# A Survey of Reliable Multicast Communication

Adrian Popescu, Doru Constantinescu, David Erman, Dragos Ilie
*Dept. of Telecommunication Systems*
*School of Engineering*
*Blekinge Institute of Technology*
*371 79 Karlskrona, Sweden*

***Abstract*— The paper reports on recent developments and challenges in reliable multicast communication, with special focus on reliable multicast communication at the application layer. The foundation of reliable multicast communication is given by several components, which are multicast communication, congestion control and error control. Our paper is providing a survey of these mechanisms in multicast environments.**

## I. Introduction

Group communication has emerged as one of the most important developments in Internet. Video conferencing, multimedia distribution, online gaming and long-distance education are today some of the most popular Internet applications, which generate large amounts of traffic. To support these applications, reliable multicast communication is a prerequisite. The purpose is to provide efficient and reliable communication services among a number of users, who are members of a multicast group.

Traditional multicast communication demands for the presence of a multicast group, together with associated facilities for reliable multicast communication, to which the users can subscribe and participate. Even though IP multicasting was introduced twenty years ago [14], it is still not widely available as an open Internet service. Problems related to per-group state maintaining, scalability, reliability, congestion control and security have been postponing the wide deployment of IP multicast.

On the other hand, other solutions have been developed for multicast service, to compensate for the above-mentioned limitations, e.g., MBone [18]. MBone provides an overlay network, which connects IP multicast capable islands by using unicast tunnel connections. Furthermore, other developments related to, e.g., video distribution and long-distance education, has further pushed the research and development of new alternative solutions for multicast, which are implemented at the application layer.

Our paper is a survey on current solutions for multicast communication as well as on solutions for the provision of reliable communication in this context. By reliable multicast communication we mean a type of multicast communication that has included facilities of error and congestion control.

The rest of the paper is as follows. Section II provides a survey of multicast communication. Section III presents some of the most important issues to be considered in reliable multicast communication. Section IV describes the main solutions existent today for reliable multicast communication. Finally, Section V concludes the paper.

This work has been done within the research project ”Routing in Overlay Networks (ROVER)”, granted in 2006 by the EuroNGI NoE.

## II. Multicast Communication

Multicast communication represents the operation of sending a packet to a group of recipients, which may be scattered throughout the Internet. A single SEND operation is used in this case to deliver copies of packets to all receivers. The source address is a unicast address, whereas the destination address is a group address of some specific type.

Unlike broadcasting, multicasting allows every member to choose whether to be part of the multicast group or not. Multicasting is a way to reduce network load and end-to-end (e2e) delay. It can be used in conjunction with caching to improve the scalability and delivery performance. Multicasting is therefore most beneficial to users that source the information as well as to ISPs and carriers. However, efficient multicast communication demands for special capabilities and specific algorithms at various layers of the protocol stack. As a minimum, a multicast service should provide several basic functionalities [7], [42]:

- Management of group membership
- Maintainance of data delivery paths
- Replication and forwarding of content
- Congestion and error control

The goal is to satisfy users, network operators and content providers.

### A. Multicast Implementation

Multicasting has been implemented at different layers in the protocol stack, i.e., physical layer, network layer and application layer [20], [27], [32]. Today, some of the most popular multicast implementations are:

- Physical layer (PHY) multicast
- IP multicast
- Application layer (AL) multicast of type Peer-to-Peer (P2P)
- AL multicast of type Overlay multicast (OM)
- AL multicast of type Waypoint multicast (WM)

*1) Physical layer multicast:* A good practice used in multicast is that hosts receive and process only packets that are addressed to them. This can be done at the link layer. This is because every packet received by a network interface causes an interrupt in the device driver. This may generate further processing at higher layers. A good solution could be in this case to use the so-called ”multicast filters”, to add multicast facilities at network interfaces, and to distribute data locally in a LAN environment [32]. This solution is known as physical layer multicast (PHY multicast). The performance of PHY multicast is however limited, especially due to the lack of flexibility. A better solution could be in this case to use a mapping between IP multicast or application layer multicast and physical layer multicast.

*2) IP multicast:* Multicast facilities can be provided at the IP level as well. The ”IP multicast” solution (fig. 1(a)) is a solution originally put forth by Steve Deering in 1989 [14]. IP multicast provides support for both efficient group management and efficient packet forwarding through the network. It is based on an open service model, which does not restrict users to create or join multicast groups. Furthermore, senders are not required to belong to a multicast group. The Internet Group Management Protocol (IGMP/IGMPv2/IGMPv3) is used in connection with IP multicasting to allow a multicast router to learn the addresses associated with networks attached to it and to allow hosts to announce interest in receiving multicast to edge routers [23]. The group management protocol is an integral part of the IP layer in all hosts and routers that support multicasting. Furthermore, other important questions are those regarding the multicast source type (e.g., Any-Source Multicast, Source Specific Multicast), multicast addressing and multicast routing (e.g., SBR, Steiner Tree, PIM) [12].

IP multicast has important advantages that significantly improve efficiency in content distribution. These are effective when the physical media is broadcast in nature, and also efficient utilization of link bandwidth and efficient content replication. Altogether, one can state that IP multicast is well suited for large-scale content distribution, especially for live, non-interactive streaming.

The openess of the model may however create serious problems with the consequence that there is a real risk that the global deployment of IP multicast may be postponed indefinitely [38]. Other important issues are related to the need to support per group state in routers (with impact on scalability), problems related to class D addresses (lack of hierarchy, limited number of addresses, long-term transition to IPv6), security problems and business-related problems (e.g., lack of standards for charging of multicast services).

*3) AL multicast:* Another solution for multicast is to provide multicast communication at the application layer (so-called AL multicast), and to use the unicast transport facilities offered by TCP and UDP. Operations related to group membership, addressing and multicast routing are implemented in this case at the application layer on the end hosts of a network. Application specific intelligence can be used to develop efficient multicast services. Consequently, the network itself is relieved of these responsibilities, and only needs to provide the basic stateless, unicast, best-effort delivery. The architecture therefore decouples multicasting from the unicast routing infrastructure, which gives important advantages in terms of ease of deployment and flexibility. However, AL multicast faces a number of challenges related to routing, efficiency, reliability and scalability, which must be solved in order to gain acceptance.

Over the last years, AL multicast has been the subject of much research and development, in spite of relative drawbacks like inherently being less efficient than IP multicast in using network resources (packet duplication on unicast links cannot be eliminated), soft QoS guarantees and increased complexity of the end system [12], [20], [27]. Furthermore, by means of cross-layer communication, the overlay network can be organized such as to provide topology-aware multicasting. Using various techniques, end hosts may collect information from IP routers to build up more efficient AL multicast networks and to reduce packet duplication.

We distinguish three categories of AL multicast, namely Peer-to-Peer (P2P) multicast, Overlay Multicast (OM) and Waypoint Multicast (WM) (fig. 1).

*4) P2P multicast:* P2P multicast means that there are only end hosts that handle the basic functions (group membership, addressing and routing), whereas in the case of OM multicast there are a number of strategically deployed overlay proxy nodes used to back-up the end hosts. P2P networking was originally designed for information sharing and messaging (e.g., Napster, Gnutella) and it offers several important advantages in terms of, e.g., self-scaling, which means that when more end hosts join the multicast group more bandwidth is supplied [20]. The price however is in terms of dependence on host bandwidth and loosely coupled relationships among the peers (with impact on QoS). Other important advantages are related to flexibility and lack of dependence on the unicast routing infrastructure. Furthermore, a specific challenge is in this case the need to handle the presence of high churn rates in P2P networks [15], [43]. An important consequence of high churn rates is that the topology is very dynamic, which makes it difficult to provide hard QoS guarantees.

*5) OM multicast:* OM multicast has an alternative architectural solution, which means that the content is distributed to proxy servers located close to end hosts. The group of proxy servers are organized into an overlay network to provide delivery service to end hosts. Better performance can be offered here in terms of, e.g., maximized bandwidth, minimized latency/jitter, improved accessibility [36], [37]. Akamai is perhaps the best example of a Content Distribution Network (CDN) provider that is using this model for video streaming delivery [1], [26]. Multicast networks are composed in this case by multiple Points of Presence (PoP) with clusters (so-called Surrogate Servers) that maintain copies of identical content, thus providing better balance between cost for content providers and QoS for customers. CDN nodes are deployed in multiple locations, placed in different backbones all over the world. They cooperate with each other, transparently moving

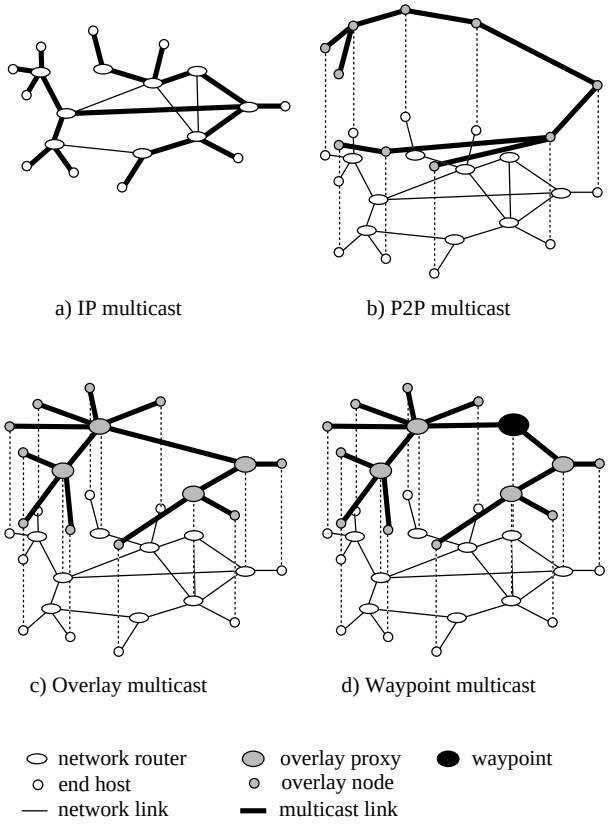


Fig. 1. Multicast architectures

content to optimize the delivery process and to provide users the most current content. The optimization process may result, e.g., in reducing the bandwidth cost, improving availability and improving QoS [36].

The client-server communication flow is replaced at OM by two communication flows, namely one between the origin server and the surrogate server and the other between the surrogate server and the client. Requests for content delivery are intelligently directed to nodes that are optimal with reference to some parameter of interest, e.g., minimum number of hops, or networks, away from the requester. However, questions related to QoS provision, content multicasting and multipath routing heavily complicate the picture.

*6) WM multicast:* An alternative solution for AL multicasting is given by the Waypoint Multicast (WM) solution, as described in [10], [20]. Waypoint nodes are specific nodes existent in a pool of common resources, which may be invoked to temporarily enter a multicast group and to provide the lacking bandwidth needed at the specific moment to support all multicast hosts. These nodes can be statically or dynamically provisioned. The behavior of a waypoint node is similar to that of an end host (as used in P2P multicast), the advantage however is given in this case by higher degree of flexibility and better resource utilization.

### *B. AL Multicast Construction*

A fundamental goal of the process of building the multicast group is to create a loop-free topology to serve, e.g., content distribution to members participating in the group. A logical distribution tree is constructed, which is rooted at the source. Depending upon the relationship among nodes, they can be partitioned into two main categories, i.e., parents and children.

The process of building the multicast group is a sophisticated one, especially in the case of AL multicast. There are a number of fundamental steps that must be considered in such a process, i.e., peer discovery, neighbor selection, parent selection and group maintenance [16].

Several performance metrics have been defined to characterize the multicast communication service and impacts on the network [17], [44]. The most important metrics are:

- Link stress, in terms of number of identical packets a physical link carries.
- Link stretch, also called relative delay penalty. This is the ratio of delay between two nodes along the overlay distribution topology to delay of the direct unicast path.
- Resource usage, in terms of the sum of the $delay*stress$ product over all participating links.
- Time to first packet, which is the time required for a new member to start receiving data after joining a group.
- Losses, which is the average number of packet losses after an ungraceful failure of a single participating node.
- Robustness to changing network conditions.
- Control overhead.

The algorithms for topology creation can be implemented in different ways, each of them with different advantages and drawbacks [17], [20]:

- Static precomputation algorithms
- Centralized algorithms, with partial or full membership knowledge
- Distributed, self-organizing algorithms, which differ in the way the topology is created (e.g., mesh first, tree first)

Desired features of such an algorithm are good performance (not much worse than IP multicast), scalability, ease of deployment, robustness (respond well to changing network conditions), quick and fair response to changing conditions of group membership and security. Today, most group construction algorithms seem to be distributed and self-organizing such as to reduce the stress on the source node and to allow for good scalability performance [20].

Another important algorithm, which takes over after the multicast group is constructed, is performance-aware adaptation of the e2e performance function to dynamics of multicast group and changing network conditions.

There are several strategies to construct AL multicast overlays [2], [17], [20], [33], [46]:

- Mesh-based overlays
- Tree-based overlays
- Multiple tree/mesh overlays
- Ring and multi-ring overlays
- Distributed hash tables

*1) Mesh-based overlays:* The mesh-based approach means that the nodes are organized in a mesh topology, where every node has knowledge of a set of other nodes, called neighbors. An important feature is that there is more than one path available for communication between an arbitrary pair of nodes and neighbors are cooperating to exchange the content according to some predefined cooperation strategy. This means that alternative paths already exist without the need to reconstruct the path between two nodes in case of negative events, e.g., path crashes. Another positive feature is that this offers advantages with respect to routing stability as well as for QoS offerings.

The main drawback of the mesh-based approach is related to difficulties in constructing loop-free forwarding paths among group members. Other drawbacks are the increase of link stress, complexity of algorithms needed for cooperation strategy as well as for chunk selection strategy [2].

*2) Tree-based overlays:* The tree-based approach means that a specific algorithm is used to build up a tree topology such that a single path is established between two arbitrary nodes. Two of the most popular algorithms are the recursive algorithm and the clustering algorithm [17]. In the case of the recursive algorithm, a newcomer node first contacts the tree root, and selects then the best node among the children of the root node with respect to some reference set of parameters. This procedure is repeated until an appropriate parent is finally selected. The clustering algorithm first creates a hierarchy of clusters, and then newcomers recursively cross it to find the appropriate cluster.

Some interesting tree-based architectures are [2], [6]:

- Linear architecture, where clients are organized in a chain with reference to the root server.
- Tree distribution with outdegree ($kTree^k$), where clients are organized in a tree with an outdegree $k$ and an interior node in the tree serves $k$ clients simultaneously.
- Forest of parallel trees ($PTree^k$), where a specific content is first split into $k$ parts, each part is then distributed over an independent tree rooted at the server, and finally the content is reconstructed at the receiver.

The tree-based approach is especially advantageous for one-to-many multicast communication, which is typical for content distribution networks. This means that, e.g., a content provider first sends the content to the root node for further distribution to multicast nodes. This is the typical communication model used in IP multicast, although larger amounts of data can be transported in the case of AL multicast. Compared to IP multicast, the AL multicast has the drawback of larger amount of resources needed to provide the multicast communication service as well as the risk of inefficient use of available resources.

*3) Multiple tree/mesh overlays:* The multiple tree/mesh approach represents an attempt to open up the bottlenecks of the above-described architectural solutions and to remove so the limitations of the mesh- and tree-based approaches [20]. The fundamental concept is to use a specific codec that generates replicated (video) streams for the same content, but at different rates, i.e., multirate video streams [13], [29]. Each of these streams can be independently decoded and the content reproduced with different QoS degrees, depending upon the specific stream.

Besides content replication, another useful concept is content decomposition. In such a case, a raw video sequence is compressed into several non-overlapping video streams (so-called "layers"), and dedicated tree/mesh topologies can be used in the multiple tree/mesh overlay to carry the specific streams [13], [29]. The receiver can selectively subscribe to a number of layers based on the resources it has, e.g., in terms of available bandwidth. QoS can in this case be improved when more streams are received and decoded together.

There are two categories of layering schemes. These are the cumulative layering and the non-cumulative layering. Cumulative layering means that one layer has the highest importance and contains the parts of content (e.g., video) with most important features. Additional layers are called enhancement layers and contain parts of content that progressively refine the quality of reconstructed content. On the other hand, non-cumulative layering means that all layers have equal importance in content reconstruction and any set of layers can be used for this purpose. The flexibility is therefore higher in the case of non-cumulative layering.

The multiple tree/mesh approach offers the advantage of reducing the impact of network and group dynamics by using decomposition and redundancy. The price is in terms of TCP non-friendliness, scalability, and responsiveness. Intensive research is therefore done to solve these problems [29].

*4) Ring and multi-ring overlays:* Another solution for group communication is ring and multi-ring overlays. These architectures have significant advantages over mesh and tree overlays in terms of reliability, survivability and security [46]. Tree- and mesh-based architectures have inherent flow and congestion control problems, especially in the case of using the traditional ACK reliability-based error control [32]. On the other hand, ring and multi-ring overlays have advantages in terms of inherent reliability and single fault tolerance. This is because of the ring-based topology itself, where packets are easily looped back to the sender. Another advantage is given by the low number of ACKs needed in this case. There are even situations where no ACKs are needed to provide a successful communication. This is because the original packets are easily looped back to the sender in the case of successful communication.

Ring and multi-ring group communication have the drawback of longer communication paths and, accordingly, larger delays and jitter. Furthermore, another possible drawback is related to scalability, but this can be improved by building up hierarchical architectures of smaller multi-rings interconnected together to replace large single rings [46].

*5) Distributed hash tables:* Distributed Hash Tables (DHTs) is an approach developed with the purpose to efficiently construct a tree such as to solve the problem of receiver scalability and efficient location of data items [8], [39]. The fundamental concept is to develop a distributed infrastructure

to provide hash-table functionality on Internet-like scales. A decentralized algorithm is used for this. Hash table semantics are exposed in this case over a multicast group of nodes. Every node may insert or retrieve a value associated with a key. Ideally, the keys and the associated values are uniformly distributed across all nodes. The exact distribution of keys and values is highly dependent on the hashing function employed in the specific DHT. The basic operations of insertion, lookup and deletion of (key, value) pairs can be performed in a DHT network. A routing algorithm is also used to allow any node to route to the node associated with a specific key.

DHTs have been shown to provide scalable routing and indexing, robustness and low latency properties. DHT is particularly advantageous for large scale distribution networks, e.g., simulation studies have shown latencies that are less than twice the IP path latency in case of networks with 260 000 nodes [39]. An important drawback is however sensibility to churn [21].

## III. Issues in Reliable Multicast Communication

Due to diverse and challenging conditions, reliable multicast communication has been shown to be a difficult task [24]. Reliable multicast communication is requested to perform well under the conditions of heterogeneity of nodes (in terms of, e.g., different processing capacities) and of the transmission channel (in terms of bandwidth, loss and delay characteristics), heterogeneity of content (static content, dynamic content and streaming media) with different characteristics and QoS requirements in distribution, and also other specific conditions (e.g., scalability, group dynamics and particular limitations in the effectiveness of caching). Appropriate protocols should be designed for error and congestion control in a multicast communication scenario, which are able to provide the requested performance in terms of, e.g., error rate, delay and fairness to competing traffic flows on shared links.

There are several issues that must be addressed by such protocols [3], [5], [16], [24], [32]:

- Where to perform loss detection in a multicast communication?
  Loss detection can be done either at the sender or at the receivers in a multicast group.
- What type of feedback message to use?
  There are two types of feedback messages that can be used, namely positive acknowledgments (ACKs) and negative acknowledgments (NACKs).
- How to send the feedback message?
  There are two alternative solutions possible in this case. These are either via unicast communication back to sender or via multicast communication to all members in the group.
- Who is responsible for retransmitting in case of corrupted data?
  In the case of multicast communication, data retransmission can be done from three different places. These are the sender, one of the receivers and one of the intermediate nodes that has a copy of the original data.
- How to correct errors?
  There are two possibilities for doing error correction. These are retransmission of corrupted data and the use of parity packets in data, i.e., the so-called Forward Error Correction (FEC) method.
- Where to perform error and congestion control in a multicast communication?
  There are two possibilities to do error and congestion control mechanisms in multicast communication. These are hop-by-hop and end-to-end mechanisms.
- What type of congestion control mechanism to use?
  Depending upon the regulating parameter, there are four mechanisms for doing congestion control in multicast environments. These are the window-based, rate-based, layer-based and local recovery-based mechanisms.
- How to do congestion control in a fair way to competing streams on shared links?
- How to develop scalable solutions for error and congestion control in multicast communication?
- How to open up the performance bottleneck related to cachability and cache consistency, which limits the effectiveness of caching?

Depending upon the specific situation and conditions for multicast communication (e.g., IP multicast, AL multicast), different solutions can be used that are suitable for the specific case. Furthermore, another important parameter that influences the mechanisms developed for error and congestion control is related to the delivery service model used in case of content distribution. There are three delivery service models considered today [31], [34]:

- Push service model
  This is a synchronous service model, where a sender initiates concurrent delivery to all receivers in the group and the receivers are supposed to be ready before the transmission begins. The goal is to minimize the synchronization between the sender and the set of receivers. Various mechanisms for session announcements, session management and receiver reports can be used in combination with this service model. The push model is particularly attractive for satellite and wireless communications.
- On-demand service model
  This service model is particularly attractive for the distribution of popular content. The content is continuously multicasted to receivers by using some specific distribution mechanism such that the receivers may join the group, download the content and leave the group whenever they want. The performance is independent of loss patterns and session joining time. The service is scalable as well, although non real-time.
- Streaming service model
  This service model is typically used for delivering of audio and video content. Streaming objects are usually much larger than Web objects and the consequence is that timeliness is more important than the transmission reliability. Delay jitter between servers and clients is

also more important than, e.g., compared to Web content delivery. Furthermore, the streaming service does not typically lend itself to caching, and the consequence is that there is a need for closer cooperation between the producers of content and the content delivery network.

## IV. Reliable Multicast Communication

The problem of reliable multicast communication refers to both IP multicast communication and AL multicast communication. Just like the case of a unicast communication that may require TCP on top of IP unicast, a multicast application may require a reliable multicast communication on top of IP multicast. Techniques similar to those used by TCP for unicast communication can be used for multicast communication as well, e.g., window-based congestion control, use of sequence numbers, positive acknowledgments. There are however significant differences, in the sense that mechanisms for reliable multicast communication should be able to handle, in a scalable manner, highly heterogeneous receivers and to cope with highly dynamic network conditions.

There are several important questions related to reliable and scalable multicast communication, like for instance:

- What is the best place for controlling network congestion, the source, the receiver or both?
- What is the most suitable regulation parameter for multicast communication, window-based or rate-based?
- What is the best implementation for congestion and error control in AL multicast communication, hop-by-hop or end-to-end?

The goal of a reliable multicast communication is to design scalable mechanisms for congestion and error control in multicast environments with similar behavior as TCP, and to allow fairness in resource sharing. Some of the most popular congestion control mechanisms used in multicast communication are window-based congestion control, layer-based congestion control, rate-based congestion control, and local recovery based protocols [3], [12], [16], [24], [32].

### *A. Window-based congestion control*

The window-based regulation has three important limitations with impact on scalability. One of them is given by the need to enforce N different window sizes and monitor the amount of outstanding Transport Protocol Data Units (TPDUs) to each of the N receivers. Furthermore, there is a real risk of acknowledgment implosion when using TCP for multicast communication where a few number of receivers experiencing high packet loss may trigger repeated retransmissions and slow down the entire multicast session. The sender is then forced to process a large number of acknowledgments from several group members only with the consequence that the sender may become a bottleneck for the whole multicast group. This also has negative impacts on scalability, and is known as the ”crying baby problem” [22]. Another important problem is related to the need for good dimensioning of the group resources such as to reduce or eliminate the risk for feedback implosion when the feedback from all receivers may overwhelm the sources and links close to source [28].

Another limitation of window-based regulation is related to fairness, i.e., the risk that other TCP sessions are driven into bandwidth starvation [12].

### *B. Layer-based congestion control*

As mentioned above, content replication and content decomposition can be used in combination with a multiple tree/mesh topology to provide multicast communication. In such a case, a raw video sequence can be compressed into several non-overlapping video streams (so-called ”layers”), and dedicated tree/mesh topologies can be used in the multiple tree/mesh overlay to carry the specific streams.

A particular feature of layer-based congestion control is that this is a receiver-based approach. This means that it is the receiver that autonomously decides whether to subscribe to the multicast group or not. Based on the available resources, the receiver may also decide on how many layers to subscribe to or to drop. Each receiver should also detect packet lost on the way to it, and to adapt the window size or nominal rate. In some specific cases, the receiver should determine the Round Trip Time (RTT) from the source as well.

Examples of implementations of layer-based congestion control are the Asynchronous Layered Coding (ALC) [24], Receiver-driven Layered Control (RLC) [45] and Layered Video Multicast Retransmission (LVMR) [30].

In spite of some difficulties (e.g., TCP non-friendliness, scalability problems), the layer-based congestion control mechanism offers advantages with reference to scalability and the heterogeneity that may exist in a multicast group, e.g., in terms of network bandwidth. Depending upon the available local resources, a client may subscribe to a particular number of layers irrespective of the other clients. This mechanism is also advantageous with reference to the heterogeneity in user behavior and to solving the fundamental conflict existent in a multicast group between the asynchronous behavior of users and the synchronous nature of multicast communication.

### *C. Rate-based congestion control*

The rate-based regulation is, in principle, a mechanism that keeps the instantaneous rate generated by the sender or received by the receiver below a specific level. The fundamental concept of enforcement of rate as the regulation parameter is identical for both cases of unicast and multicast communications. The regulation algorithms can however be different for the two cases.

This difference is especially important for the model-based case, where the feedback represents some measurement result for some parameter that is used in model calculations. Appropriate metrics for the evaluation of multicast traffic must therefore be defined [9] as well as other parameters, like the definition of fairness for rate-based regulation [41]. On the other hand, in the case of increase/decrease algorithm, the feedback simply acknowledges whether there is congestion in the network or not.

Rate-based congestion control can be partitioned into several classes, depending upon the place where the control mechanism is implemented. These are [12]:

- Source-based congestion control, where the source adjusts the transmission rate based on the information received from the multicast receivers and/or based on traffic measurements.
- Receiver-based congestion control, which is generally used in combination with layer-based multicast communication.
- Hybrid congestion control, where both the source and the receivers are participating in the congestion control mechanism by reducing the rate (at the source) and the layer subscription level (at the receivers), based upon the current network conditions.

TCP friendliness is achieved at rate-based mechanisms by forcing the transmission rate to match a throughput that is "TCP compatible", i.e., as given by the formula derived in [35]. A "TCP compatible" flow is defined as a flow that is responsive to congestion notification. Furthermore, this flow does not use, in steady state, more bandwidth than a conformant TCP flow running under comparable conditions with reference to, e.g., loss rate, RTT, packet size.

Because of the acknowledgment implosion problem associated with the window-based regulation, most of implementations for reliable multicast communication use a rate-based regulation mechanism to control and regulate the traffic [12], [16], [40].

### *D. Error control in multicast environments*

Traditionally, error control mechanisms may use several techniques, and the most popular approaches are [12], [24]:

- Automatic Repeat ReQuest (ARQ) schemes, which use acknowledgments, timers and retransmissions.
- Forward Error Correction (FEC) algorithms, which enable packet loss recovery at the destination provided that a specific number of packets are received non-corrupted.
- Error Resilient Source Coding (ERSC), which is used to conceal possible errors at the receiver.

These error control mechanisms are suitable for specific applications and they can be used in connection with TCP or UDP. Delay-insensitive multicast applications (e.g., multicast bulk data transfer) have different time delivery requirements when compared to delay-sensitive multicast applications (e.g., video distribution). For instance, in the case of multimedia distribution, reliable multicast communication means that the delivery must be done reliably but also timely. FEC-based error recovery is therefore preferable for this kind of application.

FEC erasure correction restores corrupted packets by using other redundant packets [16]. There is also another form of FEC, so-called corruption correction, which corrects a corrupted packet by using redundant information encoded within the packet. Only erasure correction is relevant to transport protocols, because unrecoverable corruption is transformed into erasure by the link or network layer.

The mathematical foundation behind FEC is linear algebra over finite fields [11], [16], [19]. The $n$ original segments are viewed as the coefficients of a polynomial function of degree $(n-1)$. Redundant segments can be generated by evaluating the polynomial function at $m$ different points. Any $n$ out of the $m$ values fully specifies the polynomial, effectively regenerating its coefficients. Two popular codes are the Reed-Solomon code and the Tornado code [16]. Furthermore, FEC-based packet recovery in the context of multicast communication can be done by inserting parity packets within a stream or across a combination of streams [26].

### *E. Reliable AL multicast communication*

Overlay networks are opening for new facilities in multicast communication, the price however can be in terms of increased latency and also the risk for lower efficiency in resource utilization. Another important issue is regarding the reliable multicast communication. Usually, this can be achieved by applying TCP on the edges of a connection. Although this is a feasible solution, the price however can be high in form of, e.g., acknowledgment implosion.

AL multicast communication opens for more possibilities to do congestion and error control in a multicast group. There are two classes of control mechanisms in AL multicast communication, which are acting on end-to-end paths and hop-by-hop paths [3], [4], [25].

End-to-end mechanisms means that congestion and error control are done on a end-to-end basis, irrespective of the number of hops. Congestion and error control mechanisms as those described above can be used in this case for reliable multicast communication.

On the other hand, hop-by-hop mechanisms means that congestion and error controls are done on a hop-by-hop basis in an AL multicast group. An end-to-end path may consist of several hops, each of which may include multiple unicast links.

Hop-by-hop reliable AL multicast communication has been shown to considerably reduce the average latency and jitter of reliable communication [4]. This approach has also the advantage that it localizes congestion and error control mechanisms to a specific subset of nodes and links in the overlay network. By this, loss recovery is localized, thus reducing the overall link stress for packet retransmissions. Another advantage is related to TCP friendliness, which can easily be implemented in this case. The overhead induced by the hop-by-hop approach have been shown to be insignificant [4].

A possible drawback could however be the difficulties in applying the hop-by-hop scheme to global Internet due to scalability and interoperability issues.

## V. CONCLUSIONS

The paper has presented a survey on current solutions for reliable multicast communication with emphasis on application layer multicast. These topics are subject for research within the research project "Routing in Overlay Networks (ROVER)", granted in 2006 by EuroNGI NoE. The main

focus of our research is on QoS-aware overlay routing for multicast environments, as a way to provide soft QoS provisioning for specific applications while retaining the best-effort Internet model. Main research questions are on overlay multicast communication, traffic measurements and modeling, QoS provisioning with multicast facilities and reliable multicast communication.

Planned future work is to develop a dedicated middleware environment, which will be used to develop new protocols for multimedia distribution over IP and to offer soft QoS guarantees for specific applications in a multicast environment. We are also planning to develop analytical and simulation models to validate our results.

## REFERENCES


[1] Akamai, http://www.akamai.com

[2] Al Hamra A. and Felber P.A., *Design Choices for Content Distribution in P2P Networks*, ACM SIGCOMM Computer Communication Review, Vol. 35, No. 5, October 2005

[3] Amir Y., Awerbuch B., Danilov C. and Stanton J., *A Cost-Benefit Flow Control for Reliable Multicast and Unicast in Overlay Networks*, IEEE/ACM Transactions on Networking, Vol. 13, No. 5, October 2005

[4] Amir Y. and Danilov C., *Reliable Communication in Overlay Networks*, IEEE International Conference on Dependable Systems and Networks (DSN03), San Francisco, CA, USA, June 2003

[5] Biersack E.W., *Where is Multicast Today?*, ACM SIGCOMM Computer Communication Review, Vol. 35, No. 5, October 2005

[6] Biersack E.W., Rodriguez P. and Felber P.A., *Performance Analysis of Peer-to-Peer Networks for File Distribution*, 5th International Workshop on Quality of future Internet Services (QofIS'04), Barcelona, Spain, September 2004

[7] Braudes R. and Zabela S., *Requirements for Multicast Protocols*, IETF RFC1458, http://www.ietf.org/rfc/rfc1458.txt, May 1993

[8] Castro M., Druschel P., Kermarrec A.-M. and Rowstron A.I.T., *Scribe: A Large-Scale and Descentralized Application-Level Multicast Infrastructure*, IEEE Journal of Selected Areas in Communications, Vol. 20, No. 8, October 2002

[9] Chalmers R.C. and Almeroth K.C., *Developing a Multicast Metric*, IEEE GLOBECOM 2000, San Francisco, CA, USA, November 2000

[10] Chu Y., Ganjam A., Ng T.S.E., Rao S., Sripanidkulchai K., Zhan J. and Zhang H., *Early Experience with An Internet Broadcast System Based on Overlay Multicast*, USENIX Annual Technical Conference, Boston, MA, USA, June 2004

[11] Comer D.E., *Internetworking with TCP/IP Principles, Protocols and Architecture*, Volume 1, Pearson Prentice Hall, 5th edition, 2006

[12] Constantinescu D., Erman D., Ilie D. and Popescu A., *Congestion and Error Control in Overlay Networks*, technical report, Blekinge Institute of Technology, Karlskrona, January 2007

[13] Cui Y. and Nahrstedt K.., *Layered Peer-to-Peer Streaming*, ACM NOSSDAV'03, Monterey, CA, USA, June 2003

[14] Deering S., *Host Extensions for IP Multicasting*, IETF RFC1112, http://www.ietf.org/rfc/rfc1112.txt, August 1989

[15] Dragos I., *Gnutella Network Traffic, Measurements and Characteristics*, Licentiate thesis, Blekinge Institute of Technology, ISBN 91-7295-084-6, April 2006

[16] Dixit S. and Wu T., *Content Networking in the Mobile Internet*, John Wiley and Sons, Inc., ISBN 0-471-46618-2, 2004

[17] El-Sayed A., Roca V. and Mathy L., *A Survey of Proposals for an Alternative Group Communication Service*, IEEE Network, Vol. 17, No. 1, January/February 2003

[18] Eriksson H., *MBONE: The Multicast Backbone*, Communications of the ACM, Vol. 37, No. 8, 1994.

[19] Forouzan B.A., *TCP/IP Protocol Suite*, 3rd edition, McGraw-Hill, 2004

[20] Ganjam A. and Zhang H., *Internet Multicast Video Delivery*, Proceedings of the IEEE, Vol. 93, No. 1, January 2005

[21] Gummadi K., Gummadi R., Gribble S., Ratnasamy S., Shenker S. and Stoica I., *The Impact of DHT Routing Geometry on Resilience and Proximity*, ACM SIGCOMM'03, Karlsruhe, Germany, August 2003

[22] Holbrook H., Singhal S. and Cheriton D.R., *Log-Based Receiver-Reliable Multicast for Distributed Interactive Simulation*, ACM SIGCOMM 1995, Cambridge, MA, USA, 1995

[23] Internet Group Management Protocol, Version 2, RFC 2236, IETF, http://www.ietf.org/rfc/rfc2236.txt?number=2236

[24] IETF Working Group Reliable Multicast Transport (rmt), http://www.ietf.org/html.charters/rmt-charter.html

[25] Kandekar K.A., *A Survey of Issues and Reliability and Congestion Control Techniques in Application Level Multicast*, Technical Report, Computer Science Dept., North Carolina State University, NC27606, USA

[26] Kontothanassis L., Sitaraman R., Wein J., Hong D., Kleinberg R., Mancuso B., Shaw D. and Stodolsky D., *A Transport Layer for Live Streaming in a Content Delivery Network*, Proceedings of the IEEE, Vol. 92, No. 9, 2004

[27] Lao L., Cui J-H., Gerla M. and Maggiorini D., *A Comparative Study of Multicast Protocols: Top, Bottom, or In the Middle?*, IEEE Global Internet Symposium (GI 2005), Miami, FL, USA, March 2005

[28] Levine B.N. and Garcia-Luna-Aceves J.J., *A Comparison of Reliable Multicast Protocols*, Multimedia Subsystems, 6(5), August 1998

[29] Li B. and Liu J., *Multirate Video Multicast over the Internet: An Overview*, IEEE Network, Vol. 17, No. 1, January/February 2003

[30] Li X., Paul S. and Ammar M.H., *Layered Video Multicast with Retransmission (LVMR): Evaluation of Hierarchical Rate Control*, IEEE INFOCOM'98, San Francisco, CA, USA, March/April 1998

[31] Luby M., Gemmell J., Vicisano L., Rizzo L. and Crowcroft J., *Asynchronous Layered Coding (ACL) Protocol Instantiation*, IETF RFC3450, http://www.ietf.org/rfc/rfc3450.txt, December 2002

[32] Matrawi A. and Lambadaris I., *A Survey of Congestion Control Schemes for Multicast Video Applications*, IEEE Communications Surveys and Tutorials, fourth quarter 2003, Vol. 5, No. 2, 2003

[33] Moen D.M., Pullen J.M. and Zhao F., *Implementation of Host-Based Overlay Multicast to Support of Web Based Services for RT-DVS*, 8th IEEE International Symposium on Distributed Simulation and Real-Time Applications (DS-RT'04), Budapest, Hungary, 2004

[34] Neumann C., Rocca V. and Walsh R., *Large Scale Content Distribution Protocols*, ACM SIGCOMM Computer Communication Review, Vol. 35, No. 5, October 2005

[35] Padhye J., Firoiu V., Towsley D.F. and Kurose J.F., *Modeling TCP Reno Performance: A Simple Model and its Empirical Validation*, IEEE/ACM Transactions on Networking, Vol. 8, No. 2, April 2000

[36] Pallis G. and Vakali A., *Insight and Perspectives for Content Delivery Networks*, Communications of the ACM, Vol. 49, No. 1, January 2006

[37] Popescu Adrian, Erman D., Ilie D., Constantinescu D. and Popescu Alexandru, *Internet Content Distribution: Developments and Challenges*, 4th Swedish National Computer Networking Workshop SNCNW2006, Luleå, Sweden, October 2006

[38] Ramalho M., *Intra- and Inter-Domain Multicast Routing Protocols: A Survey and Taxonomy*, IEEE Communications Surveys and Tutorials, first quarter 2000, Vol. 3, No. 1, 2000

[39] Ratnasamy S., Francis P., Handley M., Karp R. and Shenker S., *A Scalable Content-Addressable Network*, ACM SIGCOMM 2001, San Diego, CA, USA, August 2001

[40] Rizzo L., *PGMCC: A TCP-Friendly Single-Rate Multicast*, ACM SIGCOMM 2000, Stockholm, Sweden, August 2000

[41] Rubenstein D., Kurose J. and Towsley D., *The Impact of Multicast Layering on Network Fairness*, ACM SIGCOMM 1999, Cambridge, MA, USA, October 1999

[42] Quinn B. and Almeroth K., *IP Multicast Applications: Challenges and Solutions*, IETF RFC3170, http://www.ietf.org/rfc/rfc3170.txt, Sept. 2001

[43] Saroiu S., Gummadi P.K. and Gribble S.D., *Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts*, Multimedia Systems, Vol. 9, No. 2, August 2003

[44] Shi Y.S., *Design of Overlay Networks for Internet Multicast*, PhD thesis, Sever Institute of Technology, Washington University, Saint Louis, Missouri, USA, August 2002

[45] Vicisano L., Rizzo L. and Crowcroft J., *TCP-Like Congestion Control for Layered Multicast Data Transfer*, IEEE INFOCOM'98, San Francisco, CA, USA, March/April 1998

[46] Wang J., Yurcik W., Yang Y. and Hester J., *Multi-Ring Techniques for Scalable Battlespace Group Communications*, IEEE Communications Magazine, Vol. 43, No. 11, November 2005