Intelligent Information System Supporting
Observation, Searching and Detection for
Security of Citizens in Urban Environment

European Seventh Framework Programme
FP7-218086-Collaborative Project

D1.1 Report on the collection and analysis of user requirements

The INDECT Consortium

AGH – University of Science and Technology, AGH, Poland
Gdansk University of Technology, GUT, Poland
InnoTec DATA GmbH & Co. KG, INNOTEC, Germany
IP Grenoble (Ensimag), INP, France
MSWiA[1] - General Headquarters of Police (Polish Police), GHP, Poland
Moviquity, MOVIQUITY, Spain
Products and Systems of Information Technology, PSI, Germany
Police Service of Northern Ireland, PSNI, United Kingdom
Poznan University of Technology, PUT, Poland
Universidad Carlos III de Madrid, UC3M, Spain
Technical University of Sofia, TU-SOFIA, Bulgaria
University of Wuppertal, BUW, Germany
University of York, UoY, Great Britain
Technical University of Ostrava, VSB, Czech Republic
Technical University of Kosice, TUKE, Slovakia
X-Art Pro Division G.m.b.H., X-art, Austria
Fachhochschule Technikum Wien, FHTW, Austria

[1] MSWiA (Ministerstwo Spraw Wewnętrznych i Administracji) – Ministry of Interior Affairs and Administration. Polish Police is dependent on the Ministry

Document Information

Contract Number	*218086*
Deliverable name	*Report on the collection and analysis of user requirements*
Deliverable number	*D1.1*
Editor(s)	*Piotr Szczuko, Gdansk University of Technology* *szczuko@sound.eti.pg.gda.pl*
Author(s)	*Piotr Szczuko, Andrzej Czyżewski, Piotr Dalka, Grzegorz Szwoch, Andrzej Ciarkowski, GUT* *Piotr Romaniak, Jan Derkacz AGH* *Stanislav Ondáš, Matúš Pleva, Jozef Juhár, Eva Vozáriková, TUKE*
Reviewer(s)	*Zulema Rosborough, PSNI*
Dissemination level	*PU*
Contractual date of delivery	*M10,October 31, 2009*
Delivery date	*October 29, 2009*
Status	*v20091029* *“Final Version”*
Keywords	*INDECT, detection, events, audio, video*

This project is funded under 7th Framework Program

Table of Contents

<Please note: this section is obligatory>

(This page is left blank intentionally)

1 Executive Summary

The INDECT Project, dedicated to creation of *Intelligent information system supporting observation, searching and detection for security of citizens in urban environment* is the End-User driven enterprise. Therefore for WP1 first step is to name End-User requirements for functionality of the system, specifically for task of *Intelligent Monitoring and Automatic Detection of Threats*.

For the purpose of End-User requirements analysis an **End-User Questionnaire** was established, created with cooperation of all INDECT Project Partners. This document, Deliverable D1.1., describes **End-User Questionnaire** structure, its purpose from the point of view of WP1, outcomes of analysis of answers related to WP1 work, and preliminary specification of functionality and hardware of the system fulfilling the requirements for intelligent monitoring and automatic detection of threats.

Considering answers to questions in the **Section A Events** and **Section B Hardware and Software**, the **preliminary** specifications were made. These specifications are related to the **list of events** to be recognized and the hardware features for audio and video acquisition, processing, and storage aimed specifically at effective automatic and intelligent recognition. It should be treated as a road map for further work, and it is assumed that all of the requirements are meant to be reconsidered in a time span of INDECT Project. Final specification of these features will be provided in the following deliverables:

- **For final hardware specification**: D1.2 Report on NS and CS hardware construction (M20)
- **For final specification of event detection**: D1.4 Multimedia database documentation with analysis of recommended algorithms (M45)

It is assumed that WP1 solutions are aimed at analysis of high definition (HD[2]) video, and the algorithms are provided with direct video stream from camera with high frame rate[3], high resolution[4], low noise, and high quality compression (e.g. without colour artifacts, noise or blocking) or uncompressed frames. Fulfilling that high quality requirements is technically feasible as a result of locating the processing unit (Node Station described in Sec. 6) directly near the cameras and microphones. Acquired streams are either transmitted by wire, or short distance wirelessly, therefore high capacity connection is available, allowing high data rate of media.

The deliverable document is organized as follows. Sec. 2 presents general introduction to the task of event recognition. Sec. 4 gives description of End-User Questionnaire and its parts related to WP1. Answers to questions connected with WP1 are analyzed and discussed in Sec. 5 with distinction between event detection requirements and hardware requirements. Based on analysis outcome the initial specification of WP1 intelligent monitoring system is given, with list of audio and video events intended for detection in Sec. 6.1. Quality of Experience and Quality of Decision Making are discussed, particular assessment procedures are presented in Sec. 6.2. A new approach to media tagging, protecting and privacy management is presented utilizing Watermarking approach in Sec. 6.3. Requirements on audio and video quality are reflected in detailed specification of optimal microphones and cameras advised for usage in the system, contained in Sec. 6.4, followed by video throughput, data storage, and computational complexity analysis. Last initial proposals are related to secure communication of WP1 module with other parts of INDECT Portal, discussed in Sec. 6.5. The deliverable ends with conclusions.

[2] Video with higher parameters than standard NTSC or PAL format, see below

[3] Frame rate exceeding standard 25 frames per second

[4] Resolution higher than standard TV resolution, i.e. higher than 768x576 used for PAL or 720x480 used for NTSC

The analysis contained in the deliverable is based only on answers from one end user (Polish Police coordinated by one of the INDECT Partners - GHP) with PSNI yet to respond. PSNI input can be provided for next deliverables, e.g. **D1.2 Report on NS and CS hardware construction (M20)** and **D1.3 Document reporting on acquired results of pilot trial (M45)**.

This document has been reviewed by the members of Ethics Board appointed by the participants of the INDECT Project, with high stress on ethical issues and human rights. In that document the starting assumptions for intelligent monitoring and automatic detection of threats are presented. The acquisition, processing, transmission and storage of video signals are planned, therefore these activities should be analysed according to Directives 95/46/EC (data protection directive), 97/66/EC (processing of personal data and the protection of privacy in the telecommunications sector), and 2002/58/EC (directive on privacy and electronic communications). Throughout analysis of Directives will be contained in next deliverables by all WPs and particularly WP8 of INDECT Project.

The objective of **WP8: Security and Privacy Management** is to ensure security and privacy of processed data. The requirements of privacy and civil liberties, and the requirements and solutions for secure information transfer and storage are to be defined in WP8. First WP8 deliverable **D8.1 Specification of requirements for security and confidentiality of the system** will approach the Directives mentioned above, the final, **D8.8 Overall system security and privacy evaluation** will assess all INDECT Project components.

2 Introduction

The main purpose of the WP1 work is to develop and test new methods for automatic and intelligent detection of events related to danger, utilizing audio from microphones and video from surveillance cameras.

Current monitoring systems depends heavily on operator's attentiveness. His task is to watch numerous video monitors, in search of dangerous or untypical events, and then perform some reactive actions. Most of monitoring systems are focused on registration of the video material serving as a post-factum evidence.

Automatic event detection algorithms are meant to aid a person operating the monitoring system, allowing concurrent analysis of practically any number of audio and video streams (limited by computational power, which is easily extendable). After positive detection by the system the particular audio and video stream is presented to the operator for verification, with event occurrence replayed. Therefore the operator is focused on verification of alarms instead of inspection of limited but multiple number of streams in the same time. A new quality is achieved: effectiveness of threat detection increases.

Dangerous and untypical events are meant for automatic detection, the monitored behaviour needn't be illegal to be assessed as important one and provided to the operator for verification. For example loitering during the day in a park is normal and will not be reported, but loitering during the night on the parking lot can evolve in a car theft, therefore should be proactively detected. Similarly gatherings in some circumstances are typical, and mustn't be monitored, but during the night or in some specific locations can be related to dangerous activity. It is assumed that assessment of the situation is always provided by the person operating the system, not by a computer algorithm, and depending on the operator's decision an audio-visual material of the recorded event is stored longer and reported respectively.

Additionally by utilizing also content analysis the need for storage of continuous audio and video streams is reduced, because important events could be stored in best quality for long time, and typical activity could be represented as a lower quality recording, kept for a shorter period. Therefore privacy of bystanders is greatly improved, as their image is stored in the system as short as it is possible (the time span to be defined by End-Users).

Moreover the video algorithm can automatically protect content recognized as a private, such as faces of bystanders, car plate numbers, house windows, that can be partially obscured from the operator, and original information preserved and encoded in case of detected emergency.

Computer analysis can provide further tools – prediction of dangerous events, and detection of previously overlooked events.

First step is an identification of the most important events intended for automatic detection and selection of proper hardware for acquisition and processing of media streams. Therefore the End-User Questionnaire was established. Following sections present analysis and outcome of gathered answers.

3 Connection with other Work Packages

The WP1 is dedicated to audio and video streams processing only. Next stages, i.e. data transfer, data protection, and presentation of the multimedia to the Operators, are provided by other INDECT Work Packages, namely:

WP7 Biometrics and Intelligent Methods for Extraction and Supplying Security Information for self-organising computer network, safely transferring data provided by WP1,

WP6 Interactive Multimedia Applications Portal for Intelligent Observation System, dedicated to presentation of INDECT System results, defining use-cases, user access,

and finally **WP8 Security and Privacy Management**, evaluating all INDECT WPs for assuring data protection and privacy.

4 End-user questionnaire report

4.1 Questionnaire structure

The End-User Questionnaire is organized as follows:

Questionnaire introduction:

The questionnaire starts with general introduction of the INDECT Project, presenting the title, the scope and expected results of the project.

Sections:

The End-User Questionnaire is divided into Sections, each related to different WP of INDECT Project, namely:

- Section A Events – WP1, prepared by GUT, PSNI
- Section B Hardware and software – WP7, prepared by PSI, GUT, PUT, PSNI,
- Section C Security and Privacy – WP8, prepared by UC3M
- Section D Search Engine – WP5, prepared by AGH
- Section E INDECT Portal (Internet based intelligence gathering) – WP6, prepared by PUT,
- Section F Internet Traffic Inspection – prepared by UC3M.

Each Section starts with short description of the questionnaire objective, addressed target group and instructions on questionnaire filling procedure.

It is assumed that Team Leader (person in charge of e.g. a Department) representative of given end-user group, chooses various Sections and assigns them to his subordinates, considering their profile and competence.

End-User Category:

First task is to define the End-User category, for the answering person, that provides information on ones background and competence.

Questions:

Each question has two types of answers: limited options, e.g. Yes/No, and free text answer, a field for comments and notes related to the subject, and also suggestions or, if needed, questions for clarification.

Creating limited options questions it was assumed, that the responding person will get the grasp of idea and extend it in a free text answer. For example:

A 5. What are the most important symptoms for dangerous attempts?

- Looking around? .... □ YES

- Running with looking around repeatedly? ... □ YES

- Loitering? .... □ YES

- Standing near the door / car for too long?.... □ YES For how long? ....

- what else in your opinion?..

....

....

....

In the sample above it is obvious that not only these four symptoms are important and responder is expected to broaden that list, providing list of other factors, arising in ones expertise and experience.

Last part of every Section is dedicated to other suggestions, leaving large space for free text answer.

Persons to whom the questionnaire was targeted were asked for possibly most precise answers to the questions which are relevant to the area of their expertise and interest. Contact information for each questionnaire part was specified in case any further clarifications were necessary.

Afterward Polish academic partners (AGH, GUT, PUT) created Polish language version of the document.

Based on this the on-line version was elaborated by AGH.

4.2 Description of WP1 related questions

Section A Events is the part of the questionnaire with the strongest relation to Work Package 1 (Intelligent Monitoring and Automatic Detection of Threats). Also **Section B Hardware and Software** is related, but originates from WP7 (PSI cooperation with GUT, PUT, PSNI et al).

As Work Package 1 of INDECT is dedicated to intelligent monitoring and automatic detection of threats, therefore attached WP1 objectives description stated that the services being developed in WP1 are aimed at supporting detection of **suspicious events** by automatic and intelligent analysis of audio and video signals transmitted to monitoring systems. Consequently the questions are aimed at aggregation of a **list of the important events**, to be dealt with the developed system, and gathering what End-Users think are the most important cues of the crime, suspicious behaviour, or dangerous attempts.

For **Section A Events** the destined **target group** was defined as follows:

The questionnaire should be filled by any person related to recognition of a dangerous situation. That can be a monitoring systems operators, forces working in the field, everyone that has an experience or ability to visually detect or predict a threat.

Following questions were included in **Section A Events.** Short summary is presented here, see also the Questionnaire for reference.

A 1. What is dangerous / atypical behaviour in city streets, highways, public transport, stadiums, airport, etc.? What focus your attention in these places? Please state if that differs depending on the time of day, season, etc.

- city streets, sidewalks: a person on the road, running, laying person, falling, fighting? What type of danger can it suggest? What else in your opinion?
- highways: a person on the road, a car pulling over, driving in wrong direction, stopping abruptly, speeding? What type of danger can it suggest? What else in your opinion?
- public transport: a person sitting for more that one cycle, moving quickly, sitting/laying on the floor, left luggage? What else in your opinion?
- stadium: a person still sitting after the game, moving quickly, throwing an object, left object, going outside the stand, entering the field? What type of danger can it suggest? What else in your opinion?
- airport: a person sitting for too long, running, sitting/laying on the floor, left luggage, walking in wrong direction? What type of danger can it suggest? What else in your opinion?

A 2. What is generally dangerous / atypical behaviour (staggering, fainting, loitering)?

A 3. How would you recognize a particular person that is of following type? Is it a dress, behaviour, what type? Burglar, pickpocketeer, thief, drug dealer, drug addict, lost kid, pedophile, terrorist, hooligan, what other persons can pose a threat?

A 4. Try to describe how to recognize threat or attempts of: pulling a gun attempt, stealing a car attempt, physical attract attempt, breaking in attempt?

A 5. What are the most important symptoms for dangerous attempts? Looking around, running with looking around repeatedly, loitering, standing near the door / car for too long (for how long)? What else in your opinion?

A 6. Does the features presented below connect to an intent of the vandalism, e.g. graffiti, breaking a window, etc.? Running, hiding, holding a brick or other heavy object, holding a can (possibly spray can), lurking? What else in your opinion?

A 7. What visual and audio cues imply that a person needs help? Fainting, staggering, waving hand, shouting, covering a face with one or both hands, holding ones belly, bending forward, what is possible danger in these cases? What else in your opinion?

A 8. Which type of movement indicates dangerous event in dense crowd? Try to define type of danger: gathering in one place from all other directions, running away from a single point, disturbing current flow of a crowd? What else in your opinion?

Other suggestions: Please write here other remarks, comments, suggestions, notes you find suitable and helpful to EVENTS topics.

Also two questions from **Section B Hardware and Software** are closely related to WP1 work, namely:

B 2. Try do declare what features are sufficient for proper utilization in your work:

- video cameras: image resolution: TV standard, 1MegaPixel, other (name it); frame per second: 8 frames, 10 frames, 12 frames, 15 frames, other (name it).
- Video recording systems retention time: 24 hours, 4 days, 1 week, 1 month, other (name it).
- computer monitor resolution: 800x600, 1024x768, other (name it).
- mobile transmission devices speed: 128 kbps, 256 kbps, 512 kbps, 1Mbps, other (name it).

B 3. How important is audio channel in video monitoring systems? Did you experienced any situation that the lack of sound was a drawback of a monitoring system? Would you like to have an on-line access to the audio information in a monitoring system? Should the monitoring system automatically recognize acoustic events? What else in your opinion?

4.3 Elaboration and collection of the user requirements

One of the main objectives of INDECT is to develop a platform for the registration and exchange of operational data, acquisition of multimedia content, intelligent processing of all information and automatic detection of threats.

The developed services should support detection of suspicious events by analysis of audio and video signals transmitted to monitoring systems. That scope includes such tasks as monitoring of people (in general scope, not as individual beings), detection of abnormal behaviour, detection of threats, as well as automatic and intelligent notification of people and their protection.

The content will aid the creation of new video and audio analysis methods aimed at automatic recognition and detection of crime and threats.

INDECT addresses and answers the demands and requirements of police and secure services

For these purposes INDECT partners needed to gather from Police Services information on what in their opinion are the most important cues of the crime, suspicious behaviour, dangerous attempt.

It is of highest importance for the project to know Requirements and Expectations of Police Services who would benefit from the results of INDECT.

The idea of preparing and proceeding with a questionnaire first appeared at the meeting in Berlin, held in March 2009, at the part attended by representatives of Police Service of Northern Ireland (PSNI) and Bundeskriminalamt (BKA).

After that all Work Package leaders were asked to prepare inputs to the Questionnaire.

Based on the inputs received a combined version of the document was distributed by AGH on 13th May 2009.

Subsequently details related to the questioning of ‘End-users’ was discussed in more detail at INDECT meeting in Poznan (Poland) organized by Poznan University of Technology in May 2009. The meeting was highly represented by Officers from PSNI and General Headquarters of Police (GHP) – several Police Officers representing different departments and different kind of operational work actively participated to the meeting.

It was agreed that the final version of the questionnaire should comprise:

- Information about who is answering the questions (what is their professional focus).
- Introduction/examples to the questions to allow better understanding.
- Limited number of questions – only questions that give substantial feedback to INDECT activities should be included.
- Key terms in questionnaire have to be explained more explicitly possibly in the form of “key terms dictionary”(e.g. Viewer).

The general plan of how to proceed with the questionnaire was assumed:

- Team Leaders (Persons in charge at the Police) will decide which subsets of questions should be answered by which user category (e.g. CCTV operator, administrator of IT system, etc.).
- One contact point should be assigned for questions related to the questionnaire.
- Interviews and surveys should be done if necessary.
- Questionnaires should be verified by the authors before sending to the INDECT End-Users for filling.
- Some questions should be addressed to management level in police not end-users.
- For security reasons not all answers should be made public.
- Answers to questions should have more options like: “perhaps”, “I do not have experience/experience in the area”, “I do not know”, free text comments.

- Some questions in different questionnaire parts overlap – however this can be acceptable due to the fact that persons answering the questions can focus only on selected Questionnaire Sections.

Timing and Sequence of actions was defined. The Questionnaire version considering discussion in Poznan was released by AGH on 8th June. Comments and suggestion given by Police Partners and other colleagues were taken into account for elaboration of the "pre-final" version of the document which was distributed on 13th June.

The final version of the Questionnaire was sent to INDECT Partners on 21st July.

Polish Police (GHP) has collected answers from a few hundred of Officers who have the required expertise and experience. These copies were given to AGH in September 2009.

Further step was to introduce the information into the on-line questionnaire in order to be able to analyze the combined information in electronic form.

5 Analysis of the user requirements

Following analysis is performed for revealing the most important events intended for automatic recognition with the new generation of monitoring systems. Audio and video events are listed and their importance is rated. The ones that have the highest percentage of confirmative answers are taken into consideration for future work on automatic analysis and detection of dangerous events.

Once having the list of events an initial decision on acquisition and processing hardware is made, considering sufficient quality of audio and video signals and adequate processing power to analyze in real-time and to detect given events.

Large group of Polish Police officers took part in questionnaire filling process. Their profile is presented in Tab. 1. Considerable group did not provide their profile description.

Table 1. Responders profile

Category	**Percentage of responders in the category**
Team Leader (person in charge of e.g. A Department)	15%
	Second category
	12% on internet monitoring 17% on large area surveillance (streets, sports events) 71% not specified
Category	**Percentage of responders in the category**
Video surveillance system operator	1.5%
	Second category was not unspecified
Category	**Percentage of responders in the category**
Police officer 'in the field'	62%
	Second category
	1.5% Crime Department 1.5% Department of Investigation 4% internet monitoring 1% large and close area surveillance 17% large area surveillance (streets, sports events) 1.5% Organized Crime Department 1% prosecuting preliminary proceedings 1.5% tactics of intervention techniques 1% walking patrol 70% unspecified
Category	**Percentage of responders in the category**
IT system administrator	3.5%

	Second category
	50% internet monitoring 50% unspecified
Category	**Percentage of responders in the category**
Other:	
• An assistant in the Prevention Department	• Closed area surveillance (e.g. Railway stations)
• Criminal Investigation Department	• Unspecified
• Criminalistic laboratory specialist	• Unspecified
• Department of logistics	• Unspecified
• Economical division	• Multithread investigation
• Identification of persons (laboratory)	• Unspecified
• Independent position	• Unspecified
Category	**Percentage of responders in the category**
Not specified	12 %

5.1 Event detection requirements

Analysis of answers for questions included in **Section A Events** is presented below (Tab. 2)[5].

Table 2.: Answers for questions in Section A Events of the End-User Questionnaire

A 1. What is dangerous / atypical behaviour in city streets, highways, public transport, stadiums, airport, etc.? What focus your attention in these places? Please state if that differs depending on the time of day, season, etc.		
City streets, sidewalks		
Situation	Percentage of answered YES	What type of danger can it suggest?
Person on the road	57 %	Outside zebra
Running	38 %	At night, group of persons
Laying person	67 %	May need help, fainted, dead
Falling	33 %	
Fighting	81%	
Highways		
Situation	Percentage of answered YES	What type of danger can it suggest?

[5] Empty cells in the tables indicate that no response was gathered on the particular issue.

Person on the road	81 %	Not sober, cars can stop abruptly causing danger
A car pulling over	33 %	Driver not feeling good
Driving in wrong direction	100 %	Driver not sober, major danger
Stopping abruptly	57 %	Driver not feeling good, major danger
Speeding	48 %	
What else in your opinion?	Left lane driving, car with no lights, animals on the road	
Public transport		
Situation	Percentage of answered YES	What type of danger can it suggest?
A person sitting for more that one cycle	33 %	Theft, fainting, illness
Moving quickly	38 %	
Sitting/laying on the floor	67 %	Fainting, dead
Left luggage	81 %	Bomb, especially in crowded places
Stadium		
Situation	Percentage of answered YES	What type of danger can it suggest?
A person still sitting after the game	43 %	Illness, fainting, dead
Moving quickly	33 %	
Throwing an object	100 %	Injuring
Left object	52 %	Bomb
Going outside the stand	48 %	
Entering the field	71 %	Hooligans, disturbances
What else in your opinion?	IDs for fans, excessive grouping	
Airport		
Situation	Percentage of answered YES	What type of danger can it suggest?
A person sitting for too long	19 %	Observing, planning an attack, fainting, illness
Running	29 %	
Sitting/laying on the floor	38 %	
Left luggage	86 %	Bomb
Walking in wrong direction	38 %	

A 2. What is generally dangerous / atypical behaviour (staggering, fainting, loitering, staying for too long in a single place, repeatedly coming back to a place)?

A 3. How would you recognize a particular person that is of following type? Is it a dress, behaviour, what type?	
Burglar	Observes entrances and monitoring, loiters, nervous, untypical tools, luggage, frequent presence in the location, peeking through the window
Pickpocketeer	Observes people, holds cloth in ones hand, frequent presence in public transport nodes, doesn’t avoid crowd, a group of perpetrators is spreading, then gathers around the victim creating artificial crowd
Thief	
Drug dealer	
Drug addict	
Lost kid	Cries, bothers other people, loiters, runs without purpose, in circles
Pedophile	
Terrorist	
Hooligan	

A 4. Try to describe how to recognize threat or attempts of:	
Pulling a gun attempt	Unnatural stance, hand under the cloths, looking around
Stealing a car attempt	Long observation, loitering near cars, grabbing door handles, looking inside the car, fiddling with the lock, frequent getting back to selected car
Physical attract attempt	
Breaking in attempt	Long observation, loitering near doors, grabbing door handles, looking inside the window, fiddling with the lock, frequent getting back to the door, holds untypical tools, waiting inside a car with running engine

A 5. What are the most important symptoms for dangerous attempts?		
Situation	Percentage of answered YES	What type of danger can it suggest?
Looking around	76 %	
Running with looking around repeatedly	33 %	
Loitering	71 %	
Standing near the door / car for too long	57 %	
What else in your opinion?	Staying for too long in a single place, repeatedly coming back to a place	

A 6. Does the features presented below connect to an intent of the vandalism, e.g. Graffiti, breaking a window, etc.?		
Situation	Percentage of answered YES	What type of danger can it suggest?
Running	19 %	
Hiding	52 %	
Holding a brick or other heavy object	71 %	
Holding a can	62 %	Possibly spray can
Lurking	86 %	
A 7. What visual and audio cues imply that a person needs help?		
Situation	Percentage of answered YES	What type of danger can it suggest?
Fainting	62 %	
Staggering	62 %	
Waving hand	38 %	
Shouting	67 %	
Covering a face with one or both hands	67 %	
Holding ones belly	76 %	
Bending forward	81 %	
What else in your opinion?	Calling for help, sitting on a ground, remaining still for too long	
A 8. Which type of movement indicates dangerous event in dense crowd?		
Situation	Percentage of answered YES	What type of danger can it suggest?
Gathering in one place from all other directions	67 %	
Running away from a single point	76 %	
Disturbing current flow of a crowd	67 %	

Outcome in a form of initial system specification for event detection is presented in Sec. 6.

5.2 Hardware requirements

Also two questions from **Section B Hardware and Software** are closely related to WP1 work, that are presented below (Tab. 3)[6].

Table 3.: Answers for questions in Section A Events of the End-User Questionnaire

B 2. Try do declare what features are sufficient for proper utilization in your work:		
Video cameras: image resolution		
Option	Percentage of answered YES	Comments
TV standard	24 %	
1MegaPixel	48 %	
Other (name it)	0 %	
Video cameras: frame per second		
Option	Percentage of answered YES	Comments
8 frames	0 %	
10 frames	0 %	
12 frames	5 %	
15 frames	24 %	
24 or 25 frames	29 %	
Video recording systems retention time		
Option	Percentage of answered YES	Comments
24 hours	10 %	
4 days	0 %	
1 week	14 %	
1 month	19 %	Consistent with PSNI and ACPO[7] standard for image retention when there is no offence
3 months	5 %	
4 months	5 %	
1 year	10 %	

[6] Empty cells in the tables indicate that no response was gathered on the particular issue

[7] Association of Chief Police Officers, organisation for developing police policy in England

Mobile transmission devices speed		
Option	Percentage of answered YES	Comments
128 kbps	0 %	
256 kbps	5 %	
512 kbps	0 %	
1Mbps	71 %	
4Mbps	10 %	
10Mbps	5 %	
B 3. How important is audio channel in video monitoring systems?		
Option	Percentage of answered YES	Comments
Did you experienced any situation that the lack of sound was a drawback of a monitoring system?	86 %	
Would you like to have an on-line access to the audio information in a monitoring system?	71 %	
Should the monitoring system automatically recognize acoustic events?	71 %	

Outcome in a form of initial system specification of system hardware is presented in Sec. 6.

6 Initial specification of the WP1 intelligent monitoring system

Considering gathered answers to questions in **Section A Events** and **Section B Hardware and Software**, following **preliminary** specifications were made. These specifications are related to the list of events to be recognized and the hardware features for audio and video acquisition, processing and storage, allowing efficient media processing, suitable for automatic event detection. It should be treated as a road map for further work, and it is assumed that all of the requirements are meant to be reconsidered in a time span of INDECT Project. Final specification of these features will be provided in the following deliverables:

- **For final hardware specification**: D1.2 Report on NS and CS hardware construction (M20)
- **For final specification of event detection**: D1.4 Multimedia database documentation with analysis of recommended algorithms (M45)

The system specification can be visualized on a block diagram (Fig. 1). Node Station is the main processing unit, with sensors (cameras, microphones, or others) attached. Processing results are being sent through computer network established in WP7, to other services of INDECT Project, e.g. INDECT Portal. Sec. 6. contains description of all system elements.

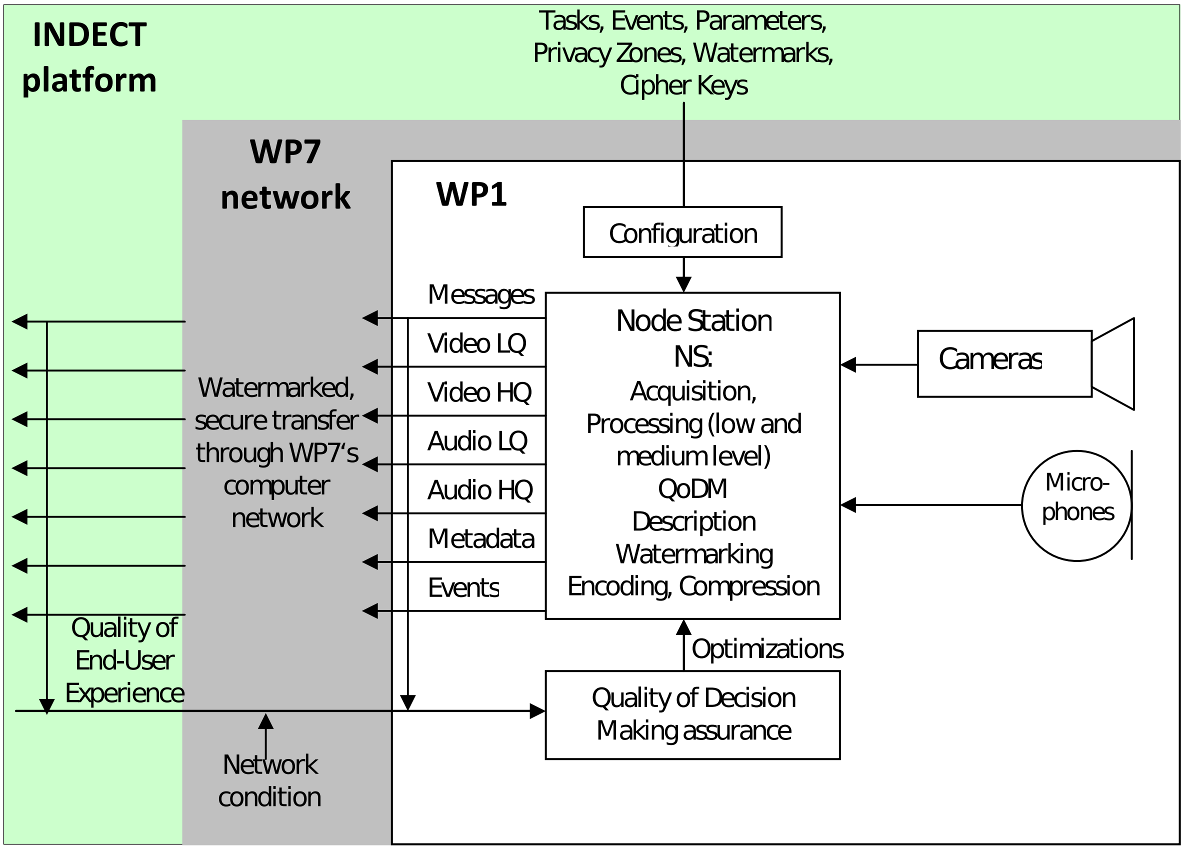
INDECT platform
Tasks, Events, Parameters, Privacy Zones, Watermarks, Cipher Keys
WP7 network
WP1
Configuration
Messages
Video LQ
Video HQ
Audio LQ
Audio HQ
Metadata
Events
Node Station NS: Acquisition, Processing (low and medium level) QoDM Description Watermarking Encoding, Compression
Cameras
Micro-phones
Watermarked, secure transfer through WP7's computer network
Optimizations
Quality of Decision Making assurance
Quality of End-User Experience
Network condition

Figure 1. Block diagram of Node Station functionality in WP1 and its connection to WP7 and INDECT Portal

Particular elements of the figure are presented as a preliminary system specification in following sections, listed below:

Table 4. Elements of Node Station functionality

Element of the system	Section number	Section contents
Processing	**Sec. 6.1**	**System functionality for automatic event detection:** - visual events detection - audio events detection and crowd mood detection
Events		
Quality of End-User Experience	**Sec.6.2**	**System functionality for Quality of Experience optimization**
Network condition		
Quality of Decision Making assurance	**Sec. 6.2.2**	**Quality of Decision Making**: - visual factors influencing QoDM for video events detection - audio factors influencing QoDM for audio events detection.
Optimizations		
Watermarking	**Sec. 6.3**	**System functionality for Watermarking**
NS	**Sec. 6.4**	**System hardware**: - NS, - cameras, - microphones, - video throughput, - data storage, - computational complexity
Cameras		
Microphones		
Video LQ and HQ		
Audio LQ and HQ		
Secure transfer	**Sec. 6.5**	Secure communication framework - secure communication framework and data storage, - general-purpose secure communication, - secure multimedia communication.

6.1 System functionality for automatic event detection

6.1.1 Visual events detection

Based on End-User Questionnaire and numerous meetings and discussions with End-Users, WP1 Partners have established a **list of events** intended for automatic detection by video algorithms developed in Work Package 1 (Tab. 5). That list contains simple events that will be processed inside WP1, and move complex constituted of basic ones, which requires a fusion of data form multiple Node Stations, communication with databases, or other input, and will be developed in **WP7, Task 7.1 Monitoring of phenomena in the environment and of people behaviour in urban areas for detection and prevention of situations with increased probability of danger.**

Table 5. List of events intended for automatic detection by video algorithms

General events		
Type / Location	**Event**	**Comments**
Simple events	*Perimeter / intrusion detection*	formal definition: system detects object recognized as a person within bounds of defined region
	Wrong way movement detection	formal definition: system detects object recognized as a person crossing boundary line in specified direction
	Detection of camera competence	formal definition: system detects loss of image quality and evaluates image features for detection of the cause. System distinguishes: out of focus, partial/total obscuration, under- and overexposure, camera dislocation.
Advanced events	*Left object / removed object*	formal definition: new still object appears in /disappears from the scene, system detects object recognized as a person whose presence is related to appearance / disappearance of new object
Behavioural event (WP7, Task 7.1)	*Loitering*	formal definition: system tracks object recognized as a person for given period of time, if the object remains in the scene for a time longer then defined threshold, and/or leaves scene but reappears numerous times in a defined period of time, then event occurs
	Gatherings	formal definition: system detects and counts new objects recognized as a person, if the number of objects exceeds defined threshold then event occurs
Behavioural events, dangerous attempts (WP7, Task 7.1)	*Breaking in*	formal definition: system tracks object recognized as a person, frequent presence in the location is detected. Also following visual cues are present: loitering near doors, grabbing door handles, looking inside the window, fiddling with the lock, frequent getting back to the door
	Stealing a car	formal definition: system tracks object recognized as a person, frequent presence in the location is detected. Also following visual cues are present: loitering near cars, grabbing door handles, looking inside the cars, fiddling with the lock, frequent getting back to selected car

Events related to locations		
Type / Location	**Event**	**Comments**
City streets, sidewalks, highways	*person on the road*	formal definition: system detects object recognized as a person within bounds of defined region * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object location is incorrectly detected * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio
	running	formal definition: system detects object recognized as a person moving with abnormally high velocity * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object velocity is incorrectly determined * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio
	falling person	formal definition: system detects object recognized as a person which is rapidly changing its orientation from vertical to horizontal (related to the floor); optionally connected with rapid change of velocity to 0 * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object orientation (or orientation change) is incorrectly determined, object velocity change is incorrectly determined * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio
Type / Location	**Event**	**Comments**

Public transport	*a person sitting for more than one cycle*	formal definition: system detects object recognized as a person which remains in the field of view longer than specified period of time * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object tracking is incorrectly determined * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio
	moving quickly	formal definition: system detects object recognized as a person moving with abnormally high velocity * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object velocity is incorrectly determined * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio
	left luggage	formal definition: system detects object recognized as a person whose area is splitting into 2 different objects, one of which remains still for specified period of time, while the other one is receding farther than specified limit from the still one * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object tracking is incorrectly determined * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio
Type / Location	**Event**	**Comments**
Airport	*a person sitting for too long*	formal definition: system detects object recognized as a person which remains in the field of view longer than specified period of time * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object tracking is incorrectly determined * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio

	running	formal definition: system detects object recognized as a person moving with abnormally high velocity * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object velocity is incorrectly determined * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio
	left luggage	formal definition: system detects object recognized as a person whose area is splitting into 2 different objects, one of which remains still for specified period of time, while the other one is receding farther than specified limit from the still one * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object tracking is incorrectly determined * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio
	walking in wrong direction	formal definition: system detects object recognized as a person crossing boundary line in specified direction * benchmarks: * success: definition conditions are achieved * false positive: object of different class is interpreted as a person, object area is incorrectly determined * false negative: event occurs but is not detected * validity measures: success/total ratio, false positive ratio

Complex events constituted of WP1 basic events, analyzed in Task 7.1		
Type / Location	**Event**	**Comments**
Behavioural event (WP7, Task 7.1)	*Loitering*	formal definition: system tracks object recognized as a person for given period of time, if the object remains in the scene for a time longer then defined threshold, and/or leaves scene but reappears numerous times in a defined period of time, then event occurs
	Gatherings	formal definition: system detects and counts new objects recognized as a person, if the number of objects exceeds defined threshold then event occurs
Behavioural event, dangerous attempt (WP7, Task 7.1)	*Breaking in*	formal definition: system tracks object recognized as a person, frequent presence in the location is detected. Also following visual cues are present: loitering near doors, grabbing door handles, looking inside the window, fiddling with the lock, frequent getting back to the door
	Stealing a car	formal definition: system tracks object recognized as a person, frequent presence in the location is detected. Also following visual cues are present: loitering near cars, grabbing door handles, looking inside the cars, fiddling with the lock, frequent getting back to selected car

6.1.2 Audio events detection and crowd mood detection

It is assumed that following audio events should be automatically recognized by the event detection system (Tab. 6).

Table 6. List of events intended for automatic detection by audio algorithms

Audio events		
Type / Location	**Event**	**Comments**
Audio event, streets	*Calling for help in some European languages*	Speech-based audio event
	Screaming	Non-speech audio events
	Broken glass	Non-speech audio events
	Explosions	Non-speech audio events
	Shooting	Non-speech audio events
Audio event, streets, stadiums	*Crowd tendency*	Non-speech audio events

Detection of audio events is an important part of the surveillance system. Audio events, could have very different properties, therefore for further work should be categorized, each of category requiring other analysis and recognition approach. Here audio events are divided into two fundamental groups:

- Speech-based audio events
- Non-speech audio events

Speech-based audio events

This group of events consists of all events, which are produced in a form of **spoken words and phrases** and relates to threats, violence or dangerous situations. As the main items of this group can be considered:

1. calling for help
2. warning shouts
3. profanities, vulgarisms

Detection of such utterances could be improved by recognition of emotions and prosody features extraction.

Non-speech audio events

Non-speech audio events can be divided into several groups:

1. Inarticulate sounds belonging to (coming from) a person:

crying, screaming, fans shout, etc.

2. Sounds belong to mobile objects /cars, trams, planes/

traffic accidents, alarms and honks

3. Sounds accompanying threats or abnormal behaviour
broken glass (show-windows, bottles), explosions, pyrotechnics, shooting,

4. Audio events produced by crowd of people

Audio events produced by crowd is a special group of sounds, which can indicate threats. A special attention should be dedicated to crowd tendency, which can indicate increase of negative emotions.

5. Other sounds

sounds of battle/fighting

Heterogeneity of above audio events require utilizing different approaches. Therefore it is considered to use several detectors in parallel setting. Very important role have the types of features, which will be extracted. Widely used parameterization (MFCC, LPC) for *speech recognition* could not be efficient *for recognition of non-speech audio events*. So there will be a group of alternative feature vectors proposed for testing the non-speech events detection. Therefore we also propose using of several feature-extraction units in parallel configuration (Fig 2.).

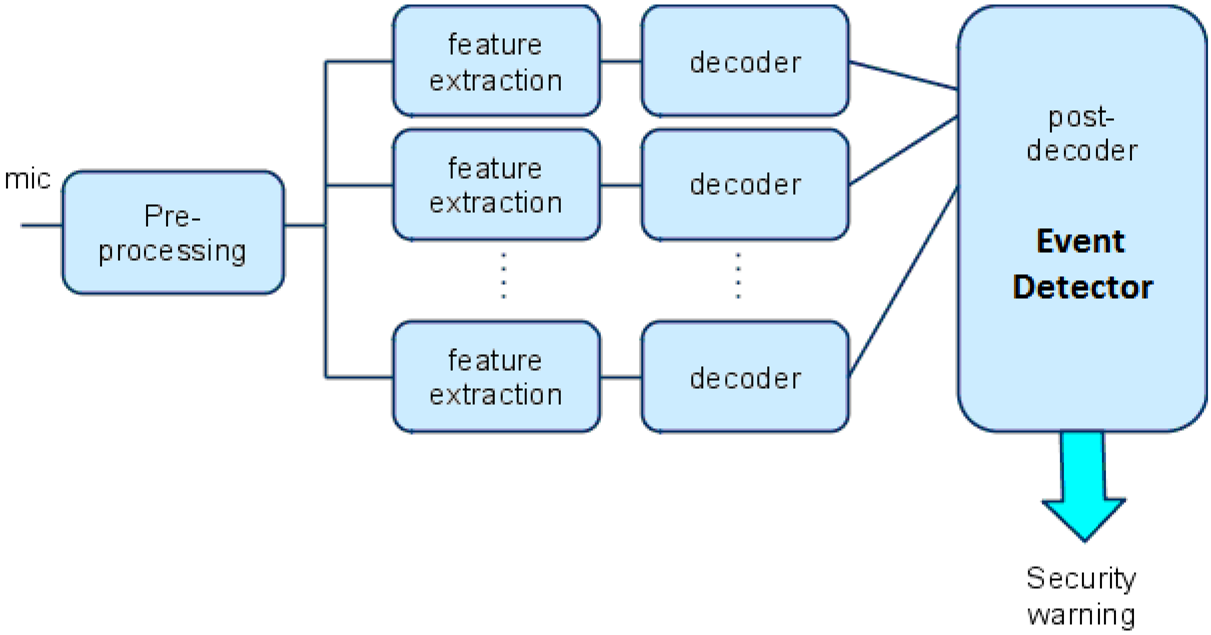

Figure 2. The principal scheme of audio event recognition system

6.2 System functionality for Quality of Experience and Quality of Decision Making

Quality of Experience methodology is employed for providing three important aspects of system functionality:

1. Defining a measurable requirements factors on an automatic recognition of events during testing of created algorithms for assurance of decision quality,

2. Optimizing system/algorithms parameters on-the-fly for assuring quality of audio and video media

3. Guaranteeing sufficient conditions for media processing. All algorithms have their limitations related to lowest possible quality for which they can deliver proper results, e.g. resolution and frame rate of video, and sound sampling frequency, bit resolution and number of audio channels.

6.2.1 System functionality for Quality of Experience optimization

Important aspect affecting usefulness and effectiveness of video monitoring systems is availability of transmission medium. Unfortunately, wide-band (or dedicated) networks allowing seamless transmission of high resolution video are scarce. Moreover, currently there are no existing mechanisms to assure quality of service parameters adequate for live video transmission.

On the other hand, ubiquitous wireless links give desirable flexibility in installation of monitoring systems. Considered technologies are wireless LAN networks (IEEE 802.11 standard) and 3rd generation cellular networks (UMTS). The cellular technology takes advantage over WLAN in a high reliability, coverage of all agglomeration areas, and inherent security mechanisms.

In video monitoring systems it is necessary to assure acceptable quality, regardless weather conditions, lighting conditions, and transport medium load (in case of public links shared with other traffic). Term "acceptable" cannot be firmly defined as it strongly depends on considered scenario. For example, video monitoring system that does not allow to identify a person (quality too low to recognize a face) is not very useful but can provide sufficient quality for intrusion detection.

Presented discussion suggests that important feature of next generation monitoring system will be a set of mechanisms designed to assure acceptable Quality of Experience for video sequences. The following elements are essential to accomplish this task:

- Quality optimization and assurance procedures. First step towards quality optimization is an assessment and identification of the quality degradation roots. This information can be utilized in order to perform preventive actions and compensate/eliminate its harmful influence. These actions will strongly depend on considered scenario and include video coding parameters adjustment (each video stream has to be compressed and encoded before transmission) and capture device parameters adjustment.

- No-reference and full-reference video quality metrics capable of real-time assessment. No-reference scenario is essential in case of acquisition-related artifacts measurement where reference sequence is unavailable. Real-time assessment requirement is essential for live monitoring systems.

In typical video monitoring systems there are a few points along video delivery path where video quality degradation may occur. Fig. 3 presents video delivery chain with four point responsible for quality degradation marked.

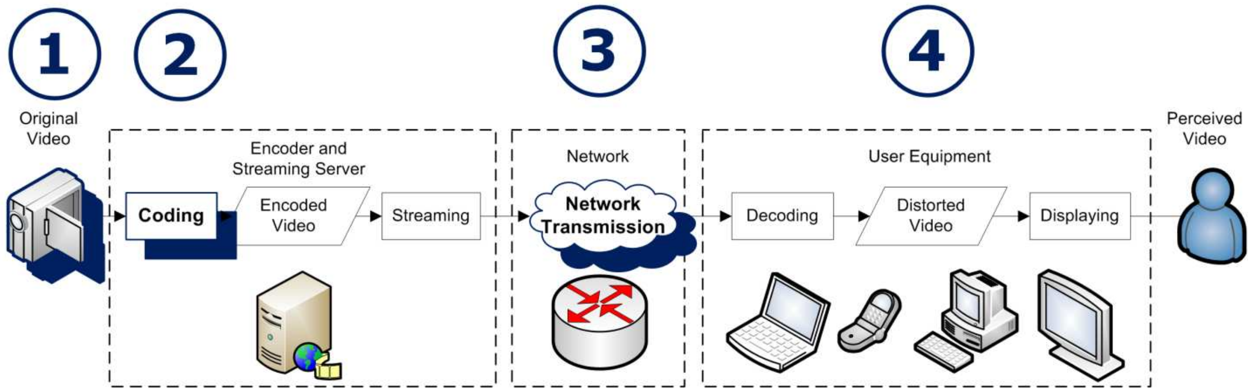

Figure 3. Video delivery chain.

One can distinguish following factors:

1. Video acquisition (inherent for User-Generated content)
    a. Noise
    b. Out-of-focus
    c. Over/Under-exposure
2. Lossy compression
    a. Quantization domain
    b. Spatial domain
    c. Temporal domain
3. Network transmission
    a. Artifacts cause by a packet loss
4. Application/scenario specific parameters
    a. Provided with user's responses

All mentioned factors affecting perceived video quality should be addressed in comprehensive examination of quality degradation. One important factor, essential for monitoring systems, proposed in the INDECT Project, is video quality degradation related to chosen watermarking procedure. This is another parameter to be controlled in quality optimization and assurance task.

The ultimate goal of video monitoring systems is to provide meaningful visual information regarding urban areas. It is aimed at targets recognition and detection. This is very different from entertainment video services where the ultimate goal is to provide End-User with the highest possible Quality of Experience. For video that is used to perform a specific task, it may not be appropriate to rate the quality of the video according to a subjective scale such as absolute category rating (ACR) described in recommendation ITU-T P.910. For that reason a new recommendation for video used in target recognition tasks was created recently.

Assessment methods for evaluating the quality of one-way video used for **target recognition tasks** are described in recommendation ITU-T P.912 "*Subjective video quality assessment methods for recognition tasks*". Described approach will be utilized during WP1 work for quality assurance. Term "target" refers to an object in the video sequence that the viewer needs to identify, e.g.: face, object, number. TRV (Target Recognition Video) is a video sequence used to accomplish specific goal through ability to recognize targets. There are three categories of target recognition tasks:

- Person identification (including facial recognition)
- Object identification
- Alphanumeric identification

The recommendation defines the following terms:

- DC (Discrimination Class): 1 of 4 levels of visual discrimination at which the target can be analyzed:
    - Elements of action – in very broad and general sense, identification of series of events that took place
    - Target presence – recognition/detection of the presence or absence of valid targets.
    - Target characteristics – recognition of unique characteristics of the target (e.g., markings, scars, tattoos, dents, colour).
    - Target positive recognition – recognition of specific instance of the target (e.g., recognition of a person, a specific object, or an exact alpha-numeric sequence).
- SG (Scenario Group): collection of scenes of same scenario, with very slight differences between scenes

For video material used for performing a specific task (contrary to quality of multimedia material meant for entertainment) it might not be appropriate to rate its quality in subjective scale. The goal of test methods for TRV are as follows:

- To assess ability of viewer to recognize appropriate information in video,
- Regardless of viewer's perceived quality of viewing experience

Methods to assess quality level of TRV, avoiding ambiguity and personal preferences:

- Reducing subjective factors, and
- Measuring ability of participant to perform task

The ITU-T P.912 defines three methods for the subjective experiment design. The first called "multiple choice method" is appropriate for all DC levels and target categories. Video is shown above the list of verbal labels representing possible answers. Since it uses fixed multiple choices it eliminates any possible ambiguity or subjectivity and allows for more accurate measurements. The second one, "single answer method", is dedicated only for non-ambiguous answers to ID question and is appropriate for alphanumeric scenarios. In the last one, “timed task method”, viewer might be asked to watch for particular action or object to be recognized in video clip. Timer button should be pushed whenever the viewer perceives that target.

Depending on nature of task, TRV test methods to be used either in real time, without ability to freeze or rewind, or for non real-time analysis. The experiment should mimic real world application of the material.

Scenes presented in the test should contain targets consistent with application under study. Individual scene might be replaced by set of scenes containing multiple variations. This is called a scenario group (SG). An example scenario could be a walking person carrying an object. The SG would consist of many shots of the same scenario but with different people and objects and varying quality conditions.

6.2.2 Quality of Automatic Decision Making

For algorithm that are developed in WP1 it is assumed that a metric of Quality of Decision Making exists and should be utilized to assess algorithm effectiveness and sufficient, minimal, media parameters for processing. When developed, such metrics should be used during testing of created algorithms for assurance of decision quality.

Analogous strategy should be provided for real-time processing of target media, during Field Tests and Pilot Trials. Proper metrics should be used for guaranteeing sufficient conditions for media processing, as all algorithms have their limitations related to lowest possible quality for which they can deliver proper results. In case of insufficient media quality two measures can be taken: automatic optimization of acquisition parameters (e.g. camera iris, exposure, frame rate), or automatic generation of the detailed report for operator, serving as a guidance for maintenance (i.e. describing in details what media parameters determine insufficient quality).

Methodology of Quality of Decision Making will be described and tested in dedicated deliverable: **D1.3 Document reporting on acquired results of pilot trial (M45)**.

It is assumed that WP1 solutions are aimed at analysis of high definition video, and the algorithms are provided with direct video stream from camera with high frame rate, high resolution, low noise video and high quality compression (e.g. without colour artifacts, noise or blocking). It is technically feasible as a result of locating processing unit directly near the cameras and microphones. Acquired streams are either transmitted by wire, or short distance wirelessly, therefore high capacity connection is available, allowing high data rate of media.

6.2.2.1 Visual factors influencing QoDM for video events detection

Video streams analyzed in the surveillance system are acquired with digital cameras. Their specification is presented in Sec. 6.4.1. Video streams must meet certain criteria in order to make event detection possible. These criteria include **resolution**, **frame rate** and **colour**.

Generally, utilizing the higher image **resolutions** the more precise results are achieved. However, higher image resolution requires more computational power and in some conditions may not provide any profits. Image resolutions determine the minimum size of objects that can be automatically detected in the system. Therefore the greater distance of the camera from the scene and the greater its angle of view, the higher image resolution should be used. For a typical setup utilizing classical camera, the resolution approx. 720x576 is enough. In case of wide-angle cameras megapixel resolutions (e.g. 1600x900) are required.

Number of **frames per second** have to be chosen according to the “dynamics” of the analysed scene. The faster objects move and the closer their distance to the camera the greater frame rate is required. It is estimated that 15 frames per second is enough to analyze typical video surveillance streams. Greater frame rate may provide higher accuracy at the cost of higher computational complexity.

Video streams must contain **colour** images with high colour depth (typically 24 bits per pixel). Black and white images are not sufficient for effective moving object detection, e.g. objects of different colour but having the same brightness might look identical to video analysis system based on black and white images only. Furthermore, practically all digital cameras currently available in the market are colour ones.

While acquiring consecutive video frames also other parameters have important impact on QoDM: **white balance, exposure, focus**. **White balance** influences colour characteristics of the image, and if not set as fixed, then any lighting changes (setting sun, clouds, switching of indoor light) will result in shifting of all colours towards bluish or reddish one. That can dramatically affect video analysis algorithms, as in various application is treated as a change in whole scene.

Improper **exposure** can lead to over- or underexposure of lightest or darkest areas, and as a result some range of dynamics is permanently lost. Then that area is represented as a set of pixels with equal value instead of a varying (and interesting for analysis) ones.

Lack of **focus** also leads to data loss, namely details in the image. Small elements of the image are merged, and blurred into single one. To some extent that is reversible with dedicated processing, but generally should be avoided.

All mentioned factors are of great importance for automatic detection of events in video stream. Therefore it is mandatory to fulfil these requirements first choosing a camera model and then assembling video monitoring system.

Recommended technical parameters of the camera are discussed in Sec. 6.4.1.

6.2.2.2 Audio factors influencing QoE for audio events detection

Ambient noises

Audio input of the surveillance system in outdoor environment contains noise of the ambient, e.g.: music, sounds produced by the abnormal weather conditions (such as strong rain, thunder storms, wind), trams, trains, buses etc. Each acoustic event can be very negatively affected by these noises. This disadvantage can be partially compensated by the background model which is designed for particular ambient. The continuous adaptation of such model to changing conditions can bring further improvements.

Microphone location and characteristic

Critical factor for the success of the system for detecting and classifying audio events is the distance from the microphone location from detected audio events.

Directional characteristics of microphones can have great impact on sound event detection. Omnidirectional one will acquire sounds from all direction, cardioidal will favour single direction, suppressing sounds approaching from other directions, and super-cardioidal is even more directional. Depending on application, one can use various setups from single microphone to multi-microphone array. That aspect will be discussed thoroughly in deliverable: **D1.2 Report on NS and CS hardware construction (M20)**. Here it is initially indicated that a type of microphone is an important factor, which cannot be easily changed after establishing monitoring infrastructure, and should be carefully considered.

Sound recording

Digital form of the sound wave is characterized with sampling frequency and bit resolution. A wave is represented as a consecutive samples, F per second, where F is sampling frequency, influencing precision of sound description over time. Each sample is stored with limited bit resolution, influencing precision of the sample value, and its dynamics. Typically for sounds present in human environment it is assumed that sufficient recording can be made with at least 44.1kHz sampling frequency and 16bit resolution. These values correspond to highest frequency that can be recorded equal to 22.05kHz (10 percent above typical human hearing) and dynamics of 96dB, interpreted as a range between loudest and softest sound that can be recorded with given resolution. Higher values require dedicated microphones and analog to digital converters. Lower values can influence effectiveness of sound recognition. During algorithm development the recording parameters will be tested against recognition accuracy for wide spectrum of environmental sounds and sound events of interest of WP1.

6.3 System functionality for Watermarking and sensitive and private data protection

Intelligent monitoring systems can be equipped with mechanisms that allow control of content access with more granularity than it is possible in contemporary solutions. Digital watermarking techniques allow to embed additional information inside the multimedia content itself without revealing visually that some information is hidden. Utilizing watermarking, any digital data can be hidden in the image or video sequence, introducing only additional noise, unnoticeable to some extent, depending on amount of hidden data. Therefore **sensitive information** recorded by surveillance cameras (etc., **human faces**, license plates of the vehicles) can be **occluded** while retaining original information encoded in the remaining part of the image as a watermark. **Authorized personnel** can use dedicated decoders that are able to reconstruct the original appearance of the images. This allows to distribute multimedia files more safely as audience not equipped with a secret key is able to see only an occluded version of the images.

Digital watermarking can also be used to provide information trust mechanisms. It is possible to embed content authorship data and content authentication data into the multimedia files. System data, camera type, location, date, etc. can also be transmitted as a watermark. Moreover utilizing watermark an image tampering operations can be detected, and then the modified region can be indicated or partially reconstructed.

The dialogue between INDECT research team and the End Users confirms that there exist interest in incorporating the mentioned techniques in the developed intelligent monitoring system. The research on digital watermarking algorithms is in progress and the results will be available as a report - deliverable D5.2 “Report on developed high capacity and fast Watermarking System with application for authentication multimedia contents and search purposes”. The results will be incorporated in WP1 as a protection, authentication and privacy measure.

Detection of some events requires collection and storage for short period of time the personal features (numeric parameters related to visual appearance instead of storing the photo or video itself) which enable to recognise persons in various images obtained from cameras. It can be assumed that the privacy is protected, because based on these numeric values the photo cannot be recreated, however, suitable protection of these parametric data will be considered: technique of hiding data in a Watermark, cryptography, or other forms of steganography.

6.4 System hardware

The INDECT concept of the multimedia platform assumes the elaboration of a distributed system whose principal element is an autonomous Node Station. This automatic data acquisition station will be used to acquire data, signals, and images from the surveyed area, then to pre-process the data intelligently and transmit the gathered information to the remote servers. It will cooperate with cameras, sensors, and microphones located within the range of its operation through wired or wireless connections, and it will pass the collected and partially processed information through the gateway of a computer network. The distributed data processing system, provided with huge computational power and a vast repository of knowledge connected also to a spatial information system, will be programmed in a way that will allow the automatic detection of events that could pose a potential threat to security and safety

The NS can be equipped with megapixel, wide angle, fixed cameras or moving PTZ cameras as well as microphones and speakers. It is assumed that WP1 solutions are aimed at analysis of high definition video, and the algorithms are provided with direct video stream from camera with high frame rate, high resolution, low noise video and high quality compression (e.g. without colour artifacts, noise or blocking). It is technically feasible as a result of locating processing unit directly near the cameras and microphones. Acquired streams are either transmitted by wire, or short distance wirelessly, therefore high capacity connection is available, allowing high data rate of media.

The video and audio data are analysed by NS and alerts accompanied with Metadata (i.e. text description, geo-location, time and date, etc.) are sent to the Central and mobile terminals by any available network. The live audio and video streams can be transcoded in NS for adaptation to different transmission medias, and terminals. Analysis algorithms in the NS are designed to communicate with other Stations, for detection and tracking of particular objects (cars, persons) in large areas covered by number of NSs. All communication is performed through Central Server for backup, storage, and control. Databases are distributed among NSs but the significant data are also backed up in Central database along with streams (audio and video).

The NS is developed in Work Package 1, as a multifunctional platform for acquisition and processing of audio and video streams. WP1 is dedicated to creation of algorithms for NS for audio and video processing and automatic intelligent detection of threats. Next in WP7 that NS is to be incorporated within INDECT Platform and the communication protocols and streaming procedures for NS/Platform are to be created.

6.4.1 Cameras

Cameras are the most important sensors used in the surveillance system. Video streams acquired from the cameras will be analysed automatically and therefore must be of high quality (high enough to allow effective analysis).

The surveillance system will use fixed cameras (with constant field of view) and pan-tilt-zoom (PTZ) cameras (where field of view may be moved and zoomed). Fixed cameras will be used for automatic image analysis and for event detection. Then, if necessary, PTZ cameras will provide a detailed view of an object of interest and allow to track its movement. Therefore requirements for fixed and PTZ cameras are slightly different. They are presented in Tab. 7 and 8 and are consistent with video acquisition parameters.

Table 7: Required values of parameters for fixed cameras

Camera parameter	Optimal value	Minimum value	Comment
Image resolution	1600x1200	640x480 720x576	Megapixel cameras provide more detailed image but its analysis requires much more computational power
Frame rate	25 fps	15 fps	The faster objects move in the camera field of view the faster frame rate is required
Maximum angle of view	> 70°	> 50°	The greater angle of view the larger area may be covered by a camera
Autofocus	Yes	No	Autofocus is turned off after the camera setup
Camera type	Digital	Digital	Analogue cameras are not compatible with the surveillance system
Video compression	MPEG-4 Part 2 or Part 10, MJPEG	MJPEG	MJPEG compression will be used for automatic video analysis
Day/night functionality	Yes	No	Working in low luminance conditions is not necessary if a camera is mounted in a well lit area

Table 8: Required values of parameters for PTZ cameras

Camera parameter	Optimal value	Minimum value	Comment
Image resolution	640x480 720x576	640x480 720x576	Mega-pixel resolutions will be substituted with greater optical zoom – the same image quality and less computational power required
Frame rate	25 fps	15 fps	Higher frame rate is especially useful while smaller angles of view (higher zoom factors) are used
Autofocus	Yes	Yes	Autofocus is adjusted every time the camera moves
Maximum optical zoom	>30x	> 20x	The greater optical zoom the more detailed image of distant objects may be provided
Pan and tilt	Pan: 360° Tilt: 180°	Pan: 360° Tilt: 180°	Covers the entire area around the camera

Camera type	Digital	Digital	Analogue cameras are not compatible with the surveillance system
Video compression	MPEG-4 Part 2 or Part 10, MJPEG	MJPEG	MJPEG compression will be used for automatic image analysis
Day/night functionality	Yes	No	Working in low luminance conditions is not necessary if a camera is mounted in a well lit area

Surveillance cameras meeting optimal requirements are universal devices that will provide good quality images in every possible conditions. However, in some circumstances, they may be substituted with less expensive cameras meeting minimal requirements without significant loss of image quality for viewing and automatic recognition tasks.

Stereo Vision and Infrared Cameras

Additional to the standard equipment **alternative setups** will be investigated, one including an additional camera. This camera will be mounted in a horizontal distance of 10 – 30cm to the standard camera. That setup allows the measurement of distances of objects to the camera and to each other. The distance from the camera to the object can be found with the method of "disparity map". Therefore the accuracy and robustness of the object detection system can be improved. The camera set (stereoscopic camera) has to be calibrated during assembly and can be used in the field as a unit.

Stereo cameras are commercially available as unit but also standard surveillance cameras can be assembled in such a way, introducing distance measuring and more effective object detection in next generation of monitoring system developed in WP1.

Another setting will use an additional infrared camera, which is aligned with the standard camera providing a second picture with infrared contents. This picture gives additional information on object temperature and will improve the robustness of the object detection.

Infrared cameras are offered as standard products by several companies, and numerous models provide standard resolution.

6.4.2 Microphones

Microphones are essential part of modern multimodal monitoring. Many dangerous events or threats have audio cues present, and moreover some situations cannot be detected utilizing video analysis only. End-User responses reveal strong need for audio monitoring and presence of sound in new generation of surveillance systems.

The WP1 Node Station will use various microphone setups depending on application. Some environments are more reflective and reverberant, such as closed spaces, and indoors, contrary to open spaces, and need other approach in sound acquisition and analysis. One can imagine situations when direction of approaching sound is crucial (need for locating sound source). In these conditions large number of microphones forming so called microphone array can be utilized. If accompanied with specialized processing, that setup should provide sufficient directivity information.

These strategies will be researched and tested during system development in WP1. Final outcome will be presented in deliverable **D1.2 Report on NS and CS hardware construction (M20)**. Results of practical evaluation of elaborated strategies will be presented in **D1.3 Document reporting on acquired results of pilot trial (M45)**.

Table 9: Required values of parameters for microphones and recording

Microphone / recording parameter	Optimal value	Minimum value	Comment
Directional characteristics	Omnidirectional, cardioidal, super-cardioidal	n/a	Decision on directivity depends on application
Number of microphones	Microphone array comprising omnidirectional microphones	Small number of directional microphones,	Decision on number of microphones depends on application
Sampling frequency	96kHz	44.1kHz	
Resolution	24bit	16bit	
Dynamic range	144dB	92dB	

6.4.3 Video throughput

One of the functionalities of video surveillance system is the possibility to transmit video streams between various nodes of the system (e.g. from the Node Station with a camera to the Central Server and to the operator desk). There are two types of video streams that will be transmitted in the system. The first type is used directly by the video analysis algorithms and must be characterized by the superior quality and therefore a very high bit rate. The second type of a video stream is a reference one, used for presentation to system operators. The quality of this stream might be lower and may change over the time. Both types of streams are summarized in Tab. 10.

Table 10: Video streams used in the surveillance systems and their projected throughputs

Parameters	Video stream used for image analysis	Reference video stream
Video compression	MJPEG	MPEG-4
Bit rate for standard resolution	~15 Mbit/s	350 kbit/s – 2 Mbit/s
Bit rate for megapixel resolution	~70 Mbit/s	1 Mbit/s – 8 Mbit/s
Possible transmission channel	Wired only	Wired and short-range wireless (Wi-Fi) for very good quality, mobile wireless (UMTS) for adequate quality

A video stream used for video analysis is compressed with MJPEG algorithm. It is less efficient than the MPEG-4 codec, but in the same time it is less complicated and is better suited for real-time analysis (e.g. it introduces a lower delay). Projected bitrates vary depending on video resolution and frame rate from 15 Mbit/s (640x480 and 15 fps) to 70 Mbit/s (1600x900 and 25 fps). Such a throughput is practically possible only with a wired transmission channel. Therefore automatic image analysis will be performed on-place (in a Node Station close to a camera). Alternatively, video streams might be sent to a computational centre with high-bandwidth transmission channels (i.e. optical fibre ones).

A reference video stream will be compressed with MPEG-4 algorithm. Its quality may vary over time depending on local conditions, transmission media, weather condition for wireless networks, etc. Very good video quality may be obtained with bitrates from 2 Mbit/s (for standard resolution video streams) to 8 Mbit/s (for high resolution). The lowest bit rate providing adequate video quality is equal to approx. 350 kbit/s (for standard resolution) or 1 Mbit/s (for high resolution). Very good video quality may be easily achieved with wired communication channels. In case of wireless communication, only short-range protocols such as IEEE 802.11b/g (Wi-Fi) provide required throughput. For adequate video quality mobile network transmission protocols (e.g. UMTS) are enough. It must be stressed that all video streams will be stored locally in the Node Stations and any recorded sequence of a very good quality may be sent offline (i.e. not in real-time) to a system operator, on demand.

6.4.4 Data storage

The storage server needs to store the video streams from the cameras as well as the results of the video content analysis (which will be referred to as metadata in this document). The amount of the disk space needed for data storage (both the video and the metadata) depends on a number of factors which are discussed below.

The proposed system includes nodes whose task is to record the communications. The purposes of this recording range from logging messages for system debugging or reasoning reconstruction, storing events for data mining up to storage of multimedia material together with accompanying metadata as investigation or trial evidence. It is possible to distinguish two basic kinds of data: XML structured data and multimedia data/metadata. Storage of XML data can be easily and effectively performed using specialized XML database. Practical example of such storage system is XDB database, however there are also other similar solutions.

First approach to storage of multimedia data would be based on serializing raw RTP packets payload into the file, thus avoiding complex transcoding otherwise necessary if data was serialized into any popular container format like AVI, OGG etc. This way it is also possible to multiplex the media content frames with metadata frames, making them directly available. Each packet payload should be preceded by header specifying packet size and RTP payload number, consulted with global file header containing payload dictionary based on SDP model used by Jingle protocol. Headers would be used for efficient navigation within the stream.

Video storage

The image frames obtained from the cameras of the monitoring system will be received in a video stream, which is encoded and stored to a file. For video encoding, MPEG4 codec will be used. The size of the data stored on the server depends only on the bitrate set in the codec – higher bitrate means higher disk space usage. At the same time, the quality of the recorded video depends on the following factors:

- bitrate of the codec – higher bitrate means higher video quality; however, if the encoded video needs to be sent through the network, the bitrate is limited by the network capability;
- size of the camera frame (camera resolution) – increasing the frame size when the bitrate is not changed decreases the video quality;
- number of frames per second – increasing the number of frames also decreases the video quality if the bitrate remains unchanged.

As a consequence, all three parameters mentioned above need to be taken into account when the required storage space is estimated. For the purpose of this document, three profiles of video quality are created (Tab. 11). A megapixel camera (1600 × 900 max resolution) was used because this type of camera is recommended in the proposed monitoring system.

Metadata storage

The results of the video content analysis will be sent to the storage server for each camera frame, in a form of binary data packet. These metadata include a list of detected events, data on moving objects, etc. The size of the metadata depends on the number of events and the number of moving objects detected in the frame. The main part of the metadata (about 80%) will be occupied by the image marking the position of the moving objects inside the frame, therefore the increase of the metadata size when the number of detected objects and events increases will not be linear. The estimated average size of the metadata for one camera frame is 5 KB.

Storage space requirements

In Tab. 11, the estimated disk space needed for storage of the encoded video stream and the metadata from a single camera are calculated for three profiles of video quality. Additionally, the estimated length of the recording (in days) that may fit on the one terabyte disk (provided that no other data is stored on it) is given.

The strategy of the archiving (how many days should the recordings be kept on the disk) depend on the system requirements. Usually, the recorded material should be stored for at least 30 days, in some less important situations – for 14 days. In Tab. 11, the required storage space for each quality profile is calculated for both these strategies. It can be seen that 1 TB disk is sufficient for storing the material from one camera for 30 days when the low or the medium quality profile is used, while the high quality profile requires at least 1.4 TB disk.

Table 11. Estimation of the storage space required for recording the video stream and the metadata from the single megapixel camera

Profile	Low quality	Medium quality	High quality
Frame size (pixels)	800 × 450	1600 × 900	1600 × 900
Frames per second	5	10	15
Bitrate (Mbps)	0.5	1.5	4.0
Video recording storage (GB/day)	5.27	15.82	42.19
Metadata storage (GB/day)	0.41	0.82	1.23
Total storage (GB/day)	5.68	16.64	43.42
Recording time of 1 TB disk (days)	180.28	61.54	23.58
Storage required for 14 days (GB)	79.52	232.96	607.88
Storage required for 30 days (GB)	170.40	499.20	1302.60

The calculations presented here are valid for a single camera. If the data from more than one camera (of the same type) will be stored on the same server, the storage requirements for each camera have to be summed up.

Audio storage

Audio stream is not that demanding as a video stream. Estimated storage space for two typical audio formats are presented in Tab. 12.

Table 12. Estimation of the storage space required for recording of three types of audio streams

Profile	**44.1kHz, 16bit, mono**	**96.0kHz, 24bit, mono**	**96.0kHz, 24bit, mono**
	Uncompressed lossless WAV		**Compressed lossy mp3, 256kbps**
Bitrate (Mbps)	0.6	2.2	0.25
Audio recording storage (GB/day)	7.1	23.1	2.6
Recording time of 1 TB disk (days)	144	44	388
Storage required for 14 days (GB)	100	324	37
Storage required for 30 days (GB)	213	695	79

6.4.5 Computational complexity

Real-time video image analysis in a surveillance system is a very complex and computationally expensive task. Practically, any available resources may be easily saturated by video image processing. The more computational power is available the more accurate results of video analysis are. Tab. 13 presents projected video analysis performance depending on input image resolution, number of processing threads and internal accuracy of analysis.

Table 13: Projected video analysis performance (in frames per second)*

Parameters	High accuracy	Optimal accuracy
1600x900, 1 thread	3.4 fps	9.4 fps
1600x900, 2 threads	6.2 fps	17.1 fps
1600x900, 4 threads	8.9 fps	24.5 fps
720x576, 1 thread	12.5 fps	36.4 fps
720x576, 2 threads	23.4 fps	68.4 fps
720x576, 4 threads	32.3 fps	88.5 fps

*simulations were performed on a state-of-the-art computer equipped with one quad-core 2500 MHZ processor

The minimum frame rate sufficient for effective event detection is estimated to be 15 frames per second. Such a performance is easily achieved for standard resolution video streams while using high accuracy video processing settings. However, the required performance for megapixel resolutions is obtained only with optimal (and less accurate) settings. Therefore only the most powerful hardware components should be used for image analysis. Such components may already be found in the market, but they are very expensive. However it is projected that in 5 years (INDECT Project duration) hardware components sufficient for high accuracy video analysis will be commonly available for a reasonable price.

6.5 Secure communication framework

This section describes the data exchange formats and protocols proposed for use for communications between WP1 entities, providing suitable **level of security** for data and media transmission. The choice of specific solutions has been determined by the availability of open standards related to discussed application and their functional suitability. Careful **security analysis** has been also performed in order to determine whether the specific solution grants sufficient protection for sensitive data, which are processed within the system. This factor affected the decision-making process towards the solutions which are mature and well-established within the industry, as this increases the chance that the design is free from "child-age" defects and minimizes the chance of discovering security exploits.

The system should be able to support two types of communications traffic:

- Interactive, low volume data exchange capable of transporting arbitrary, self-describing data

- Real-time multimedia streaming for high volume on-demand traffic

Although there are protocols capable of supporting both types of traffic, such solutions are suboptimal, because of their vastly differing real-time characteristics. Therefore it was decided to use two separate communications protocols.

6.5.1 General-purpose secure communication

Each entity connected to the system should be able to receive and send general-purpose data to any other entity. This imposes the requirement on the addressability of the entities. It was decided that addressing based on IP number is not sufficient for the purpose, as it is possible to run several services based on the same machine, therefore additional logical addressing was required to uniquely identify each communication entity. Moreover, the chosen solution must allow for easy communications with nodes which are located behind firewalls or NATs. These requirements may be easily fulfilled by using system with central server acting as a "bus controller". This ensures that all the connections are initiated by the clients connecting to the public server, so the traffic will not be blocked by firewall or NAT. The server also acts as a "hub" for the purpose of addressing connected clients, so that the connecting entities need not know physical address of any other entity they wish to communicate with. The actual communications routing is done by the server based on the logical addresses, what greatly enhances the flexibility and scalability of the system.

The connection between clients and the server should be secure. As a general framework for **secure transport layer TLS 1.0** should be used, which is currently most advanced and reviewed solution. TLS is extensible framework, which allows using arbitrary public-key and symmetric ciphers for the purpose of securing the communications channel. The specific choice of used ciphersuite at this point is not required and should be a matter of discussion with WP8 representatives.

The analysis of aforementioned requirements leads to the choice of XMPP protocol as a medium for inter-entity general purpose communications. XMPP protocol is based on the exchange of pieces of XML documents, so called "stanzas". This feature makes for its unlimited extensibility – the data may be formed into arbitrarily complex structures, provided they conform to the XML Schema. XMPP protocol includes also support for entity addressing, as each connecting node is assigned an address very similar to e-mail address. The communication is secure – server may be configured to enforce TLS1.0 channel protection, it is also possible to perform mutual authentication of client and server. The protocol has built-in support for node and service discovery, so that it is possible to create service repositories. The XMPP community developed a number of protocol extensions.

Some notable examples include XMPP-SOAP bindings, which allow for easy encapsulation of SOAP messages within XMPP "stanzas", which is very useful for the purpose of constructing XMPP-SOAP gateway enabling publishing of some services running within the system as Web Services. This feature is an enabler for interoperability with the other parts of the INDECT project, which may choose SOAP/WSDL as their primary means of communications.

Other features of XMPP protocol which are useful in the scope of WP1 system include:

- Presence – each XMPP node may publish information about its state, which is automatically broadcast by the server to the interested parties
- Message routing rules – each "stanza" produced by any node may include additional information describing which nodes it is addressed to, how should it be processed in case the specific node is unavailable etc.
- File transfer – XMPP includes numerous mechanisms for file transfer, including both in-band and out-of-band methods. This enables for efficient transport of data, which is not easily representible as XML.
- Publish/Subscribe – any node is capable of publishing or subscribing to arbitrary type of information, identified by XML namespace and/or publishing node address. Subscribing nodes automatically receive notifications upon publication update.
- Text messaging – the primary purpose XMPP was invented for, is useful in the context of communication between operators of the system.

It is worth noting, that within the scope of WP1 system XMPP protocol would be used as a transport layer, on top of which a service layer is proposed, where XMPP "stanzas" are used as a means of delivering information and invoking published functionality. Therefore, the XMPP server should not be identified with security system server, which (if any) would function as a service deployed within the XMPP network.

An important feature of XMPP protocol, which greatly influences the way aforementioned multimedia traffic is delivered, is its Jingle extension, which implements multimedia signalling protocol on top of XMPP. This way it is possible to use XMPP network directly as a means of controlling out-of-band multimedia sessions described in subsequent session.

6.5.2 Secure multimedia communication

XMPP protocol is typically used with TCP streams with **TLS security layer**. Such design together with additional XML encapsulation make it highly inefficient in terms of latency and overhead, when it comes to the transport of real-time, high-volume data, such as multimedia streaming. In order to efficiently transport such densely-packetized traffic, typically protocols based on UDP are used, which thanks to lower overhead allows for much better bandwidth utilization. Industry standard pertaining this use is RTP protocol, commonly used in applications like VoIP telephony and Internet TV or radio. RTP provides payload encapsulation and identification, packet numbering and synchronization and has provisions for exchange of various statistics relating to delivered streams. There is also an extension to the core protocol called SRTP which adds support for payload **encryption** and **signing**, adding **channel protection**. However, RTP by itself does not include any mechanisms for establishing and negotiating the session. For this purpose so called signalling protocols are used, such as SIP, RTSP or aforementioned XMPP/Jingle. The choice of Jingle signalling allows to exchange information regarding the multimedia session in a **secure**, **authenticated connection**. This information includes parameters like IP addresses and UDP port numbers of communicating endpoints, definition of media formats used within session and security data for the configuration of block cipher used for data encryption.

The use of RTP poses similar difficulty as described earlier when it comes to use of firewalls and NAT devices. However, it should be noted that there exist two approaches to the problem, both of which are applicable. The first one is the use of so-called Interactive Connectivity Establishment methodology, designed specifically to aid in establishing RTP session in such complex environments. The second solution is the use of proxy, which is less efficient than ICE, however in the context of described system may be still better option because it is possible to integrate the proxy with centrally-located media recording service.

Similarly to general-purpose communications, it is not required at this stage to identify the best ciphersuite for use with SRTP encryption. The proposed solution is flexible, so it is possible to switch to any other encryption algorithm easily. The actual cipher should be chosen based on discussion with WP8 representatives. The communications framework architecture is presented in Fig. 4.

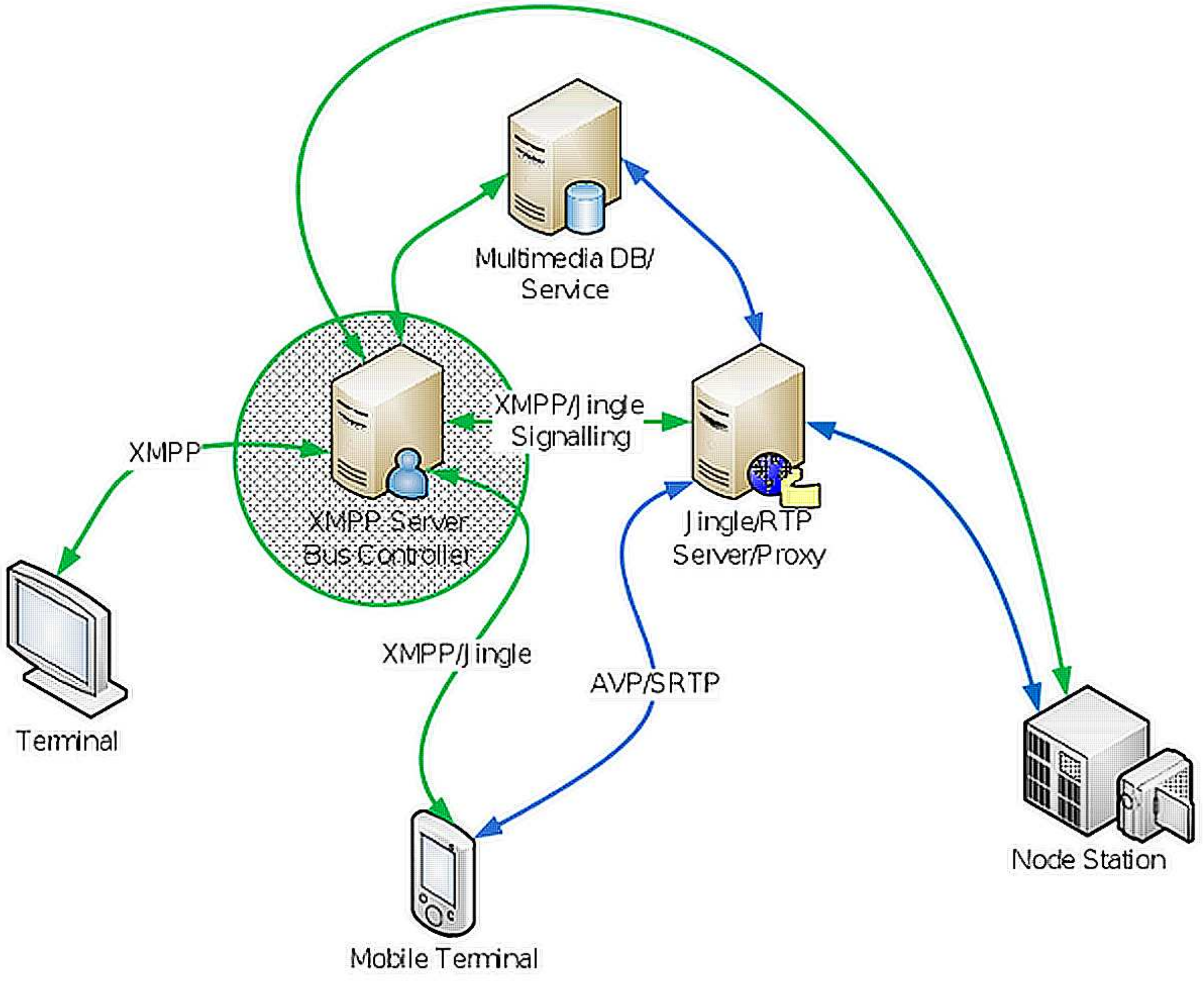

Figure 4. Communications framework architecture

7 Conclusions

For WP1 the first step is to gather End-User requirements helping to define functionality of the system, specifically for task related to automatic detection of events. For that purpose the End-User Questionnaire was established, created with cooperation of all INDECT Project Partners.

The objective of the End-User Questionnaire, and outcomes of the analysis were presented, resulting in preliminary specification of the functionality for event detection and hardware specification.

Automatic event detection algorithms that will be developed in WP1 are meant to aid a person operating the monitoring system, allowing concurrent analysis of practically any number of audio and video streams (limited by computational power, which is easily extendable). The operator work will be verification of alarms instead of inspection of multiple number of streams in the same time, resulting in increase of effectiveness of threat detection.

Other added values were discussed, such as reduction of storage space, automatic protection of content recognized as a private, prediction of dangerous events, and detection of previously overlooked events.

Gathered and analysed End-User Questionnaires leads to definition of list of the events that are intended for automatic recognition with the WP1 event detection module.

Content of the document should be treated as a road map for further work, and it is assumed that all of the requirements are meant to be reconsidered in a time span of INDECT Project. Final specification of these features will be provided in the following deliverables:

- For final hardware specification: D1.2 Report on NS and CS hardware construction (M20)
- For final specification of event detection: D1.4 Multimedia database documentation with analysis of recommended algorithms (M45).

Document Updates

Version[8]	Date[9]	Updates and Revision History[10]	Author
v20090720	20/07/2009	First version of Table of Contents	Piotr Szczuko, GUT
v20090907	07/09/2009	Section 6.2 added	Piotr Romaniak, AGH
v20090930	30/09/2009	Section 6.2 extended Section 6.3 added	Piotr Romaniak, AGH
v20091006	06/10/2009	Sections by co-authors added, extended and corrected	Piotr Szczuko, GUT
v20091006	06/10/2009	Final Sections added, extended and corr.	Piotr Szczuko, GUT
v20091008	08/10/2009	Stable Version, released for internal review	Piotr Szczuko, GUT
v20091029	29/10/2009	Final Version	Piotr Szczuko, GUT

[8] In form of "vYYYYMMDD"; Version number and edition should correspond to the actual document name conventions.

[9] In form of "DD/MM/YYYY"

[10] Attach as appendix document reviews when appropriate; describe also the current status of the document e.g. "released for internal review", "released for comments from partners"