), when a new node, N_n, intends to join a multicast group, it sends a join request to the source including a proposal for a preferred latency to its prospective parent, L_p. N_n, as in 7(b), also sends a proposal for a requested bandwidth to its prospective parent, B_r.

Fig. 7(c) shows a *SN* responding with the best latency, L_b, for the parents that it can assign to N_n. This value is the immediate one less than or equal to L_b, if any, else, the smallest value is assigned. In this manner, the algorithm just scans a few rtt's instead of all the existing rtt's. Applying the pathChirp mechanism, in Fig. 7(d), SN also estimates the available bandwidth for the requesting node, B_a, and in Fig. 7(e), the SN measures the remaining out-degrees and compares with the requesting node's requirements. Finally, in Fig. 7(h), the adjustment procedure is applied such that the NGS value for a parent is greater than the individual NGS values for the children and the routine to adjust the nodes positions is applied to comply with the inequality $NGS_{SN} > NGS_{N1} > NGS_{N2} > NGS_{N3} > NGS_{N11} > NGS_{N21} > NGS_{N31} > NGS_{N12} > NGS_{N22} > NGS_{N32} > NGS_{N13} > NGS_{N23} > NGS_{N33} >$ … and so on.

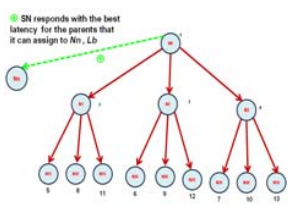
The Node Gain Score (NGS), in Fig. 7(f), is calculated as a function of B_a, B_r, L_b, and L_p, such that $NGS \propto [(B_a - B_r), 1/B_r, (L_b - L_p), 1/ L_p)]$ and that the nearest unique integers are allocated to each node in a tree of overlay multicasting group. As shown in Fig. 7(g), based on the NGS values, a parent node is selected according to the algorithm and based on the maximum NGS heap form of tree with load balancing put into account.



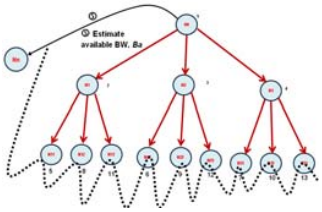
(a) N_n, sends a join request to source including a proposal for a preferred latency to its prospective parent, L_p.



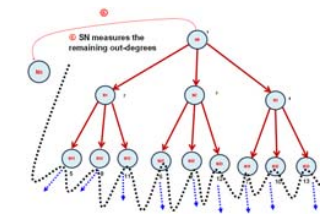
(b) N_n also sends a proposal for a requested bandwidth to its prospective parent, B_r.



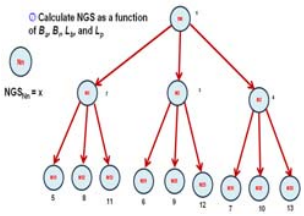
(c) *SN* responding with the best latency, L_b, for the parents that it can assign to N_n.



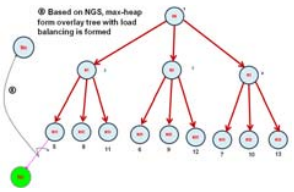
(d) SN estimates the available bandwidth for the requesting node, B_a.



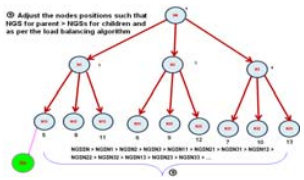
(e) SN measures the remaining out-degrees and compares with the requesting node's requirements.



(f) The Node Gain Score (NGS), in (f), is calculated as a function of B_a, B_r, L_b, and L.



(g) Based on the NGS values, a parent node is selected according to the algorithm.



(h) The adjustment is done such that the NGS value for a parent is greater than the children's.

Fig. 7 (a) – (h): Overlay membership based on NGS values

Talking about the NGS-based overlay multicast tree maintenance, it is sad, but performance wise paying, that the tree based on bandwidth-latency products should be rebuilt whenever a group member's end-to-end delay exceeds the delay bound as set with a threshold. In addition, we also need to reconstruct the max-heap of the overlay if the available NGS value, based on the BDR and LDR, is less than the threshold set. This is mainly because if the NGS value, and hence the BDR and LDR decrease and the current number of children is not changed, then we can surely experience undesired node congestion. Likewise, if end-to-end delay set is exceeded, a member reinitiates the group join procedure with emphasis that delay should be checked to be within the agreed delay range.

V. SIMULATION SETUP, RESULTS, AND INFERENCE

5.1 Simulation Setup

Transit-stub graph model topologies (GT-ITM) were considered for the ns-2 tool to create topologies of 1,000 nodes and up to 128 group members. The assignment of random link delays of 5 - 20ms and link capacities of between 20 and 100MB was done. During the process, the available bandwidths were measured by pathChirp scheme. The evaluation measures as compared to BASE, HMTP, and TBCP were chosen as packet loss ratio against link capacity traffic, maximum link stress, minimum control overhead, and end-to-end delay.

5.2 Simulation Results and Inference

There is, of course, a trade off between adopting this newly designed scheme and sticking to the already available traditional schemes. The comparison simulations of the NGS based mechanism with the schemes like BASE, HMTP, and TBCP gave the results as described in Figs. 8 through 11. Fig. 8 shows the comparison results of BASE, HMTP, and TBCP systems and the NGS based overlay constructed for the packet loss ratios. It can be clearly seen that the proposed NGS mechanism leads to less packet than its counterparts.

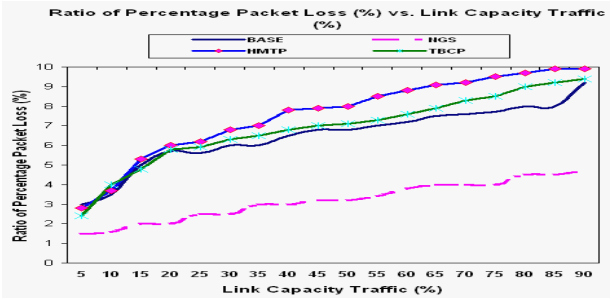


Fig. 8 Ratio of percentage packet loss (%) vs. Link capacity traffic (%)

The discrepancy, in favor of the NGS based bandwidth-latency product max-heap overlay, increases for the best as the number of nodes increases. It seems to have reduced the packet loss tendency which could have otherwise been a must when a simple BASE or other traditional mechanisms would have been used. Similarly, in terms of maximum link stress, the NGS based mechanism seems to do better with the lowest values, as depicted in Fig. 9. For the case of the NGS scheme, the maximum link stress remains with a sparingly increasing change when the group size gets bigger.

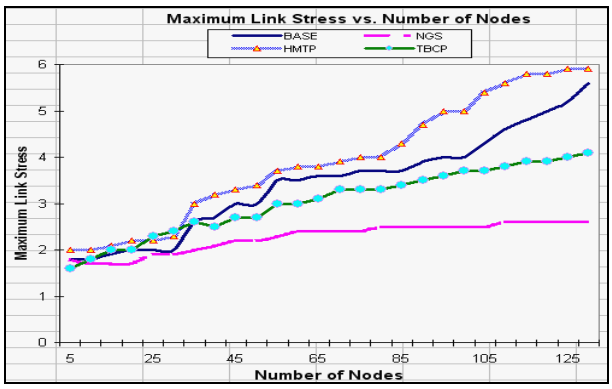


Fig. 9 Maximum link stress vs. Number of nodes

In contrast, the BASE scheme does not tolerate bigger groups in the sense of maximum link stress. It is, again, clear that the NGS mechanism fits better than the BASE and other mechanisms, especially for a growing group of overlay multicasting.

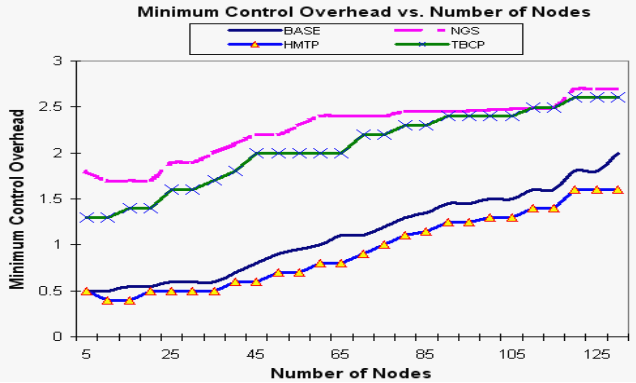


Fig. 10 Minimum control overhead vs. Link capacity traffic

In terms of minimum control overhead needed to maintain service, Fig. 10 shows that NGS based max-heap overlay does not perform better than BASE and other mechanisms, but the bigger the group, the closer the NGS's values to the BASE's and other mechanisms'.

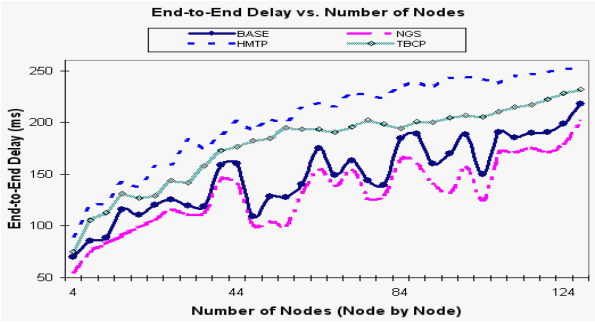


Fig. 11 End-to-end delay vs. Number of nodes

In addition, the trade-off between the packet loss and the control overhead pays enough as far as the whole system performance is concerned in enhancing media delivery. Fig. 11 indicates that the end-to-end delays encountered by BASE and NGS systems are almost overlapping with NGS based scheme doing quite better at lower values especially when the number of nodes becomes higher. The values are somehow unpredictable due to the nature of the NGS calculations based on the available bandwidths and other parameters. For the topology chosen, HMTP seems to have the worst performance in terms of end to end delays while TBCP comes second.

VI. CONCLUSION, DISCUSSION AND FUTURE DIRECTION

Instead of just considering a general delay metric, the newly proposed NGS, bandwidth-latency products based architecture adapts the available bandwidth metric, the delay metric, the relative rtt metric with respect to a suggested/preferred/proposed rtt, the best delay metric as suggested by the source root to individual members. This makes NGS to enhance the performance by evenly loading the tree and recursively distributing the members with closer NGS on each branch of the network and, hence, reducing the possibility of path congestion or group members congestion by making each group member independently check the congestion situation, considers its position with respect to the source root and other members, and then dynamically adjusts its own variables according to varied traffic environments and its position to quickly reach the source node in case of failure. The NGS awareness, via the available bandwidth, the requested bandwidth, the proposed latency from the member, and the assigned parent's best available latency have been demonstrated as a very important metric in establishing an overlay which largely reduces the induced packet loss, end-to-end delays, and maximum link stress. It is, however, unavoidable with this scheme to reduce control overhead comparing to other schemes and, therefore, it is needed to device an incorporated mechanism that will maintain the achieved performance with reduced control overhead units.

A node may experience a lot of pre-join routine check-ups and measurements before being assigned a parent for the hope that the stability after being accepted to join will be reliable. Secondly, a node may choose a parent with RTT greater than a minimum available as it just scan for a node with RTT < its proposal. However, a node saves time to scan all the rtt's which, sometimes, give the same results as above or may be less performance. Finally, a node may join at a parent that is not the best for it to maintain max-heap NGS form of tree, but a node saves from the possibility of having un-balanced tree. Since the available bandwidth estimation time may affect the immediate member join, there is a need to device a mechanism such that the estimation is done quicker and adaptively. There should be a mechanism to keep a history of the available paths such that there should be no need to re-estimate the available bandwidth. A mechanism is also needed to reduce the overhead while maintaining other better performances.

ACKNOWLEDGMENTS

This work was partly supported by TWAS-AAS-Microsoft Research through the 2009 Microsoft Award for Young Scientists and the South African National Research Foundation through the University of Johannesburg's Research and Innovation Division (RID).

REFERENCES

[1] K. Kim.: Bandwidth Dependent Overlay Multicast Scheme: In International Conference on Communication Systems, 2006. ICCS 2006. 10th IEEE Singapore, IEEE 2006.
[2] T. M. T. Kwan et al.: On Overlay Multicast Tree Construction and Maintenance: In International Conference on Collaborative Computing: Networking, Applications and Worksharing, IEEE 2005.
[3] S. Zhu et al: Efficient Security Mechanisms for Overlay Multicast Based Content Delivery: In Computer Communications 30 (2007), Elsevier B.V., pp 793–806.
[4] V. J. Ribeiro et al., "pathChirp: Efficient Available Bandwidth Estimation for Network Paths," In Proceedings of PAM (Passive and Active Measurement Workshop), Apr. 2003.
[5] Y. Jiang et al: A Hierarchical Overlay Multicast Network: In 2004 IEEE International Conference on Multimedia and Expo (ICME), IEEE 2004.
[6] M. Hosseini et al: End System Multicast Routing for Multi-party Videoconferencing Applications: In Computer Communications 29 (2006), 2005 Elsevier B.V., pp. 2046–2065.

[7] S. S. Wang et al: Fast End-to-End Available Bandwidth Estimation for Real-Time Multimedia Networking: In 8th Workshop on Multimedia Signal Processing, IEEE 2006.
[8] M. Jain et al: End-to-End Available Bandwidth: Measurement Methodology, Dynamics, and Relation with TCP Throughput: In SIGCOMM 2002.
[9] Melander et al: A New End-to-End Probing and Analysis Method for Estimating Bandwidth Bottlenecks: In Proceedings of IEEE Globecom, Nov. 2000.
[10] O. Dolejs et al: Optimality of the Tree Building Control Protocol: In Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, CSREA Press, 2002.
[11] L. Mathy et al: An Overlay Tree Building Control Protocol: In J. Crowcroft and M. Hofmann (Eds.): NGC 2001, LNCS 2233, pp. 76-87.
[12] Abderrahim Benslimane: Multimedia Multicast on the Internet, British Library Cataloguing-in-Publication Data.
[13] Doru Constantinescu: Overlay Multicast Networks: Elements, Architectures and Performance: Publisher: Blekinge Institute of Technology Printed by Printfabriken, Karlskrona, Sweden 2007, ISBN 978-91-7295-125-9.
[14] A. Kim: OPNET Tutorial: OPNET Modeler (OPNET Technologies), March 7, 2003.
[15] K. Kim et al: A Novel Overlay Multicast Protocol in Mobile Ad Hoc Networks: Design and Evaluation: IEEE Trans. on Vehicular Tech., Vol. 54, No. 6, Nov. 2005, pp. 2094–2101.
[16] K. Fall et al: The ns Manual (formerly ns Notes and Documentation): The VINT Project: A Collaboration between researchers at UC Berkeley, LBL, USC/ISI, and Xerox PARC, June 22, 2007.
[17] A. Eswaradass et al: Network Bandwidth Predictor (NBP): A System for Online Network performance Forecasting: In Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid (CCGRID'06), IEEE 2006.
[18] A. Habib et al: Incentive Mechanism for Peer-to-Peer Media Streaming: 12th IEEE International Workshop on Quality of Service, IWQOS 2004.
[19] J. A. Strauss: Choosing Internet Paths with High Bulk Transfer Capacity, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2001), September 2002.
[20] Y. Cui et al: Max-Min Overlay Multicast: Rate Allocation and Tree Construction: In Tech. Rep. UIUCDCS-R2003-2373/UILU-ENG-2003-1760, 2003.
[21] K. K. To et al: Parallel Overlays for High Data-rate Multicast Data Transfer: In Computer Networks 51 (2007), Elsevier B.V., pp 31–42.

Omar F Hamad (SM'1997-M'2009) graduated PhD from Multimedia Data Communications Lab in the School of Electronics and Computer Engineering at Chonnam National University – Korea – in February, 2008. He has, since July 2002, been a Telecommunications Engineering Lecturer at the University of Dar es Salaam. He is now an IEEE Member and he has been IEEE Student/Graduate Member since 1997. He is a member of NEPAD Council and a Technical Committee of ICTe Africa Conferences and Workshops and other international conferences. His research interest is in the fields of Bandwidth Calculus in Overlay Multicast, Multimedia Systems, RDMA, FTTH Technologies, Multimedia Delivery over PLC Networks, and Telecommunications Systems. He is now a Post Doctoral Fellow at the University of Johannesburg, South Africa.

Tshilidzi Marwala is currently a Fellow of the CSIR and the Executive Dean of the Faculty of Engineering and the Built Environment at the University of Johannesburg. He holds a Bachelor of Mechanical Engineering with a Magna Cum Laude from Case Western Reserve University, a Master of Engineering from the University of Pretoria, a PhD in Computational Intelligence from University of Cambridge, and was a post-doctoral research associate at the Imperial College of Science Technology and Medicine of the University of London. He has successfully completed a Program for Leadership Development at Harvard University. In year 2006-2007, he was a visiting fellow at Harvard University and in year 2007-2008, he was a visiting fellow at the University of Cambridge. Prof. Marwala has received over 41 awards; has published over 170 articles in refereed international journals, conference proceedings and book chapters and has successfully supervised more than 30 postgraduate students at masters and PhD levels and has collaborated with more than 44 national as well as international researchers. His research interests include the application of computational intelligence to engineering, computer science, finance, social science and medicine. His work has been featured in magazines such as Time Magazine, New Scientist and ACM Tech News.