多頻群播通訊服務的控制架構、協定與機置 (II)

Control Architecture, Protocol, and Mechanism for Multiple-Stream Multicsat Communication Services (II)

計劃編號：NSC-88-2219-E-006-012

執行期限：88 年 8 月 1 日至 89 年 7 月 31 日

主持人：黃崇明 成功大學 資訊工程學系

Email: huangcm@locust.csie.ncku.edu.tw

1. 中英文摘要

本子計劃目的是在達成於網際網路上進行多方的 (multiparty) 即時 (real-time) 的互動式數位廣播電台以及多頻 (Multiple-Stream) 媒體資料播放, 其中將利用多頻群播播放系統中的資料同步機置 (synchronization mechanism), 以達成較為平順的多頻群播播放服務品質 (QoS), 以及較高的網路擴展性 (Scalability)。並可以讓使用者利用視窗式介面環境來方便地制定 (build) 所需要的多頻群播播放程序。本子計劃的將完成下列 4 個目標：

(1) 互動式數位廣播電台網路架構與系統 (Network Architecture and Software for Interactive Digital Radio System)

(2) 多頻群播網路架構和協定 (Multiple-Stream Multicast Network Architecture and Protocol)

(3) 多頻群播播放系統 (Multicast Multi-stream Presentation System)

(4) 多頻群播環境建制器 (Multiple-Stream Multicast Environment Builder)

第一年完成之成果：具有叩應 (call-in) 功能的互動式數位廣播電台網路架構與系統的設計、建立、測試，和效能測試，以及階層式主從多頻群播系統之網路架構與同步控制模式設計和效能測試與評估。第二年已 完 為 階 層 式 主 從 (Client→Virtual Server→Physical Server) 多頻群播以內容為導向與以時間為導向的階層式主從多頻群播播放同步模式 (model) 與系統。第三年預期完成成果：MP3 播放環境之互動式數位廣播電台系統之設計、建立，與測試，多頻群播同步控制機置 (Multiple-Stream Multicast Synchronization Control Mechanism) 與多頻群播環境建制器 (Multiple-Stream Multicast Environment Builder)之設計、建立，與測試，並與多頻群播網路架構以及多頻群播播放系統整合。最後再與其它子計劃之系統完成整合。

關鍵字：群播、多頻、同步模式、同步與播放控制

The Internet explosion is driving the need for multicasting multiple-stream multimedia presentations, e.g., media distribution and virtual classroom. The jitter phenomenon over Internet always disturbs the orchestration of multimedia presentations. Multimedia synchronization is one of the important techniques to compensate for jitter anomalies. However, the characteristics of multiple media streams combining the multicast delivery complicate the multimedia synchronization problem. Therefore, a formal synchronization definition and specification model that can accurately specify temporal relations among multiple media streams has to be firstly devised. In this report, we (i) identify the primary issues for ensuring smooth multiple-stream multimedia presentations in the multicast environment and (ii) propose a formal definition mechanism to specify related attributes of a multimedia presentation. Based on the proposed model, a communication engine for multicasting multiple streams is developed. The communication engine (i) can achieve smooth multimedia presentations and (ii) is the middleware of multicasting multiple-stream multimedia applications.


1 Network and system architecture

The proposed multicast multimedia network, which is called Multicast Multimedia Communication Network (M^3CN), is a two level hierarchical architecture that spans a distributed environment. The M^3CN consists of a WAN and a lot of LANs that are attached with the WAN. Each LAN is composed of a local Multicast MultiMedia Server (M^3 server) and clients. An M^3 server transmits media units to hosts via LAN or (and) WAN. Clients of a presentation group present the same multimedia resource simultaneously and maybe
scattered over different LANs.

In M\textsuperscript{CN}, the concept of a “virtual server” is adopted. A virtual server receives media units from the “physical server”, which owns the presentation resource, and re-transmits them to end clients. The virtual server is a local server in a LAN and compensates for WAN’s anomalies by means of pre-depositing some media units and having corresponding synchronization schemes. The concept of virtual servers can simplify the overhead of synchronization control in clients because WAN's asynchronous anomalies are compensated for and media streams are synchronized at virtual servers. Clients become simpler and low-end, e.g. a Set-Top-Box, a diskless networking PC, or a networking TV.

Figure 1: System architecture of Mcast.

Figure 1 depicts the system architecture of Mcast. The specification tool is the authoring system. The temporal control mechanism is composed of the presentation information file, the code generator, and the multicast presentation system. An author uses the specification tool to specify temporal and spatial attributes of media and to author his multimedia presentation.

The presentation information file contains a set of data structures that record the spatial and temporal attributes of the corresponding multimedia presentation. Based on the presentation information file, the code generator generates C codes for part of the synchronization control. The Synchronizer and Actors to compose the multicast presentation system. A Synchronizer controls inter-media synchronization and presentation among media streams. An Actor transmits and receives media units of a medium stream and controls the medium flow. That is, an Actor controls intra-medium synchronization and presentation.

The multicast presentation system is composed of the physical server system (PSS), the virtual server system (VSS) and the client system (CS). The PSS retrieves media units from media bases, and multicasts media units to virtual servers according to the information schedule file. Through the PSS, a system manager can specify a communication configuration that contains the multicast group address, communication socket ports, and media files. The VSS receives media units from the PSS and stores media units in media buffers temporarily. According to the presentation schedule, the VSS re-multicasts media units to clients with the synchronization control. With the help of the VSS, temporal anomalies induced by the WAN's transmission can be compensated. The CS receives media units and achieves a smooth presentation adopting the synchronization and presentation control.

In brief, the steps of preparing a multicast multiple-stream multimedia presentation using Mcast are as follows. (i) A user constructs the desired presentation schedule using the specification tool. (ii) The code generator generates C codes for part of synchronization/presentation control according to the presentation schedule. (iii) The system manager compiles the generated C codes with the kernel of Mcast to generate the complete executable C codes for the desired multimedia presentation. (iv) The complete executable C codes contain the physical server code, the virtual server code and the client code.

2 The formal definition of multimedia presentations

A formal definition mechanism should be proposed to specify related temporal attributes in a multimedia presentation. An authoring system can adopt the formal definition mechanism to have users to specify what their multimedia presentations look like.

A multimedia presentation $MP$ is defined as $MP = \{MS, TS\}$, where $MS = \{ms_1, ms_2, ..., ms_n\}$ represents the set of $n$ involved media streams and $TS = \{p-stage_1, p-stage_2, ..., p-stage_n\} = \{p-section_1, p-section_2, ..., p-section_n\}$ represents the presentation temporal schedule. A presentation temporal schedule consists of many presentation stages $\{p-stage_i\}$, where $i = 1...k$. Each $p-stage_i$ contains $x_i$ presentation sections $\{p-section_{i1}, p-section_{i2}, ..., p-section_{in}\}$, where $1 \leq i \leq m$ and $x_1+x_2+...+x_n = m$. In other words, $p-stage_i = \{p-section_{i1}, p-section_{i2}, ..., p-section_{in}\}$. 

![Diagram of Mcast system architecture](image-url)
p-stage\(_i\) = \{p-section\(_{i+1}\), p-section\(_{i+2}\), ... , p-section\(_{i+x}\)\} ...

Media streams involved in a presentation stage are a subset of MS\(_i\) defined as A-MS(p-stage\(_i\)) = \{A\(_i\)(ms\(_i\))\}, ... , A\(_i\)(ms\(_n\))\}. In stage \(i\), if medium stream ms\(_i\) has media units to be presented, A\(_i\)(ms\(_i\)) = 1. \(1 \leq x \leq n\), if medium stream ms\(_i\) has no media units to be presented, A\(_i\)(ms\(_i\)) = 0, \(1 \leq x \leq n\).

A presentation section represents that some media objects have temporal relationships, e.g. the start relation. Thus, one medium object's presentation in a section depends on another medium object's presentation status. It is not necessary that a medium stream always has media units presented throughout an entire section. In Figure 2, the text stream has nothing to present between \(t_2\) and \(t_3\). The time period that a medium stream has nothing to present is called an idle segment. Idle segments can be defined as i-segment(p-stage\(_i\), ms\(_i\)) = \{t\(_i\), t\(_j\)\}, where A\(_i\)(ms\(_i\)) = 1 and ms\(_i\) has nothing to present from time point t\(_i\) to time point t\(_j\).

For example, the text medium has an idle segment \(\delta_i\) from \(t_2\) to \(t_1\) in section 2 of stage 1.

Temporal relationships of two media objects can be formally defined as \(x(t_1) \oplus y(t_2)\), where x and y denote medium objects, \(t_1\) is the display time period of x, \(t_2\) is the display time period of y, \(\oplus\) denotes the type of the temporal relationship, and \(t_1\) and \(t_2\) describe the front, tail, or gap time interval for the corresponding temporal relationship. (Parameters \(t_1\) and \(t_2\) are optional. If there is no front, tail, or gap time period, nothing has to be specified.)

![Temporal Relationship Diagram](image)

Figure 3: Possible temporal relationships between two media objects and their corresponding formal definitions.

![Temporal Relationship Diagram](image)

Figure 3 depicts the corresponding formal definitions of the seven temporal relationships. For example, the formal definition “x\(t_1\) during y\(t_2\)” denotes that the medium object x, which display time period is \(t_2\), has the during relationship with the medium object y, which display time period is \(t_2\) and the corresponding front and tail time periods between the media objects x and y are \(t_1\) and \(t_2\) respectively.

Based on the proposed formal definition, temporal relationships for involved media objects in a presentation stage is defined as TR(p-stage\(_i\)) = \(O_p(t_1) \oplus O_q(t_2)\). For \(i = 1, 2, ... , n\), where media objects \(O_p\) and \(O_q\), 1 \(\leq j \leq i\), has some temporal relationship in presentation stage \(i\).

Each presentation stage is associated with a master stream. The master stream of a stage is formally defined as M(p-stage\(_i\)) = \{ms\(_i\)\}, where \(1 \leq j \leq n\) and A\(_i\)(ms\(_i\)) = 1

3 Presentation control schemes

In order to compensate for jitter anomalies, one can adopt the blocking scheme for the audio medium and the non-blocking scheme for the video medium to achieve intra-medium synchronization. Figure 4(a) depicts an illustrated example for the video stream. When medium unit \(k\) was presented at time \(t\), the next medium unit \(k+1\) should be presented at time \(t+\theta\), where \(\theta\) is the presentation duration of a medium unit. Unfortunately, medium unit \(k+1\) does not arrive on time. Hence, medium unit \(k\) is re-presented at time \(t+\theta\) according to the non-blocking scheme. During the time of re-presenting medium unit \(k\), medium units \(k+1\) and \(k+2\) arrive before time \(t+2\theta\). At time \(t+2\theta\), the “expected” medium unit is presented. Should the “expected” medium unit be unit \(k+1\) or unit \(k+2\)? In order to solve the above problem, two presentation schemes that are considered: (i) time-oriented and (ii) content-oriented schemes.

![Presentation Control Diagram](image)

Figure 4: An example of (a) the time-oriented presentation control and (b) the content-oriented presentation control for a video stream.

If the main concern is (i) to satisfy time-related temporal relations and (ii) to keep the actual presentation time length equal to the nominal presentation time length as much as possible, the time-oriented presentation scheme can be adopted. The “expected” medium unit should be the one that is closest to the nominal one. In Figure 4(a), medium unit \(k+2\) is presented at time \(t+2\theta\) and medium unit \(k+1\) is discarded. The drawback of the time-oriented scheme is that there may be some flickers at the synchronization point when several delayed media units
are discarded at the same time.

If the main concern is to keep the completeness of a media presentation as much as possible, the content-oriented presentation scheme can be adopted. Each medium stream is assigned as much as possible. In Figure 4(b), medium unit $k+1$ is presented at time $t+20$ and medium unit $k+2$ is presented at time $t+30$. The drawbacks of the content-oriented scheme are twofold: (1) the total presentation time may become longer than the nominal presentation time; and (2) more inter-media asynchrony anomalies may exist until an inter-media synchronization is practiced.

4 Usage of Mcast

![Figure 5: An Mcast presentation example.](image)

Figure 5 shows a presentation example that contains four kinds of media streams in the client system. In Figure 5, the presentation status is depicted on the left side, and the presentation time is depicted on the right side. The user can click button “Option” to select presentation control schemes, which includes the content-oriented and the time-oriented schemes. The user can click button "Action" to assign a desired multicast group identification and communication ports for each medium stream. When a presentation is finished, the user clicks "File" button to terminate the client system.

![Figure 6: An illustration of (a) the option menu dialog and (b) the action menu dialog.](image)

Figure 6-(a) depicts the dialog of the option menu, in which the corresponding presentation scheme can be selected arbitrarily by users. Figure 6-(b) shows the dialog of group and communication attributes setup in the action menu. A user can join a multicast group by typing the group address in the "Multicast Group ID" field. Multicast communication ports for each medium stream are assigned in the corresponding fields. The "Media Server" field is used to specify the location of the server in order to setup TCP connections. The "Image Port" and "Text Port" fields are used to setup TCP communication ports for image and text stream respectively. For user's convenience, when a user fills each field, a help description is displayed at the bottom of dialog. When the user has filled all fields, the user can click the "Join" button to finish the assignment and join the associated multicast group.

5 Conclusion

This report describes the main issues of designing a synchronization model for multicasting multiple-stream multimedia presentations. The specifications of a synchronization model are essentially composed of the behavior specification and the function specification. A formal definition is proposed for the behavior specification, which concisely specifies the thirteen temporal relationships of a DMP behavior. Based on the formal definition, we have built a specification tool, which allows a user to conveniently specify spatial and temporal attributes of related media streams and generates the corresponding presentation information file. An EFSM-based (Extended Finite State Machines) formal model is proposed to precisely devise the function specifications, which specify the multimedia synchronization control. According to the EFS-based model, we have developed the temporal control mechanism of the Mcast communication engine. Mcast provides (1) the specification tool to specify presentation's appearance with a WYSIWYG way, and (2) the temporal control mechanism to achieve smooth presentations.

References