

Incorporating Global Positioning Data in Real Time P2P Audio/Video Streams for Mobile Devices

JOSÉ-VICENTE AGUIRRE¹, RAFAEL ÁLVAREZ²,
LEANDRO TORTOSA³ and JOSÉ-FRANCISCO VICENT⁴
Departamento de Ciencia de la Computación e Inteligencia Artificial
Universidad de Alicante
Campus de Sant Vicent del Raspeig, Ap. Correos 99, E-03080, Alicante
SPAIN

This work was partially supported by the Spanish grants GVPRE/2008/363

Abstract: - We propose an original method to geoposition an audio/video stream with multiple emitters that are at the same time receivers of the mixed signal. The achieved method is suitable for those cases where a list of positions within a designated area is encoded with a degree of precision adjusted to the visualization capabilities; and is also easily extensible to support new requirements. This method extends a previously proposed protocol, without incurring in any performance penalty.

Key-Words: Codification, geodesic coordinates, Multiparty, Stream, VoIP, Videoconference, P2P, Security, Pocket PC, Smart Phone, PDA

1 Introduction

Nowadays, the increase of domestic available bandwidth and computing power is making videoconferencing a reality in situations that were unthinkable a few years ago. Applications like skype[1], qik, etc., try to take advantage of this scenario and provide new features and functionality. As an example, by using videoconferencing, people residing in distant areas can perform meetings and cooperative work successfully, despite the possible geographical restrictions.

Any application where multiple users can interact by means of audio and video channels under a Peer to Peer (P2P) communication [1] [2], suffers many problems related to the available bandwidth and computing power [6] [8], which can be critical as the number of users rises. Any solution to these problems implies restrictions in the way the users interact with each other or in the quality of the communication.

The geopositioning of any type of multimedia content is a current trend that can be seen in all Internet content sharing services (picasa, qik, etc.). Being able to transmit securely the location of the audio/video stream in real time is an interesting application, especially when it does not imply any loss of quality which is already severely limited by the available resources and bandwidth.

There are several ways to encode global positioning coordinates. The geographic coordinates system uses two angular coordinates in a general spherical coordinates system used in Astronomy. The two angular

coordinates are related to the following angles, measured from the center of the Earth:

- The latitude of a point on the Earth's surface is the angle (measured in degrees) between the plane of the equator and the straight line segment that joins the point to the center of the globe.
- The longitude of any point on the Earth's surface is the measure of the angle (in degrees) between the planes that contain the point, the Earth's axis and the Greenwich Meridian (adopted as reference).

Another coordinate system is the Universal Transverse Mercator (UTM). A position on the Earth is referenced in the UTM system by the UTM zone, and the easting and northing coordinate pair. The easting is the projected distance of the position from the central meridian, while the northing is the projected distance of the point from the equator. The point of origin of each UTM zone is the intersection of the equator and the zone's central meridian.

The use of a concatenation of coordinates for the transmission of all the positions of a set of nodes could be a waste of precision or possible locations, when transmitting the information of the position from many nodes simultaneously is required and the transmission bandwidth and the data size are relevant.

We have observed certain restrictions in the underlying problem, which would provide an opportunity for reducing the amount of data transmitted and allowing it to be adapted to our voice/video transmission system [3].

In this paper, we propose a technique for adding a geopositioning signal corresponding to the N participants in a multi-party videoconferencing, so that the accuracy is tailored to the bandwidth unused by the audio/video channels, without producing delay. The main idea is to propose a lossy positioning information compression technique based on three factors: the minimum precision required, the range of probable positions and the number of bits available for geopositioning information.

2 Notation

We use the following notation in this paper:

- A devices ring is a subset of devices with modular sequential order and some characteristics in common.
- R_R is the real devices ring.
- R_C is the connected devices ring.
- N_R is the total number of machines in R_R
- N_C or N is the total number of machines in R_C .
- n is the current machine in R_C .
- I is the total number of iterations of the algorithm.
- i is the current iteration of the algorithm.
- $Mix_{y=0}^x(v_y)$ is a packet mixing function that mixes from $y=0$ to x .
- *Parallel* $\{\{Job1\} \{Job2\}\}$ denotes that jobs 1 and 2 are carried out in parallel.

3 Protocol specification

In the following, we describe the protocol specification and requirements.

3.1 Requirements

This method for embedding the geopositioning signal in an audio/video data stream in real time, although valid for many other purposes, is a part in the development of our secure communications system [3], determining the supposed initial situation with its related bounds and problem restrictions.

Firstly, the system provides an environment where the connection changes and, therefore, the number of currently connected devices with the capability to participate in an audio or audio/video communication, N_C , out of the total number of ready devices, N_R , is known and efficiently managed.

On the other hand, the system must provide mechanisms to establish and manage audio or audio/video communications in real time, ensuring the integrity, confidentiality and non-repudiation of data transmitted in the stream according to the policies specified for such communication.

Finally, the method used for mixing and transmitting all audio and video sources [3] [4] provides, depending on the terms of communication, a variable difference between the number of bits required by each package and the Maxim Transfer Unit (MTU) of the network, which can be used to transmit small amounts of data, like geopositioning. It is not optimal to send this data in new packages while the stream of data in real time is happening, since the high latency of data networks for mobile [5] networks would introduce significant delays in the communication noticeable to the human ear.

3.2 Position Coordinates Encoding

The amount of bytes required to represent the coordinates in different systems and different precision is shown in Table 1: Geographical Coordinates with second precision (approximately, and depending on the area, corresponds to a precision of about 25m) and of cents of seconds (around 0,2m); and UTM coordinates with a precision of 10m and 0,1m.

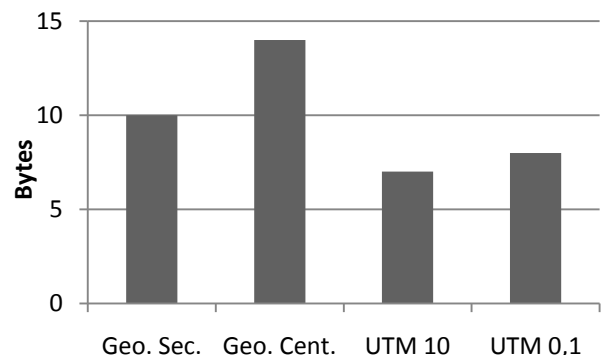


Table 1 Number of bytes required for coordinate representation.

Due to the specific requirements of the problem under study, the concatenation of all node positions could mean a waste of precision or possible locations. The different coordinate systems have been designed for global positioning; in our case we will need to position a certain number of nodes in their respective geographical locations; the further away they are located from each other, the location precision becomes less relevant while as they get closer the total possible area for location becomes smaller.

	Precision	Latitude Precision	Longitude Precision	bits Lat	bits Lon	bits Total	Bytes
Sec.	0.01	0.24 m	0.31 m	26.65	25.65	53	6.63
	0.1	2.40 m	3.08 m	23.48	22.48	46	5.75
	1	24.01 m	30.76 m	20.31	19.31	40	5.00
	10	240.08 m	307.64 m	16.98	15.98	33	4.13
Deg. Min.	1	1440.50 m	1845.83 m	14.40	13.40	28	3.50
	10	14405.00 m	18458.33 m	11.08	10.08	22	2.75
	1	86430.00 m	110750.00 m	8.49	7.49	16	2.00
	10	864300.00 m	1107500.00 m	5.17	4.17	10	1.25

Table 2 Byte and precision tradeoff for latitude and longitude in an area 40°N from the equator.

This reasoning comes from the fact that the location signal to be transmitted is meant to be a general view of the position of every node that will be represented as a set in a mobile device. Thus, precision is useless when dealing with large distances, if more precision is needed for a specific node's location then a direct query for that node's location will be performed, but this is out of the scope of this paper.

In order to achieve this, we define an action window where all nodes' location will be represented with a level of precision suitable to the visualization capabilities and free bits for transmission available. The encoding scheme is shown in Fig. 1.

In the following, we detail the different steps in the encoding.

3.2.1 Action Window

The action window is defined as an area of the Earth surface that contains all node locations. A huge level of precision is not required to define this window although it must be able to adapt to large scales (whole Earth) or small ones (a town) as required.

The action window is defined by its anchor point (upper-left corner) and a certain length. The latter is used to specify size of the square containing the area. This is done in this way to reduce the number of bytes required to represent it, avoiding the need to encode two points or any unnecessary precision for its purpose.

Observing Table 2, we see that the reason to choose a minutes precision level is because it achieves a good tradeoff between precision and the amount of bytes required.

The four remaining bits in the encoding of the anchor point are used to encode the scale in which the square side length will be specified in the next byte as shown in Table 3. In this table we can observe the 16 different ways to distribute the 8 bits of the byte representing the action window side size, that have been chosen so they offer an adequate range of maximum distances and minimum precisions.

Position	Precision and bits				Min. precision	Max. distance
	Deg	min	sec	Cent sec		
0	Precision	1	1	1	1	1°
	bits	8	0	0	0	
1	Precision	1	1	1	1	1'
	bits	0	8	0	0	
2	Precision	1	1	1	1	1"
	bits	0	0	8	0	
3	Precision	1.5	1	1	1	1.5°
	bits	8	0	0	0	
4	Precision	1	1	1	12.5	0.125"
	bits	0	0	5	3	
5	Precision	1	1	2	25	0.25"
	bits	0	0	5	3	
6	Precision	1	1	7.5	1	7.5"
	bits	0	5	3	0	
7	Precision	1	2	15	1	15"
	bits	0	5	3	0	
8	Precision	1	7.5	1	1	7.5'
	bits	5	3	0	0	
9	Precision	2	15	1	1	15'
	bits	5	3	0	0	
10	Precision	2.75	20.6	1	1	20.6'
	bits	5	3	0	0	
11	Precision	1	3.75	56	1	56"
	bits	2	4	2	0	
12	Precision	10	37.5	563	1	563"
	bits	2	4	2	0	
13	Precision	1	1	3.8	93.8	0.938"
	bits	0	2	4	2	
14	Precision	1	15	56	1406	14.06"
	bits	0	2	4	2	
15	Precision	1	15	225	1	225"
	bits	4	2	2	0	

Table 3 Different encodings for the action window side size.

3.2.2 Position List

In the proposed encoding method the precision level is meant to always be less than the maximum achievable precision. In our case of study we define the maximum precision as the one with which a coordinate can be selected within our action window, visualized in full screen on a mobile device with a maximum resolution of 640x480 pixels. The maximum precision would be a single pixel, so the maximum achievable precision would be the action window side size divided by the greatest dimension of the resolution (640).

Each node will be positioned in the action window determining its position within an n by m elements grid. Since, in order to successfully represent graphically a node, a square of 10 pixels of side will be used, we consider a grid of 64x64 elements as an adequate representation precision.

Considering the required precision, one and a half bytes (12 bits) allow concatenating the different node positions while only wasting 4 bits when the number of nodes is odd.

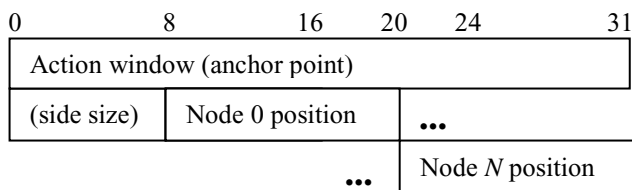


Fig. 1 Encoding bit arrangement.

3.3 Protocol Definition

The protocol proposed in previous papers [3] [4], consisted, regarding the transmission of an audio/video stream, of a stream establishment phase and of a transmission phase. The algorithm corresponding to the transmission phase is shown in Fig. 2, where it can be observed that it takes $\log_2(N)$ iterations for each transmission during the sub-mixing phase.

The following additions to the protocol have been necessary in order to incorporate geopositioning to the data stream:

- Reception of each nodes position during the stream establishment phase. With these data the node that started the communication computes, as shown later, the action window parameters.
- In the sub-mixing phase, each node encapsulates the geopositioning data that it has available on each iteration within the audio/video data.
- In case that a dynamic change of the action window parameters is required during the communication, the starting node is in charge of sending the corresponding control signal encapsulated within the stream as specified in the protocol.

```

Function TransmitVoice (VoicePacket myVoice, int numNodes, int
myPosition)
{
    N= numNodes;
    n= myPosition;
    AllPacketReceived.add ( myVoice );

    For (i=1; i < log2(N); i++)
    {
        NodeDestination = n + 2i-1;
        NodeOrigin = n - 2i-1;

        Parallel
        {
            {
                PacketReceive = receive(NodeOrigin);
                AllPacketReceived.add ( Mix (PacketReceive,
                    AllPacketReceived [i-1] ) );
            }
            {
                PacketSend = AllPacketReceived [i-1];
                Send(NodeDestination, PacketSend);
            }
        }
    }
    Float X= log2(N - 2i-1)

    If ( x≠[x] = 1 )
    {
        NodeDestination = n + 2i-1;
        NodeOrigin = n - 2i-1;

        Parallel
        {
            {
                PacketReceive = receive(NodeOrigin);
                AllPacketReceived.add(Mix(PacketReceive,
                    AllPacketReceived[i-1]));
            }
            {
                PacketSend = AllPacketReceived [ log2(N - 2i-1) ];
                Send(NodeDestination, PacketSend);
            }
        }
    }
    Else
        TransmitVoiceLastPackets (VoicePacket myVoice,
            int numNodes, int myPosition);
}
    
```

Fig. 2 Transmission phase algorithm.

3.3.1 Action Window Algorithm

The specification of the action window parameters is done with the following process:

1. Determine the minimum and maximum longitude and latitude of all node positions, obtaining two coordinates defining a window.
2. Extend the obtained area with a distance to the window frame. If not specified otherwise, this is taken as 300 meters.
3. The anchor point is obtained by rounding to the nearest position to the one calculated for the window, with a precision level of minutes. This rounding process will always try to find the

upper and left-most representable point available.

4. By subtracting the lower right corner point with the computer anchor point, we establish the greatest distance (horizontal or vertical) as the square side length.
5. To represent the square size length we take the most precise representation in Table 3 that allows reaching the required distance.
6. The action window is the one obtained after all rounding processes.

ago, then the action window is changed with a frame size corresponding to the difference between a hypothetical new action window with a frame of size zero and the previous action window.

4. After an hour has elapsed, a new action window is computed and if the encompassed area is two magnitude units smaller the it is changed with standard frame size.

3.3.2 Action Window Dynamic Change Detection Algorithm

In order to detect that a change of action window parameters is required the following procedure should be executed:

1. Determine out of all nodes, if any of them lie in the first or last row or column.
2. If any of them satisfies such criteria and the action window has not changed during the last hour, then the action window is changed according to the standard frame size.
3. If any of them satisfies this criteria and the action window has been changed less than hour

4 Results

The encapsulation of the geopositioning data produces the desired effects over the audio/video transmission in real time. Since the positioning data is transmitted using less than 50 bytes (for node sizes inferior to 30, see Fig. 5), the audio/video quality or the transmission times are not affected. Sizes over 60 bytes would provoke serious quality degradation in the video transmission without serious modifications in the stream management scheme.

In Fig. 4 we can observe that the only encoding scheme that never uses more information than what is representable is the one proposed. Since the other schemes use fixed precision, they always reach a point

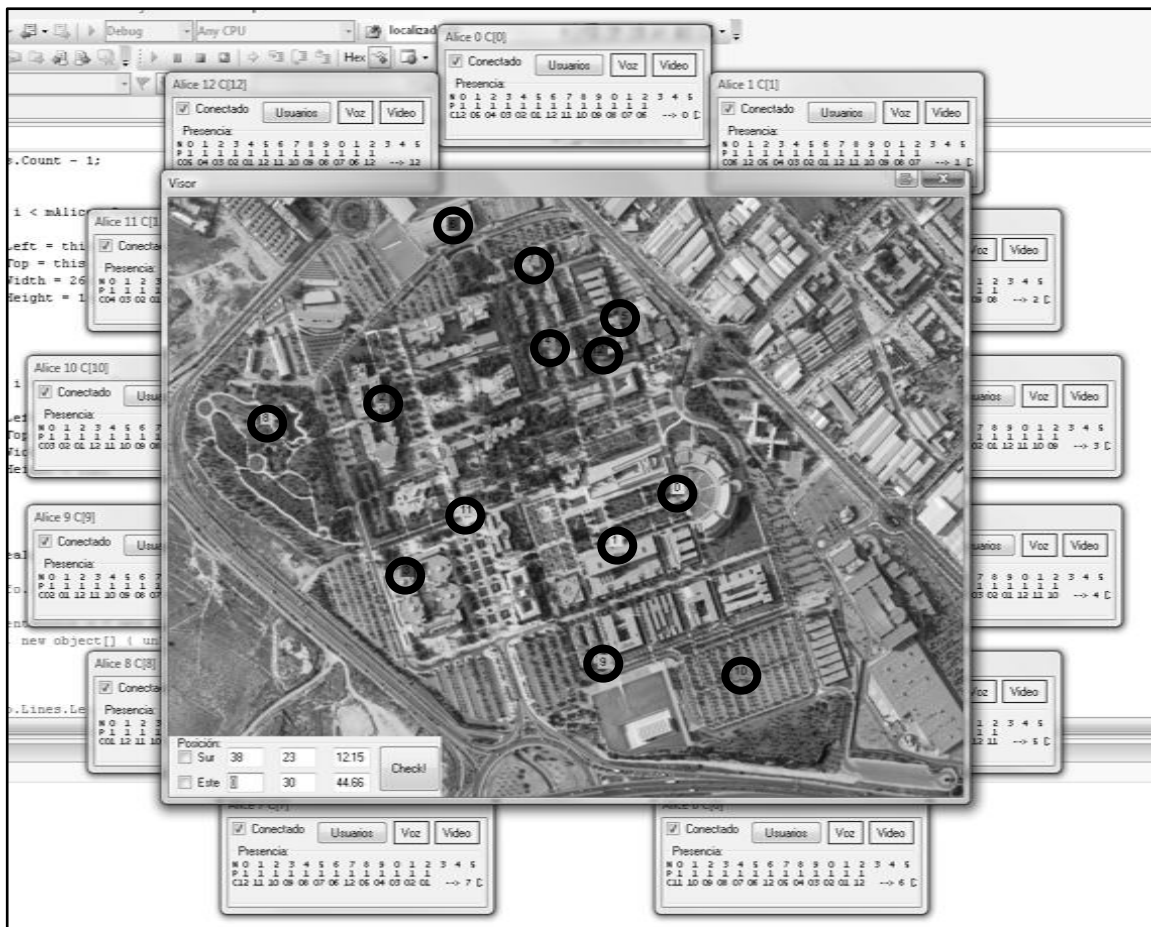


Fig. 3 Simulation with 13 Nodes.

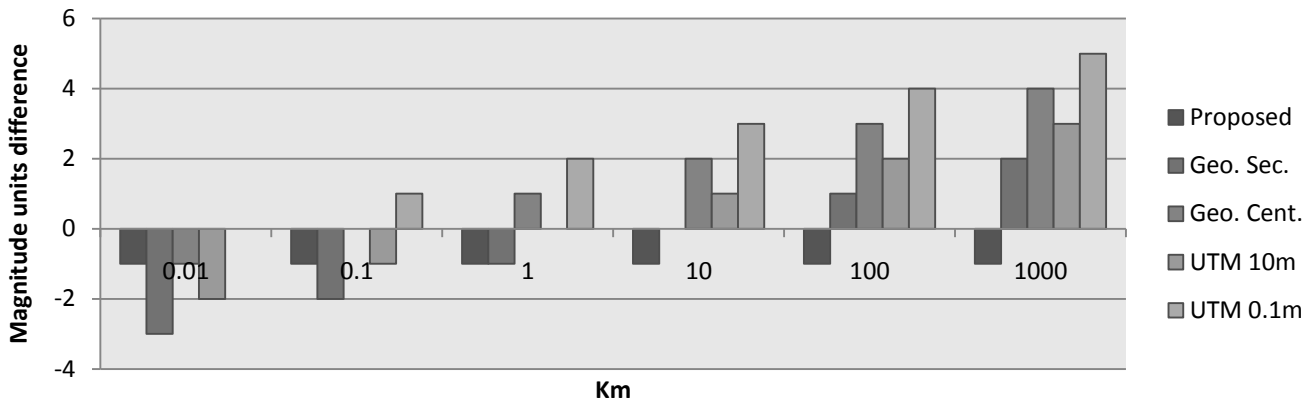


Fig. 4 Magnitude units difference between data precision and maximum achievable precision.

where more information is transmitted than what can be represented. But even in the case of dynamic precision the proposed scheme still more efficient taking up less bytes for the same precision level; as shown in Fig. 5. This size savings increase with the number of nodes.

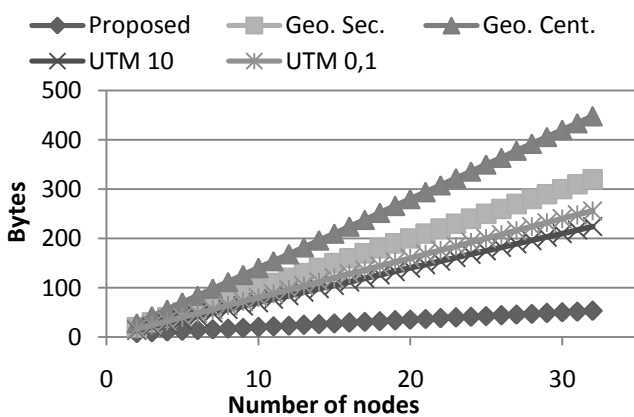


Fig. 5 Encoding sizes in bytes in relation to the number of nodes.

5 Conclusions

We have proposed an original method to geoposition an audio/video stream with multiple emitters that are, at the same time, consumers of the mixed signal.

The obtained method is suitable when a list of positions within a known area is encoded with precision tailored to the visualization capabilities of the target device. Nevertheless, it is easily adaptable to new precision requirements, as well as parameterized data precision.

The method is designed as an extension to the previously proposed protocol; adding audio/video signal geopositioning capabilities in real time (see Fig. 3) without incurring in any significant performance penalty or loss of features.

As future research, we plan to incorporate single node high precision location queries in order to represent them with the required detail during zoom operations. Also, we plan incorporating a role based security scheme to control access to the geopositioning data.

References:

- [1] Salman A., Baset and Henning Schulzrinne, An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol. *Technical Report CUCS-039-04*, Columbia University, 2004.
- [2] Yasusi Kanada, Multi-Context Voice Communication Controlled By using An Auditory Virtual Space, *2nd Int. Conference on Communication and Computer Networks*, 2004.
- [3] José-Vicente Aguirre, Rafael Álvarez, Leandro Tortosa and Antonio Zamora, Secure VoIP and Instant Messaging on Small PDA Devices, *Transactions on Computers*, Vol .5, No.1, 2005, pp 171-176.
- [4] José-Vicente Aguirre, Rafael Álvarez And Antonio Zamora, Secure Mobile P2P Videoconferencing Protocol, *Eatis'07 ACM-DL Proceedings*.
- [5] M. Catalan, C. Gomez, D. Viamonte, J. Paradells, A. Calveras, F. Barcelo, TCP/IP analysis and optimization over a precommercial live UMTS network, *IEEE Communications Society / WCNC 2005*, pp. 1503-1508.
- [6] Schulzrinne H., Casner S., Frederick R., Jacobson V. RTP: A Transport Protocol for Real-Time Applications, *IETF*, RFC 3550, 2003.
- [7] Kuhn R., Walsh J., Fries S., *Security Considerations for Voice Over IP Systems*, Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, 2004.
- [8] Xiaohui G., Zhen W., Philip S., ZonYin S., Supporting MultiParty VoiceOverIP Services with PeertoPeer Stream Processing, *Proceedings of the 13th annual ACM international conference on Multimedia*, 2005, pp. 303 - 30