Book Chapter:

PERVASIVE VIDEO SURVEILLANCE SYSTEMS OVER TCP/IP NETWORKS

L. Badia[1], A. Erta[12], U. Malesci[2]

[1]*IMT Lucca Institute for Advanced Studies, Lucca, Italy*
{l.badia, a.erta}@imtlucca.it

[2]*Fluidmesh Networks Inc., Milano, Italy*
umberto.malesci@fluidmesh.com

Tel. +39 0583 4326716

ABSTRACT

Traditional analog video surveillance systems technology has recently become inadequate to face the massive demand of security systems consisting of hundreds and sometimes thousands of cameras often deployed in hostile environments tens of miles far away from the control room. During the last few years, the rapid growth of the digital technology has produced sophisticated cameras which can directly record high-definition digital videos. The packetized video stream can be straightforwardly conveyed to the control room relaying on common IP network infrastructures. This solution results extremely flexible as the network infrastructure can be built over a wide variety of heterogeneous network technologies from the traditional Ethernet-based Local Area Networks (LANs) to the recently proposed Wireless Mesh Networks (WMNs). However, the widespread adoption of IP-based solutions for video surveillance poses serious problems in terms of required bandwidth, processing power, network security and system dependability. In this paper, we first investigate the advantages of the IP-based video surveillance systems over the traditional analog ones. Then, we describe the technical challenges and the open research issues which still lack an ultimate solution which permits to completely abandon the traditional analog technology. Finally, we propose and verify, by means of a case study, a methodology to address the design of video surveillance systems in real deployment.

KEYWORDS

Video Surveillance, Video Streaming, Wireless Mesh Networks, Multi-hop wireless networks, TCP/IP networks, Network engineering, Design guidelines.

INTRODUCTION

During the last decades, the world has become on the move. Intelligent transportation and emergency or disaster recovery facilities are more and more often integrated with remote video control. At the same time, urbanization trends combined with socio-economic changes have changed criminal and terrorism-related activities to a globalized phenomenon. As a consequence, the market for security and video surveillance systems has expanded significantly (Welsh & Ferrington, 2002). Security-system installers and integrators face several challenges in designing security and video surveillance systems that must operate in difficult and demanding settings, streaming and recording simultaneously hundreds and often thousands of video flows. In the last few years, the physical and video-security field is experiencing a massive shift from analog transmission over coaxial cables and fiber optic to digital transmission over IP networks (In-stat, 2006). In fact, until the mid-Nineties, recording was mainly performed on tapes using VHS equipments which require analog video streams as input. In the late Nineties, the majority of tape recorders have been substituted by Digital Video Recorders (DVRs) which are embedded systems that integrate hard drives with video encoding hardware. The analog video streams coming from, for example, coaxial cables to the DVR are digitized and compressed using video encoding algorithms. To this end, it is possible to subsequently transmit the video stream as a sequence of independent Joint Photographic Experts Group (JPEG) pictures, so as to realize what is informally called Motion JPEG (M-JPEG), or to utilize techniques such as Moving Pictures Expert Group (MPEG) which exploit inter-frame prediction; after this step, the stream is recorded on the hard drives.

The major drawback of both VHS- and DVR-based video surveillance systems is that the transmission from the cameras to the recording and viewing locations is kept analog, and therefore video quality is often directly affected by the distance between the control room and the cameras. Additionally, installing analog cameras in rural or even dense residential areas may not be feasible given the impossibility of laying long enough cables to reach the control room. Embracing the digital revolution from the camera to the head-end location and encoding the video stream into TCP/IP-like packets directly in the camera present multiple advantages (Sedky *et al.*, 2005; Cisco, 2006). In fact, the system designer can leverage existing networks and infrastructures irrespective of the specific medium used to convey the video stream (e.g., copper, fiber-optic, radio waves etc.). Furthermore, viewing and recording capabilities can be distributed by enhancing the cameras with integrated reporting and recording systems. However, the IP-based approach requires solving several additional issues in order to meet the requirements, in terms of security and reliability, of a traditional video surveillance system. For instance, the high bandwidth capacity required to transmit hundreds of video streams simultaneously (Koutsakis *et al.*, 2004) and the processing power needed to encode and decode multiple MPEG-4 streams (Ziliani, 2005) are definitely two main problems of IP-based systems which still lack a definitive solution. In this work, we describe the advantages and investigate the research open issues and technical challenges of the IP-based approach with respect the traditional analog systems for video surveillance. Furthermore, we analyze several existing IP-based solutions and develop a viable methodology to deploy efficient and effective video surveillance systems. Finally, we present, as a case study, a video surveillance system installed in a seaport in Europe where the above methodology has been successfully employed at the design stage.

BACKGROUND

In this section, we describe in detail the components of a video surveillance system. Specifically, we first analyze the video surveillance application scenarios. Then, we focus on the specific system components and we describe their characteristics. Finally, we investigate the peculiar features of IP-based video surveillance and review the enabling IP technologies for supporting IP video streaming.

A. Video surveillance applications

The major applications of video surveillance systems fall under the *physical* security umbrella. Every physical security system has to verify the major objectives of deterrence, detection, and verification (Smith & Robinson, 1999). The video surveillance system is usually adopted as a verification tool to reduce false positives coming out of the intrusion detection and access control systems. Typical scenarios are office buildings, banks, museums, parking lots and industrial plants. In other situations, such as municipalities and highways, video surveillance is not employed in conjunction with a detection system (e.g., perimeter protection, intrusion detection, etc.) but it is rather a stand-alone system managed by the local security/police personnel. In these specific contexts, the Closed Circuit TeleVision (CCTV) system plays the role of a deterrence, detection and verification means at the same time. During the last few years, due to the increasing adoption of intelligent video analysis software, video surveillance systems are becoming from passive verification systems to complete detection solutions. Video analysis is able to detect "alarm" situations by automatically analyzing the video streams. Examples of such applications are lost baggage detection in airports or the queue formation on highways (Wolf *et al.*, 2002; Bramberger *et al.*, 2006).

Both enterprise and government organizations are increasing the number of cameras on their premises to provide a 24 hours video-stream recording which helps managing security and liability threats at the same time. Thanks to extensive and pervasive video surveillance, a lot of organizations are nowadays able to decrease insurance premium and, thus, achieve more efficiency in asset management. For instance, casinos are investing huge budgets in video surveillance by installing systems with 2,000-5,000 cameras per casino. The CCTV systems help casinos to increase security, reduce frauds and solve many liability issues involved with gambling, thus, greatly rewarding the invested budget.

Other important applications for video surveillance involve monitoring specific industrial processes. However, these applications usually require special types of camera equipment which exhibit different technological challenges with respect to those for large scale physical security. For example, production plans in the pharmaceutical industry employ very high frame rate cameras (100+ frame per second) and hyperspectral imaging to monitor and certify their processes (Hamilton & Lodder, 2002). However, these special industrial applications are not the primary focus of this paper. Therefore, in the following, we will mainly refer to the physical security applications and their research challenges.

B. Video surveillance systems components

Every IP-based CCTV system can be divided into 5 main components:

1. Video Capturing
2. Video Encoding
3. Video Transmission
4. Video Monitoring & Management
5. Video Recording

1) Video capturing

The video capturing component of any CCTV system is the camera, whose core is the Charged Couple Device (CCD) sensor that converts light signals in electrical signals. Traditional analog CCDs are now very often replaced with Digital Signal Processing (DSP)-based CCDs. Also, cameras usually have an analog output with a Bayonet Neill-Concelman (BNC) connector for a coaxial cable.

2) Video encoding

As the majority of cameras installed today have digital DSP-based CCDs but still analog BNC output, a separate encoding device is required in order to streams video over an IP network. Video encoders (or video-server) are basically analog-to-digital converters that encode an analog video stream coming from a coaxial cable into a stream of IP packets usually compressed with MPEG-4 or M-JPEG. Video encoders usually have a BNC input and an IEEE 802.3 Ethernet input/output. To control motorized Pan-Tilt-Zoom (PTZ) cameras, many encoders are also equipped with a serial output which reports the telemetry readings of the electric engine. Recently, the increasing success of IP cameras, where video capturing and video encoding components are integrated into the same device, has reduced the need for separate video-encoders (Cai *et al.,* 2003).

3) Video transmission

In traditional video surveillance systems, the video stream was conveyed towards the control room through analog transmission on coaxial cables, Unshielded Twisted Pair (UTP) cables or multi-mode and single-mode fiber optic. However, in the last decade, mid to large scale video surveillance systems are migrating to wired IP-based network, and, more recently, to wireless infrastructures thanks to their increasingly lower installment and management cost (Norris *et al.*, 2004).

IP networks. As the narrow waist of the Internet, the IP protocol itself is helping also the video surveillance world in integrating different transmission technologies together. Moreover, packet switching networks allows multiple video streams to be conveyed on the same cable by sharing the available transmission capacity and leveraging the infrastructure costs with other data, video and voice flows.

Transmission Media: Copper, Fiber, Wireless. In-building IP-networks are mainly based on copper and fiber as transmission medium. Conversely, large outdoor environments are usually covered by Wide Area Networks (WANs). WANs are usually provided by large Internet Service Providers (ISPs) through high capacity leased lines or wireless technologies. A major challenge in using leased-lines and WAN technology for video surveillance is due to the high bandwidth required per video stream that dramatically bring up the leasing and operating costs.

4) Video monitoring and management

The video monitor and management system is the user interface that allows the operator to select different video streams and watch both real time and recorded video. Video monitoring and management systems are often software-based and run on generic workstation hardware. Very often the ultimate bottleneck of large video surveillance systems is the processing power required by the video monitoring workstation to decode the large number of video streams acquired. Due to the heavy hardware requirements, many custom solutions at a hybrid hardware-plus-software level are available to reduce issues related to inappropriate or insufficient general purpose hardware/software.

5) Video recording

The video recording hardware works very often in conjunction with the video management system and sometimes run even on the same server. The video recording component is usually referred to as Network Video Recorder (NVR) and tends to be a generic server with high hard disk capacity or connected to storage arrays, like Storage Area Network (SAN) or Network Attached Storage (NAS) technologies (Gemmell *et al.*, 1995). Usually, the NVR does not require extremely high processing capabilities because it does not decompress the video streams but it simply indexes the video and records the video on the network storage system.

C. Video surveillance over IP

Based on the above discussion, we can identify the following key advantages of video streaming over IP network with respect to traditional solutions:

- leveraging existing networks and infrastructures;
- distributing intelligence, recording and viewing capabilities;
- enhancing system flexibility in expanding the network with different transmission medium (e.g., copper, fiber-optic, radio waves etc.).

We now discuss the above key points in more detail.

Providing connectivity for multiple and diverse applications has always been among the philosophical basis of the Internet and IP networks in general. An IP network is multi-purpose in its nature (Clark, 1988). Thus, on the one hand it is possible to use existing off-the-shelf technology for the needs of our service, even though they are rather specialized, i.e., instantiating the transport flows to convey video streams. On the other hand, existing network structures deployed for data or Internet-related applications can come in handy when video flows generating from security cameras need to be transmitted across a building or even across an entire city. Moreover, thanks to the presence of large Wide Area Networks (WANs) and/or the Internet, remote monitoring of cameras hundreds of miles away is allowed without requiring any dedicated infrastructure. Finally, using IP flows is preferable from the point of view of data retrieval and storage.

From an architectural perspective, the migration of a video security system toward IP networks means that intelligence and capabilities move from the control room towards the cameras (Norris *et al.*, 2004). Compared to a traditional analog CCTV camera, an IP camera is a more complex embedded computing system, where the optical, video, encoding, and transmission components are integrated. Modern IP cameras very often include also a Web-server which allows the user to display the video stream simply through a Web-browser. Moreover, many IP cameras nowadays implement advanced artificial intelligent algorithms that make them control how and when to stream the video (Bramberger *et al.*, 2006).

Since we can regard an IP network as a distributed infrastructure that can be accessed at any point, viewing and recording can be also performed in a distributed manner. Specifically, we can have multiple viewing stations in different physical locations, each of those with different information access privileges. Additionally, each camera can locally record the video stream, both to save bandwidth and to increase the reliability and the resiliency of the overall security system.

Finally, migrating to IP networks allows for selecting the most appropriate transmission technology for every video stream and transparently switching transmission media along the path that the video stream takes to reach the control room. In this way, the IP protocol becomes also the common denominator along the path of every video streams (Saltzer *et al.*, 1984).

However, it must be observed that the usage of an all-IP structure, distributing the processing capability all over the network, can also lead to increased costs and vulnerability of the networks. For what concerns the cost increase, this is still clearly compensated by the advanced features. Given the increasing concerns about physical security, it is reasonable to assume that this additional expenditure is tolerated as long as it goes with an improved service. Still, concerns may occur about the increased safety protection required by more expensive all-IP terminals against physical damage or vandalism. From the point of view of network security and vulnerability, the challenges are different and, at the same time, more intriguing as they mainly involve the network management itself. This point is even stronger for wireless networks, which pose additional issues to the developers, as will be discussed, together with some design guidelines, in the next subsection.

D. Migration and extension to wireless networks for video surveillance systems

The concept of utilizing flexible wireless interconnections in a dynamic fashion is not itself new. Already in the early Nineties, ad hoc networking became popular thanks to the diffusion of notebook computers, open-source software, and viable communication equipments based on radio frequency (RF) and infrared (Ramanathan & Redi, 2002). The concept of ad hoc network, already established for military application, moved to commercial civilian scenarios. Within 10 years, Wireless Fidelity (Wi-Fi) networks started to spread, creating a revolution in Internet service provisioning. With the increasing popularity and rising demand for more Wi-Fi connectivity, implementation problems arose. In fact, Wi-Fi typically requires extensive wired infrastructure to access the backhaul network, which is often expensive to provide and easy to damage, thus violating the security of the information delivery. To this end, the wireless mesh network (WMN) paradigm recently appeared as a valid alternative to wired connection, offering an easy and economical means to provide broadband wireless connectivity (Nandiraju *et al.*, 2007). In place of an underlying wired backbone, a WMN forms a wireless backhaul network, thus obviating the need for extensive cabling. However, WMNs are based on multihop communication to form a connected network. It is well known that multihop links are often limited in throughput and capacity in practical scenarios. For this reason we will employ the following guidelines in creating a WMN to use in video surveillance applications.

First of all, we focus on tree based solutions, because the video flow must be sent to a single data gateway, corresponding to the control room terminal. This means that we can exploit existing solutions for data collection and routing available for ad hoc and sensor networks (Al-Karaki & Kamal, 2004). However, differently from sensor networks which are commonly assumed to comprise hundreds of nodes, our video surveillance wireless network is much smaller, thus simplifying the management. Moreover, we also know the network topology in advance, which enables significant simplifications in the management of wireless communication modes. In fact, we can think of distributing a centralized management policy so as to avoid inefficiencies related to distributed and/or random access. Another guideline we will use is to limit as much as possible multihop relaying. As a rule of thumb, WMN do not work well in practice if they have to relay the traffic for more than three hops (Liese *et al.*, 2006). This limitation is relevant for our case study, since it is also reflected in bandwidth limitations.

A well known hurdle that wireless networks have to face, which very often is the ultimate limiting factor of their capacity, is the wireless interference phenomenon (Jain *et al.*, 2005). In particular, as the wireless medium is inherently broadcast, signals propagate in every direction and are virtually audible by every other terminal in the network. Apart from being a security concern, which will be discussed in the following, this also represents an inefficiency element from the propagation point of view. However, since video surveillance systems mostly consist of static terminals, we can think of using directional antennas. This, beyond decreasing interference, also improves link reliability (Ko *et al.*, 2000). However, as will be shown in the following, also radio bridges realized with directional antennas may become unreliable, especially in the presence of physical obstacles. Actually, the existence itself of the video surveillance system is justified by the desire of properly monitoring the presence of moving exogenous objects. To properly react to this case and more in general to guarantee improved reliability, we will make use of terminals with multiple antennas (usually two) so as to exploit the antenna diversity principle. In other words, we have an extremely low probability that *both* antennas fail at the same time.

For all these reasons, our scenario includes both classic wireless network characteristics but also original aspects. For what concerns network and medium access issues, it can be regarded as a standard WMN, with all the implied advantages in terms of ease of deployment and low costs. However, all the special elements described above allow solving, with limited cost increase, well known problems of wireless networks and designing a reliable system.

TECHNICAL AND RESEARCH CHALLENGES

Despite the aforementioned key advantages, the TCP/IP technology and the related IP-based networks were not intended to support the stringent requirements in terms of bandwidth and system reliability which are required by a security system. Therefore, important points must be taken into account and several issues, often in contrast with each other, arise when employing this technology in the field video surveillance:

- great amount of network bandwidth (already high even for few cameras);
- high processing capability required to encode, decode and record multiple video streams;
- network security protocols, algorithms and policies for guaranteeing the privacy and authenticity of the video streams;
- encoding algorithms designed for live streaming which offer bandwidth efficiency, low processing power requirements and whose output can be used as evidence in legal trials;
- mechanisms to detect and react to Denial of Service (DoS) attacks and guarantee service even in faulty situations.

Due to the numbers of streams involved, video-streaming and, in particular, video surveillance is among the most critical application in terms of bandwidth requirements. Every video camera usually requires between 1 and 10 Mb/s for a high quality stream with a frame rate of 25 frames per second and an image resolution of 640×480 pixels (Koutsakis *et al.*, 2005). Common enterprise and municipal networks are not designed to support tens or even hundreds of streams because the majority of currently installed Local Area Networks (LANs) only provides a maximum (theoretical) bandwidth of 10/100 Mb/s. Note that a mid-size shopping mall might easily need between 100 and 300 cameras. While for data transmission purposes a common 100 Mb/s LAN is adequate for most shopping malls, to stream high resolution video flows a 1 Gb/s network is barely sufficient and a 10 Gb/s infrastructure is clearly advisable.

There are three major approaches and techniques to cope with the bandwidth requirements of video security applications. First, *over-provisioning* of network resources can be introduced at the network design stage. Deploying high bandwidth networks by means of 1 Gb/s Ethernet standard should be a common solution for LANs. Backbones at 10 Gb/s will be necessary to guarantee a good interconnection among the LAN segments. Second, video *compression ratio* can be increased by using more efficient video compression algorithms. However, in general, the higher the compression ratio, the slower the algorithm to decode the video and, thus, the higher the computational power needed. Finally, implementing smart and efficient context-aware video *adaptation algorithms* (Ziliani, 2005) in the camera which dynamically switch from the idle to video transmission state when situations of interest are detected and vice versa.

Quite interestingly, in many video security systems, the real bottleneck is not bandwidth but the processing complexity. Although the migration to differential video encoding algorithms (e.g., MPEG-4) has greatly decreased the bandwidth required per video stream, the more complex compression schemes have dramatically increased the processing requirements and, therefore, the costs of hardware capable of simultaneously decoding multiple high resolution video streams.

As video surveillance is mostly used in crime prevention, network security, privacy and detection of faulty situations are extremely important topics (Welsh & Farrington, 2002). Furthermore, guaranteeing the video surveillance service to be always available is much more important than for other types of services. Specifically, the video surveillance infrastructure must be robust against DoS attacks and promptly report any problem (attack or fault) to the control room, along with as detailed as possible information about the location where the issue arose. Susceptibility to DoS is an intrinsic problem of any service provisioning system where events must be processed to determine their validity. As with prank telephone calls or ringing of door bells of old times, an effective means of preventing DoS attacks from occurring lies in identification of the attacker. This is also the only fundamental solution, given the intrinsic susceptibility of such service provisioning systems to DoS. If the physical source of DoS traffic can be identified, then at the very least the invaded network element can be isolated or shut down and, in some instances, the attacker’s identity can be further traced back. In our video surveillance system, it may be thought of dealing with DoS by means of both proactive and reactive countermeasures. In particular, route-based packet filtering solutions (Park & Lee, 2001) may be used to this end. Indeed, observe that our WMN insulated from the external networks, the only gateway being the control room, which is assumed to be reliable enough. Thus, intrusion from external network can be easily identified and filtered at the control room terminals. Malicious data sent to the camera through the wireless links can instead be counteracted by means of the surveillance system itself. In fact, as discussed previously, directional grid antennas are used in the WMN, which are not easy to jam or maliciously disturb unless an external device is put within line of sight of transmitters and receivers. However, this would be likely detected by the surveillance system itself, thus enabling the detection of the intrusion and possibly also identifying the intruder.

VIDEO ENCODING AND COMPRESSION

The most common types of video encoding employed in video surveillance systems are M-JPEG and MPEG-4 (Halshall, 2001). M-JPEG codifies each frame as a JPEG picture, without exploiting any inter-frame prediction. In this manner, the resulting video stream consists of independent frames. The advantages of this approach are related to the low processing power required in compressing and decompressing the video streams. Moreover, M-JPEG is often chosen when trial evidence is the objective of the recording: M-JPEG can be divided up into separate frames that can be analyzed independently of the previous and subsequent frames. Also, since JPEG frames are themselves compressed, M-JPEG format can also be adjusted to save bandwidth.

However, with M-JPEG there is no exploitation of the inherent correlation between different frames. For this reason, differential encoding algorithms, in particular the one belonging to a MPEG format, are increasingly replacing M-JPEG due to their higher bandwidth efficiency, as they take advantage of the static nature of the scene shot during the routine cameras operations. Note also that efficient transcoding techniques are available to translate MPEG flows into M-JPEG in an efficient manner; in this way also MPEG flows can fulfill the requirement of deriving static pictures from the video flow, to be used as trial evidence.

In the following, we will focus in detail on the most recent one, MPEG-4, since to-date the majority of CCTV systems runs MPEG-4 or similar differential encoding algorithms. MPEG-4 (Fitzek & Reisslein, 2001) extends the applicability of the initial standard MPEG-1 and the more recent format MPEG-2, which is specifically targeted at high-definition television (HDTV) applications. In particular, MPEG-4 provides support for both very low bit rate encoding and for 3D content and complex video/audio objects. In this way, MPEG-4 can be said to extend the provisioning of efficient coding for video flows covering the whole range from very low bit rates up to HDTV and beyond.

Analogously to MPEG-1 and MPEG-2, MPEG-4 decomposes the video flow exploiting its redundancy by means of prediction mechanisms. However, differently from MPEG-1 and MPEG-2, the unit of representation is not called a *frame*, but rather a *video object* (VO). In the simplest case, for low bit rates, the whole scene may be a single VO. Each VO is layered into video object layers (VOLs), for example it is possible to have one base layer and several enhancement layers. VOLs are ordered sequences of snapshots in time, referred to as video object planes (VOPs). For each VOP the encoder processes the shape, motion, and texture characteristics.

The shape information is encoded by bounding the VO with a rectangular box and then dividing the bounding box into macro-blocks (MBs). Each MB is classified differently according to its position and then shape coded. The texture coding is similar to frame-based standards such as MPEG-1 or H.263, using VOPs as frames, which are classified as intracoded (I), forward predicted (P) or bidirectionally predicted (B). In an I VOP the absolute texture values in each MB are coded using Discrete Cosine Transform (DCT), whose coefficients are quantized and coded with variable length. In P VOPs each MB is predicted from the closest match in the preceding I (or P) VOP using motion vectors. In B VOPs each MB is predicted both from the previous and the subsequent I (or P) VOPs. The prediction errors are transformed with DCT, and the coefficients are quantized, and coded with variable length. The I, P, and B VOPs are finally arranged in a periodic pattern referred to as a group of pictures (GoP). A typical GoP structure is IBBPBBPBBPBB. However, the standard leaves open the possibility of using different sequences. Note also that using a GoP structure entirely consisting of I VOP will make the MPEG-4 be-

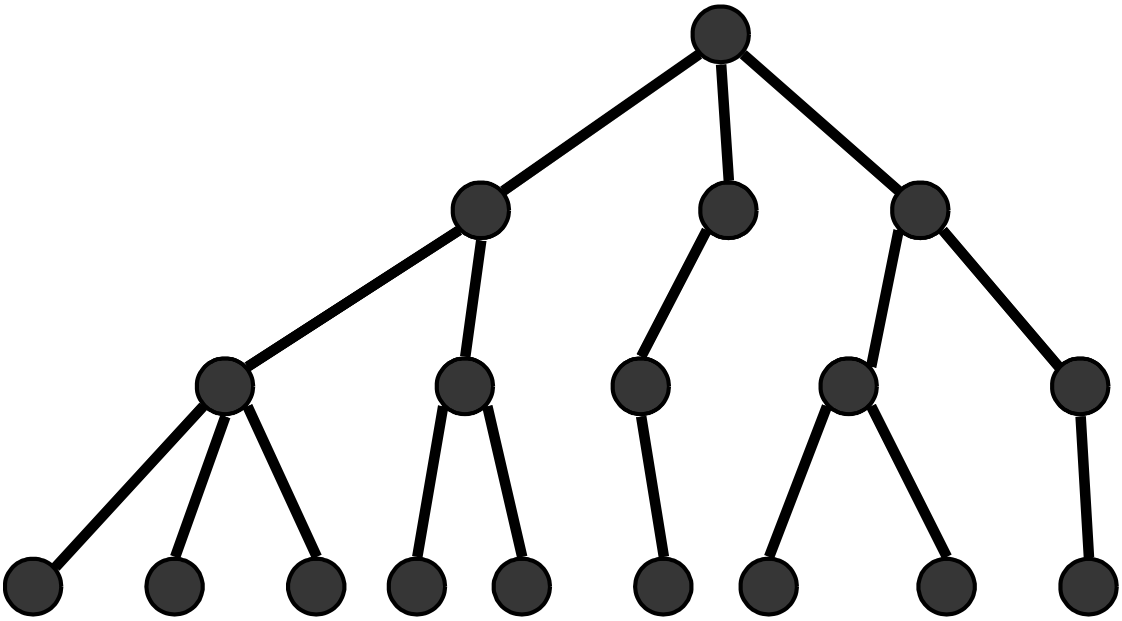

Figure 1. IP Video surveillance network topology.

come a mere variant of the aforementioned M-JPEG technique. More in general, as the standard includes the possibility of regulating the GoP structure, we can fine-tune the system by trading the transmission efficiency, which increases if the exploitation of predictive VOP is pushed further, with the accuracy of VOPs taken individually, which is highest when no predictive VOP are used at all.

Additionally, MPEG-4 provides a number of error resilience and error concealment features to counteract the frequent transmission errors typical in wireless communication; the interested reader is referred to the textbook by Halshall (2001) for further details.

TRAFFIC ENGINEERING AND CENTRALIZED NETWORK DESIGN

Video traffic in a video surveillance network has a very different characterization with respect to the typical data traffic running on a common TCP/IP network. First of all, traffic in a CCTV system tends to maintain a quasi-constant rate rather than showing a bursty characterization as that of the Internet (Molina *et al.*, 2000). In fact, every video encoder streams flows towards the control room 24 hours a day, 7 days a week. The rate of each video stream is constant with the M-JPEG encoding whereas it is variable for MPEG-4. However, even though the latter usually exhibits a highly variable bit-rate with an average and maximum peak rate in common video application, the static nature of the acquired video images of a CCTV system makes the MPEG-4 video rate quasi-constant over time. Note that this consideration is also exploited for developing Object Detection Algorithms (ODA) which are able to recognize motion activity within video frames (Nascimento & Marques, 2006).

While a statistical approach based on the average and peak rate is appropriate to design a network for data and Internet applications, in video surveillance applications, network and resource provisioning should consider the requirements of constant rate flows. Additionally, video frame loss and network congestion are definitely not acceptable for a security system. Therefore, the network must be designed assuming the worst-case scenario in terms of bandwidth requirement of every video stream.

The typical network topology of a video surveillance system is reported in Fig. 1 which represents a tree-shaped graph where edges are communication links and vertices can be considered as Ethernet switches or IP routers.

The root of the topology tree represents the control room where all the video-traffic flows converge. The video streams coming from the leaves of the tree can reach the control room if, at the network design stage, the following condition is enforced. Each sub-tree is connected to its parent node through a link whose capacity is *sufficient* to cover the bandwidth demand of the entire sub-tree. In other words, the following equation must hold:

$$B_{cr}^{k} = \sum_{i \in C_k} b_{i,\max} \qquad (1)$$

where B_{cr}^{k} is the capacity available at the link connecting the sub-tree C_k to its parent node and $b_{i,max}$ is the maximum rate required by every video stream present in the sub-tree C_k. Therefore, the overall capacity required in the control room LAN will be $B_{cr} = \sum_{k=1}^{n} B_{cr}^{k}$ where n is the number of root sub-trees.

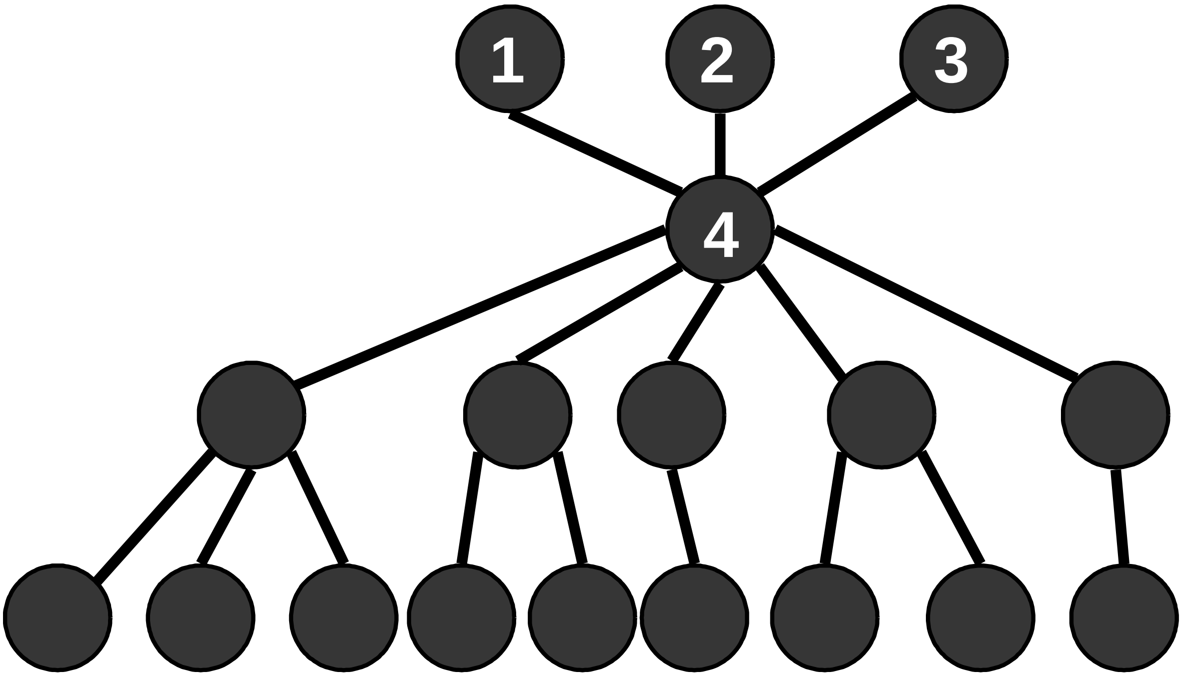

Figure 2. IP network topology with network devices representation.

More precisely, the tree representation of Fig. 1 can be modified as in Fig. 2, where nodes 1, 2 and 3 depict the devices present in the control room, i.e., video monitoring, management and recording, node 4 depicts the control network and/or the control room Ethernet Switch/Router and all the other nodes are cameras or network devices on the field.

We assume that every camera is continuously transmitting a video stream at constant rate. Video streams are usually sent as multicast UDP traffic in order to reduce the bandwidth demand when multiple video surveillance devices (viewing, recording, etc.) need to receive the same stream. However, in CCTV systems, each camera can generate more than one video stream at the same time because often the user wants to view the live streaming at a higher resolution and frame rate compared to the recording stream so as to save hard disk space. In this case we must consider the rate of each video stream to compute the total bandwidth requirements.

Advanced video surveillance cameras and recording software implement features that allow for dynamic changes in the bandwidth and/or frame-rate and resolution of each video stream. For example, when clicking or zooming on a particular video stream, the camera will automatically increase its frame rate and resolution. In other cases, the camera itself increases the frame rate and the resolution when motion in a particular area of interest is detected by performing automatic video-analysis on the video stream in real time (see for example Pelco Endura cameras, at http://www.pelco.com/endura). These advanced features make the bandwidth provisioning more complex and should be considered when designing any video surveillance system.

DISTRIBUTED INTELLIGENCE NETWORK DESIGN

A radically different approach to network design involves moving intelligence away from the control room towards the edges of the network, to the cameras themselves. Several researchers and cameras manufactures are trying to implement video analytics algorithms able to determine whether or not the image is of interest for the operator directly inside the camera encoding chip (Wolf *et al.*, 2002). With this approach, video streams are sent towards the control room only in case an alarm is generated by the camera. The most common version of this distributed architecture involves video motion detection algorithms that are able to determine if there is any movement in the scene shot by the camera. Moreover, advanced techniques perform a detailed analysis of the video stream, searching for special unusual video patterns. For example, by setting borders and virtual fences to delimit the interest scene, an alarm can be generated as soon as any object crosses these virtual borders. Different algorithms are able to determine movement of specific objects, for example if bags are left unattended, or paintings are removed, generating appropriate alarms. Bringing this on-alarm only approach to the extreme, the network becomes idle most of the time and video streams are sent only in case of alarm. This approach, although extremely powerful, is usually not implemented in real life projects, mainly because video analytic algorithms are still at their infancy phase and not totally reliable.

Another distributed intelligence approach is to move video recording functionalities to the edges of the network through integration with the cameras. In this case, the operator decides which video streams are sent to the control room for live viewing. The network is never completely idle but it is designed to transmit only a subset of the entire video streams available. The major drawback of the distributed recording approach derives from the technical challenges and the costs for installing devices with high

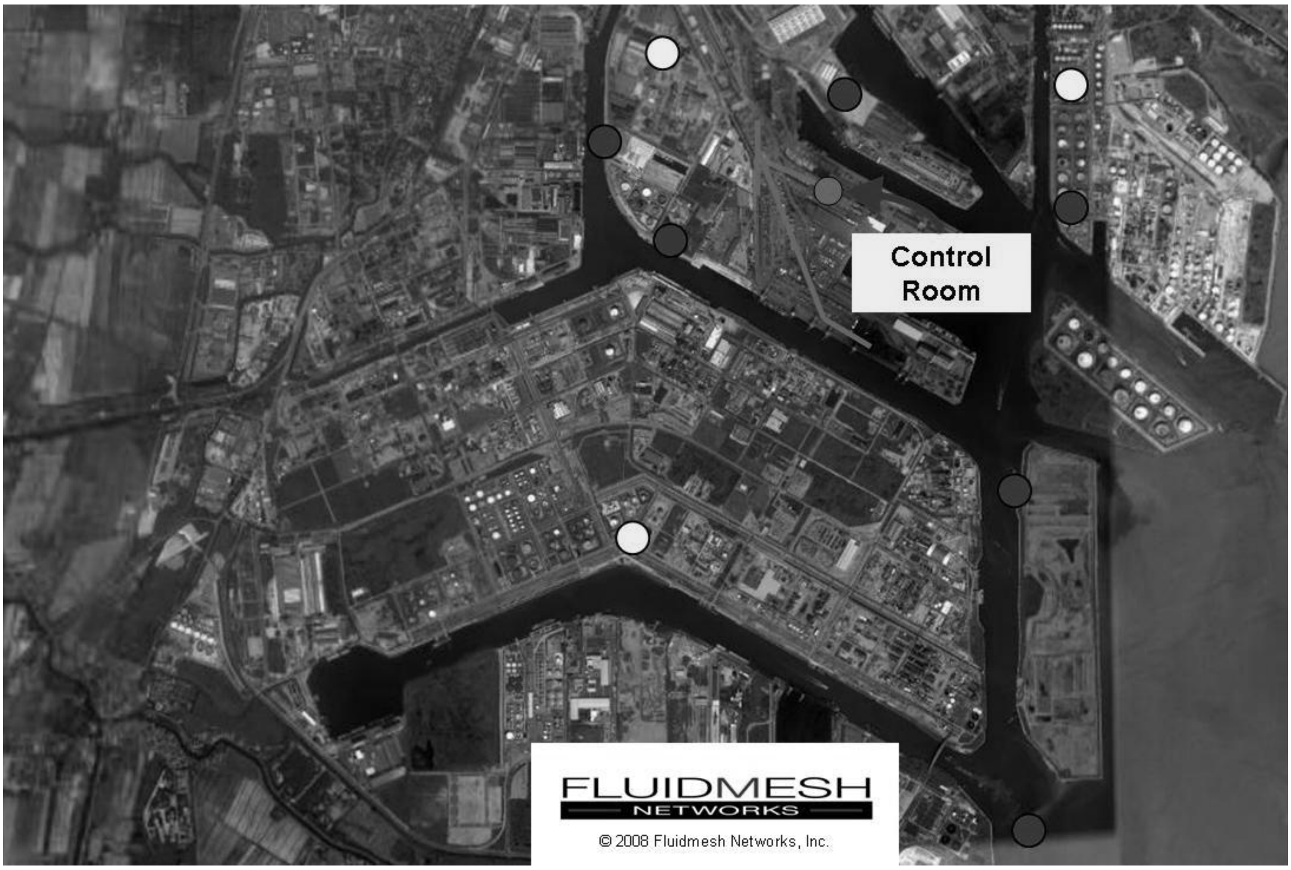

Figure 3. Seaport map and cameras installation points.

recording capacity, like hard drives, in outdoor environments, where the physical conditions and maintenance can be difficult and expensive.

REAL-LIFE CASE STUDY: VIDEO SURVEILLANCE IN A LARGE INDUSTRIAL SEAPORT

We now describe the technical challenges and design decisions made to implement a large-scale video surveillance systems for a large industrial seaport in Europe.[1] The seaport manages about 300.000 containers and more than 1.000.000 passengers per year. The purpose of the video surveillance system is mainly to comply with the international regulations regarding counter-terrorism measure that every port with international traffic must follow. The set of these regulations is often referred to as International Ship and Port Facility Security Code (ISPS). [2]

A map of the seaport is reported in Fig. 3. The seaport required a set of motorized PTZ cameras to monitor the entrances, the main canals and the premises in general. The port facility was initially

[1] All the material in this section is by courtesy of Fluidmesh Network, Inc.
[2] More details can be found on IMO – International Maritime Organization website. http://www.imo.org.

equipped with a fiber optic infrastructure (the blue line in Fig. 3) connecting several buildings to a central location. However, the fiber asset was not widespread and extensive enough to be used directly to cover all the locations which are represented by the red and yellow points in Fig. 3. Moreover, many cameras had to be installed on docks and on buoys at the entrance of the port. Therefore, the existing fiber-based network is extended with a wireless mesh, by means of three parallel mesh networks operating at 2.4 GHz and 5 GHz as shown in Fig. 4. The wireless mesh network employs wireless mesh routers, namely *Fluidmesh 2200*, produced by Fluidmesh Network, Inc., which mount 2 independent radios on board. Due to the distances involved between the mesh routers, directional grid antennas were installed. The use of multi-radio wireless mesh routers is not specifically used here with the goal of increasing capacity. In fact it has been shown, e.g., by Munawar & Ward (2005) that the mere use of dual-interface nodes does not necessarily increase the capacity of the WMN. Actually, dual interfaces can sometimes even lower the throughput due to higher number of transmitters and subsequent increase of interference. However, for our specific application scenario where losses should be kept at a minimum, and in this sense the use

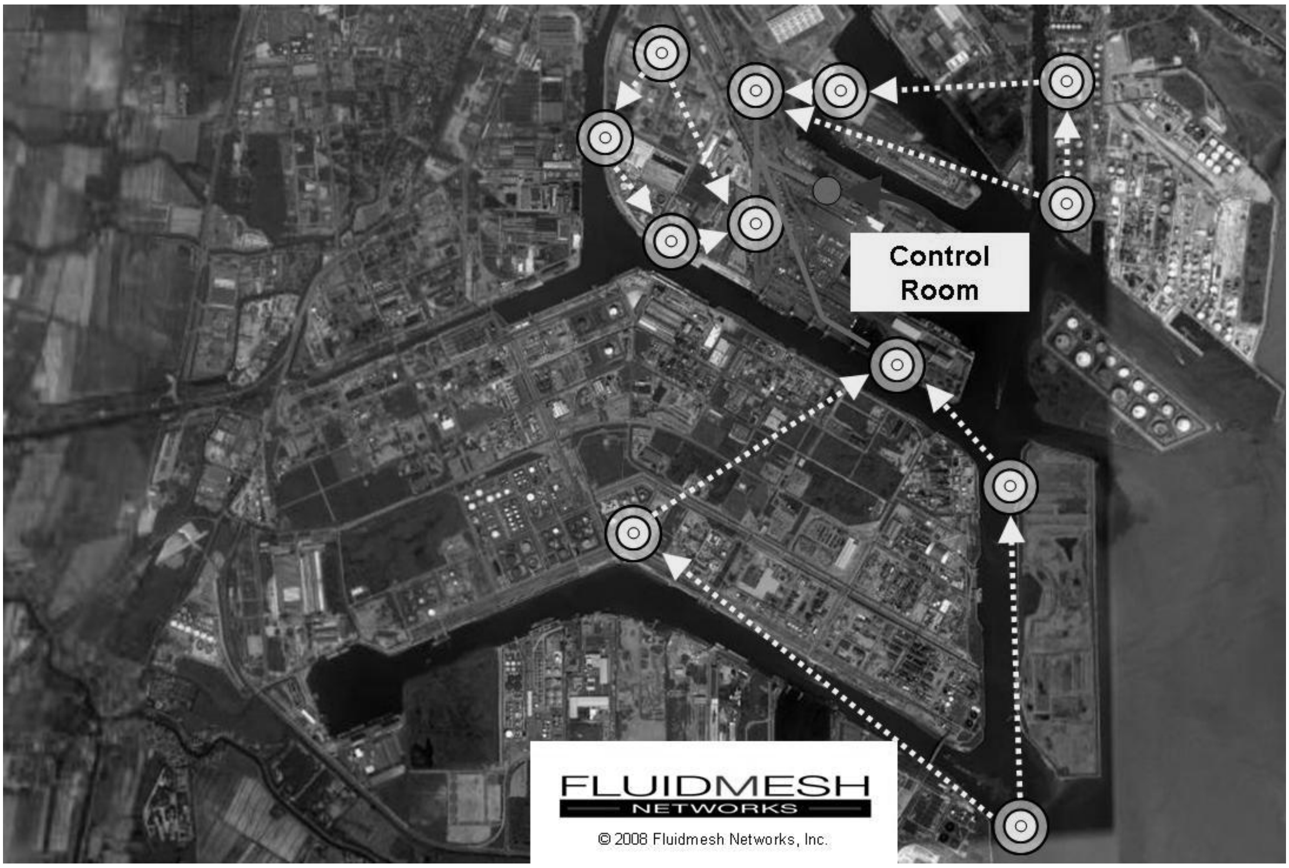

Figure 4. Seaport video surveillance system.

of multi-radio wireless mesh routers offers an advantage with respect to traditional point-to-point wireless bridges, which comes from the *redundancy* of wireless links. Specifically, this solution solves the problem of weak links, as it guarantees that the video stream is not interrupted in case of wireless link failure due to wireless interference or link quality degradation since multiple paths towards the control room are provided. Temporary link unavailability or quality degradation are actually rather frequent in this installation as ships can suddenly cut off the clear line of sight of a link when moving along the canals as documented by Fig. 5. A dedicated path selection algorithm automatically selects the best path based on the actual quality of the links. Importantly, this solution implements a homogenous video surveillance system by means of heterogeneous network technologies, i.e., fiber optics and wireless mesh networks.

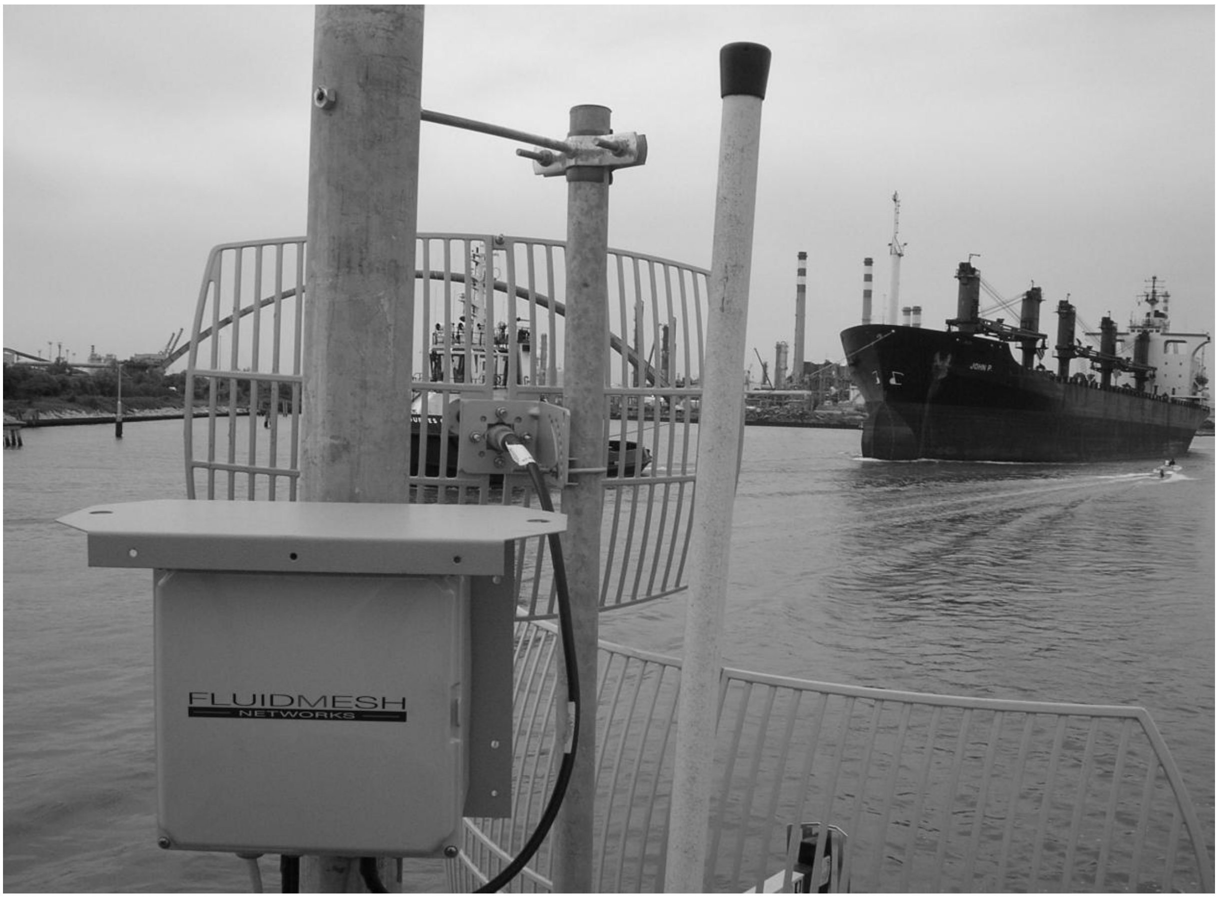

Figure 5. Temporary link unavailability.

With respect to video capturing and video encoding, the solution involved 7 PTZ and 6 fixed cameras with analog output, encoded through MPEG-4 video encoders. Every PTZ camera can be controlled from the control room and it is able to pan tilt and zoom through a telemetry serial input. The telemetry information is carried out by TCP segments. Every video encoder sends one single stream of video packets in multicast mode. Both the live viewing and the recording are performed with the same video stream. The bandwidth requirement per video stream is between 1 Mb/s and 1.5 Mb/s depending on the action in the scene shot and the noise in the image.

A major challenge in any seaport environment is the harsh atmosphere that can damage the electronics of cameras and transceivers, decreasing the Mean Time To Failure (MTTF) of every device by increasing the wear in particular of moving parts. Therefore, the cameras deployed in this specific project are pressurized cameras manufactured by Pelco. Every cameras, both PTZ and fixed, has an IP67-rated enclosure pressurized using nitrogen in order to avoid, at the same time, any leakage of the seaport aggressive atmosphere and the formation of humidity within the enclosure. In a cameras system, humidity can

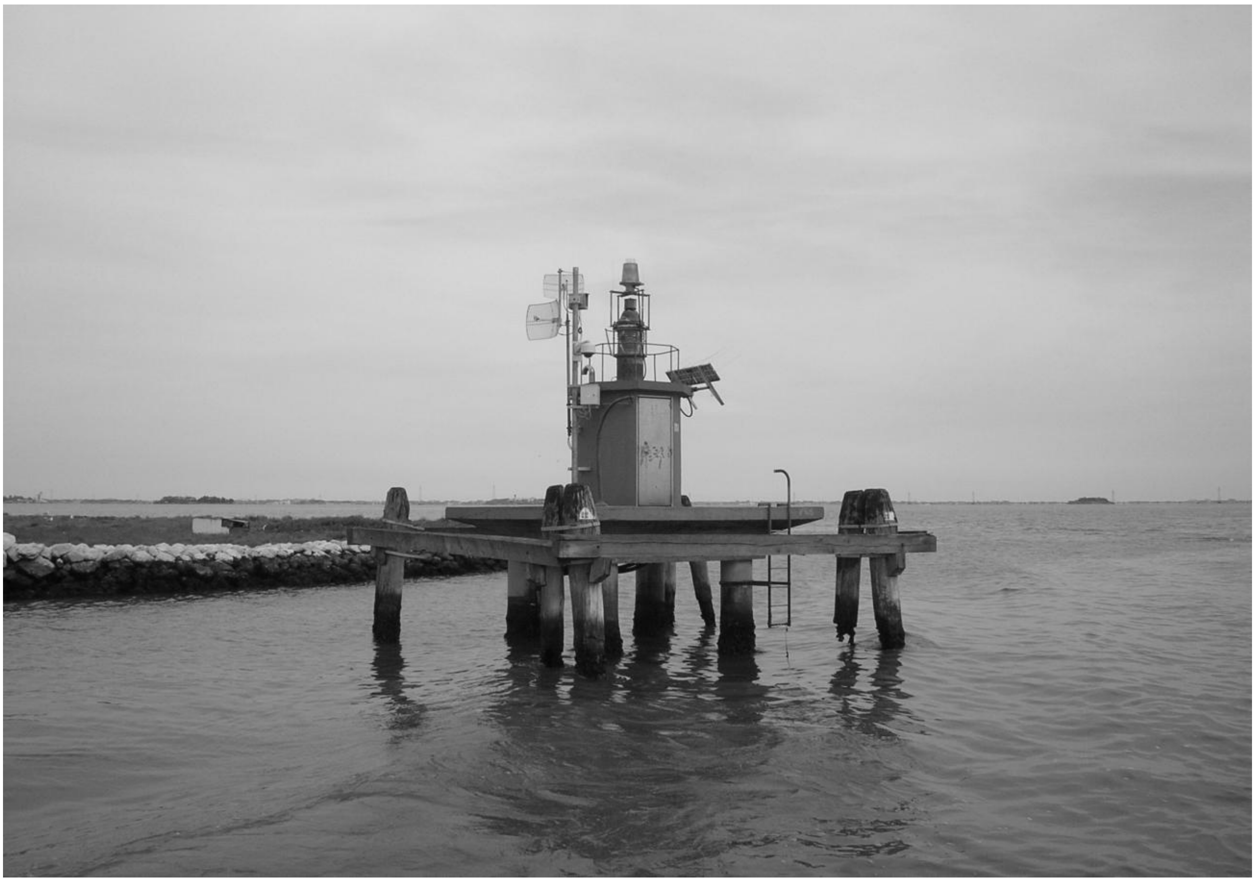

Figure 6. Fluidmesh 2200 wireless mesh router with two directional grid antennas and a PTZ camera.

be very dangerous because the formation of fog within the enclosure can dramatically impair the field of view of the camera. All the external components of the enclosure are manufactured with iron specifically treated for marine use. The pressurized enclosure is extremely important in case of motorized PTZ cameras that present multiple electric servos to change the orientation of the cameras because moving parts are particularly subject to wear that can be accelerated by the direct contact with aggressive atmosphere.

Fluidmesh wireless devices are protected by polycarbonate enclosure IP68 rated. The IP68-rate polycarbonate enclosure, beyond protecting the electronics by any water or humidity leakage, offers also a very good level of protection from vandalism: it is elastic and robust enough to withstand direct attacks like hits with hammers or stones. The IP68 rating guarantees no water leakage not only in case of water splash but also in case of prolonged submersion under water up to a depth of 1 m.

The control room mainly consists of an integrated server that acts, at the same time, as video-management and video-recording unit. A separate workstation is connected to the server and works as a video-monitoring device. Due to the fairly limited number of video-streams involved in this implementation, the video-monitoring unit is able to decompress simultaneously all the 13 video-streams displaying them in real time on multiple monitors in the control room. The operator in the control room is able to move the PTZ cameras focus through a joystick and to retrieve the recordings on the server.

Since each camera is expected to stream video at a rate between 1 Mb/s and 1.5 Mb/s, each stream is conservatively assigned with 2 Mb/s. Each red point in Fig. 3 represents one PTZ camera while each yellow point represents two fixed cameras installed on the same pole.

The yellow circles of Fig. 4 represents a Fluidmesh 2200 dual-radio wireless mesh router with two directional antennas connected. In order to reduce interference, the two radio chips of every mesh router operate on two different frequency bands, i.e., at 2.4 GHz and 5 GHz. Fig. 6 shows a sea beacon where the Fluidmesh 2200 with two grid antennas and a PTZ camera are visible. Each link in the deployment is set to a different 20 MHz wide channel and is thus independent of the other links. From a network design standpoint, we assumed that each link modulates at a *fixed* and *conservative* data rate of 18 Mb/s. Based on that, we designed the system so that the maximum offered load injected in each wireless link does not exceed 14-15 Mb/s. Additionally, due to the unreliable and unpredictable nature of the wireless links, we exploit a heuristic to determine the maximum number of cameras N_{max} in the wireless mesh network. The objective of the heuristic is to avoid congestion in the network while, at the same time, minimizing the number of mesh nodes (gateways) connected to the fiber backbone:

$$N_{\max} = \gamma \min_{\ell}(C_{\ell}) / R_{\max} \qquad (2)$$

C_ℓ is the available capacity (in Mb/s) of a link in the wireless mesh, $R_{\max}$ is the maximum rate required by a camera and γ is a tunable parameter within the range [0.5, 1]. The latter is used to adjust the number of cameras depending on the expected wireless links conditions. For example, setting γ to 0.75 can be found to be a good compromise between performance and cost.

The heuristic is based on the assumption that a percentage of packets will be retransmitted multiple times because either the packet itself or the packet acknowledgment is lost due to channel errors. According to the above heuristic, the maximum number of cameras per wireless mesh results equal to 5.25.

Therefore, to deliver to the control room the total of 13 cameras, we used three parallel mesh networks with independent gateways. In order to verify the capacity required in the fiber backbone, each mesh loop can be considered as a sub-tree whose maximum traffic load is computed according to equation (1) . In this specific case, each sub-tree has a maximum traffic load towards the fiber backbone of about 10 Mb/s and therefore the total capacity required in the fiber backbone would be 30 Mb/s which is definitely reasonable for a fiber network. The video surveillance system has been installed at the beginning of 2006 and it has been continuously operated and extensively used. Despite the harsh and hostile environmental conditions due to the continuous exposure to salt water, no maintenance interventions was necessary thanks to the water-proof polycarbonate enclosure of the Fluidmesh wireless mesh routers and the pressurized cameras. From 2006, more than 300 Terabytes of video data have been streamed to the control room for recording. The large commercial activity of the sea-port, both in terms of goods and passengers, took remarkable advantage from the video surveillance system installation. The seaport security administrators soon decided to employ the latter in conjunction with the standard detection systems to identify anomalous conditions or dangerous situations beyond simple monitoring. In fact, despite the variable and unreliable nature of the wireless links, the network resulted highly reliable with a 99.8% average operational time per year. Three main factors guarantee this high reliability as can be drawn by the above overall discussion. First, the redundancy provided by the multiple wireless paths available from any camera to the control room is efficiently exploited by the optimized Fluidmesh routing algorithm. Second, our simple methodology for the network bandwidth provisioning is effective in coping with the video traffic demand which can vary over time. Finally, the careful selection of material and electronic devices which are specifically suited for harsh outdoor environments contribute to absolutely increase the MTTF of the entire system.

CONCLUSIONS

The rising security concerns will accelerate the demand for continuous video-monitoring and the number of cameras installed in urban areas, private building, airport, seaports and industrial facilities will keep growing. On the one hand, as the number of cameras increases, the technical challenges in managing this enormous amount of information will keep researchers and practitioners busy in developing viable solutions to sustain this growth. Different solutions have been pushed forward affecting the most various system components. However, these solutions are often in contrast with each other. For example, innovative compression algorithms will decrease the bandwidth required per video stream but, at the same time, will likely increase the processing power required for compressing and decompressing the video. On the other hand, users will keep asking for higher resolution in order to obtain more and more detailed video frames. Beyond network bandwidth and processing power, other important design constraints and research topic involve video analytics and automatic video analysis which aim at decreasing the labor component in video monitoring. Due to the continuous demand for more video streams and higher resolution, large scale video surveillance will probably be among the applications that will drive most of the demand for more processing capabilities and bandwidth pushing the Moore’s law limits in the coming decade. In this chapter, all these aspects are analyzed in detail, especially focusing on video surveillance over TCP/IP. In this context, a complete review of the major IP-based technologies is carried out. Additionally, we presented a simple methodology to design and dimension an IP-based video surveillance system. Based on a video traffic characterization, we developed a set of “rules of thumb” to verify whether the video surveillance system will be able to support the overall bandwidth demand. Finally, we described, as a case study, a video surveillance installation in a seaport in Europe which combines a preexisting fiber-optic asset with the wireless mesh networks technology. The above methodology is extended and applied to dimension the network in presence of wireless links.

REFERENCES

Al-Karaki, J.N. & Kamal, A. E. (2004). Routing techniques in wireless sensor networks: a survey. *IEEE Wireless Communications*, 11 (6), 6-28.

Bramberger, M., Doblander, A., Maier, A., Rinner, B., & Schwabach, H. (2006). Distributed embedded smart cameras for surveillance applications. *IEEE Computer*, 39 (2), 68–75.

Cai, X., Ali, F.H., & Stipidis, E. (2003). MPEG-4 over local area mobile surveillance system. In *Proceedings of the IEE Symposium on Intelligence Distributed Surveillance Systems*, 15/1-15/3.

Cisco (2006). Cisco Systems IP Network-Centric Video Surveillance. White paper.

Clark, D. D. (1988). The design philosophy of the DARPA internet protocols. *ACM SIGCOMM Computer Communication Review*, 18 (4), 106-114.

Fitzek, F.H.P. & Reisslein, M. (2001) MPEG-4 and H.263 video traces for network performance evaluation. *IEEE Network*, 15 (6), 40-54.

Freeman, J. P. (2001). 2001 report on the closed circuit TV & video surveillance market. In A. Laurin, Impressive CCTV growth but analog technology lags behind, Axis Company Leaflet.
Available at: http://www.axis.com/documentation/whitepaper/video/2460_article.pdf

Gemmell, D.J., Vin, H.M., Kandlur, D.D., Venkat Rangan, P., & Rowe, L.A. (1995). Multimedia storage servers: a tutorial. *IEEE Computer*, 28 (5), 40-49.

Halsall, F. (2001). Multimedia Communications: Applications, Networks, Protocols, and Standards. Reading, MA: Addison-Wesley.

Hamilton, S. & Lodder, R. (2002). Hyperspectral imaging technology for pharmaceutical analysis. In *Proceedings of the Society of Photo-Optical Instrumentation Engineers Conference*, 4626, 136-147.

In-Stat (2006). In-Sights: Video Surveillance Systems on the Move to IP. Industry Report.

Jain, K., Padhye, J., Padmanabhan, V. N., & Qiu, L. (2005). Impact of Interference on Multi-Hop Wireless Network Performance. *Springer Wireless Networks*, 11 (4), 471-487.

Ko, Y., Shankarkumar, V., & Vaidya, N.H. (2000). Medium access control protocols using directional antennas in adhoc networks. In *Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, 1, 13-21.

Koutsakis, P., Psychis, S., & Paterakis, M. (2005). Integrated wireless access for videoconference from MPEG-4 and H.263 video coders with voice, E-mail, and web traffic. *IEEE Transactions on Vehicular Technology*, 54 (5), 1863-1874.

Liese, S., Wu, D., & Mohapatra, P. (2006). Experimental characterization of an 802.11b wireless mesh network. In *Proceedings of the 2006 ACM international conference on Wireless communications and mobile computing (IWCMC)*, 587-592.

Molina, M., Castelli, P., & Foddis, G. (2000). Web traffic modeling exploiting TCP connections' temporal clustering through HTML-REDUCE. *IEEE Network*, 14 (3), 46-55.

Munawar, M.A. & Ward, P.A.S. (2005). Are two interfaces better than one? In *Proceedings of the IEEE International Conference on Wireless And Mobile Computing, Networking And Communications (WiMob)*, 2, 119-125.

Nandiraju, N., Nandiraju, D., Santhanam, L., He, B., Wang, J., & Agrawal, D.P. (2007). Wireless Mesh Networks: Current Challenges and Future Directions of Web-In-The-Sky. *IEEE Wireless Communications*, 14 (4), 79-89.

Nascimento, J. C. , & Marques, J. S. (2006). Performance Evaluation of Object Detection Algorithms for Video Surveillance. *IEEE Transactions on Multimedia*, 8 (4), 761-774.

Norris, C., McCahill, M., & Wood, D. (2004). The Growth of CCTV: a global perspective on the international diffusion of video surveillance in publicly accessible space. Surveillance & Society. *CCTV Special*. 2(2/3), 376-395.

Park, K. & Lee, H. (2001). On the effectiveness of route-based packet filtering for distributed DoS attack prevention in power-law internets. *ACM SIGCOMM Computer Communication Review*, 31 (4), 15-26.

Ramanathan, R. & Redi, J. (2002). A brief overview of ad hoc networks: challenges and directions. *IEEE Communications Magazine*, 40 (5), 20-22.

Sedky, M.H., Moniri, M., & Chibelushi, C.C. (2005). Classification of smart video surveillance systems for commercial applications. In *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 638-643.

Smith, C.L. & Robinson, M. (1999). The understanding of security technology and its applications. In *Proceedings of the IEEE International Carnahan Conference on Security Technology*, 26-37.

Welsh, B., & Farrington, D. (2002). Crime prevention effects of closed circuit television: a systematic review. London: Home Office Research, Development and Statistics Directorate.

Wolf, W., Ozer, B., & Lv, T. (2002). Smart cameras as embedded systems. *IEEE Computer*, 35 (9), 48–53.

Ziliani, F. (2005). The importance of 'scalability' in video surveillance architectures. In *Proceedings of the IEE International Symposium on Imaging for Crime Detection and Prevention (ICDP)*, 29-32.

Terminology

Video Surveillance – It corresponds to the use of video cameras to transmit signal to a specific, limited set of monitors. It is often used for monitoring and crime prevention in sensitive areas such as banks, casinos, airports, seaports, military installations and convenience stores. Note that even though wireless links may be employed, they are not intended for a broadcast audience.

Video Streaming – This term refers to a continuous exchange of data, which can be monitored by the receiver while its transmission is ongoing, over a communication network. In particular, video surveillance pictures require an efficient streaming in order to actuate crime prevention and realize the basic functions of deterrence, detection and verification.

Wireless Mesh Network (WMN) – It is a communication network, where terminals are connected via radio to routers which are in turn interconnected via multi-hop wireless links. Its structure is entirely wireless, thus making WMNs especially applicable where cable deployment is difficult or too expensive, or the absence of cables is even recommended for security reasons.

Multi-hop wireless networks – Since the control room can be far from the area where surveillance is performed, remote control may be realized by employing multi-hop networks. This implies that the radio nodes belonging to the WMN need special procedures to work in harmony with each other and enable dedicated communications.

TCP/IP networks – This corresponds to the realization of the Internet structure over the network of interest. In particular, TCP/IP implies a layered structure for the network, which is hence able to provide an upper-layer service (in this case, video streaming) by means of lower-layer data exchange, in particular for what concerns network routing taking place on wireless multi-hop links.

Network engineering – This term implies the design of hardware and software solutions to implement a network structure, for what concerns both information exchange and physical creation of links. In particular, for video streaming this corresponds to enabling a multi-hop communication whose routes and content are predictable, yet there are several limiting factors (bandwidth, complexity) to take into account.

Design guidelines – In the chapter we identify several practical rules to use in the development of hardware and software solutions. In particular, we deal with network hierarchy, multi-hop routing and bandwidth dimensioning. Finally, we also envisioned some design choices such as multiple antennas as related to link diversity.