Chapter 11

Scalable Virtual Environments

11.1. Introduction

Virtual environments are multi-user network applications in which the users (or participants) can move in a virtual world, communicate with other participants and interact with their environment. We call them large scale virtual environments when the number of participants is very large (dozens, maybe hundreds of thousands). Network games are an example of such applications.

While moving inside the virtual world, the participants meet and can communicate with each other by sharing various types of data flows such as their characteristics (coordinates, capability), text messages and audio or video data flows. If the network supports a multipoint transmission service, the application can benefit from it to increase the efficiency of the communication module. By means of the native multipoint, data sending to several participants can be carried out in a transparent fashion for the participant, with only one sending operation, provided the communications do not require reliable transmissions.

The multipoint extensions of the IP protocol defined by Steve Deering in the RFC 1112 standard make it possible to provide a multipoint transmission service in connectionless mode towards a "group" of participants.

Chapter written by Walid DABBOUS and Thierry TURLETTI.

The concept of group is very important because it makes it possible to have a "logical" meeting point in the network. It is as if the "world" were transformed into an open communication space where we can easily establish group communications with participants who are not pre-identified but have locally specified their interest in these communications. In this model called ASM (*any source multicast*), the network supports the establishment of multipoint broadcast trees as well as the packet relay towards the interested participants.

Since the ASM model did not experience a universal deployment because of an important protocol complexity and of the absence of an interesting economical model for the Internet service providers, another model has been considered by the IETF engineers. This is the SSM (*source specific multicast*) which is proposed in the hope of a better deployment. In this model, the very "elegant" concept of multipoint group which is used as a logical meeting point disappears. The SSM designers actually consider that the worldwide establishment of group communications is not feasible in the actual context of technology. Unlike the ASM model which enables any user of the network to send data to the group, in SSM there is only one communication channel that comes from a well identified source.

A priori, the ASM model seems to be an appropriate model for virtual environments: the participants need to communicate with other participants that they do not know beforehand. The concept of group in this model makes it possible to do that implicitly. However, a major problem still needs to be sorted out: the participants are not interested in all the exchanges in the virtual world and it is essential to add a filtering mechanism of the data received by each participant. If everyone was represented by a unique multipoint group, the participants would then be overwhelmed by all the exchanges in the world. A solution consists of dividing the world into zones, each represented by a distinct multipoint group. A participant who moves into the virtual world and who arrives in a given zone subscribes to the multipoint group corresponding to that zone. This makes it possible to decrease the amount of information *a priori* "useless" received by the participant. The zones thus correspond to the main interests of the participants. The more participants, the more necessary this approach is. It is therefore essential in the case of large scale virtual environments (or LSVE) for which the traffic control is a critical need. Several other issues must then be considered: what size to choose for the zones? How to manage the heterogenity of participants in terms of capability and different communication needs? What are the limits of the system? etc. A communication protocol for LSVE based on ASM should be able to answer all these questions. A big part of this chapter will be dedicated to the description of the SCORE-ASM communication structure that we designed and implemented for an application of a large scale virtual environment.

As far as the SSM service is concerned, at first sight it seems less suitable than the ASM service for the LSVE applications. In fact, the absence of the concept of logical meeting point at the network service level requires the introduction of signaling in order to identify and link the “neighboring” participants. In fact, there is no longer “free” way to communicate with the neighbors (when the participants are in the same zone). A communication protocol for LSVE based on SSM should therefore contain additional mechanisms, which in particular leads to more signaling in the network from the communication module. However, we will see that the possibility of filtering information coming from a specific participant (non-existent in the ASM model) opens more interesting prospects for the setting up of a large scale traffic control for participants who have different capabilities and different needs. The communication structure for LSVE based on SSM could therefore be more sophisticated than the one based on ASM, but it would make it possible to provide enhanced services for this type of applications. It is therefore suitable for this approach to study the compromise between the additional signaling generated by the absence of a logical meeting point in the SSM model and the best management of data traffic by this model. In this chapter, we will briefly describe the basic principles to design a communication structure for LSVE based on the SSM model.

The rest of the chapter is organized as follows: section 11.2 introduces the characteristics and the requirements of large scale virtual environments (LSVE). Section 11.3 describes the limitations of multipoint routing for the support of LSVE applications. Sections 11.4 and 11.5 describe the communication protocols for LSVE, based on the ASM and SSM models. Finally, section 11.6 concludes the chapter.

11.2. Specificities of the LSVE

11.2.1. *Scalability*

In the LSVE, each user is both information source and recipient. We commonly refer to this type of applications as “many to many”, opposed to the “one to many” applications which only contain one source and several recipients. This type of application is complex because it requires the simultaneous connection of a large number of users and each of them is often the source of several data flows with different contents and types. However, since at any moment each participant interacts only with a limited number of participants, it is therefore only interested in receiving the traffic related to them. When the number of participants increases, the percentage of information that a participant wishes to receive decreases (if we consider all the information sent by all the participants). The transmission of all the data packets sent by each participant to all the other participants can have disastrous consequences at several levels, pertaining to the quality perceived by the users. Inside the network, this can lead to a saturation of memory buffers at the router level and to more congestion on the transmission links. At the host machine level, the effects can lead to an overflow of the queues and a useless use of CPU resources. Besides, the use of a reliable point-to-point transmission protocol between the users (such as TCP) can only make the phenomenon grow in case of loss of packets, by forwarding data which are often useless.

11.2.2. *Interactivity*

A LSVE is formed of a set of entities controlled by interaction rules. For example, two participants face-to-face in the virtual world must be able to observe each of theirs actions. However, the current technology does not make it possible to process the information and to instantly communicate it. In the best case, the optic fiber networks reach 2/3 of the light speed, which gives for example a 140 ms minimum delay between Europe and Australia. This physical delay plus the processing time at the router level must be considered in the development of interactive broadcast applications. An example of inconvenience caused by a too long delay is noticeable in the satellite phones where the forwarding delay is 250 ms. The DIS (*distributed interactive simulation*) standard [SOC 95] recommends to set this interaction delay to 150 ms at the maximum. The companies that market games broadcasted in a network often consider that the quality of the game starts degrading as soon as this delay goes over 200 ms. Consequently, mechanisms of transmission control that are adapted to various needs of interactivity must be implemented. Their role is to hide the effects of the network on the application in order to maintain the interactivity.

11.2.3. *Heterogenity*

The LSVE can contain a very big number of participants who are distributed on several Internet domains. This characteristic sparks off a very strong heterogenity between the participants. This heterogenity firstly comes from the various capabilities of the links that connect the participants to the Internet, as well as various powers of calculation of the host machines. However, this heterogenity can also be expressed in terms of amount of requested details, consistency level and type of preferred data.

This high heterogenity between the users' preferences must therefore be considered by the LSVE so that the resources that they have are used as efficiently as possible, quantitatively as well as qualitatively. The mechanisms that are used must therefore find the right compromise amongst the state added in the routers, the bandwidth used and the satisfaction of the final user. Each participant of an LSVE thus can be characterized by two parameters that are its own: its main interests and its capability to receive and process in real-time the information that it receives. Consequently, the communication structure for LSVE has to consider the heterogenity amongst the participants for these two parameters and assign the resources that it has as well as possible in order to improve the quality of the information that each participant receives.

11.2.4. *Consistency*

Since they are broadcast interactive applications, the LSVE must provide a consistent perception of the virtual world at any moment to each participant. The needs for a given LSVE in terms of consistency are most often related to its needs in terms of interactivity. However, the Internet is a non reliable heterogeneous network, and each participant overcomes the effects of the network in a variable, uneven and unpredictable fashion (delay, jitter and packet losses). Thus, each participant receives a subset of various packets transmitted in the network, each packet being itself received with different delays. Another consequence of the variable delays of the network between participants leads to the fact that the order of reception of the packets by the participants does not always correspond to the order in which theses packets were sent. But some events must be transmitted in an ordered manner, in order to respect the causal order [BAL 96].

11.2.5. *Reliability*

According to the type of data transmitted among the participants of an LSVE, different transport protocols can be used. Some data can be transmitted with various streams, adapting themselves to the available resources at a given moment in the network (congestion control) or at the recipient level (stream control). The audio or video messages transmitted by the participants are a good example. However, the LSVE, because of their complexity, almost always need to enable a reliable communication on certain types of data. In the virtual worlds, this is notably the case during critical interactions, which are often brief, at the time of collisions, for example. The exchange of text messages between participants, via the use of a shared white board, for example, can be carried out only if a quasi-full reliability is guaranteed. However, the Internet is a best effort network and does not offer any guarantee on the quality of transmissions. There is therefore no guarantee on the fact that the messages will be delivered to the participants (and even less in the right order) and there is also no commitment regarding the transmission delays. Besides, the real-time characteristic of the LSVE most of the time prohibits the use of a forwarding protocol at the transport layer level. In fact, the exchange of acknowledgement messages and the forwarding delays most often make the forwarded information obsolete.

11.3. Multipoint limitations

The research work conducted so far in the field of scalability in the LSVE clearly highlighted the necessity to use multipoint in the development of a possible communication structure, capable of meeting the needs that are specific to this type of application. From then on, several proposals based on multipoint have been made. Sometimes, some of them only use one multipoint group [GAU 99]. Other proposals recommend the unlimited use of multipoint, regardless of the problems and limitations that are inherent in this type of Internet communication. In this chapter, we are going to try and identify these problems and these limits in order to obtain the broadest possible view of the networking issue of LSVE on the Internet.

11.3.1. *Routing*

A first limitation of the use of several multipoint groups is the direct result of multipoint routing protocols. They make it possible to build multipoint trees connecting the members of each group with each other. For each of the groups, all the routers present on the related multipoint tree must maintain a set of states. These states are stored in the routing and tracking tables of the multipoint traffic present within the routers. They specify the output ports from which the packets sent in the group will have to be duplicated. However, it would not be very realistic to consider these tables as infinite capability resources. Besides, the intra-domain multipoint routing protocols make it possible to build two types of trees: source based trees (DVRMP, MOSPF, PIM-DM, PIM-SM), and shared trees (CBT, PIM-SM). In the case of source based trees, one state per source and per group must be kept in each router. However, in the case of shared trees, only one state per group is kept in the routers, regardless of the number of sources present in the group. But in the LSVE, each participant is both source and recipient (but the grouping techniques of the participants can imply that a participant is a member of several groups without being a sender in each of these groups). Consequently, a large number of multipoint groups would involve a substantial cost within the network.

A second limitation in the use of multipoint appears when we consider the dynamic property of the LSVE. In fact, during a session, the participants have to subscribe or unsubscribe to/from multipoint groups. These two operations lead to an additional cost in terms of processing at the router level and in terms of bandwidth within the network. An IGMP subscribing or unsubscribing message is transmitted between the host machine of the participant and its first multipoint router, connected to its sub-network. This information is then forwarded (e.g. in a PIM message) to the first junction of the multipoint tree (if there is no other participant present in the group among the other users of the sub-network). The effect of this forwarding is an additional cost of processing at the level of each intermediate router. However, the extent of this additional cost strongly depends on the topology of the existing multipoint tree. The further the junction is, the higher the additional cost is. Besides, some multipoint routing protocols such as DVRMP are based on a periodical flooding of the messages in the network, in order to update the multipoint tree. This traffic has of course a cost, and is a new argument in support of the limitation of the number of multipoint groups in the LSVE.

11.3.2. *Subscriptions and unsubscriptions latency*

Another problem in the use of IP multipoint must also be considered. This is the latency related to subscription, i.e. the time between the sending of an IGMP subscription request and the possible reception of the first packet going around the group. The use of a large number of multipoint groups in the communication structure for LSVE necessarily implies a higher frequency of arrivals and departures in each of the groups.

In the previous section on routing, we saw that the consequence of the arrival of a new participant in a multipoint group is the update of some present states within the routers between the participant and the closest junction on the related multipoint tree. This update not only has an impact on the bandwidth and the processing at the router levels, but also in terms of subscription latency. However, the LSVE are interactive applications with time constraints that are generally strong depending on the application. In all the cases, this latency time must be considered and “hidden” from the final user in some way or another. The interactive property of the LSVE must in fact be kept during the whole session and regardless of the mobility of the participants in the virtual environment. This latency time can take different values during the session and according to the participants. In fact, if another participant present in the same local network has already subscribed to the multipoint group, then the subscription will be effective almost instantly. However, a participant connected in Australia and wishing to join a session to which only Europeans are subscribed will experience a latency time for the subscription that can last up to several hundreds of milliseconds.

Latency related to unsubscription is also a problem that must be specified: if some techniques [RIZ 98] enable a member of a group to quickly unsubscribe using the router that is specified on its sub-network, it would be unrealistic to imagine that this operation is instantaneous at the multipoint tree level. According to the authors of [KAS 99] concerning the use of several multipoint groups in the reliable multipoint transport, the router can wait for several seconds before forwarding this unsubscription message to the next tree router. This implies that even when an isolated participant (at the physical level) unsubscribes from a multipoint group, the sub-network on which the host machine of the participant is, continues to receive the packets sent in the group. Consequently, an unsubscription operation at the level of a participant enables the participant to quickly filter the traffic that it receives and thus save the resources of its host machine, without saving the network bandwidth (not immediately anyway).

However, the issue of subscription latency can have effects that are a lot more disastrous to the application than the issue of unsubscription latency. The only way that makes it possible to solve this problem is to force the application to anticipate the subscription to the multipoint groups which are susceptible to interest it after a time equal to subscription latency.

11.4. SCORE-ASM

This section presents SCORE-ASM [LET 99], [LET 04], a multi-agent communication protocol for large scale virtual environments that is based on the ASM model. Defining a communication structure for LSVE is especially complex due to the number of parameters to be considered and the requirements for this type of applications which are significant. Once these needs and these parameters are clearly identified, the next step is to gather the tools that are necessary to elaborate a solution capable of providing an overall satisfactory answer to the LSVE. That is how multipoint appears to be a tool that is essential to solve the scalability problem of LSVE because it makes it possible to implement information filtering techniques. The objective of SCORE-ASM is to enable the virtual environments to be deployed on the Internet, for several thousands of heterogeneous participants. SCORE-ASM carries out a data filtering at the transport layer level assuming, however, that each user is capable of receiving and sending from multipoint traffic. In order to select the information that it wishes to receive, each participant must dynamically subscribe to and unsubscribe from multipoint groups, depending on its interest for the content of the transmit data in each group.

This section is organized in the following manner. First of all, in order to be able to judge the efficiency of multipoint to solve the LSVE scalability problems, we define an overall assessing metric of the grouping techniques of the participants within the multipoint groups. Then, we describe the role of the agents as well as the various types of communication in SCORE-ASM. Finally, we give an overview of the connection mechanism of a participant to the virtual environment (VE), of the subscription update mechanism and of the VE clipping algorithm into cells.

11.4.1. *Assessment of the additional cost related to the use of multipoint*

For the final user, the ideal thing is to receive no useless information. However, this situation is not very realistic if we consider the cost of multipoint and, consequently, the limited number of multipoint groups that the application can use. It is therefore important to enable the users to filter the incoming traffic, by drastically reducing the amount of useless traffic. If we consider the concept of interest within virtual environments defined by a zone of interest surrounding each participant, we reach the following observation: the more chances a virtual entity has to interact with a participant, the more important the interest that a participant has for the traffic generated by this entity is. Consequently, we can refine the objective previously defined by trying to limit as much as possible the useless traffic by means of the resources that we have (the multipoint groups), while enabling the users to gradually filter this traffic. Thus, the useless traffic filtered in priority represents the traffic generated by the other participants with whom the probability to interact is the lowest. On the other hand, the participants have limitations in the bandwidth and CPU resources available. This leads to a maximum number of data flows (or participants) that they can receive and process, and therefore to a maximum size of the zone of interest.

For all these reasons, SCORE-ASM uses the satisfaction metric S defined by the following formula:

$$S = U_r/min(U_t, C)$$

where U_r represents the payload rate that is received and processed. The payload represents the traffic sent by the participants located inside the zone of interest of a typical participant. U_t represents the total payload rate, received or not received (limitation of the capacity of the incoming link), processed or not processed (limitation of the power of calculation). Finally, C represents the capability of the participant to receive and process the information that it receives in real-time. C therefore represents the maximum information rate that the participant is capable of handling, limited by its Internet connection and by its power of calculation.

We therefore consider here that when useless incoming traffic causes no disturbance to the quality perceived at the level of the application by the final user, its satisfaction is maximal ($S = 1$). In addition, if for a given participant, $U_r = C$, then S is also maximal. In fact, by definition, $U_r \leq U_t$. Considering the previous equality, $C \leq U_t$, therefore $min(U_t, C) = C = U_r$, and consequently $S = 1$, even when the payload has not been completely received. This property is justified by the fact that no communication protocol can enable a participant to receive an information rate that is higher than its maximum capacity.

This satisfaction metric is necessary to establish the best compromise between the payload rate received at the level of the participant, the added state in the network and the multipoint group. The use of this satisfaction metric does not make it possible to adapt to the weakest recipient (in terms of network connection or computing power), but to maximize the satisfaction of the participant who has the lowest S value. This approach is often referenced as “max-min equity” [BER 87].

11.4.2. *The role of the agents*

In SCORE-ASM, the *agents* are machines or processes placed at various locations in the network (for example, in the LAN of a campus, hosted by one or several ISP, etc.). The administrators of LSVE have the responsibility to gradually deploy these agents in various strategic points of the network, in this instance, close to potential customers.

One of the key points is that these agents do not receive any data that the participants exchange with each other. The use of the term “agent” rather than “server” is therefore not fortuitous. The agents are not involved in any way in the state calculations, in the synchronization between participants or in the centralization of the emitted traffic and its redistribution to the concerned participants.

The role of the agents is to dynamically define zones inside the VE, considering the participants’ distribution, and to calculate appropriate cell sizes according to the various densities of participants in these zones (see Figure 11.1). The task of the agents is also to periodically assess the satisfaction of each participant, by considering their respective capability, the size of their zone of interest, as well as the density of participants in the zone in which they are. Once these satisfactions are calculated, the agents can calculate new zones again (or on the contrary aggregate some of them) and modify the cell size in the zones where the least satisfied participants are.

Thus, the role of the agents is to carry out a dynamic clipping of the VE into cells of different sizes and to relate each of these cells to a different multipoint group. Finally, it must inform the participants of the set of the multipoint groups to which it must subscribe in order to be able to communicate with its neighbors.

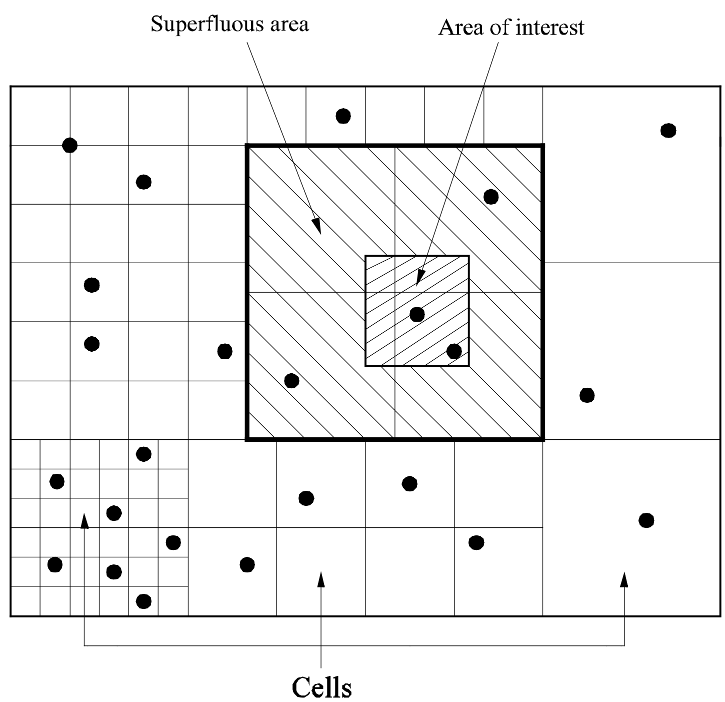

Figure 11.1. *Various cell sizes depending on the density of participants*

11.4.2.1. *Association of multipoint cells-groups*

The association of a cell of the virtual environment with a multipoint group can be divided into two steps. Firstly, the accurate identification and assignment of the cell in the virtual environment: its size, location, the virtual region in which it is as well as the neighboring communicating cells. The second step consists of associating a unique multipoint IP address to this cell. The result obtained after these two steps is called *mapping information* in the rest of the chapter.

The virtual environment is previously statically clipped into several big parts called *start zones*. These start zones are initially defined considering the intrinsic nature of the VE (room, floor, walls, etc.). Each start zone is then dynamically clipped into zones during the session; a zone being by definition a sub-part of the VE inside which all cells have the same size.

In order to facilitate the dynamic clipping of the start zones into zones, each start zone is statically pre-clipped into *zone units*, whose size is the minimum size that the zones can have. Thus, a cell is identified by the following three pieces of information:

– the zone in which it is;

– its position within this zone;

– the size of the cells in this zone.

11.4.2.2. *Assignment of multipoint groups*

Before any participant's connection, the VE is pre-clipped into start zones and a multipoint group is related to each of these start zones. During a session, four successive operations are carried out by the agents:

– clipping of each start zone into several zones, depending on the distribution of the participants in the start zone;

– calculation of the appropriate cell size for each zone;

– division of each zone into cells, depending on the size of the cell previously calculated, and assignment of a multipoint IP address to each cell of each zone;

– broadcasting of the information to the concerned participants so that they can subscribe to the multipoint groups related to the cells located in their neighborhood.

11.4.3. *Communications in SCORE-ASM*

SCORE-ASM is a communication protocol whose first objective is to make the deployment of the LSVE scalable on the Internet, while considering the heterogenity of its users. In order to meet this objective in a realistic manner, the multipoint groups act as traffic filtering tools for the participants. To each group is attached a cell of the VE such as a "label" providing information about the content of the traffic going around it. The participants are responsible for the state calculation, the data synchronization and the subscriptions to/unsubscriptions from the multipoint groups.

On top of this first level of distributed structure, a second level of communication is added through the agents. Their task consists of associating cells of different size to the multipoint groups in order to get even closer to the objective previously stated. However, the presence of the agents adds a level of complexity to the communication protocol. The agents must transmit the mapping information to the participants. In fact, they are going to take the final decision and decide when to subscribe to or unsubscribe from the multipoint groups. Now, the use of several data groups is precisely motivated by the need not to transmit the whole traffic to all the participants and therefore to make this type of application scalable.

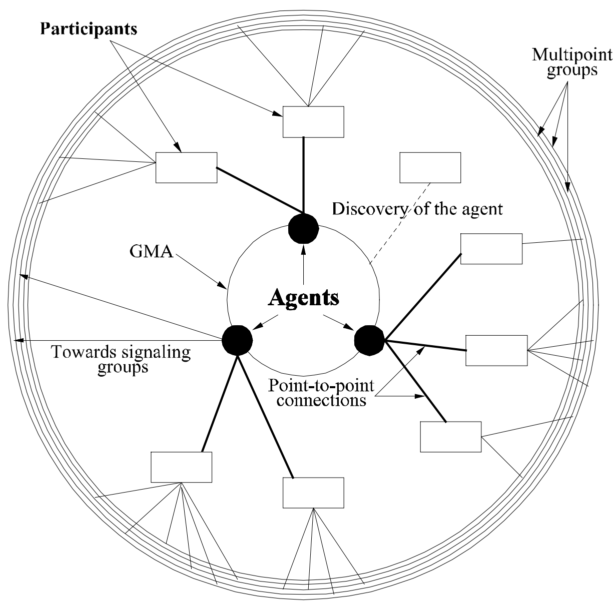

Figure 11.2. *A double distributed multi-agents protocol*

SCORE-ASM uses the concept of virtual clipping, based on the concepts of cell, zone unit, zone and start zone, in order to enable the participants not only to carry out a filtering on the incoming data traffic but also a filtering on the incoming signaling traffic (i.e. mapping information). While the multipoint groups on which the data generated by the participants circulate are related to the cells, signaling multipoint groups are related to the start zones. This association is carried out statically at this time. The connection between the start zones and the signaling group addresses is therefore known by all the participants as soon as they connect to the VE. The additional cost related to the use of these groups is negligible compared to the one related to the data groups.

In Figure 11.2, the different communication levels of the protocol appear:

– each participant subscribes to a number n of multipoint groups, with $n \geq 1$, but emits its traffic only in one group;

– each participant is connected to a unique agent, via a UDP point-to-point connection;

– the agents communicate with each other in a unique multipoint group which they are always subscribed to: the agent multipoint group (*AMG*);

– the new participants who want to connect emit "*Hello*" packets in the AGM group.

11.4.3.1. *Communication between participants*

The participants do not know *a priori* the list of participants connected to the LSVE. No point-to-point connection is established between the participants. Each participant is responsible for its subscription to the multipoint groups on which the traffic that circulates is interesting. In order to define the list of groups which a participant must subscribe to at a given moment t, the following parameters are necessary:

– the zone of interest of the participant;

– the list l_1 of the cells which meet the zone of interest of the participant. Each cell is identified by the zone in which it is and by its position in this zone;

– the list l_2 of cells that can meet the zone of interest of the participant in a time that is equal to the subscription latency and considering its velocity V;

– the mapping information for each of the cells that belong to the lists l_1 and l_2 thus established;

– the satisfaction S of the participant at the time $t - 1$.

The establishment of lists l_1 and l_2 is carried out dynamically by each participant during its displacements in the VE. List l_2 is especially important because each participant must anticipate the subscription to the multipoint groups in order to hide the latency related to the subscription.

11.4.3.2. *Participants-agent communication*

The participants are led to subscribe to two categories of multipoint groups:

– *data groups* which are related to each of the cells that meet their zone of interest. Let us note that a participant transmits data in only one multipoint group, i.e. the multipoint group related with the cell in which it is. It only behaves like a recipient for all the other data groups which it is subscribed to;

– *signaling groups* which are related to the start zones and in which the participants are only recipients.

The agents send the mapping information for the whole start zone in the signaling group that is related to it. This information is sent periodically in each of the start zones (with a period $P = P_{start\_zone}$) and contains the mapping information for all the zones that belong to the start zone (i.e. the cell size for each zone and the addresses of the related multipoint groups). The agents share the responsibility of emission in the various start zones.

For each of these groups, the participant must anticipate the emission of its subscription message considering its speed and the subscription latency. For the signaling groups, the period P_{start_zone} is also considered so that the participants can receive the mapping information before their main interests meet new cells that belong to new start zones.

Besides subscribing to signaling groups only as a recipient, each participant is connected in point-to-point mode to its agent via the UDP transport protocol. This connection is established between each participant and its closest agent.

Finally, SCORE-ASM uses a flow control mechanism between the participants and their respective agents. The participants have to send the control messages at a minimum rate to their agent, so that they can quickly detect a disconnection and update the number of participants in the virtual world and its various zones. Depending on the number of connected participants, the agents inform the participants about the minimum period P_{min} between two control messages. This period is transmitted with the mapping information sent by the agents in the signaling groups related to the start zones. At each emission of a signaling message, each participant resets a timer. If the participant did not reach a new zone unit during a time equal to a period fixed by the agents, it sends a *keep-alive* packet to its agent and resets its timer again. This very short packet only contains the identification number of the participant as well as its current zone unit. Depending on the rate of the incoming control traffic and their capability to handle the workload that is given to them, the agents can dynamically modify the value of P_{min} and thus control the incoming rate.

11.4.3.3. *Communication between agents*

SCORE-ASM must meet the scalability constraints that appear during the large scale VE deployment. To this end, it has been designed in a twofold distributed manner. On the one hand, the state calculation and the synchronization are carried out in a distributed way between the participants. On the other hand, the control traffic emitted by the participants does not converge to only one agent, but is distributed among several agents. In this section, only the communication part among the agents is presented. Section 11.4.6 explains in more detail the use of this information by the agents.

However, and in order to enable the agents to calculate the various densities of participants in the different zone units of the virtual world, the agents must in turn exchange a certain number of information. In fact, each agent knows only the number of participants who are connected to it as well as the position (the zone unit) in which they are. In order to exchange this information, the agents communicate in multipoint using a specific group that is dedicated to them: the AMG group. Four pieces of information are used by the agents in order to carry out the clipping of the VE into zones and cells:

– the total number of participants connected to the VE. This information enables each agent to calculate the average number of participants in each multipoint group;

– the number of participants in each zone defined at a time *t* by the agents in the VE. This information makes it possible to assess the density of participants in each of the zones;

– the number of participants in each zone unit. This information enables the agents to reform new zones while always attempting to get closer to a uniform distribution of participants in each zone;

– the list of zone units in which the participants with the weakest satisfaction *S* are. Thus, the agents have the possibility to define new zones and can calculate a smaller cell size in the zones where the unsatisfied participants are.

11.4.4. *Connection to the virtual world*

Before starting a session, we assume that each participant already downloaded the description of the local virtual world on its machine. This description includes the location of start zones and zone units which form these start zones. However, the only known multipoint address is the address of the agent multipoint group (the AMG group). No mapping information is communicated before the connection starts.

In order to start a logon procedure in the VE, a new participant must firstly be connected to an agent. To do this, the participants discover their agent by sending "Hello" packets within the AMG group. As soon as the participant retrieves the IP address of an agent, it can initiate a logon procedure to the virtual world, possibly going through an authentication procedure. After this procedure, the participant receives from its agent the list of multipoint addresses that are related to the start zones and its position in the VE. The participant can then listen in multipoint to the signaling group related to its start zone, in order to retrieve the mapping information that it is interested in and start to move in the VE.

Commentaire [hs1] : Is this discussing a human or not?
Answer : it is the software implementation launched by a human or an automated "robot"

11.4.5. *Subscriptions update mechanism*

This mechanism must enable the participants to update their subscriptions to the multipoint groups considering the new mapping information. It must meet the following conditions:

– keep the interactivity between the participants during the update;

– avoid any sudden increase in the incoming data traffic at the participant level, which would be due to a possible duplication of the packets in the old and new multipoint groups;

– limit the additional cost in terms of signaling traffic related to this mechanism.

The following operations must be carried out in order to update the subscriptions:

– as soon as a participant receives new mapping information, it must subscribe to the new multipoint groups related to the new cells that meet its zone of interest;

– we have already explained that the agents periodically broadcast the mapping information regarding the related start zones in the multipoint signaling groups. However, when the agents decide to change the mapping information in a given start zone, they must temporarily increase their emission rate in the related signaling group;

– the participant waits to have received *n* new pieces of mapping information before emitting in the new multipoint groups. However, if the participant starts to receive data in one of the new multipoint groups, it emits in turn in the new groups and stops emitting in the old ones;

– when a participant does not receive anymore packets in an old multipoint group since a time $T_{no\text{-}receive}$, it unsubscribes from this group.

It is important to note here that the participants subscribe to the new multipoint groups before proceeding to an effective change of emission groups. Two reasons are at the origin of this operation: firstly, waiting that all the concerned participants have received the new mapping information before proceeding to the changes. Secondly, waiting that the new trees of multipoint transmission have the time to establish themselves in the network, so that the firsts packets emitted in the related groups are broadcasted to all the members. This mechanism also makes it possible to synchronize the effective change of emission groups between the participants, even if packets containing the mapping information have been lost before waiting for some participants.

11.4.6. *Clipping algorithm*

The clipping algorithm is used by all the agents. It aims at dynamically defining zones and calculating for each of them an appropriate cell size considering the density of participants per zone as well as their satisfaction.

During the whole session, the agents periodically calculate the average density of participants per multipoint group by dividing the number of participants connected by the number of multipoint groups available for the application. This density is called *reclipping threshold* in the rest of the chapter. As arrivals, departures and displacements of participants occur in the VE, the agents calculate the densities of participants by cell in each of the zones.

The clipping algorithm is carried out in three successive steps:

– a first calculation makes it possible to define a cell size in each zone by considering only the distribution of participants in the VE (and not their satisfaction). This calculation is carried out by comparing the density of participants per cell in each zone with the reclipping threshold;

– in a second step, the least satisfied participants are identified, as well as their distribution in the VE. If a concentration of unsatisfied participants is detected in only one part of a zone, this zone is clipped into two new zones in order to isolate these participants; otherwise, the zone remains unchanged. A smaller cell size is then calculated in the zones containing the least satisfied participants. These participants will then be able to carry out a better approximation of their zone of interest. They will be able to reduce their superfluous traffic which will increase their satisfaction;

– finally, a third operation is periodically executed but less frequently than the two previous ones. During this operation, the agents can decide to aggregate two contiguous zones if the cell size is identical in these two zones and if they belong to the same start zone.

11.4.7. *Conclusions regarding SCORE-ASM*

The limitation of the number of available multipoint groups by the application is of course independent of SCORE-ASM. It can have detrimental consequences on the application because, given that the number of available multipoint groups is limited, this implicitly restrains the number of cells in the VE. Thus, when the density of participants per cell exceeds a certain threshold, the agents are no longer capable of improving the satisfaction of theirs participants (no other communication protocol could pretend to do that). However, various solutions can be contemplated in order to keep the quality perceived by the LSVE users. All these solutions propose a way of limiting the density of participants in the VE:

– having an "extensible" VE, whose size adapts depending on the number of users, so that the average density of participants in the VE always remains lower than a certain threshold;

– limiting the maximum number of users connected to the VE and building a VE that is big enough so that this density does not exceed a maximum threshold. This solution would effectively make it possible to have a large number of users connected, but we cannot use here the term scalability;

– using protocols such as those defined in MAAA [KUM 98], so as to enable a dynamic allocation of multipoint groups in SCORE-ASM when the density of participants exceeds a maximum threshold. However, this solution only solves a part of the problem by enabling the participants to reduce the average rate of incoming superfluous traffic. In fact, beyond a certain threshold, the density of participants would be such that even getting perfectly close to its zone of interest by means of very small cells, the incoming rate of a participant would exceed its capability.

The works on SCORE-ASM have been the subject of a thesis defended in December 2000 [LET 00]. The limited deployment of the ASM model and the emergence of the SSM model led us to study a specific communication structure for the SSM-based LSVEs.

11.5. SCORE-SSM

The deployment of LSVE applications is based on the large scale availability of the IP multipoint on the Internet. In the absence of a universal support of the ASM model, it is necessary to review the structure in order to adapt it to the existing technologies of group communication. The SSM model has been the subject of studies within the IETF with the objective to propose a multipoint service which can be deployed at the inter-domain level. We have therefore proposed a new communication structure which makes it possible to meet the requirements of the LSVE applications above a network supporting the SSM service. In the rest of this chapter we are going to briefly present these works which have been subject to a thesis defended in July 2004 [BAR 04] and which can be summarized in the following contributions:

– the elaboration of the SCORE-SSM communication structure: we have designed a communication structure for the SSM multipoint based VEs. This structure uses a two level dynamic filtering mechanism which enables the efficient broadcasting of data. The first filtering level (topological filtering) ensures the scalability of signaling and it is implemented by a hierarchy of agents. The second filtering level (individualized filtering) ensures the scalability of the data received by the participants and it is carried out through individualized control for each participant;

– the realization of a large number of experiments that prove the feasibility and assess the performances of the SCORE-SSM structure by comparing it to SCORE-ASM. Even if it is widely accepted that the ASM model is the best approach for LSVEs, the results obtained show that a communication structure based on SSM makes it possible to obtain performances that are as good as a communication structure based on ASM, with a signaling additional cost which is quite reduced and that can even bring further advantages such as “finer” granularity for filtering;

– the implementation of SCORE-SSM in NetBSD and its integration in V-Eye [PAR 04], which is a virtual world application developed by the Planète project at the INRIA of Sophia-Antipolis. This gave us the possibility to analyze the efficiency of the SCORE-SSM structure in real conditions.

The rest of this section is organized as follows: we are going to firstly explain the problematic and the choice of design that we have made for the SCORE-SSM communication protocol, then we are going to briefly present this protocol that we propose to use for group communication in LSVE. We will then conclude by presenting a series of prospects.

11.5.1. *Problematic*

In this section we will explain the hypotheses that we have chosen to design the SCORE-SSM communication protocol:

– the native SSM multipoint transmission service provided by the network: we use SSM here as a model of data distribution and as a filtering mechanism at the network level. The support of native multipoint service is crucially useful to obtain a communication protocol that scales up (i.e. in which a large number of users can be simultaneously connected, the users do not receive many inessential data, signaling is reduced, etc.). SSM being a "restrained" version of multipoint network, the communication protocol for LSVE based on SSM that provides the same services and functionalities to the final users will be more complex than a communication module based on ASM;

– heterogenity: the communication protocol must consider the heterogenity of the participants, both in terms of available bandwidth and power of calculation;

– filtering and scalability: the communication structure must efficiently integrate specific filtering requests of the participants. The participants need mechanisms to dynamically specify their main interest in a zone of the VE. It must also be possible to simultaneously support as many participants as possible without degrading the communication quality of the participants already connected in the VE;

– the real-time requirements of the multimedia communications between participants: the audio and video flows have real-time constraints that require reduced communication periods and a certain available band width. This imposes the use of fast data routing mechanisms in order to provide a good quality of service to the participants.

11.5.2. *Choice of design*

These various needs led to the following choice of design for the SCORE-SSM protocol:

– the use of a distinct SSM channel for each type of data flow emitted by the participants. Besides, since the participants may have heterogeneous interests, they have the choice to dynamically modify the area of interest of each type of data flow received;

Commentaire [hs2] : Is this correct? Yes.

– a solution based on an individualized filtering that gives the participants the possibility of expressing their specific interests and which is based on a topological filtering to distribute the global information of the VE in smaller zones. The topological filtering is hierarchical. It is about clipping the virtual world in smaller areas, each of these areas being controlled by an agent. This operation can be repeated in order to divide the areas into smaller zones, depending on the local density of participants. The hierarchical filtering is supposed to increase the scalability of the communication model by considering the characteristics of the participants (the size of a zone is determined depending on the capabilities of the participants) and the characteristics of the agents (each agent controls a region of the VE depending on its possibilities). The individualized filtering requires the previous determination of the neighbors, i.e. of the participants located in the area of interest of this participant, in order to be able to choose the type of flow needed from each of them.

– the exchange of multimedia flow such as audio and video. To this end, we consider that the use of SSM (based on the construction of a broadcast tree having the source of data as a root) makes it possible to reduce the communication periods to the minimum, because the participants receive the data from their preferred sources via the shortest path.

11.5.3. *SCORE-SSM structure*

We consider VEs in which the participants exchange data only with their closest neighbors; the communication held in other parts of the VE does not interest them. There is therefore no "*many to many*" communications but rather a large number of "*many few to few*" communications. In order to carry out such a model of communication, each participant must filter the data sent by the remote participants. In what follows, we will present the main characteristics of the SCORE-SSM structure: filtering of the inessential data, heterogenity of the participants and use of the SSM model.

11.5.3.1. *Filtering*

SCORE-SSM proposes a two-level filtering structure. Firstly, the VE is subdivided into disjoint zones, depending on the position and the density of the participants in the VE. A participant receives the approximate position of the other participants located in the zones that overlap with its main interests. Secondly, the participants calculate their closest neighbors among the participants retained by the first filtering level. The first filtering level of SCORE-SSM – called topological filtering – requires a signaling between agents. In order to improve the scaling of the communications, the VE is clipped into zones and all the participant displacements in a specific zone are recorded by an agent. However, so that the agents share the same view of the virtual world, an information exchange or signaling with each other is necessary. By "signaling", we mean here all the network traffic necessary to establish the communication between the participants (do not forget that there is no logical meeting in SSM and the neighbor mapping in the same zone requires an additional "signaling" traffic). Each agent is responsible for the signaling in one or several zones of the VE. Hence, the number of agents required to cover the virtual world depends on the number of participants connected to the VE at a given time. The agents decide on the clipping of the VE into zones. A dynamic clipping into zones depending on the distribution of the participants in the VE leads to a better approximation of the participant's main interest zones. The second filtering level – which we call individualized filtering – can be carried out by using SSM as data distribution model. Within a zone, a participant can choose a "sub-set" of participants considered as representing its closest neighbors with which the participant decides to communicate depending on its main interests. It en subscribes to the data flow emitted by these neighbors only. The use of SSM thus provides the tools to implement an individualized filtering by enabling the participants to communicate only with their "closest" neighbors, without receiving the data flows from the participants considered as "remote". This is not possible with the model ASM because the participant has to receive everything which is emitted in the multipoint group independently of the sources; with the ASM model, the only possible filtering level is the topological filtering. SSM also gives the possibility to the participants to receive or not a specific flow from a specific neighbor, given that each participant uses a different SSM channel for each type of data flow that it ansmits. This very accurate distinction between the data flow transmitted in the VE is at the basis of the individualized filtering of the data received by the participants. It enables each participant to very finely express its interest to receive or not the data flows from the other participants. Thus, it is worthwhile to use SSM, but at the expense of complex filtering mechanisms which should be implemented by the communication protocol in order to detect the neighbors (and the types of data flow they exchange) with whom a participant communicates. For a detailed comparison of the performances of the two structures (SCORE-ASM and SCORE-SSM), see [BAR 04].

11.5.3.2. *Heterogenity and multimedia flow*

SCORE-SSM gives each participant connected to the VE the possibility to communicate with a certain number of neighbors depending on its capabilities and its main interests. The quality of service perceived by a user does not depend *a priori* on the capabilities of the other users.

We can classify the participants according to different capabilities in terms of communication flow transmitted (transmission capability) or received (reception capability). In the VE, each participant must transmit its position and can transmit text messages. Besides, certain participants can transmit audio and/or video streams. These streams determine the transmission capabilities of a participant. Since the participants are capable of transmitting and receiving different types of streams, their reception capabilities are defined depending on the types of stream that a participant can receive. The reception capabilities of a participant consist of several zones of interest that have different sizes: the *circle of presence*, the *audio circle* and the *visual circle*:

– the circle of presence corresponds to the visual field of the participant: the participant can detect the displacements of the participants located at a distance which is smaller than a given value, called its visual radius. Besides, the participant is capable of receiving the text messages sent by the participants located in its visual field;

– the audio circle of the participant corresponds to the listening area (all the participants located at a distance that is smaller than the audio radius can be heard);

– the visual circle of the participant corresponds to the vision area (all the participants located at a distance that is smaller than the visual radius can be seen).

Each participant has two additional parameters: the maximum number of audio and video streams that it is capable of receiving simultaneously. These two values are upper limits that help refine the capabilities of a participant. The audio circle and the video circle are calculated from these values: the larger the maximum number of incoming audio stream, the more it is possible to enlarge the audio circle of this participant. For example, the value of the audio radius must be reduced when the number of participants who send audio streams to the neighborhood of a participant is higher than the maximum number of audio streams that it is capable of receiving. This traffic control makes it possible to reduce the packet loss observed by a participant due to an excessive overload of the network resources by the audio and video streams received. Intuitively, the video circle is included in the audio circle, which in turn is included in the circle of presence: in fact, the stream rates follow an inverse variation curve: the rate of the video stream is higher than the one of the audio streams, which is higher than the rates of position streams and text messages.

11.5.3.3. *Correspondence with the network multipoint*

The variation of the density of participants in different regions of the VE is measured by the agents and can trigger a new reclipping of the virtual world into zones. We recall that each participant uses a different multipoint channel for each type of data that it transmits: a channel for its position, one for the textual data, one for the audio stream and one for the video stream. In order to receive the data of the participants located in its zone of interest, a participant must subscribe to the multipoint channels corresponding to its closest neighbors and depending on its interests for the text, the audio and the video. We consider that all the participants use the same IP multipoint group address, G_{pos}, for the position stream. Likewise, a unique multipoint group address is used for the audio streams G_{audio} and another address G_{video} for the video streams. This makes it possible to reduce the additional cost of signaling and to improve the performances of SCORE-SSM. In fact, the use of the SSM model can generate large routing tables for the routing of multipoint packets on the Internet. The issue of aggregation of the entries (*S*, *G*) in the routing tables is always subject to research. The routing tables that have large sizes increase the problems of scalability as well as the packet transmission delays, because of longer research times. By using an addressing system that is easy to aggregate, we can limit the scalability problems, as it is the case here, by using a unique multipoint group address per type of data streams, only the source address is different.

11.5.4. *Prospects regarding SCORE-SSM*

The multipoint is still not always deployed on the Internet. However, most of the existing virtual worlds are proprietary VEs and have been built with a very specific purpose (military training, online games, etc.). In our study, we considered virtual environments in which the participants move, meet other participants and communicate with each other through multimedia streams. This type of application looks like an "enhanced" version of the well-known IRC service (*Internet relay chat*).

Commentaire [hs3] : Is this OK? Yes.

The SCORE-SSM structure that we have presented sets an individualized filtering at the level of the network by using the network mechanisms during deployment (IGMPv3 is delivered along with the last versions of Windows, Linux and *BSD system software). A solution at the network layer level is more efficient in terms of exchanged traffic in the network compared to the filtering solutions at the application level.

We have implemented SCORE-SSM in C++ above the operating systems NetBSD, Linux and Windows. The network communication module (library), an "agent" and a "client" application that uses this SCORE-SSM library have also been integrated in an LSVE application called V-Eye [PAR 04], which has been developed by the Planète project at the INRIA Sophia-Antipolis.

To end this chapter, we present here some prospects for future work related to the SCORE-SSM communication structure:

– LSVE must be able to meet the "persistence" and "scalability" conditions, i.e. the VE must be functional at any time and it must be possible to enlarge it. Of course, the addition of new domains, new participants or new agents, should not stop the current progress of the virtual world for the participants that are in it;

– the LSVE must be able to operate even without the presence of the multipoint service everywhere in the network. We can imagine designing a communication protocol that combines the native network multipoint with the applicative multipoint for the participants who do not have one. We can use a known technique in the *peer-to-peer* networks to discover who the neighbors of a participant in the virtual world are and use the native network multipoint to send the data streams to the various multipoint channels. In this context, the agents can be considered as specific nodes of the network that play a similar role as the *super nodes* in the peer-to-peer networks;

– the use of the SSM model as a data distribution model assumes that it is possible to have packet losses due to the use of UDP as a transport protocol for multipoint. A reliable transmission mechanism can be added for flows that do not tolerate packet losses.

11.6. Final comment

It is very likely that the large scale VEs experience a short-term rise on the Internet. In fact, their use is not restricted either to the field of military training or to network games. We can easily imagine a virtual music or movies shop, in which the users move around like in a real shop. They thus have the possibility of going through the shelves, viewing the movie trailers, listening to an audio clip, asking advice to a seller or other customers before buying some items. We saw here that these applications can well benefit from the underlying distribution model (be it ASM or SSM) in order to provide the best service to the users.

11.7. Bibliography

[BAL 96] BALDONI R., MOSTEFAOUI A., M. RAYNAL M., "Causal delivery of message with real-time data in unreliable networks", *Journal of Real-Time Systems*, vol. 10, p. 1-18, 1996.

[BAR 04] BARZA L., "Une Architecture de Communication pour Environnements Virtuels Distribués à Grande Échelle basée sur les canaux multipoint", PhD Thesis, 2004.

[BER 87] BERTSEKAS D., GALLAGER R., *Data Networks*, Chapter 6, p. 524-529, Prentice-Hall, 1987.

[GAU 99] GAUTIER L., KUROSE K., DIOT C., "End-to-end Transmission Control Mechanisms for Multiparty Interactive Applications on the Internet", in *Proceedings IEEE INFOCOM*, 1999.

[KAS 99] KASERA S. K., HJALMTYSSON G., TOWSLEY D., KUROSE J., "Scalable Reliable Multicast using Multiple Multicast Channels", in *IEEE/ACM Transaction on Networking*, 1999.

[KUM 98] KUMARY S., RADOSLAVOV P., THALER D., ALAETTINOGLU C., ESTRIN D., HANDLEY M., "The MASC/BGMP Architecture for Inter-domain Multicast Routing", *Proceedings of ACM SIGCOMM*, 1998.

[LET 99] LÉTY E., TURLETTI T., "Issues in Designing a Communication Architecture for Large-Scale Virtual Environments", in *Proceedings the 1st International Workshop on Networked Group Communication*, Pise, 1999.

[LET 00] LETY E., "Une architecture de communication pour environnements virtuels distribués à grande échelle sur l'Internet", PhD Thesis, 2000.

[LET 04] LÉTY E., TURLETTI T., BACCELLI F., "SCORE: a Scalable Communication Protocol for Large-Scale Virtual Environments", in *IEEE/ACM Transactions on Networking*, 2004.

[PAR 04] PARMENTELAT T., GOURDON A., TURLETTI T., KUROSE J., "A Very Large Environment for Multimedia Conferencing", *INRIA Technical Report No RT-0296*, 2004 (available from http://www-sop.inria.fr/planete/software/V-Eye).

[RIZ 98] RIZZO L., "Fast group management in IGMP", *Proceedings of Hipparch Workshop*, 1998.

[SOC 95] IEEE COMPUTER SOCIETY, *IEEE Standard for Interactive Distributed Simulation*, Number Std 1278.2, 1995.