Empirical Studies of Multimedia Semantic Models for Multimedia Presentations

Shu-Ching Chen and R. L. Kashyap*
School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN 47907-1285, U.S.A.

Abstract

Empirical studies of an abstract semantic model, augmented transition network (ATN), with Object Composition Petri Net (OCPN) to model multimedia presentations are performed in this paper. An ATN consists of a set of states and directed arcs and uses a regular expression as its input. The advantages to use a regular expression are its simplicity and ease of modification. Simulation experiments to compare ATN and OCPN are performed in this paper. The results show that ATN requires fewer nodes and arcs to represent a multimedia presentation than OCPN does. These results indicate ATN handles on-line multimedia presentations more efficient and requires less precious main memory space.

Key words: Multimedia Presentations, Augmented Transition Network (ATN), Regular Expression

1 Introduction

In this paper, a detailed comparison of ATN [3] with Object Composition Petri Net (OCPN) [5] is shown. OCPN is based on the logic of temporal intervals and Timed Petri-Nets. Multimedia objects are organized by the presentation sequence. OCPN augments the conventional petri net model with time duration and resource utilization on the places in the net. Many later abstract semantic models are based on a petri-net [2], [4], [6]. All these models use nodes and arcs to connect the media streams to form a multimedia presentation. Therefore, the numbers of nodes and arcs are essential for the multimedia browsing and searching. Since latter petri-net semantic models are similar to OCPN, OCPN is chosen to compare with ATN in this paper.

Conditions and actions in ATNs are used to control the quality-of-service (QoS) and synchronization, and can be separated into several smaller tables based on the input symbols. These separate tables can be put into secondary storage memory or other remote storages and be loaded into main memory when needed in any real-time presentation. Therefore, ATNs will not occupy a great deal of precious main memory space. Also, when the number of nodes increases in OCPN, the searching time and the complexity increase too. The searching time increases since OCPN is a left to right model and any searching needs to begin from the leftmost node. For example, if a user wants to fast forward to a certain point then the searching time will increase when too many nodes need to be traversed. The complexity increases because media streams are assigned to state nodes which connected by arcs. The number of arcs increases when the number of media streams increases. Hence, users have difficulty in understanding a presentation sequence under so many nodes and arcs. From above, we know that the numbers of nodes and arcs play important parts in ATN and OCPN. Simulation experiments to compare ATN and OCPN based on different numbers of media streams are performed in this study to show that ATN requires fewer nodes and arcs to represent a multimedia presentation than OCPN does.

The organization of this paper is as follows. Section 2 discusses how to use regular expressions and ATNs to model multimedia presentations. In section 3, simulations to compare ATN with OCPN are conducted. Conclusions are presented in section 4.

2 An Example of Multimedia Presentation Using regular expression and ATN

In this section, an example to illustrate how to use regular expressions and ATNs to model a multimedia presentation is demonstrated. The details of how to use regular expressions and ATNs are shown in [3].

*Partially supported by National Science Foundation under contract 9619812-IRI and the office of Naval Research under contract N00014-91-J-4126.

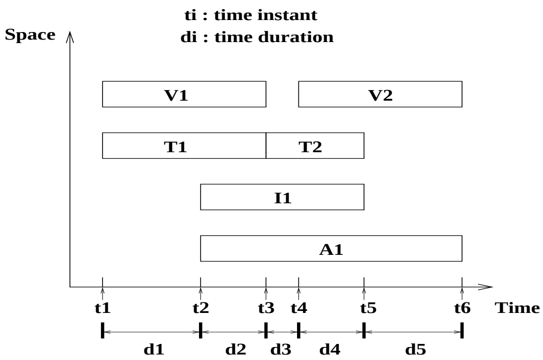

Figure 1: Timeline for Multimedia Presentation. t1 to t6 are the time instances. d1 is time duration between t1 and t2 and so on.

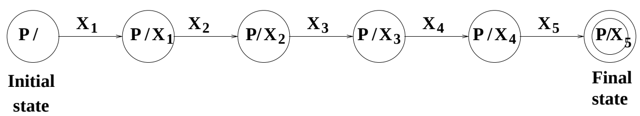

Figure 2: Augmented Transition Network for Multimedia Presentation.

2.1 Regular Expression

Figure 1 is a traditional timeline to show the temporal relations among media streams. In this example, the regular expression string is:

$$(V_1\&T_1)(V_1\&T_1\&I_1\&A_1)(T_2\&I_1\&A_1)(V_2\&T_2\&I_1\&A_1)(V_2\&A_1)$$

In this input example, at time t1, input symbol X_1 $(V_1\&T_1)$ is read and contains V_1 (video stream 1) and T_1 (text 1) which start to play at the same time and continue to play. At time t2, I_1 (Image 1) and A_1 (Audio 1) begin and overlap with V_1 and T_1. The delay time for I_1 and A_1 to display is equal to d1 and does not need to be specified in the regular expression explicitly since the regular expression is read from left to right and I_1 and A_1 will display when the input symbol X_1 is processed which takes the same time as delay for I_1 and A_1. This process continues until all the input symbols are read.

2.2 Augmented Transition Network

An ATN goes from left to right. The presentation shown in Figure 2 goes from the initial state (state P/) to the final state (state P/X_5). When the presentation begins, the transition goes from state P/ to state P/X_1 with the input symbol X_1 and the next transition goes from state P/X_1 to state P/X_2 with the input symbol X_2 $(V_1\&T_1\&I_1\&A_1)$. The presentation continues until the final state is reached.

ATNs can maintain QoS and synchronization by permitting a sequence of conditions and actions to be specified on each arc [3]. Each media stream contains a feature set $\mathcal{F}$ which has all the control information related to the media stream. The definition and the meaning of each element are defined as follows :

Definition 1: Suppose there are n media streams appeared in the input symbols. Each media stream has a feature set together with it.

$\mathcal{F}_i$ = {tentative_starting_time, tentative_ending_time, starting_frame, ending_frame, window_position_X, window_position_Y, window_size_width, window_size_height, priority} where $i = 1 \ldots n$.

The meaning of each element is illustrated below :

- **tentative_starting_time :** the original media stream desired starting time.
- **tentative_ending_time :** the original media stream desired ending time.
- **starting_frame :** the starting video frame number.
- **ending_frame :** the ending video frame number.
- **window_position_X :** the horizontal distance from the upper left corner of the computer screen.
- **window_position_Y :** the vertical distance from the upper left corner of the computer screen.
- **window_size_width :** the window size width of the media stream.
- **window_size_height :** the window size height of the media stream.
- **priority :** the display priority if several media streams are to be displayed concurrently.

3 Empirical Studies of comparing ATN and OCPN Models for Multimedia Presentations

In OCPN, each *place* (circle) contains the required presentation resource (device), the time required to output the presentation data, and spatial/content information. Each place is represented by a state node in the OCPN model. The *transitions* (bars) in the net

Table 1: Media type combinations at different media stream numbers for **case study 1** and **case study 2**

Experiment number	Total number of **media stream**	Number of **Text** streams		Number of **Image** streams		Number of **Audio** streams		Number of **Video** streams	
		Case 1	Case 2	Case 1	Case 2	Case 1	Case 2	Case 1	Case 2
1	25	9	10	9	5	3	5	4	5
2	50	12	20	15	10	11	10	12	10
3	100	21	40	28	20	29	20	22	20
4	1000	270	400	266	200	234	200	230	200
5	2000	497	800	522	400	481	400	500	400
6	3000	740	1200	764	600	762	600	734	600
7	4000	987	1600	1030	800	998	800	985	800
8	5000	1249	2000	1281	1000	1252	1000	1218	1000

indicate points of synchronization and the places processing. While in ATN, the state nodes do not store information. The information is stored in the condition and action table instead. Thirteen temporal relations were proposed in [1]. Based on these temporal relations, two case studies are performed to compare ATN with OCPN. These two case studies are as follows:

- **Case study 1:** different temporal relation combinations
- **Case study 2:** only *meets* temporal relation combination

In case study 1, arbitrary combinations of the thirteen temporal relations of media streams are used. While in case study 2, only *meets* temporal relation is considered, for example a slide presentation. Under case study 2, all the five types of media streams are displayed in each interval. They all have the same starting and ending times as shown in Figure 3. Figure 3 is part of a multimedia presentation in case study 2 and it contains 15 media streams with three intervals.

We want to compare numbers of nodes and arcs under these two case studies.

3.1 Experimental Parameters

We used random number generators to generate eight multimedia presentations that contain 25, 50, 100, 1000, 2000, 3000, 4000, and 5000 media streams. Four media types – text, image, audio, and video – are studied here. The combinations of media types for case studies 1 and 2 are shown in Table 1. For example, when the media streams number is 1000 in case study 1, there are 270 text streams, 266 image streams,

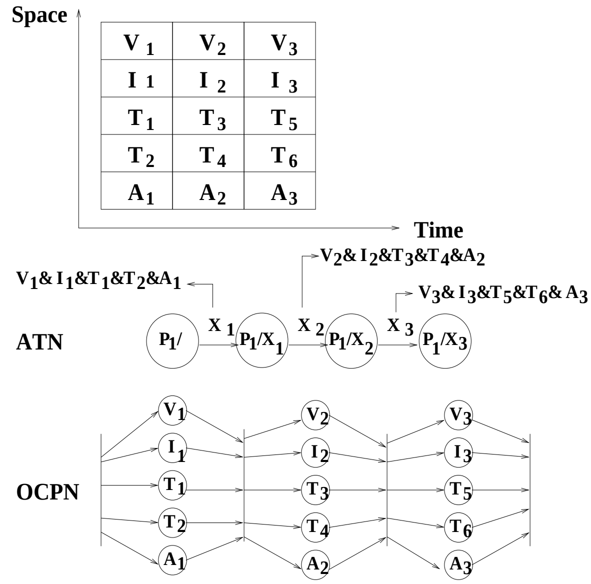

Figure 3: A multimedia presentation contains three intervals and each interval contains five media streams display at the same time which the starting and ending times are the same.

Table 2: Comparison of the numbers of **nodes** between ATN and OCPN

Experiment number	Number of media streams	Number of **nodes** in ATN		Number of **nodes** in OCPN	
		Case 1	Case 2	Case 1	Case 2
1	25	32	6	38	25
2	50	55	11	69	50
3	100	104	21	132	100
4	1000	1017	201	1305	1000
5	2000	1990	401	2607	2000
6	3000	3001	601	3919	3000
7	4000	4003	801	5214	4000
8	5000	5010	1001	6536	5000

Table 3: Comparison of the numbers of **arcs** between ATN and OCPN

Experiment number	Number of media streams	Number of **arcs** in ATN		Number of **arcs** in OCPN	
		Case 1	Case 2	Case 1	Case 2
1	25	62	10	76	50
2	50	108	20	138	100
3	100	206	40	264	200
4	1000	2032	400	2610	2000
5	2000	3978	800	5214	4000
6	3000	6000	1200	7838	6000
7	4000	8004	1600	10428	8000
8	5000	10018	2000	13072	10000

234 audio streams, and 230 video streams. Each media stream has its tentative starting time and ending time so the duration is obtained. In case study 2, each duration contains one video, image, audio streams and two text streams and each media stream has the same starting and ending times in each duration. We want to compare the numbers of nodes and arcs needed under ATN and OCPN approaches.

3.2 Results

Tables 2 and 3 compare the number of nodes and number of arcs between ATN and OCPN approaches under two case studies. From these two tables, it can be seen that ATN needs fewer nodes and arcs than OCPN in eight experiments under both case studies.

In case study 1, the number of nodes needed in ATN actually is very close to the number of media streams. The number of arcs needed in ATN is about the double of the number of media streams. However, since each node in OCPN has an incoming arc and an outgoing arc, the number of arcs is actually twice the number of nodes. When the media stream number increases, the difference between the number of nodes and the number of arcs increases, too. In case study 2, under different numbers of media streams, OCPN needs about 5 times more number of nodes than ATN does. This tells us ATN is much better than OCPN in case study 2 situation. The reason is that ATN creates a state node for each interval and OCPN needs to create a state node for each media stream in each interval. When comparing case studies 1 and 2 in Tables 2 and 3, we can see that the difference between the numbers of nodes and arcs is bigger in case study 2.

From the above results, we know that when a multimedia presentation contains more media streams, ATN needs fewer nodes and arcs than OCPN does. Therefore, ATN needs less memory space and less searching time as number of media streams increases than OCPN. An example of searching is to fast forward to a particular time point and display. All the nodes and arcs between the current time point and the target time point need to be traversed. In this situation, ATN performs better than OCPN since ATN consists of fewer nodes and arcs than OCPN. Moreover, since ATN contains fewer nodes and arcs, it provides a clearer view of the presentation than OCPN. Therefore, ATN is easier to manage, understand, and construct.

4 Conclusions

Some simulations are performed in this study. The results show that ATN needs fewer nodes and arcs than OCPN at different numbers of media streams. This makes ATN handle real-time multimedia presentations with fewer main memory space. Also, any editing of the original presentation sequence is easier because fewer numbers of nodes and arcs need to be dealt with.

An OCPN creates a node for each media stream which makes OCPN too complicated to understand when the number of media streams increases. On the other hand, an ATN uses input symbols to represent media streams displayed at the same time. This feature makes ATN simpler to manage, easier for users to understand and less main memory to handle real-time multimedia presentations.

References

[1] J. F. Allen, "Maintaining Knowledge About Temporal Intervals," *Commun. ACM*, vol. 26, pp. 832-843, Nov. 1983

[2] Yahya Y. Al-salqan and Carl K. Chang, "Temporal Relations and Synchronization Agents," *IEEE Multimedia*, pp. 30-39, Summer 1996.

[3] Shu-Ching Chen and R. L. Kashyap, "Temporal and Spatial Semantic Models for Multimedia Presentations," *1997 International Symposium on Multimedia Information Processing*, Academia Sinica, Taipei, Taiwan, pp. 441-446, Dec. 1997.

[4] Hui-Jung Chang, Tai-Yuan Hou, Shi-Kuo Chang, "The Management and Application of Teleaction Objects," *ACM Multimedia Systems Journal* (1995) Volume 3, pp. 228-237, November 1995.

[5] T.D.C. Little and A. Ghafoor, "Synchronization and Storage Models for Multimedia Objects," *IEEE J. Selected Areas in Commun.*, Vol. 9, pp. 413-427, Apr. 1990.

[6] H. Thimm and W. Klas, "δ-Sets for Optimized Relative Adaptive Playout Management in Distributed Multimedia Database Systems," *IEEE 12th International Conference on Data Engineering*, New Orleans, Louisiana, pp. 584-592, 1996.