ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Computer Networks xxx (2006) xxx–xxx

Computer Networks

www.elsevier.com/locate/comnet

A survey on wireless multimedia sensor networks

Ian F. Akyildiz *, Tommaso Melodia, Kaushik R. Chowdhury

Broadband and Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology,
Atlanta, GA 30332, United States

Received 11 March 2006; received in revised form 6 August 2006; accepted 5 October 2006

Abstract

The availability of low-cost hardware such as CMOS cameras and microphones has fostered the development of Wireless Multimedia Sensor Networks (WMSNs), i.e., networks of wirelessly interconnected devices that are able to ubiquitously retrieve multimedia content such as video and audio streams, still images, and scalar sensor data from the environment. In this paper, the state of the art in algorithms, protocols, and hardware for wireless multimedia sensor networks is surveyed, and open research issues are discussed in detail. Architectures for WMSNs are explored, along with their advantages and drawbacks. Currently off-the-shelf hardware as well as available research prototypes for WMSNs are listed and classified. Existing solutions and open research issues at the application, transport, network, link, and physical layers of the communication protocol stack are investigated, along with possible cross-layer synergies and optimizations.

Keywords: Wireless sensor networks; Multimedia communications; Distributed smart cameras; Video sensor networks; Energy-aware protocol design; Cross-layer protocol design; Quality of service

1. Introduction

Wireless sensor networks (WSN) [22] have drawn the attention of the research community in the last few years, driven by a wealth of theoretical and practical challenges. This growing interest can be largely attributed to new applications enabled by large-scale networks of small devices capable of harvesting information from the physical environment, performing simple processing on the extracted data and transmitting it to remote locations. Significant results in this area over the last few years have ushered in a surge of civil and military applications. As of today, most deployed wireless sensor networks measure scalar physical phenomena like temperature, pressure, humidity, or location of objects. In general, most of the applications have low bandwidth demands, and are usually delay tolerant.

More recently, the availability of inexpensive hardware such as CMOS cameras and microphones that are able to ubiquitously capture multimedia content from the environment has fostered the

---

* Corresponding author. Tel.: +1 404 894 5141; fax: +1 404 894 7883.

E-mail addresses: ian@ece.gatech.edu (I.F. Akyildiz), tommaso@ece.gatech.edu (T. Melodia), kaushikc@ece.gatech.edu (K.R. Chowdhury).

1389-1286/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2006.10.002

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

development of Wireless Multimedia Sensor Networks (WMSNs) [54,90], i.e., networks of wirelessly interconnected devices that allow retrieving video and audio streams, still images, and scalar sensor data. With rapid improvements and miniaturization in hardware, a single sensor device can be equipped with audio and visual information collection modules. As an example, the Cyclops image capturing and inference module [103], is designed for extremely light-weight imaging and can be interfaced with a host mote such as Crossbow's MICA2 [4] or MICAz [5]. In addition to the ability to retrieve multimedia data, WMSNs will also be able to store, process in real-time, correlate and fuse multimedia data originated from heterogeneous sources.

Wireless multimedia sensor networks will not only enhance existing sensor network applications such as tracking, home automation, and environmental monitoring, but they will also enable several new applications such as:

- *Multimedia surveillance sensor networks*. Wireless video sensor networks will be composed of interconnected, battery-powered miniature video cameras, each packaged with a low-power wireless transceiver that is capable of processing, sending, and receiving data. Video and audio sensors will be used to enhance and complement existing surveillance systems against crime and terrorist attacks. Large-scale networks of video sensors can extend the ability of law enforcement agencies to monitor areas, public events, private properties and borders.
- *Storage of potentially relevant activities*. Multimedia sensors could infer and record potentially relevant activities (thefts, car accidents, traffic violations), and make video/audio streams or reports available for future query.
- *Traffic avoidance, enforcement and control systems*. It will be possible to monitor car traffic in big cities or highways and deploy services that offer traffic routing advice to avoid congestion. In addition, smart parking advice systems based on WMSNs [29] will allow monitoring available parking spaces and provide drivers with automated parking advice, thus improving mobility in urban areas. Moreover, multimedia sensors may monitor the flow of vehicular traffic on highways and retrieve aggregate information such as average speed and number of cars. Sensors could also detect violations and transmit video streams to law enforcement agencies to identify the violator, or buffer images and streams in case of accidents for subsequent accident scene analysis.
- *Advanced health care delivery*. Telemedicine sensor networks [59] can be integrated with 3G multimedia networks to provide ubiquitous health care services. Patients will carry medical sensors to monitor parameters such as body temperature, blood pressure, pulse oximetry, ECG, breathing activity. Furthermore, remote medical centers will perform advanced remote monitoring of their patients via video and audio sensors, location sensors, motion or activity sensors, which can also be embedded in wrist devices [59].
- *Automated assistance for the elderly and family monitors*. Multimedia sensor networks can be used to monitor and study the behavior of elderly people as a means to identify the causes of illnesses that affect them such as dementia [106]. Networks of wearable or video and audio sensors can infer emergency situations and immediately connect elderly patients with remote assistance services or with relatives.
- *Environmental monitoring*. Several projects on habitat monitoring that use acoustic and video feeds are being envisaged, in which information has to be conveyed in a time-critical fashion. For example, arrays of video sensors are already used by oceanographers to determine the evolution of sandbars via image processing techniques [58].
- *Person locator services*. Multimedia content such as video streams and still images, along with advanced signal processing techniques, can be used to locate missing persons, or identify criminals or terrorists.
- *Industrial process control*. Multimedia content such as imaging, temperature, or pressure amongst others, may be used for time-critical industrial process control. *Machine vision* is the application of computer vision techniques to industry and manufacturing, where information can be extracted and analyzed by WMSNs to support a manufacturing process such as those used in semiconductor chips, automobiles, food or pharmaceutical products. For example, in quality control of manufacturing processes, details or final products are automatically inspected to find defects. In addition, machine vision systems can detect the position and orientation of parts of the product to be picked up by a robotic arm. The integration of machine vision

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

systems with WMSNs can simplify and add flexibility to systems for visual inspections and automated actions that require high-speed, high-magnification, and continuous operation.

As observed in [37], WMSNs will stretch the horizon of traditional monitoring and surveillance systems by:

- *Enlarging the view*. The Field of View (FoV) of a single fixed camera, or the Field of Regard (FoR) of a single moving pan-tilt-zoom (PTZ) camera is limited. Instead, a distributed system of multiple cameras and sensors enables perception of the environment from multiple disparate viewpoints, and helps overcoming occlusion effects.
- *Enhancing the view*. The redundancy introduced by multiple, possibly heterogeneous, overlapped sensors can provide enhanced understanding and monitoring of the environment. Overlapped cameras can provide different views of the same area or target, while the joint operation of cameras and audio or infrared sensors can help disambiguate cluttered situations.
- *Enabling multi-resolution views*. Heterogeneous media streams with different granularity can be acquired from the same point of view to provide a multi-resolution description of the scene and multiple levels of abstraction. For example, static medium-resolution camera views can be enriched by views from a zoom camera that provides a high-resolution view of a region of interest. For example, such feature could be used to recognize people based on their facial characteristics.

Many of the above applications require the sensor network paradigm to be rethought in view of the need for mechanisms to deliver multimedia content with a certain level of quality of service (QoS). Since the need to minimize the energy consumption has driven most of the research in sensor networks so far, mechanisms to efficiently deliver application level QoS, and to map these requirements to network layer metrics such as latency and jitter, have not been primary concerns in mainstream research on classical sensor networks.

Conversely, algorithms, protocols and techniques to deliver multimedia content over large-scale networks have been the focus of intensive research in the last 20 years, especially in ATM wired and wireless networks. Later, many of the results derived for ATM networks have been readapted, and architectures such as Diffserv and Intserv for Internet QoS delivery have been developed. However, there are several main peculiarities that make QoS delivery of multimedia content in sensor networks an even more challenging, and largely unexplored, task:

- *Resource constraints*. Sensor devices are constrained in terms of battery, memory, processing capability, and achievable data rate [22]. Hence, efficient use of these scarce resources is mandatory.
- *Variable channel capacity*. While in wired networks the capacity of each link is assumed to be fixed and pre-determined, in multi-hop wireless networks, the attainable capacity of each wireless link depends on the interference level perceived at the receiver. This, in turn, depends on the interaction of several functionalities that are distributively handled by all network devices such as power control, routing, and rate policies. Hence, capacity and delay attainable at each link are location dependent, vary continuously, and may be bursty in nature, thus making QoS provisioning a challenging task.
- *Cross-layer coupling of functionalities*. In multi-hop wireless networks, there is a strict interdependence among functions handled at all layers of the communication stack. Functionalities handled at different layers are inherently and strictly coupled due to the shared nature of the wireless communication channel. Hence, the various functionalities aimed at QoS provisioning should not be treated separately when efficient solutions are sought.
- *Multimedia in-network processing*. Processing of multimedia content has mostly been approached as a problem isolated from the network-design problem, with a few exceptions such as joint source-channel coding [44] and channel-adaptive streaming [51]. Hence, research that addressed the content delivery aspects has typically not considered the characteristics of the source content and has primarily studied cross-layer interactions among lower layers of the protocol stack. However, the processing and delivery of multimedia content are not independent and their interaction has a major impact on the levels of QoS that can be delivered. WMSNs will allow performing multimedia in-network processing algorithms on the raw data. Hence, the QoS required at the application level will be delivered by means of a combination of both cross-layer optimization of the

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

communication process, and in-network processing of raw data streams that describe the phenomenon of interest from multiple views, with different media, and on multiple resolutions. Hence, it is necessary to develop application-independent and self-organizing architectures to flexibly perform in-network processing of multimedia contents.

Efforts from several research areas will need to converge to develop efficient and flexible WMSNs, and this in turn, will significantly enhance our ability to interact with the physical environment. These include advances in the understanding of energy-constrained wireless communications, and the integration of advanced multimedia processing techniques in the communication process. Another crucial issue is the development of flexible system architectures and software to allow querying the network to specify the required service (thus providing abstraction from implementation details). At the same time, it is necessary to provide the service in the most efficient way, which may be in contrast with the need for abstraction.

In this paper, we survey the state of the art in algorithms, protocols, and hardware for the development of wireless multimedia sensor networks, and discuss open research issues in detail. In particular, in Section 2 we point out the characteristics of wireless multimedia sensor networks, i.e., the major factors influencing their design. In Section 3, we suggest possible architectures for WMSNs and describe their characterizing features. In Section 4, we discuss and classify existing hardware and prototypal implementations for WMSNs, while in Section 5 we discuss possible advantages and challenges of multimedia in-network processing. In Sections 6–10 we discuss existing solutions and open research issues at the application, transport, network, link, and physical layers of the communication stack, respectively. In Section 11, we discuss cross-layer synergies and possible optimizations, while in Section 12 we discuss additional complementary research areas such as actuation, synchronization and security. Finally, in Section 13 we conclude the paper.

2. Factors influencing the design of multimedia sensor networks

Wireless Multimedia Sensor Networks (WMSNs) will be enabled by the convergence of communication and computation with signal processing and several branches of control theory and embedded computing. This cross-disciplinary research will enable distributed systems of heterogeneous embedded devices that sense, interact, and control the physical environment. There are several factors that mainly influence the design of a WMSN, which are outlined in this section.

- *Application-specific QoS requirements*. The wide variety of applications envisaged on WMSNs will have different requirements. In addition to data delivery modes typical of scalar sensor networks, multimedia data include *snapshot* and *streaming multimedia* content. Snapshot-type multimedia data contain event triggered observations obtained in a short time period. Streaming multimedia content is generated over longer time periods and requires sustained information delivery. Hence, a strong foundation is needed in terms of hardware and supporting high-level algorithms to deliver QoS and consider application-specific requirements. These requirements may pertain to multiple domains and can be expressed, amongst others, in terms of a combination of bounds on energy consumption, delay, reliability, distortion, or network lifetime.
- *High bandwidth demand*. Multimedia content, especially video streams, require transmission bandwidth that is orders of magnitude higher than that supported by currently available sensors. For example, the nominal transmission rate of state-of-the-art IEEE 802.15.4 compliant components such as Crossbow's [3] MICAz or TelosB [6] motes is 250 kbit/s. Data rates at least one order of magnitude higher may be required for high-end multimedia sensors, with comparable power consumption. Hence, high data rate and low-power consumption transmission techniques need to be leveraged. In this respect, the ultra wide band (UWB) transmission technique seems particularly promising for WMSNs, and its applicability is discussed in Section 10.
- *Multimedia source coding techniques*. Uncompressed raw video streams require excessive bandwidth for a multi-hop wireless environment. For example, a single monochrome frame in the NTSC-based *Quarter Common Intermediate Format* (QCIF, 176×120), requires around 21 Kbyte, and at 30 frames per second (fps), a video stream requires over 5 Mbit/s. Hence, it is apparent that efficient processing techniques for

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

lossy compression are necessary for multimedia sensor networks. Traditional video coding techniques used for wireline and wireless communications are based on the idea of reducing the bit rate generated by the source encoder by exploiting source statistics. To this aim, encoders rely on *intra-frame* compression techniques to reduce redundancy within one frame, while they leverage *inter-frame* compression (also known as *predictive encoding* or *motion estimation*) to exploit redundancy among subsequent frames to reduce the amount of data to be transmitted and stored, thus achieving good rate-distortion performance. Since predictive encoding requires complex encoders, powerful processing algorithms, and entails high energy consumption, it may not be suited for low-cost multimedia sensors. However, it has recently been shown [50] that the traditional balance of complex encoder and simple decoder can be reversed within the framework of the so-called *distributed source coding*, which exploits the source statistics at the decoder, and by shifting the complexity at this end, allows the use of simple encoders. Clearly, such algorithms are very promising for WMSNs and especially for networks of video sensors, where it may not be feasible to use existing video encoders at the source node due to processing and energy constraints.

- *Multimedia in-network processing*. WMSNs allow performing multimedia in-network processing algorithms on the raw data extracted from the environment. This requires new architectures for collaborative, distributed, and resource-constrained processing that allow for filtering and extraction of semantically relevant information at the edge of the sensor network. This may increase the system scalability by reducing the transmission of redundant information, merging data originated from multiple views, on different media, and with multiple resolutions. For example, in video security applications, information from uninteresting scenes can be compressed to a simple scalar value or not be transmitted altogether, while in environmental applications, distributed filtering techniques can create a time-elapsed image [120]. Hence, it is necessary to develop application-independent architectures to flexibly perform in-network processing of the multimedia content gathered from the environment. For example, IrisNet [93] uses application-specific filtering of sensor feeds at the source, i.e., each application processes its desired sensor feeds on the CPU of the sensor nodes where data are gathered. This dramatically reduces the bandwidth consumed, since instead of transferring raw data, IrisNet sends only a potentially small amount of processed data. However, the cost of multimedia processing algorithms may be prohibitive for low-end multimedia sensors. Hence, it is necessary to develop scalable and energy-efficient distributed filtering architectures to enable processing of redundant data as close as possible to the periphery of the network.
- *Power consumption*. Power consumption is a fundamental concern in WMSNs, even more than in traditional wireless sensor networks. In fact, sensors are battery-constrained devices, while multimedia applications produce high volumes of data, which require high transmission rates, and extensive processing. While the energy consumption of traditional sensor nodes is known to be dominated by the communication functionalities, this may not necessarily be true in WMSNs. Therefore, protocols, algorithms and architectures to maximize the network lifetime while providing the QoS required by the application are a critical issue.
- *Flexible architecture to support heterogeneous applications*. WMSN architectures will support several heterogeneous and independent applications with different requirements. It is necessary to develop flexible, hierarchical architectures that can accommodate the requirements of all these applications in the same infrastructure.
- *Multimedia coverage*. Some multimedia sensors, in particular video sensors, have larger sensing radii and are sensitive to direction of acquisition (directivity). Furthermore, video sensors can capture images only when there is unobstructed line of sight between the event and the sensor. Hence, coverage models developed for traditional wireless sensor networks are not sufficient for pre-deployment planning of a multimedia sensor network.
- *Integration with Internet (IP) architecture*. It is of fundamental importance for the commercial development of sensor networks to provide services that allow querying the network to retrieve useful information from anywhere and at any time. For this reason, future WMSNs will be remotely accessible from the Internet, and will therefore need to be integrated with the IP

architecture. The characteristics of WSNs rule out the possibility of all-IP sensor networks and recommend the use of application level gateways or overlay IP networks as the best approach for integration between WSNs and the Internet [138].

- *Integration with other wireless technologies*. Large-scale sensor networks may be created by interconnecting local "islands" of sensors through other wireless technologies. This needs to be achieved without sacrificing on the efficiency of the operation within each individual technology.

3. Network architecture

The problem of designing a *scalable network architecture* is of primary importance. Most proposals for wireless sensor networks are based on a flat, homogenous architecture in which every sensor has the same physical capabilities and can only interact with neighboring sensors. Traditionally, the research on algorithms and protocols for sensor networks has focused on *scalability*, i.e., how to design solutions whose applicability would not be limited by the growing size of the network. Flat topologies may not always be suited to handle the amount of traffic generated by multimedia applications including audio and video. Likewise, the processing power required for data processing and communications, and the power required to operate it, may not be available on each node.

3.1. Reference architecture

In Fig. 1, we introduce a reference architecture for WMSNs, where three sensor networks with different characteristics are shown, possibly deployed in different physical locations. The first cloud on the left shows a single-tier network of homogeneous video sensors. A subset of the deployed sensors have higher processing capabilities, and are thus referred to as *processing hubs*. The union of the processing hubs constitutes a distributed processing architecture. The multimedia content gathered is relayed to a *wireless gateway* through a multi-hop path. The gateway is interconnected to a *storage hub*, that is in charge of storing multimedia content locally for subsequent retrieval. Clearly, more complex architectures for distributed storage can be implemented when allowed by the environment and the

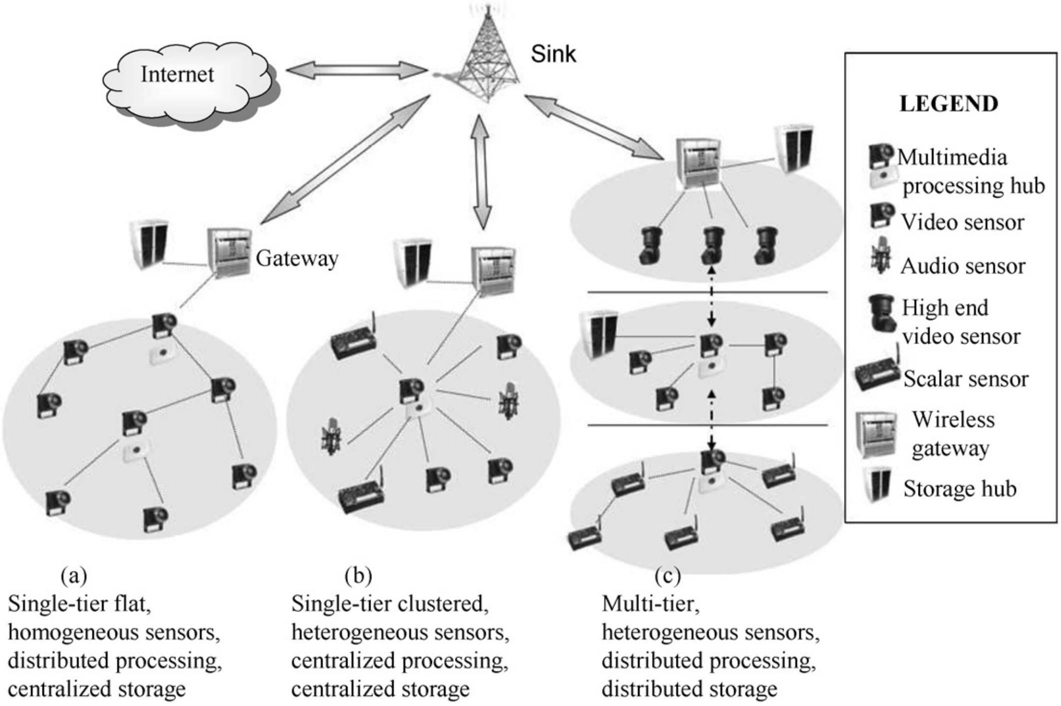

Fig. 1. Reference architecture of a wireless multimedia sensor network.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

application needs, which may result in energy savings since by storing it locally, the multimedia content does not need to be wirelessly relayed to remote locations. The wireless gateway is also connected to a central *sink*, which implements the software front-end for network querying and tasking. The second cloud represents a single-tiered clustered architecture of heterogeneous sensors (only one cluster is depicted). Video, audio, and scalar sensors relay data to a central clusterhead, which is also in charge of performing intensive multimedia processing on the data (processing hub). The clusterhead relays the gathered content to the wireless gateway and to the storage hub. The last cloud on the right represents a multi-tiered network, with heterogeneous sensors. Each tier is in charge of a subset of the functionalities. Resource-constrained, low-power scalar sensors are in charge of performing simpler tasks, such as detecting scalar physical measurements, while resource-rich, high-power devices are responsible for more complex tasks. Data processing and storage can be performed in a distributed fashion at each different tier.

3.2. Single-tier vs. multi-tier sensor deployment

One possible approach for designing a multimedia sensor application is to deploy homogeneous sensors and program each sensor to perform all possible application tasks. Such an approach yields a flat, single-tier network of homogeneous sensor nodes. An alternative, multi-tier approach is to use heterogeneous elements [69]. In this approach, resource-constrained, low-power elements are in charge of performing simpler tasks, such as detecting scalar physical measurements, while resource-rich, high-power devices take on more complex tasks. For instance, a surveillance application can rely on low-fidelity cameras or scalar acoustic sensors to perform motion or intrusion detection, while high-fidelity cameras can be woken up on-demand for object recognition and tracking. In [68], a multi-tier architecture is advocated for video sensor networks for surveillance applications. The architecture is based on multiple tiers of cameras with different functionalities, with the lower tier constituted of low-resolution imaging sensors, and the higher tier composed of high-end pan-tilt-zoom cameras. It is argued, and shown by means of experiments, that such an architecture offers considerable advantages with respect to a single-tier architecture in terms of scalability, lower cost, better coverage, higher functionality, and better reliability.

3.3. Coverage

In traditional WSNs, sensor nodes collect information from the environment within a pre-defined *sensing range*, i.e., a roughly circular area defined by the type of sensor being used.

Multimedia sensors generally have larger sensing radii and are also sensitive to the direction of data acquisition. In particular, cameras can capture images of objects or parts of regions that are not necessarily close to the camera itself. However, the image can obviously be captured only when there is an unobstructed line-of-sight between the event and the sensor. Furthermore, each multimedia sensor/camera perceives the environment or the observed object from a different and unique viewpoint, given the different orientations and positions of the cameras relative to the observed event or region. In [118], a preliminary investigation of the coverage problem for video sensor networks is conducted. The concept of sensing range is replaced with the camera's *field of view*, i.e., the maximum volume visible from the camera. It is also shown how an algorithm designed for traditional sensor networks does not perform well with video sensors in terms of coverage preservation of the monitored area.

4. Multimedia sensor hardware

In this section, we review and classify existing imaging, multimedia, and processing wireless devices that will find application in next generation wireless multimedia sensor networks. In particular, we discuss existing hardware, with a particular emphasis on video capturing devices, review existing implementations of multimedia sensor networks, and discuss current possibilities for *energy harvesting* for multimedia sensor devices.

4.1. Enabling hardware platforms

High-end pan-tilt-zoom cameras and high resolution digital cameras are widely available on the market. However, while such sophisticated devices can find application as high-quality tiers of multimedia sensor networks, we concentrate on low-cost, low-energy consumption imaging and processing devices that will be densely deployed and provide detailed

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

visual information from multiple disparate viewpoints, help overcoming occlusion effects, and thus enable enhanced interaction with the environment.

4.1.1. Low-resolution imaging motes

The recent availability of CMOS imaging sensors [61] that capture and process an optical image within a single integrated chip, thus eliminating the need for many separate chips required by the traditional charged-coupled device (CCD) technology, has enabled the massive deployment of low-cost visual sensors. CMOS image sensors are already in many industrial and consumer sectors, such as cell phones, personal digital assistants (PDAs), consumer and industrial digital cameras. CMOS image quality is now matching CCD quality in the low- and mid-range, while CCD is still the technology of choice for high-end image sensors. The CMOS technology allows integrating a lens, an image sensor and image processing algorithms, including image stabilization and image compression, on the same chip. With respect to CCD, cameras are smaller, lighter, and consume less power. Hence, they constitute a suitable technology to realize imaging sensors to be interfaced with wireless motes.

However, existing CMOS imagers are still designed to be interfaced with computationally rich host devices, such as cell phones or PDAs. For this reason, the objective of the Cyclops module [103] is to fill the gap between CMOS cameras and computationally constrained devices. Cyclops is an electronic interface between a CMOS camera module and a wireless mote such as MICA2 or MICAz, and contains programmable logic and memory for high-speed data communication. Cyclops consists of an imager (CMOS Agilent ADCM-1700 CIF camera), an 8-bit ATMEL ATmega128L microcontroller (MCU), a complex programmable logic device (CPLD), an external SRAM and an external Flash. The MCU controls the imager, configures its parameters, and performs local processing on the image to produce an inference. Since image capture requires faster data transfer and address generation than the 4 MHz MCU used, a CPLD is used to provide access to the high-speed clock. Cyclops firmware is written in the nesC language [48], based on the TinyOS libraries. The module is connected to a host mote to which it provides a high level interface that hides the complexity of the imaging device to the host mote. Moreover, it can perform simple inference on the image data and present it to the host.

Researchers at Carnegie Mellon University are developing the CMUcam 3, which is an embedded camera endowed with a CIF Resolution (352 × 288) RGB color sensor that can load images into memory at 26 frames per second. CMUcam 3 has software JPEG compression and has a basic image manipulation library, and can be interface with an 802.15.4 compliant TelosB mote [6].

In [41], the design of an integrated mote for wireless image sensor networks is described. The design is driven by the need to endow motes with adequate processing power and memory size for image sensing applications. It is argued that 32-bit processors are better suited for image processing than their 8-bit counterpart, which is used in most existing motes. It is shown that the time needed to perform operations such as 2-D convolution on an 8-bit processor such as the ATMEL ATmega128 clocked at 4 MHz is 16 times higher than with a 32-bit ARM7 device clocked at 48 MHz, while the power consumption of the 32-bit processor is only six times higher. Hence, an 8-bit processor turns out to be slower and more energy-consuming. Based on these premises, a new image mote is developed based on an ARM7 32-bit CPU clocked at 48 MHz, with external FRAM or Flash memory, 802.15.4 compliant Chipcon CC2420 radio, that is interfaced with mid-resolution ADCM-1670 CIF CMOS sensors and low-resolution 30 × 30 pixel optical sensors.

The same conclusion is drawn in [81], where the energy consumption of the 8-bit Atmel AVR processor clocked at 8 MHz is compared to that of the PXA255 32-bit Intel processor, embedded on a Stargate platform [10] and clocked at 400 MHz. Three representative algorithms are selected as benchmarks, i.e., the cyclic redundancy check, a finite impulse response filter, and a fast Fourier transform. Surprisingly, it is shown that even for such relatively simple algorithms the energy consumption of an 8-bit processor is between one and two orders of magnitude higher.

4.1.2. Medium-resolution imaging motes based on the Stargate platform

Intel has developed several prototypes that constitute important building platform for WMSN applications. The Stargate board [10] is a high-performance processing platform designed for sensor, signal processing, control, robotics, and sensor network applications. It is designed by Intel and produced by Crossbow. Stargate is based on Intel's PXA-255 XScale 400 MHz RISC processor, which

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

is the same processor found in many handheld computers including the Compaq IPAQ and the Dell Axim. Stargate has 32 Mbyte of Flash memory, 64 Mbyte of SDRAM, and an on-board connector for Crossbow's MICA2 or MICAz motes as well as PCMCIA Bluetooth or IEEE 802.11 cards. Hence, it can work as a wireless gateway and as a computational hub for in-network processing algorithms. When connected with a webcam or other capturing device, it can function as a medium-resolution multimedia sensor, although its energy consumption is still high, as documented in [80]. Moreover, although efficient software implementations exist, XScale processors do not have hardware support for floating point operations, which may be needed to efficiently perform multimedia processing algorithms.

Intel has also developed two prototypal generations of wireless sensors, known as Imote and Imote2. Imote is built around an integrated wireless microcontroller consisting of an 8-bit 12 MHz ARM7 processor, a Bluetooth radio, 64 Kbyte RAM and 32 Kbyte FLASH memory, as well as several I/O options. The software architecture is based on an ARM port of TinyOS. The second generation of Intel motes has a common core to the next generation Stargate 2 platform, and is built around a new low-power 32-bit PXA271 XScale processor at 320/416/520 MHz, which enables performing DSP operations for storage or compression, and an IEEE 802.15.4 ChipCon CC2420 radio. It has large on-board RAM and Flash memories (32 Mbyte), additional support for alternate radios, and a variety of high-speed I/O to connect digital sensors or cameras. Its size is also very limited, 48 × 33 mm, and it can run the Linux operating system and Java applications.

4.2. Energy harvesting

As mentioned before, techniques for prolonging the lifetime of battery-powered sensors have been the focus of a vast amount of literature in sensor networks. These techniques include hardware optimizations such as dynamic optimization of voltage and clock rate, wake-up procedures to keep electronics inactive most of the time, and energy-aware protocol development for sensor communications. In addition, *energy-harvesting* techniques, which extract energy from the environment where the sensor itself lies, offer another important mean to prolong the lifetime of sensor devices.

Systems able to perpetually power sensors based on simple COTS photovoltaic cells coupled with supercapacitors and rechargeable batteries have been already demonstrated [64]. In [96], the state of the art in more unconventional techniques for energy harvesting (also referred to as *energy scavenging*) is surveyed. Technologies to generate energy from background radio signals, thermoelectric conversion, vibrational excitation, and the human body, are overviewed.

As far as collecting energy from background radio signals is concerned, unfortunately, an electric field of 1 V/m yields only $0.26\ \mu W/cm^2$, as opposed to $100\ \mu W/cm^2$ produced by a crystalline silicon solar cell exposed to bright sunlight. Electric fields of intensity of a few volts per meter are only encountered close to strong transmitters. Another practice, which consists in broadcasting RF energy deliberately to power electronic devices, is severely limited by legal limits set by health and safety concerns.

While thermoelectric conversion may not be suitable for wireless devices, harvesting energy from vibrations in the surrounding environment may provide another useful source of energy. Vibrational magnetic power generators based on moving magnets or coils may yield powers that range from tens of microwatts when based on microelectromechanical system (MEMS) technologies to over a milliwatt for larger devices. Other vibrational microgenerators are based on charged capacitors with moving plates, and depending on their excitation and power conditioning, yield power on the order of 10 μW. In [96], it is also reported that recent analysis [91] suggested that $1\ cm^3$ vibrational microgenerators can be expected to yield up to $800\ \mu W/cm^3$ from machine-induced stimuli, which is orders of magnitude higher than what provided by currently available microgenerators. Hence, this is a promising area of research for small battery-powered devices.

While these techniques may provide an additional source of energy and help prolong the lifetime of sensor devices, they yield power that is several orders of magnitude lower as compared to the power consumption of state-of-the-art multimedia devices. Hence, they may currently be suitable only for very-low duty cycle devices.

4.3. Examples of deployed multimedia sensor networks

There have been several recent experimental studies, mostly limited to video sensor networks.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

Panoptes [46] is a system developed for environmental observation and surveillance applications, based on Intel StrongARM PDA platforms with a Logitech webcam as a video capture device. Here, video sensors are high-end devices with Linux operating system, 64 Mbyte of memory, and are networked through 802.11 networking cards. The system includes spatial compression (but not temporal), distributed filtering, buffering, and adaptive priorities for the video stream.

In [35], a system whose objective is to limit the computation, bandwidth, and human attention burdens imposed by large-scale video surveillance systems is described. In-network processing is used on each camera to filter out uninteresting events locally, avoiding disambiguation and tracking of irrelevant environmental distractors. A resource allocation algorithm is also proposed to steer pan-tilt cameras to follow interesting targets while maintaining awareness of possibly emerging new targets.

In [69], the design and implementation of SensEye, a multi-tier network of heterogeneous wireless nodes and cameras, is described. The surveillance application consists of three tasks: object detection, recognition and tracking. The objective of the design is to demonstrate that a camera sensor network containing heterogeneous elements provides numerous benefits over traditional homogeneous sensor networks. For this reason, SensEye follows a three-tier architecture, as shown in Fig. 2. The lowest tier consists of low-end devices, i.e., MICA2 Motes equipped with 900 MHz radios interfaced with scalar sensors, e.g., vibration sensors. The second tier is made up of motes equipped with low-fidelity Cyclops [103] or CMUcam [107] camera sensors. The third tier consists of Stargate [10] nodes equipped with webcams. Each Stargate is equipped with an embedded 400 MHz XScale processor that runs Linux and a webcam that can capture higher fidelity images than tier 2 cameras. Tier 3 nodes also perform gateway functions, as they are endowed with a low data rate radio to communicate with motes in tiers 1–2 at 900 MHz, and an 802.11 radio to communicate with tier 3 Stargate nodes. An additional fourth tier may consist of a sparse deployment of high-resolution, high-end pan-tilt-zoom cameras connected to embedded PCs. The camera sensors at this tier can be used to track moving objects, and can be utilized to fill coverage gaps and provide additional redundancy. The underlying design principle is to map each task requested by the application to the lowest tier with sufficient resources to perform the task. Devices from higher tiers are woken up on-demand only when necessary. For example, a high-resolution camera can be woken up to retrieve high resolution images of an object that has been previously detected by a lower tier. It is shown that the system can achieve an order of magnitude reduction in energy consumption while providing comparable surveillance accuracy with respect to single-tier surveillance systems.

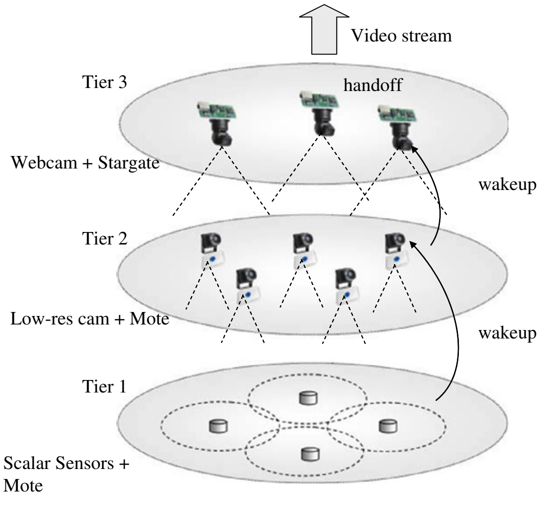

Fig. 2. The multi-tier architecture of SensEye [69].

In [80], experimental results on the energy consumption of a video sensor network testbed are presented. Each sensing node in the testbed consists of a Stargate board equipped with an 802.11 wireless network card and a Logitech QuickCam Pro 4000 webcam. The energy consumption is assessed using a benchmark that runs basic tasks such as processing, flash memory access, image acquisition, and communication over the network. Both steady state and transient energy consumption behavior obtained by direct measurements of current with a digital multimeter are reported. In the steady state, it is shown that communication-related tasks are less energy-consuming than intensive processing and flash access when the radio modules are loaded. Interestingly, and unlike in traditional wireless sensor networks [99], the processing-intensive benchmark results in the highest current requirement, and transmission is shown to be only about 5% more energy-consuming than reception. Experimental results also show that delay and additional amount of energy consumed due to transitions (e.g., to go to sleep mode) are not negligible and must be accounted for in network and protocol design.

IrisNet (Internet-scale Resource-Intensive Sensor Network Services) [93] is an example software platform to deploy heterogeneous services on WMSNs. IrisNet allows harnessing a global, wide-area sensor network by performing Internet-like queries on this infrastructure. Video sensors and scalar sensors are spread throughout the environment, and collect potentially useful data. IrisNet allows users to perform Internet-like queries to video sensors and other data. The user views the sensor network as a single unit that can be queried through a high-level language. Each query operates over data collected from the global sensor network, and allows simple Google-like queries as well as more complex queries involving arithmetic and database operators.

The architecture of IrisNet is two-tiered: heterogeneous sensors implement a common shared interface and are called *sensing agents* (SA), while the data produced by sensors is stored in a distributed database that is implemented on *organizing agents* (OA). Different *sensing services* are run simultaneously on the architecture. Hence, the same hardware infrastructure can provide different sensing services. For example, a set of video sensors can provide a parking space finder service, as well as a surveillance service. Sensor data is represented in the Extensible Markup Language (XML), which allows easy organization of hierarchical data. A group of OAs is responsible for a sensing service, collects data produced by that service, and organizes the information in a distributed database to answer the class of relevant queries. IrisNet also allows programming sensors with filtering code that processes sensor readings in a service-specific way. A single SA can execute several such software filters (called senselets) that process the raw sensor data based on the requirements of the service that needs to access the data. After senselet processing, the distilled information is sent to a nearby OA.

We have recently built an experimental testbed at the Broadband and Wireless Networking (BWN) Laboratory at Georgia Tech based on currently off-the-shelf advanced devices to demonstrate the efficiency of algorithms and protocols for multimedia communications through wireless sensor networks.

The testbed is integrated with our scalar sensor network testbed, which is composed of a heterogeneous collection of imotes from Intel and MICAz motes from Crossbow. Although our testbed already includes 60 scalar sensors, we plan to increase its size to deploy a higher scale testbed that allows testing more complex algorithms and assess the scalability of the communication protocols under examination.

The WMSN-testbed includes three different types of multimedia sensors: low-end imaging sensors, medium-quality webcam-based multimedia sensors, and pan-tilt cameras mounted on mobile robots.

Low-end imaging sensors such as CMOS cameras can be interfaced with Crossbow MICAz motes. Medium-end video sensors are based on Logitech webcams interfaced with Stargate platforms (see Fig. 3).

The high-end video sensors consist of pan-tilt cameras installed on an Acroname GARCIA

Fig. 3. Stargate board interfaced with a medium resolution camera. Stargate hosts an 802.11 card and a MICAz mote that functions as a gateway to the sensor network.

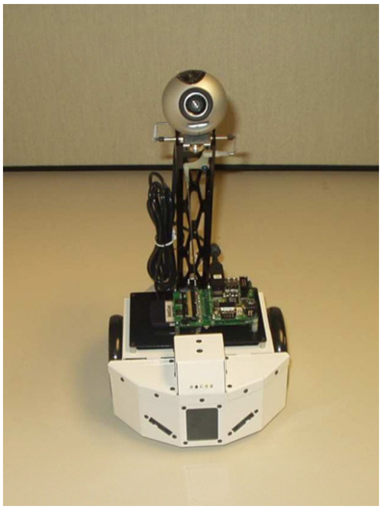

Fig. 4. Acroname GARCIA, a mobile robot with a mounted pan-tilt camera and endowed with 802.11 as well as Zigbee interfaces.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

Fig. 5. GARCIA deployed on the sensor testbed. It acts as a mobile sink, and can move to the area of interest for closer visual inspection. It can also coordinate with other actors and has built-in collision avoidance capability.

robotic platform [1], which we refer to as actor, and shown in Fig. 4. Actors constitute a mobile platform that can perform adaptive sampling based on event features detected by low-end motes. The mobile actor can redirect high-resolution cameras to a region of interest when events are detected by lower-tier, low-resolution video sensors that are densely deployed, as seen in Fig. 5.

The testbed also includes storage and computational hubs, which are needed to store large multimedia content and perform computationally intensive multimedia processing algorithms.

5. Collaborative in-network processing

As discussed previously, collaborative in-network multimedia processing techniques are of great interest in the context of a WMSN. It is necessary to develop architectures and algorithms to flexibly perform these functionalities *in-network* with minimum energy consumption and limited execution time. The objective is usually to avoid transmitting large amounts of raw streams to the sink by processing the data in the network to reduce the communication volume.

Given a source of data (e.g., a video stream), different applications may require diverse information (e.g., raw video stream vs. simple scalar or binary information inferred by processing the video stream). This is referred to as *application-specific querying and processing*. Hence, it is necessary to develop expressive and efficient querying languages, and to develop distributed filtering and in-network processing architectures, to allow real-time retrieval of useful information.

Similarly, it is necessary to develop architectures that efficiently allow performing data fusion or other complex processing operations *in-network*. Algorithms for both inter-media and intra-media data aggregation and fusion need to be developed, as simple distributed processing schemes developed for existing scalar sensors are not suitable for computation-intensive processing required by multimedia contents. Multimedia sensor networks may require computation-intensive processing algorithms (e.g., to detect the presence of suspicious activity from a video stream). This may require considerable processing to extract meaningful information and/or to perform compression. A fundamental question to be answered is whether this processing can be done on sensor nodes (i.e., a flat architecture of multi-functional sensors that can perform any task), or if the need for specialized devices, e.g., *computation hubs*, arises.

In what follows, we discuss a non-exhaustive set of significative examples of processing techniques that would be applicable distributively in a WMSN, and that will likely drive research on architectures and algorithms for distributed processing of raw sensor data.

5.1. Data alignment and image registration

Data alignment consists of merging information from multiple sources. One of the most widespread data alignment concepts, image registration [137], is a family of techniques, widely used in areas such as remote sensing, medical imaging, and computer vision, to geometrically align different images (reference and sensed images) of the same scene taken at different times, from different viewpoints, and/or by different sensors:

- *Different Viewpoints (Multi-view Analysis)*. Images of the same scene are acquired from different viewpoints, to gain a larger 2D view or a 3D representation of the scene of interest. Main applications are in remote sensing, computer vision and 3D shape recovery.
- *Different times (multi-temporal analysis)*. Images of the same scene are acquired at different times. The aim is to find and evaluate changes in time in the scene of interest. The main applications are in computer vision, security monitoring, and motion tracking.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

- *Different sensors (multi-modal analysis)*. Images of the same scene are acquired by different sensors. The objective is to integrate the information obtained from different source streams to gain more complex and detailed scene representation.

Registration methods usually consist of four steps, i.e., *feature detection*, *feature matching*, *transform model estimation*, and *image resampling and transformation*. In feature detection, distinctive objects such as closed-boundary regions, edges, contours, line intersections, corners, etc. are detected. In feature matching, the correspondence between the features detected in the sensed image and those detected in the reference image is established. In transform model estimation, the type and parameters of the so-called mapping functions, which align the sensed image with the reference image, are estimated. The parameters of the mapping functions are computed by means of the established feature correspondence. In the last step, image resampling and transformation, the sensed image is transformed by means of the mapping functions.

These functionalities can clearly be prohibitive for a single sensor. Hence, research is needed on how to perform these functionalities on parallel architectures of sensors to produce single data sets.

5.2. WMSNs as distributed computer vision systems

Computer vision is a subfield of artificial intelligence, whose purpose is to allow a computer to extract features from a scene, an image or multidimensional data in general. The objective is to present this information to a human operator or to control some process (e.g., a mobile robot or an autonomous vehicle). The image data that is fed into a computer vision system is often a digital image, a video sequence, a 3D volume from a tomography device or other multimedia content. Traditional computer vision algorithms require extensive computation, which in turn entails high power consumption.

WMSNs enable a new approach to computer vision, where visual observations across the network can be performed by means of distributed computations on multiple, possibly low-end, vision nodes. This requires tools to interface with the user such as new querying languages and abstractions to express complex tasks that are then distributively accomplished through low-level operations on multiple vision nodes. To this aim, it is necessary to coordinate computations across the vision nodes and return the integrated results, which will consist of metadata information, to the final user.

In [102], the proposed Deep Vision network performs operations including object detection or classification, image segmentation, and motion analysis through a network of low-end MICA motes equipped with Cyclops cameras [103]. Information such as the presence of an intruder, the number of visitors in a scene or the probability of presence of a human in the monitored area is obtained by collecting the results of these operations. Deep Vision provides a querying interface to the user in the form of declarative queries. Each operation is represented as an attribute that can be executed through an appropriate query. In this way, low-level operations and processing are encapsulated in a high-level querying interface that enables simple interaction with the video network. As an example, the vision network can be deployed in areas with public and restricted access spaces. The task of detecting objects in the restricted-access area can be expressed as a query that requests the result of object detection computations such as

SELECT *Object*,*Location*
REPORT = 30
FROM *Network*
WHERE *Access* = *Restricted*
PERIOD = 30.

The above query triggers the execution of the object detection process on the vision nodes that are located in the restricted-access areas in 30 s intervals.

6. Application layer

The functionalities handled at the application layer of a WMSN are characterized by high heterogeneity, and encompass traditional communication problems as well as more general system challenges. The services offered by the application layer include: (i) providing *traffic management and admission control functionalities*, i.e., prevent applications from establishing data flows when the network resources needed are not available; (ii) performing *source coding* according to application requirements and hardware constraints, by leveraging advanced multimedia encoding techniques; (iii) providing *flexible and efficient system software*, i.e., operating systems and middleware, to export services for higher-layer

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

applications to build upon; (iv) providing *primitives for applications to leverage collaborative, advanced in-network multimedia processing techniques*. In this section, we provide an overview of these challenges.

6.1. Traffic classes

Admission control has to be based on QoS requirements of the overlying application. We envision that WMSNs will need to provide support and differentiated service for several different classes of applications. In particular, they will need to provide differentiated service between real-time and delay-tolerant applications, and loss-tolerant and loss-intolerant applications. Moreover, some applications may require a continuous stream of multimedia data for a prolonged period of time (*multimedia streaming*), while some other applications may require event triggered observations obtained in a short time period (*snapshot multimedia content*). The main traffic classes that need to be supported are:

- *Real-time, Loss-tolerant, Multimedia Streams*. This class includes video and audio streams, or multi-level streams composed of video/audio and other scalar data (e.g., temperature readings), as well as metadata associated with the stream, that need to reach a human or automated operator in real-time, i.e., within strict delay bounds, and that are however relatively loss tolerant (e.g., video streams can be within a certain level of distortion). Traffic in this class usually has high bandwidth demand.
- *Delay-tolerant, Loss-tolerant, Multimedia Streams*. This class includes multimedia streams that, being intended for storage or subsequent offline processing, do not need to be delivered within strict delay bounds. However, due to the typically high bandwidth demand of multimedia streams and to limited buffers of multimedia sensors, data in this traffic class needs to be transmitted almost in real-time to avoid excessive losses.
- *Real-time, Loss-tolerant, Data*. This class may include monitoring data from densely deployed scalar sensors such as light sensors whose monitored phenomenon is characterized by spatial correlation, or loss-tolerant snapshot multimedia data (e.g., images of a phenomenon taken from several multiple viewpoints at the same time). Hence, sensor data has to be received timely but the application is moderately loss-tolerant. The bandwidth demand is usually between low and moderate.
- *Real-time, Loss-intolerant, Data*. This may include data from time-critical monitoring processes such as distributed control applications. The bandwidth demand varies between low and moderate.
- *Delay-tolerant, Loss-intolerant, Data*. This may include data from critical monitoring processes, with low or moderate bandwidth demand that require some form of offline post processing.
- *Delay-tolerant, Loss-tolerant, Data*. This may include environmental data from scalar sensor networks, or non-time-critical snapshot multimedia content, with low or moderate bandwidth demand.

QoS requirements have recently been considered as application admission criteria for sensor networks. In [97], an application admission control algorithm is proposed whose objective is to maximize the network lifetime subject to bandwidth and reliability constraints of the application. An application admission control method is proposed in [28], which determines admissions based on the added energy load and application rewards. While these approaches address application level QoS considerations, they fail to consider multiple QoS requirements (e.g., delay, reliability, and energy consumption) simultaneously, as required in WMSNs. Furthermore, these solutions do not consider the peculiarities of WMSNs, i.e., they do not try to base admission control on a tight balancing between communication optimizations and in-network computation. There is a clear need for new criteria and mechanisms to manage the admission of multimedia flows according to the desired application-layer QoS.

6.2. Multimedia encoding techniques

There exists a vast literature on multimedia encoding techniques. The captured multimedia content should ideally be represented in such a way as to allow reliable transmission over lossy channels (*error-resilient coding*), using algorithms that minimize processing power and the amount of information to be transmitted. The main design objectives of a coder for multimedia sensor networks are thus:

- *High compression efficiency*. Uncompressed raw video streams require high data rates and thus consume excessive bandwidth and energy. It is necessary to achieve a high ratio of compression to effectively limit bandwidth and energy consumption.

- *Low complexity*. Multimedia encoders are embedded in sensor devices. Hence, they need to be *low complexity* to reduce cost and form factors, and *low-power* to prolong the lifetime of sensor nodes.
- *Error resiliency*. The source coder should provide robust and error-resilient coding of source data.

To achieve a high compression efficiency, the traditional broadcasting paradigm for wireline and wireless communications, where video is compressed once at the encoder and decoded several times, has been dominated by predictive encoding techniques. These, used in the widely spread ISO MPEG schemes, or the ITU-T recommendations H.263 [11] and H.264 [2] (also known as AVC or MPEG-4 part 10), are based on the idea of reducing the bit rate generated by the source encoder by exploiting source statistics. Hence, *intra-frame* compression techniques are used to reduce redundancy within one frame, while *inter-frame* compression (also known as *predictive encoding* or *motion estimation*) exploits correlation among subsequent frames to reduce the amount of data to be transmitted and stored, thus achieving good rate-distortion performance. Since the computational complexity is dominated by the motion estimation functionality, these techniques require complex encoders, powerful processing algorithms, and entail high energy consumption, while decoders are simpler and loaded with lower processing burden. For typical implementations of state-of-the-art video compression standards, such as MPEG or H.263 and H.264, the encoder is 5–10 times more complex than the decoder [50]. It is easy to see that to realize low-cost, low-energy-consumption multimedia sensors it is necessary to develop simpler encoders, and still retain the advantages of high compression efficiency.

However, it is known from information-theoretic bounds established by Slepian and Wolf for lossless coding [117] and by Wyner and Ziv [130] for lossy coding with decoder side information, that efficient compression can be achieved by leveraging knowledge of the source statistics at the decoder only. This way, the traditional balance of complex encoder and simple decoder can be reversed [50]. Techniques that build upon these results are usually referred to as *distributed source coding*. Distributed source coding refers to the compression of multiple correlated sensor outputs that do not communicate with each other [131]. Joint decoding is performed by a central entity that receives data independently compressed by different sensors. However, practical solutions have not been developed until recently. Clearly, such techniques are very promising for WMSNs and especially for networks of video sensors. The encoder can be simple and low-power, while the decoder at the sink will be complex and loaded with most of the processing and energy burden. The reader is referred to [131,50] for excellent surveys on the state of the art of distributed source coding in sensor networks and in distributed video coding, respectively. Other encoding and compression schemes that may be considered for source coding of multimedia streams, including JPEG with differential encoding, distributed coding of images taken by cameras having overlapping fields of view, or multi-layer coding with wavelet compression, are discussed in [90]. Here, we focus on recent advances on low complexity encoders based on Wyner–Ziv coding [130], which are promising solutions for distributed networks of video sensors that are likely to have a major impact in future design of protocols for WMSNs.

The objective of a Wyner–Ziv video coder is to achieve lossy compression of video streams and achieve performance comparable to that of inter-frame encoding (e.g., MPEG), with complexity at the encoder comparable to that of intra-frame coders (e.g., Motion-JPEG).

6.2.1. Pixel-domain Wyner–Ziv encoder

In [14,15], a practical Wyner–Ziv encoder is proposed as a combination of a pixel-domain intra-frame encoder and inter-frame decoder system for video compression. A block diagram of the system is reported in Fig. 6. A regularly spaced subset of frames is coded using a conventional intra-frame coding technique, such as JPEG, as shown at the bottom of the figure. These are referred to as *key frames*. All frames between the key frames are referred to as *Wyner–Ziv frames* and are intra-frame encoded but inter-frame decoded. The intra-frame encoder for Wyner–Ziv frames (shown on top) is composed of a quantizer followed by a Slepian–Wolf coder. Each Wyner–Ziv frame is quantized and blocks of symbols are sent to the Slepian–Wolf coder, which is implemented through rate-compatible punctured turbo codes (RCPT). The parity bits generated by the RCPT coder are stored in a buffer.

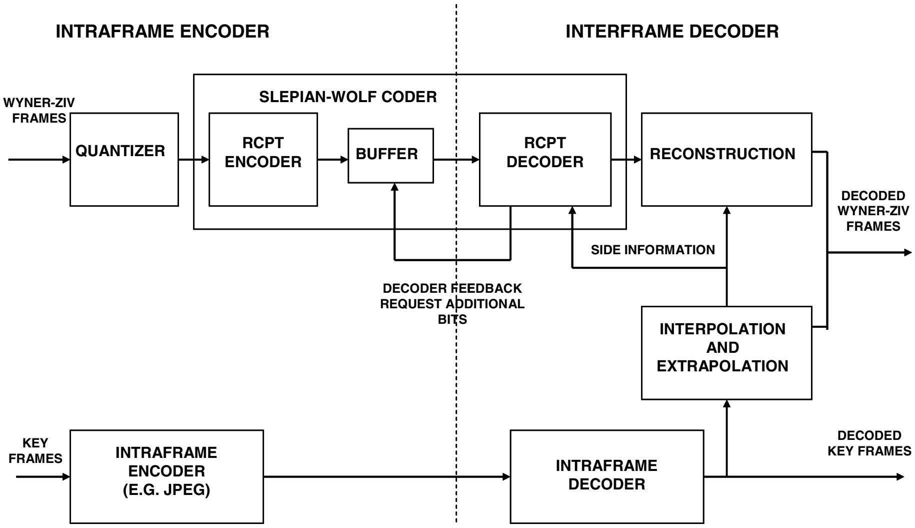

Fig. 6. Block diagram of a pixel-domain Wyner–Ziv encoder [14].

A subset of these bits is then transmitted upon request from the decoder. This allows adapting the rate based on the temporally varying statistics between the Wyner–Ziv frame and the side information. The parity bits generated by the RCPT coder are in fact used to "correct" the frame interpolated at the decoder. For each Wyner–Ziv frame, the decoder generates the side information frame by interpolation or extrapolation of previously decoded key frames and Wyner–Ziv frames. The side information is leveraged by assuming a Laplacian distribution of the difference between the individual pixels of the original frame and the side information. The parameter defining the Laplacian distribution is estimated online. The turbo decoder combines the side information and the parity bits to reconstruct the original sequence of symbols. If reliable decoding of the original symbols is impossible, the turbo decoder requests additional parity bits from the encoder buffer.

Compared to predictive coding such as MPEG or H.26X, pixel-domain Wyner–Ziv encoding is much simpler. The Slepian–Wolf encoder only requires two feedback shift registers and an interleaver. Its performance, in terms of peak signal-to-noise ratio (PSNR), is 2–5 dB better than conventional motion-JPEG intra-frame coding. The main drawback of this scheme is that it relies on online feedback from the receiver. Hence it may not be suitable for applications where video is encoded and stored for subsequent use. Moreover, the feedback may introduce excessive latency for video decoding in a multi-hop network.

6.2.2. Transform-domain Wyner–Ziv encoder

In conventional source coding, a source vector is typically decomposed into spectral coefficients by using orthonormal transforms such as the Discrete Cosine Transform (DCT). These coefficients are then individually coded with scalar quantizers and entropy coders. In [13], a transform-domain Wyner–Ziv encoder is proposed. A block-wise DCT of each Wyner–Ziv frame is performed. The transform coefficients are independently quantized, grouped into coefficient bands, and then compressed by a Slepian–Wolf turbo coder. As in the pixel-domain encoder described in the previous section, the decoder generates a side information frame based on previously reconstructed frames. Based on the side information, a bank of turbo decoders reconstructs the quantized coefficient bands independently. The rate-distortion performance is between conventional intra-frame transform coding and conventional motion-compensated transform coding.

A different approach consists of allowing some simple temporal dependence estimation at the encoder to perform rate control without the need for feedback from the receiver. In the PRISM scheme [100], the encoder selects the coding mode based on the frame difference energy between the current frame and a previous frame. If the energy of the difference is very small, the block is not encoded. If the block difference is large, the block is intra-coded. Between these two situations, one of different encoding modes with different rates is selected. The rate estimation does not involve motion com-

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

pensation and hence is necessarily inaccurate, if motion compensation is used at the decoder. Further, the flexibility of the decoder is restricted.

6.3. System software and middleware

The development of efficient and flexible system software to make functional abstractions and information gathered by scalar and multimedia sensors available to higher layer applications is one of the most important challenges faced by researchers to manage complexity and heterogeneity of sensor systems. As in [66], the term *system software* is used here to refer to operating systems, virtual machines, and middleware, which export services to higher-layer applications. Different multimedia sensor network applications are extremely diverse in their requirements and in the way they interact with the components of a sensor system. Hence, the main desired characteristics of a system software for WMSNs can be identified as follows:

- Provides a high-level interface to specify the behavior of the sensor system. This includes semantically rich querying languages that allow specifying what kind of data is requested from the sensor network, the quality of the required data, and how it should be presented to the user;
- Allows the user to specify application-specific algorithms to perform in-network processing on the multimedia content [47]. For example, the user should be able to specify particular image processing algorithms or multimedia coding format;
- Long-lived, i.e., needs to smoothly support evolutions of the underlying hardware and software;
- Shared among multiple heterogeneous applications;
- Shared among heterogeneous sensors and platforms. Scalar and multimedia sensor networks should coexist in the same architecture, without compromising on performance;
- Scalable.

There is an inherent trade-off between degrees of flexibility and network performance. Platform-independence is usually achieved through layers of abstraction, which usually introduce redundancy and prevent the developer from accessing low-level details and functionalities. However, WMSNs are characterized by the contrasting objectives of optimizing the use of the scarce network resources and not compromising on performance. The principal design objective of existing operating systems for sensor networks such as TinyOS is high performance. However, their flexibility, inter-operability and reprogrammability are very limited. There is a need for research on systems that allow for this integration.

We believe that it is of paramount importance to develop efficient, high level abstractions that will enable easy and fast development of sensor network applications. An abstraction similar to the famous Berkeley TCP sockets, that fostered the development of Internet applications, is needed for sensor systems. However, differently from the Berkeley sockets, it is necessary to retain control on the efficiency of the low-level operations performed on battery-limited and resource-constrained sensor nodes.

As a first step towards this direction, Chu et al. [34] recently proposed Sdlib, a sensor network data and communications library built upon the nesc language [48] for applications that require best-effort collection of large-size data such as video monitoring applications. The objective of the effort is to identify common functionalities shared by several sensor network applications and to develop a library of thoroughly-tested, reusable and efficient nesC components that abstract high-level operations common to most applications, while leaving differences among them to adjustable parameters. The library is called Sdlib, Sensor Data Library, as an analogy to the traditional C++ Standard Template Library. Sdlib provides an abstraction for common operations in sensor networks while the developer is still able to access low-level operations, which are implemented as a collection of nesC components, when desired. Moreover, to retain efficiency of operations that are so critical for sensor networks battery lifetime and resource constraints, Sdlib exposes policy decisions such as resource allocation and rate of operation to the developer, while hiding the mechanisms of policy enforcement.

6.4. Open research issues

- While theoretical results on Slepian–Wolf and Wyner–Ziv coding exist since 30 years, there is still a lack of practical solutions. The net benefits and the practicality of these techniques still need to be demonstrated.
- It is necessary to fully explore the trade-offs between the achieved fidelity in the description of the phenomenon observed, and the resulting energy consumption. As an example, the video

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

distortion perceived by the final user depends on source coding (frame rate, quantization), and on channel coding strength. For example, in a surveillance application, the objective of maximizing the event detection probability is in contrast with the objective of minimizing the power consumption.

- As discussed above, there is a need for high-layer abstractions that will allow fast development of sensor applications. However, due to the resource-constrained nature of sensor systems, it is necessary to control the efficiency of the low-level operations performed on battery-limited and resource-constrained sensor nodes.
- There is a need for simple yet expressive high-level primitives for applications to leverage collaborative, advanced in-network multimedia processing techniques.

7. Transport layer

In applications involving high-rate data, the transport layer assumes special importance by providing end-to-end reliability and congestion control mechanisms. Particularly, in WMSNs, the following additional considerations are in order to accommodate both the unique characteristics of the WSN paradigm and multimedia transport requirements.

- *Effects of congestion*. In WMSNs, the effect of congestion may be even more pronounced as compared to traditional networks. When a bottleneck sensor is swamped with packets coming from several high-rate multimedia streams, apart from temporary disruption of the application, it may cause rapid depletion of the node's energy. While applications running on traditional wireless networks may only experience performance degradation, the energy loss (due to collisions and retransmissions) can result in network partition. Thus, congestion control algorithms may need to be tuned for immediate response and yet avoid oscillations of data rate along the affected path.
- *Packet re-ordering due to multi-path*. Multiple paths may exist between a given source-sink pair, and the order of packet delivery is strongly influenced by the characteristics of the route chosen. As an additional challenge, in real-time video/audio feeds or streaming media, information that cannot be used in the proper sequence becomes redundant, thus stressing on the need for transport layer packet reordering.

We next explore the functionalities and support provided by the transport layer to address these and other challenges of WMSNs. The following discussion is classified into (1) TCP/UDP and TCP friendly schemes for WMSNs, and (2) application-specific and non-standardized protocols. Fig. 7 summarizes the discussion in this section.

7.1. TCP/UDP and TCP friendly schemes for WMSNs

For real-time applications like streaming media, the User Datagram Protocol (UDP) is preferred over TCP as *timeliness* is of greater concern than

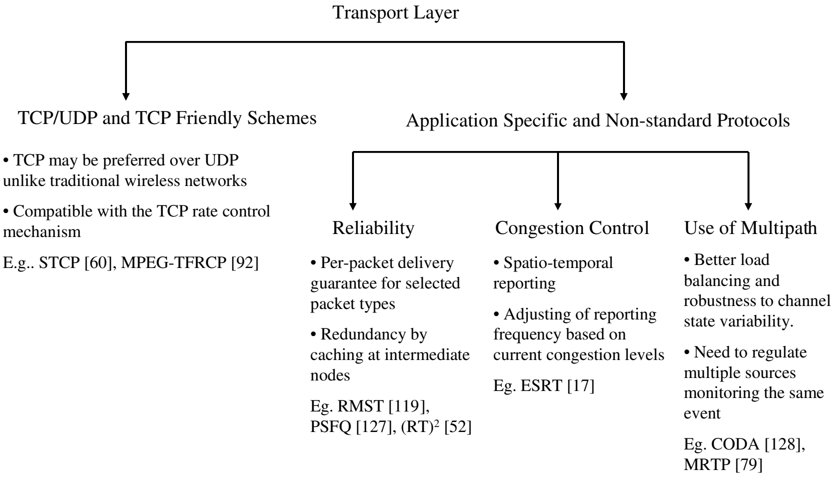

Fig. 7. Classification of existing transport layer protocols.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

reliability. However, in WMSNs, it is expected that packets are significantly compressed at the source and redundancy is reduced as far as possible owing to the high transmission overhead in the energy-constrained nodes. Under these conditions, we note the following important characteristics that may necessitate an approach very different from classical wireless networks.

- *Effect of dropping packets in UDP*. Simply dropping packets during congestion conditions, as undertaken in UDP, may introduce discernable disruptions in the order of a fraction of a second. This effect is even more pronounced if the packet dropped contains important original content not captured by inter-frame interpolation, like the Region of Interest (ROI) feature used in JPEG2000 [12] or the I-frame used in the MPEG family.
- *Support for traffic heterogeneity*. Multimedia traffic comprising of video, audio, and still images exhibits a high level of heterogeneity and may be further classified into periodic or event driven. The UDP header has no provision to allow any description of these traffic classes that may influence congestion control policies. As a contrast to this, the *options* field in the TCP header can be modified to carry data specific information. As an example, the Sensor Transmission Control Protocol (STCP) [60] accommodates a differentiated approach by including relevant fields in the TCP header. Several other major changes to the traditional TCP model are proposed in the round trip time estimation, congestion notification, the packet drop policy and by introducing a reliability driven intimation of lost packets.

We thus believe that TCP with appropriate modifications is preferable over UDP for WMSNs, if standardized protocols are to be used. With respect to sensor networks, several problems and their likely solutions like large TCP header size, data vs. address centric routing, energy efficiency, amongst others, are identified and solutions are proposed in [42]. We next indicate the recent work in this direction that evaluates the case for using TCP in WMSNs.

- *Effect of jitter induced by TCP*. A key factor that limits multimedia transport based on TCP, and TCP-like rate control schemes, is the jitter introduced by the congestion control mechanism. This can be, however, mitigated to a large extent by playout buffers at the sink, which is typically assumed to be rich in resources. As an example, the MPEG-TFRCP (TCP Friendly Rate Control Protocol for MPEG-2 Video Transfer) [92] is an equation-based rate control scheme designed for transporting MPEG video in a TCP-friendly manner.
- *Overhead of the reliability mechanism in TCP*. As discussed earlier, blind dropping of packets in UDP containing highly compressed video/audio data may adversely affect the quality of transmission. Yet, at the same time, the reliability mechanism provided by TCP introduces an end-to-end message passing overhead and energy efficiency must also be considered. Distributed TCP Caching (DTC) [43] overcomes these problems by caching TCP segments inside the sensor network and by local retransmission of TCP segments. The nodes closest to the sink are the last-hop forwarders on most of the high-rate data paths and thus run out of energy first. DTC shifts the burden of the energy consumption from nodes close to the sink into the network, apart from reducing network wide retransmissions.
- *Regulating streaming through multiple TCP connections*. The availability of multiple paths between source and sink can be exploited by opening multiple TCP connections for multimedia traffic [94]. Here, the desired streaming rate and the allowed throughput reduction in presence of bursty traffic, like sending of video data, is communicated to the receiver by the sender. This information is used by the receiver which then measures the actual throughput and controls the rate within the allowed bounds by using multiple TCP connections and dynamically changing its TCP window size for each connection.

TCP protocols tailor-made for wireless sensor networks is an active research area with recent implementations of the light-weight Sensor Internet Protocol (SIP) [78] and the open source uIP [42] that has a code size of few Kbyte. However, a major problem with the TCP-based approach in wireless networks is its inability to distinguish between bad channel conditions and network congestion. This has motivated a new family of specialized transport layer where the design practices followed are entirely opposite to that of TCP [122], or stress on a particular functionality of the transport layer, like reliability or congestion control.

7.2. Application specific and non-standard protocols

Depending on the application, both reliability and congestion control may be equally important functionalities or one may be preferred over the other. As an example, in the CYCLOPS image capturing and inference module [103] designed for extremely light-weight imaging, congestion control would be the primary functionality with multiple sensor flows arriving at the sink, each being moderately loss-tolerant. We next list the important characteristics of such TCP incompatible protocols in context of WMSNs.

7.2.1. Reliability

Multimedia streams may consist of images, video and audio data, each of which merits a different metric for reliability. As discussed in Section 7.1, when an image or video is sent with differentially coded packets, the arrival of the packets with the ROI field or the I-frame respectively should be guaranteed. The application can, however, withstand moderate loss for the other packets containing differential information. Thus, we believe that reliability needs to be enforced on a per-packet basis to best utilize the existing networking resources. If a prior recorded video is being sent to the sink, all the I-frames could be separated and the transport protocol should ensure that each of these reach the sink. Reliable Multi-Segment Transport (RMST) [119] or the Pump Slowly Fetch Quickly (PSFQ) protocol [127] can be used for this purpose as they buffer packets at intermediate nodes, allowing for faster retransmission in case of packet loss. However, there is an overhead of using the limited buffer space at a given sensor node for caching packets destined for other nodes, as well as performing timely storage and flushing operations on the buffer. In a heterogeneous network, where real-time data is used by actors as discussed in Section 4.3, the Real-time and Reliable Transport $(RT)^2$ protocol [52] can be used that defines different reliability constraints for sensor–actor and actor–actor communication.

7.2.2. Congestion control

The high rate of injection of multimedia packets into the network causes resources to be used up quickly. While typical transmission rates for sensor nodes may be about 40 kbit/s, indicative data rates of a constant bit rate voice traffic may be 64 kbit/s. Video traffic, on the other hand, may be bursty and in the order of 500 kbit/s [136], thus making it clear that congestion must be addressed in WMSNs. While these data generation rates are high for a single node, multiple sensors in overlapped regions may inject similar traffic on sensing the same phenomenon. The Event-to-Sink Reliable Transport (ESRT) protocol [17] leverages the fact that spatial and temporal correlation exists among the individual sensor readings [125]. The ESRT protocol regulates the frequency of event reporting in a remote neighborhood to avoid congestions in the network. However, this approach may not be viable for all sensor applications as nodes transmit data only when they detect an event, which may be a short duration burst as in the case of a video monitoring application. The feedback from the base-station may hence not reach in time to prevent a sudden congestion due to this burst.

7.2.3. Use of multi-path

We advocate the use of multiple paths for data transfer in WMSNs owing to the following two reasons:

- A large burst of data (say, resulting from an I-frame) can be split into several smaller bursts, thus not overwhelming the limited buffers at the intermediate sensor nodes.
- The channel conditions may not permit high data rate for the entire duration of the event being monitored. By allowing multiple flows, the effective data rate at each path gets reduced and the application can be supported.

The design of a multiple source-sink transport protocol is challenging, and is addressed by the COngestion Detection and Avoidance (CODA) protocol [128]. It allows a sink to regulate multiple sources associated with a single event in case of persistent network congestion. However, as the congestion inference in CODA is based on queue length at intermediate nodes, any action taken by the source occurs only after a considerable time delay. Other solutions include the Multi-flow Real-time Transport Protocol (MRTP) [79] that does not specifically address energy efficiency considerations in WMSNs, but is suited for real-time streaming of multimedia content by splitting packets over different flows. MRTP does not have any mechanism for packet retransmission and is mainly used for real-time data transmission and hence, reliability can be an issue for scalar data traffic.

7.3. Open research issues

In summary, the transport layer mechanisms that can simultaneously address the unique challenges posed by the WMSN paradigm and multimedia communication requirements must be incorporated. While several approaches were discussed, some open issues remain and are outlined below:

- *Trade-off between reliability and congestion control*. In WMSN applications, the data gathered from the field may contain multimedia information such as target images, acoustic signal, and even video captures of a moving target, all of which enjoy a permissible level of loss tolerance. Presence or absence of an intruder, however, may require a single data field but needs to be communicated without any loss of fidelity. Thus, when a single network contains multimedia as well as scalar data, the transport protocol must decide whether to focus on one or more functionalities so that the application needs are met without an unwarranted energy expense. The design of such a layer may as well be modular, with the functional blocks of reliability and/or congestion control being invoked as per network demands.
- *Real-time communication support*. Despite the existence of reliable transport solutions for WSN as discussed above, none of these protocols provide real-time communication support for the applications with strict delay bounds. Therefore, new transport solutions which can also meet certain application deadlines must be researched.
- *Relation between multimedia coding rate and reliability*. The success in energy-efficient and reliable delivery of multimedia information extracted from the phenomenon directly depends on selecting appropriate coding rate, number of sensor nodes, and data rate for a given event [125]. However, to this end, the event reliability should be accurately measured in order to efficiently adapt the multimedia coding and transmission rates. For this purpose, new reliability metrics coupled with the application layer coding techniques should be investigated.

8. Network layer

The network layer addresses the challenging task of providing variable QoS guarantees depending on whether the stream carries time-independent data like configuration or initialization parameters, time-critical low rate data like presence or absence of the sensed phenomenon, high bandwidth video/audio data, etc. Each of the traffic classes described in Section 6.1 has its own QoS requirement which must be accommodated in the network layer.

Research on the network layer becomes important from the standpoint of supporting multimedia applications constrained by lack of global knowledge, reduced energy, and computational ability of the individual nodes.

We next discuss the existing research directions for the network layer functionalities of addressing and routing, while stressing on their applicability to delay-sensitive and high bandwidth needs.

8.1. Addressing and localization

In the case of large WMSNs like IrisNet [29], it is required that the individual nodes be monitored via the Internet. Such an integration between a randomly deployed sensor network and the established wired network becomes a difficult research challenge. The key problem of global addressing could be solved by the use of IPv6 in which the sensor can concatenate its cluster ID with its own MAC address to create the full IPv6 address. However, the 16-byte address field of IPv6 introduces excessive overhead in each sensor data packet. There are several other schemes that assign unique network-wide IDs (see [95] and references therein) or leverage location information to create an address-free environment but they, however, run the risk of incompatibility with the established standards of the Internet. Location information is a key characteristic of any sensor network system. The ability to associate localization information to the raw data sampled from the environment increases the capability of the system and the meaningfulness of the information extracted. Localization techniques for WMSNs are unlikely to differ substantially from those developed for traditional sensor networks, which are reviewed in [111]. Moreover, WMSNs will most likely leverage the accurate ranging capabilities that come with high bandwidth transmissions (such as UWB techniques, as described in Section 10).

8.2. Routing

Data collected by the sensor nodes needs to be sent to the sink, where useful information can be

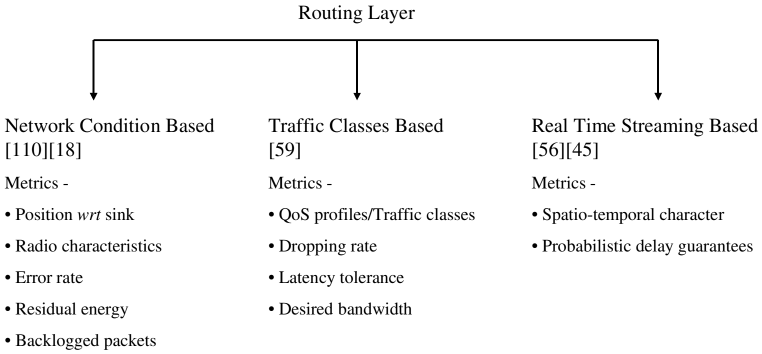

Fig. 8. Classification of existing routing protocols.

extracted from it. Comprehensive surveys of the major routing schemes existing in the literature are presented in [19,23]. The concerns of routing in general differ significantly from the specialized service requirements of multimedia streaming applications. As an example, multiple routes may be necessary to satisfy the desired data rate at the destination node. Also, different paths exhibiting varying channel conditions may be preferred depending on the type of traffic and its resilience to packet loss. We next discuss the various approaches to routing in WMSNs while maintaining that protocols may incorporate features from more than one of the following classes. Majorly, they can be classified into routing based on (i) network conditions that leverage channel and link statistics, (ii) traffic classes that decide paths based on packet priorities, and (iii) specialized protocols for real-time streaming that use spatio-temporal forwarding. Fig. 8 provides a classification of existing routing protocols and summarizes the discussion in this section.

8.2.1. QoS routing based on network conditions

Network conditions include interference seen at intermediate hops, the number of backlogged flows along a path, residual energy of the nodes, amongst others. A routing decision based on these metrics can avoid paths that may not support high bandwidth applications or introduce retransmission owing to bad channel conditions.

The use of image sensors is explored in [110], in which visual information is used to gather topology information that is then leveraged to develop efficient geographic routing schemes. A weighted cost function is constructed that takes into account position with respect to the base station, backlogged packets in the queue, and remaining energy of the nodes to decide the next hop along a route. This approach involves an overhead in which nodes must apprise their neighbors of any changes in the cost function parameters. This work also deals with relative priority levels for event based (high bandwidth) and periodic (low bandwidth) data.

A similar scenario is considered in [18] where imaging data for sensor networks results in QoS considerations for routing, apart from the traditional goal of energy conservation. Here, the cost function evaluates the residual energy of a node, transmission energy, error rate and other communication parameters. The protocol finds a least-cost, energy efficient path while considering maximum allowed delays.

8.2.2. QoS routing based on traffic classes

Sensor data may originate from various types of events that have different levels of importance, as described in Section 6.1. Consequently, the content and nature of the sensed data also varies. As an example that highlights the need for network level QoS, consider the task of bandwidth assignment for multimedia mobile medical calls, which include patients' sensing data, voice, pictures and video data [59]. Unlike the typical source-to-sink multi-hop communication used by classical sensor networks, the proposed architecture uses a 3G cellular system in which individual nodes forward the sensed data to a cellular phone or a specialized information collecting entity. Different priorities are assigned to video data originating from sensors on ambulances, audio traffic from elderly people, and images returned by sensors placed on the body. In order to achieve this, parameters like hand-off dropping rate (HDR), latency tolerance and desired amount of wireless effective bandwidth are taken into consideration.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

8.2.3. Routing protocols with support for streaming

The SPEED protocol [56] provides three types of real-time communication services, namely, real-time unicast, real-time area-multicast and real-time area-anycast. It uses geographical location for routing and a key difference with other schemes of this genre is its spatio-temporal character, i.e., it takes into account timely delivery of the packets. It is specifically tailored to be a stateless, localized algorithm with minimal control overhead. End-to-end soft real-time communication is achieved by maintaining a desired delivery speed across the sensor network through a combination of feedback control and non-deterministic geographic forwarding. As it works satisfactorily under scarce resource conditions and can provide service differentiation, SPEED takes the first step in addressing the concerns of real-time routing in WMSNs.

A significant extension over SPEED, the MMSPEED protocol [45] can efficiently differentiate between flows with different delay and reliability requirements. MMSPEED is based on a cross-layer approach between the network and the MAC layers in which a judicious choice is made over reliability and timeliness of packet arrival. It is argued that the differentiation in reliability is an effective way of channeling resources from flows with relaxed requirements to flows with tighter requirements. Importantly, a new metric called *On-Time Reachability* is introduced which is a measure of the probability that a packet reaches its destination within required delay bounds. While current research directions make an effort to provide real-time streaming, they are still best effort services. Giving firm delay guarantees in a dynamically changing network is a difficult problem and yet is important for seamless viewing of the multimedia frames. MMSPEED takes the step towards this end by adopting a probabilistic approach but clearly, further work is needed in this area.

8.3. Open research issues

- The identification of the optimal routing metrics is a continual area of research. Most routing protocols that consider more than one metric, like energy, delay etc., form a cost function that is then minimized. The choice of the weights for these metrics need to be judiciously undertaken, and is often subject to dynamic network conditions. Thus, further work is needed to shift this decision making process and network tuning from the user end into the network.
- As the connectivity between different domains improves, end-to-end QOS guarantees are complicated by the inherent differences in the nature of the wired and wireless media. When sensed data from the field is sent via the Internet, a single routing metric is unsuitable for the entire path between source and end user. Decoupling of reliability and routing parameters at such network boundaries and a seamless integration of schemes better suited to wired or wireless domains, respectively, will need to be explored.

9. MAC layer

Owing to the energy constraints of the small, battery-powered sensor nodes, it is desirable that the medium access control (MAC) protocol enable reliable, error-free data transfer with minimum retransmissions while supporting application-specific QoS requirements. Multimedia traffic, namely audio, video, and still images can be classified as separate service classes and subjected to different policies of buffering, scheduling and transmission. The need for packet-specific differentiation is justified in the context of the following example. The new standard for the compression of still images, JPEG2000 [12], incorporates a feature called region of interest (ROI) that may be applicable to visual data sensing. It allows the allocation of greater importance to certain parts of the image which can then be coded and transmitted over a better quality link or on a priority basis. Especially relevant to systems for military surveillance or fault monitoring, such application layer features could be leveraged by the MAC by differentially treating the ROI packets.

Research efforts to provide MAC layer QoS can be classified mainly into (i) channel access policies, (ii) scheduling and buffer management, and (iii) error control. We next provide a brief description of each and highlight their support to multimedia traffic. The scope of this paper is limited to the challenges posed by multimedia traffic in sensor networks and the efforts at the MAC layer to address them. A detailed survey of MAC protocols for classical sensor networks using scalar data can be found in [67]. Fig. 9 provides a classification of relevant MAC layer functionalities and summarizes the discussion in this section.

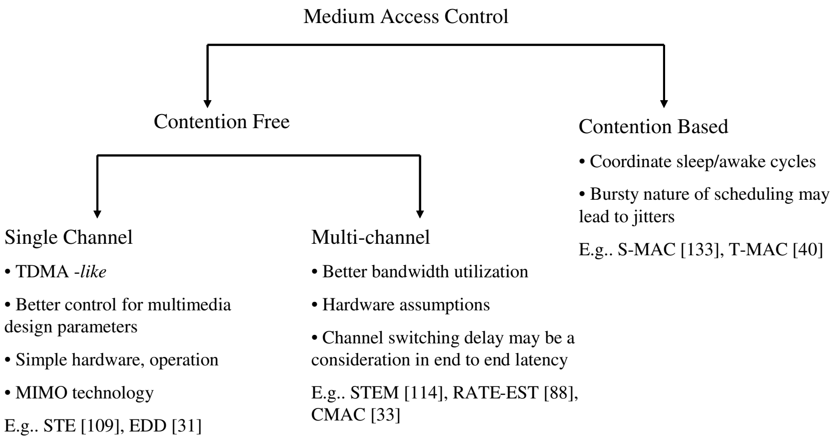

Fig. 9. Classification of protocols at the data link layer.

9.1. Channel access policies

The main causes of energy loss in sensor networks are attributed to *packet collisions* and subsequent retransmissions, *overhearing packets* destined for other nodes, and *idle listening*, a state in which the transceiver circuits remain active even in the absence of data transfer. Thus, regulating access to the channel assumes primary importance and several solutions have been proposed in the literature.

9.1.1. Contention-based protocols

Most contention-based protocols like S-MAC [133], and protocols inspired by it [40], have a single-radio architecture. They alternate between sleep cycles (low-power modes with transceiver switched off) and listen cycles (for channel contention and data transmission). However, we believe that their applicability to multimedia transmission is limited owing to the following reasons:

- The primary concern in the protocols of this class is saving energy, and this is accomplished at the cost of latency and by allowing throughput degradation. A sophisticated duty cycle calculation based on permissible end-to-end delay needs to be implemented and coordinating overlapping *listen* period with neighbors based on this calculation is a difficult research challenge.
- Coordinating the sleep–awake cycles between neighbors is generally accomplished though schedule exchanges. In case of dynamic duty cycles based on perceived values of instantaneous or time averaged end-to-end latency, the overhead of passing frequent schedules also needs investigation in light of the ongoing high data rate video/audio messaging.
- Video traffic exhibits an inherent bursty nature and can lead to sudden buffer overflow at the receiver. This problem is further aggravated by the transmission policy adopted in T-MAC [40]. By choosing to send a burst of data during the listen cycle, T-MAC shows performance improvement over S-MAC, but at the cost of monopolizing a bottleneck node. Such an operation could well lead to strong jitters and result in discontinuous real-time playback.

9.1.2. Contention-free single channel protocols

Time Division Multiple Access (TDMA) is a representative protocol of this class in which the clusterhead (CH) or sink helps in slot assignment, querying particular sensors and maintaining time schedules. We believe that such protocols can be easily adapted for multimedia transmission and highlight the likely design considerations.

- TDMA schemes designed exclusively for sensor networks [70] (and references therein) have a small reservation period (RP) that is generally contention based, followed by a contention-free period that spans the rest of the frame. This RP could occur in each frame or at pre-decided intervals in order to assign slots to active nodes taking into consideration the QoS requirement of their data streams. The length of the TDMA frames and the frequency of the RP interval are some of the design parameters that can be exploited while designing a multimedia system.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

- For real-time streaming video, packets are time constrained and scheduling policies like *Shortest Time to Extinction* (STE) [109] or *Earliest Due Date* (EDD) [31] can be adopted. Both of these are similar in principle as packets are sent in the increasing order of their respective delay tolerance but differ in respect that EDD may still forward a packet that has crossed its allowed delay bound. Based on the allowed packet loss of the multimedia stream, the dependencies between packet dropping rate, arrival rate, and delay tolerance [109] can be used to decide the TDMA frame structure and thus ensure smooth replay of data. This allows greater design choices as against [31], where the frame lengths and slot duration are considered constant.
- As sensor nodes are often limited by their maximum data transmission rate, depending upon their multimedia traffic class, the duration of transmission could be made variable. Thus variable TDMA (V-TDMA) schemes should be preferred when heterogeneous traffic is present in the network. Tools for calculating the minimum worst-case delay in such schemes and algorithms for link scheduling are provided in [38]. As real-time streaming media is delay bounded, the link-layer latency introduced in a given flow due in order to satisfy data rate requirements of another flow needs to be analyzed well when V-TDMA schemes are used.

9.1.2.1. MIMO technology. The high data rate required by multimedia applications can be addressed by spatial multiplexing in MIMO systems, that use a single channel but employ interference cancellation techniques. Recently, *virtual* MIMO schemes have been proposed for sensor networks [62], where nodes in close proximity form a cluster. Each sensor functions as a single antenna element, sharing information and thus simulating the operation of a multiple antenna array. A distributed compression scheme for correlated sensor data, that specially addresses multimedia requirements, is integrated with the MIMO framework in [63]. However, a key consideration in MIMO based systems is the number of sensor transmissions and the required signal energy per transmission. As the complexity is shifted from hardware to sensor coordination, further research is needed at the MAC layer to ensure that the required MIMO parameters like channel state, desired diversity/processing gain are known to both the sender and receiver at an acceptable energy cost.

9.1.2.2. Open research issues.

- While TDMA schedules within a cluster can be easily devised, the problem is more involved when individual CHs are not in direct range of the sink, thus necessitating inter-cluster multi-hop communication. An acceptable, non-overlapping slot assignment for all neighboring clusters needs to be derived in a distributed manner requiring coordination between them at the set-up phase. This problem has been shown to be NP-complete [57] by reduction to an instance of graph coloring and the development of efficient heuristics is an open issue.
- The effect of clock drift is pronounced if the slot duration is small and rigid time synchronization is required for best performance [73] (and references therein). Network scalability is another important area of research and the TDMA schedules must be able to accommodate high node densities that are characteristic of sensor networks. As channel capacity in TDMA is fixed, only slot durations or number of slots in a frame may be changed keeping in mind the number of users and their respective traffic types.
- Bounds on unequal slot/frame lengths for differentiated services should be decided by the allowed per-hop delay (and consequently end-to-end delay). Schedules, once created, should also be able to account for a dynamically changing topology due to nodes dying off or new ones being added.

9.1.3. Contention-free multi-channel protocols

Clearly, existing data rates of about 40 kbit/s and 250 kbit/s supported by the MICA2 and MICAz motes are not geared to support multimedia traffic. Along with improving hardware and thus increasing cost, an alternate approach is to efficiently utilize the available bandwidth. By using multiple channels in a spatially overlapped manner, existing bandwidth can be efficiently utilized for supporting multimedia applications. We observe that Berkeley's third generation MICA2 Mote has an 868/916 GHz multi-channel transceiver [4]. In Rockwell's WINS nodes, the radio operates on one of 40 channels in the ISM frequency band, selectable by the controller [16]. We next outline the design parameters that could influence MAC design in multi-channel WMSNs.

- Recent research has focused on a two-transceiver paradigm in which the main radio (MR) is supplemented by the presence of a wake-up radio (LR) having similar characteristics [88,114,115] or a simple low-energy design [53] that emits a series of short pulses or a busy tone. The LR is used for exchanging control messages and is assigned a dedicated channel. In high bandwidth applications, like streaming video, the use of a separate channel for channel arbitration alone does not allow best utilization of the network resources.
- We propose that WMSNs use in-band signalling, where the same channel is used for both data and channel arbitration [33]. While such protocols undoubtedly improve bandwidth efficiency, they introduce the problem of distinct channel assignment and need to account for the delay to switch to a different channel [72], as its cumulative nature at each hop affects real-time media.
- Existing compression techniques like JPEG and MPEG cause size variations of the captured video frame depending upon the movement and compression rate within a given frame, and the subsequent following frames. Thus, the arrival packet rate (assuming packet size constant) may change over time and, we believe, enforcing wake-ups based on the packet backlog is a better design approach as compared to those with static timer based schemes. A dual approach is followed in [88], but this, however, raises questions of receiver side synchronization as knowledge of the sender's buffer state and its consequent scheduling instant is unknown at the receiver end.

9.1.3.1. Open research issues.

- Multi-channel protocols utilize bandwidth better, and thus may perform favorably in cases of applications demanding high data rate. The results obtained in [72], leave an open question on whether switching delay can be successfully hidden with only one interface per node. If this is possible, it may greatly simplify sensor design while performing as well as a multi-channel, multi-interface solution. Also, the sleep–awake schedule of the radios should be made dynamic in order to accommodate the varying frame rate for video sensors.
- Recently, the cognitive radio paradigm [21,55] has drawn significant interest in which the radio is aware of its surroundings, learns from it and adapts in order to give the best possible functionality to the user. By dynamically sharing spectrum with other services, unused frequency bands can be utilized ensuring optimum bandwidth usage. The release of new spectrum in the 5-GHz U-NII band has spurred on interest in realizing a practical system for multimedia applications. However, porting it to a low-power sensor node and developing controlling mechanisms for channel hand-off is an area that is yet to be explored.
- Multi-channel protocols are not completely collision free as is seen in the case of control packet collision [33,88,114]. All available channels cannot be assumed to be perfectly non-overlapping, as is seen in the case of 802.11b based WLANs [89]. This may necessitate dynamic channel assignment, taking into account the effect of adjacent channel interference, in order to maintain the network QoS.

9.2. Scheduling

MAC layer scheduling in the context of WMSNs differs from the traditional networking model in the sense that apart from choosing the queueing discipline that accounts for latency bounds, rate/power control and consideration of high channel error conditions needs to be incorporated. We believe that an optimal solution is a function of all of these factors, appropriately weighted and seamlessly integrated with a suitable channel access policy described in Section 9.1.

In order to generate optimal schedules that minimize both power consumption and the probability of missing deadlines for real-time messages, PARM [24] integrates the EDD metric described in Section 9.1.2 into an energy consumption function. While significant performance improvements are demonstrated, this work needs to be extended for large-scale networks that are typically envisaged for WMSNs.

Queueing at the MAC layer has been extensively researched and several schemes with varying levels of complexity exist. Of interest to multimedia applications is the development of schemes that allow a delay bound and thus assure smooth streaming of multimedia content. E^2WFQ [101], a variant of the established weighted fair queuing (WFQ) discipline, allows adjustments to be made to the energy-latency-fidelity trade-off space. Depending upon

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

the current residual energy in the network, it is possible to adapt the scheme for greater energy savings, albeit at the cost of a small, bounded increase in worst-case packet latency.

Given the bursty nature of voice and video data, queueing disciplines are needed that can accommodate sudden peaks, as well as operate under localized channel errors. Extending WFQ, the Wireless Packet Scheduling (WPS) presented in [76], addresses the concerns of delay and rate-sensitive packet flows thus making it suitable for multimedia traffic. WPS, however, assumes that the channel error is fully predictable at any time and its practical implementation shows marked deviations from the idealized case in terms of worst-case complexity. This work is suitable for single-hop sensor-sink communication and multi-hop forwarding issues are not explored.

Network calculus [27,36] is a theory for deterministic queueing systems that allows assignment of service guarantees by traffic regulation and deterministic scheduling. Through tools provided by network calculus, bounds on various performance measures, such as delay and queue length, at each element of the network can be derived and thus QoS of a flow can be specified. *Arrival*, *Departure* and *Service curves* reflect the constraints that flows are subjected to within a network. The calculus relies on *Min-plus algebra*, in which addition and multiplication are replaced by minimum and addition, respectively, to operate on these curves. Current network calculus results have been mostly derived for wired networks, and assume static topologies and fixed link capacity, which are clearly unreasonable assumptions in sensor networks. We believe that extending network calculus results to WMSNs is a challenging but promising research thrust, likely to produce important advancements in our ability to provide provable QoS guarantees in multi-hop networks.

9.3. Link-layer error control

Streaming of real-time multimedia data over a sensor network is particularly challenging due to the QoS requirements of a video/audio stream and the unreliability of the wireless medium. For example, for good quality video perception a frame loss rate lower than 10^{-2} is required. This constitutes a hard task since the wireless channel is highly unreliable, mostly caused by multi-path fading and shadowing at the physical layer, and by collisions or co-channel interference at the MAC layer. Two main classes of mechanisms are traditionally employed to combat the unreliability of the wireless channel at the physical and data link layer, namely *forward error correction* (FEC) and *automatic repeat request* (ARQ), along with hybrid schemes. ARQ mechanisms use bandwidth efficiently at the cost of additional latency. Hence, while carefully designed selective repeat schemes may be of some interest, naive use of ARQ techniques is clearly infeasible for applications requiring real-time delivery of multimedia content.

An important characteristic of multimedia content is *unequal importance*, i.e., not all packets have the same importance for correct perceptual reconstruction of the multimedia content. Moreover, multimedia data are usually *error-tolerant*, so even if some errors are introduced, the original information may still be reconstructed with tolerable distortion. Therefore, an idea that has been used effectively consists of applying different degrees of FEC to different parts of the video stream, depending on their relative importance (*unequal protection*). For example, this idea can be applied to layered coded streams to provide graceful degradation in the observed image quality in presence of error losses, thus avoiding so-called ‘‘cliff’’ effects [54].

In general, delivering error-resilient multimedia content and minimizing energy consumption are contradicting objectives. For this reason, and due to the time-varying characteristics of the wireless channel, several joint source and channel coding schemes have been developed, e.g. [44], which try to reduce the energy consumption of the whole process. Some recent papers [77,135] even try to jointly reduce the energy consumption of the whole process of multimedia content delivery, i.e., jointly optimize source coding, channel coding, and transmission power control. In these schemes, the image coding and transmission strategies are adaptively adjusted to match current channel conditions by exploiting the peculiar characteristics of multimedia data, such as unequal importance of different frames or layers. However, most of these efforts have originated from the multimedia or coding communities, and thus do not jointly consider other important networking aspects of content delivery over a multi-hop wireless networks of memory-, processing- and battery-constrained devices.

In [126], a cross-layer analysis of error control schemes for WSNs is presented. The effects of multi-hop routing and of the broadcast nature of wireless communications are investigated to model the energy consumption, latency and packet error rate performance of error control schemes. As a result, error control schemes are studied through a cross-layer analysis that considers the effects of routing, medium access, and physical layer. This analysis enables a comprehensive comparison of FEC and ARQ schemes in WSNs. FEC schemes are shown to improve the error resiliency compared to ARQ. In a multi-hop network, this improvement can be exploited by reducing the transmit power (*transmit power control*) or by constructing longer hops (*hop length extension*) through channel-aware routing protocols. The analysis reveals that, for certain FEC codes, hop length extension decreases both the energy consumption and the end-to-end latency subject to a target packet error rate compared to ARQ. Thus, FEC codes are an important candidate for delay-sensitive traffic in WSNs. On the other hand, transmit power control results in significant savings in energy consumption at the cost of increased latency.

A different approach to link-layer reliability is proposed through the use of erasure codes [65]. In sensor networks where message passing is minimized, such a scheme significantly reduces the need for retransmissions. It allows recovering of m original messages by receiving any m out of n code words. In lossy conditions, the number of such code words generated can be optimized against the energy expense in their transmission to ensure greater reliability. In multimedia applications, where the loss of a few frames may be tolerated, shorter code words may be used. When detailed, higher resolution images are needed, regenerating the lost information through these codes may be preferred over interpolation of the missing data at the receiver end. It should be noted that this approach works best for static reliability requirements. Dynamically changing code lengths also increases the packet size and its overall effect on other factors like congestion cannot be trivially estimated.

9.3.1. Open research issues

- There is a need to develop models and algorithms to integrate source and channel coding schemes in existing cross-layer optimization frameworks. The existing schemes mostly consider point-to-point wireless links, and neglect interference from neighboring devices and multi-hop routes.
- Since multimedia data is usually error-tolerant, new packet dropping schemes for multimedia delivery have to be delivered, that selectively drop packets that will not impact the perceived quality at the end user.
- Energy-constrained sensor networks naturally call for Selective Repeat ARQ techniques. However, this can introduce excessive latency. There is a need to study trade-offs between the degree of reliability required (i.e., acceptable packet error rate) and the sustainable delay at the application layer.
- Coordinating link and transport layer error recovery schemes is a challenge that remains to be addressed. In order to ensure that buffer over-flow conditions do not occur, mechanisms that detect increased MAC level contention and regulate data generation rate could be implemented.
- In-network storage schemes [26], in which data is stored within the sensor nodes itself and accessed on demand, may further reduce available buffer capacity of the sensor nodes. Thus allocation of optimum buffer sizes truly merits a cross-layer approach, spanning application layer architecture to existing channel conditions seen at the PHY layer.

9.4. Multimedia packet size

In wireless networks, the successful reception of a packet depends upon environmental factors that decide the bit error rate (*BER*) of the link.

Packet length clearly has a bearing on reliable link level communication and may be adjusted according to application delay sensitivity requirements. The Dynamic Packet Size Mechanism (DPSM) scheme [124] for wireless networks follows an additive increase, multiplicative decrease (AIMD) mechanism to decide the packet length, analogous to the congestion control performed by TCP at the transport layer. As an example, if a packet fails the checksum, the sender is intimated and the subsequent packets are sent with a multiplicative decrease in length. However, the problems associated with burst-error channel and the adaptation of video quality have not been addressed in this work. Grouping smaller packets together in order to reduce contention has been explored in Packet Frame Grouping (PFG) [123] and PAcket Concatenation (PAC) [134]. Originally devised for 802.11-like protocols, here the header overhead is shared by the frames. In PFG, the individual frames

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

may be addressed to different senders and requires per-frame ACKs while PAC requires buffering as all frames need to have a common destination. Depending upon the information content of the frame and the channel conditions, variable length forward error-correcting codes (FEC) can be used to reduce the effects of transmission errors at the decoder. The trade-off between the increase of packet length due to the additional parity bits and energy constraints is evaluated in [108], where FEC is shown to perform better than retransmissions.

10. Physical layer

In this section, we discuss the applicability of the UWB transmission technique, which we advocate over other technologies such as Zigbee, as the most suitable choice for multimedia sensors.

10.1. Ultra wide band communications

The ultra wide band (UWB)[1] technology has the potential to enable low-power consumption, high data rate communications within tens of meters, characteristics that make it an ideal choice for WMSNs.

UWB signals have been used for several decades in the radar community. Recently, the US Federal Communications Commission (FCC) Notice of Inquiry in 1998 and the First Report and Order in 2002 [9] inspired a renewed flourish of research and development efforts in both academy and industry due to the characteristics of UWB that make it a viable candidate for wireless communications in dense multi-path environments.

Although UWB signals, as per the specifications of the FCC, use the spectrum from 3.1 GHz to 10.6 GHz, with appropriate interference limitation, UWB devices can operate using spectrum occupied by existing radio services without causing interference, thereby permitting scarce spectrum resources to be used more efficiently. Instead of dividing the spectrum into distinct bands that are then allocated to specific services, UWB devices are allowed to operate overlaid and thus interfere with existing services, at a low enough power level that existing services would not experience performance degradation. The First Report and Order by the FCC includes standards designed to ensure that existing and planned radio services, particularly safety services, are adequately protected.

There exist two main variants of UWB. The first, known as Time-Hopping Impulse Radio UWB (TH-IR-UWB) [105], and mainly developed by Win and Scholtz [129], is based on sending very short duration pulses (in the order of hundreds of picoseconds) to convey information. Time is divided into frames, each of which is composed of several chips of very short duration. Each sender transmits one pulse in a chip per frame only, and multi-user access is provided by pseudo-random time hopping sequences (THS) that determine in which chip each user should transmit. A different approach, known as Multi-Carrier UWB (MC-UWB), uses multiple simultaneous carriers, and is usually based on Orthogonal Frequency Division Multiplexing (OFDM) [25].

MC-UWB is particularly well suited for avoiding interference because its carrier frequencies can be precisely chosen to avoid narrowband interference to or from narrowband systems. However, implementing a MC-UWB front-end power amplifier can be challenging due to the continuous variations in power over a very wide bandwidth. Moreover, when OFDM is used, high-speed FFT processing is necessary, which requires significant processing power and leads to complex transceivers.

TH-IR-UWB signals require fast switching times for the transmitter and receiver and highly precise synchronization. Transient properties become important in the design of the radio and antenna. The high instantaneous power during the brief interval of the pulse helps to overcome interference to UWB systems, but increases the possibility of interference from UWB to narrowband systems. The RF front-end of a TH-IR-UWB system may resemble a digital circuit, thus circumventing many of the problems associated with mixed signal integrated circuits. Simple TH-IR-UWB systems can be very inexpensive to construct.

Although no sound analytical or experimental comparison between the two technologies is available to our knowledge, we believe that TH-IR-UWB is particularly appealing for WMSNs for the following reasons:

---

[1] The FCC defines UWB as a signal with either a fractional bandwidth of 20% of the center frequency or 500 MHz (when the center frequency is above 6 GHz). The FCC calculates the fractional bandwidth as $2(f_H - f_L)/(f_H + f_L)$ where f_H represents the upper frequency of the −10 dB emission limit and f_L represents the lower frequency limit of the −10 dB emission limit [105].

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

- It enables high data rate, very low-power wireless communications, on simple-design, low-cost radios (carrierless, baseband communications) [129].
- Its fine delay resolution properties are appropriate for wireless communications in dense multi-path environment, by exploiting more resolvable paths [113].
- Provides large processing gain in presence of interference.
- Provides flexibility, as data rate can be traded for power spectral density and multi-path performance.
- Finding suitable codes for THS is trivial (as opposed to CDMA codes), and no assignment protocol is necessary.
- It naturally allows for integrated MAC/PHY solutions; [87]. Moreover, interference mitigation techniques [87] allow realizing MAC protocols that do not require mutual temporal exclusion between different transmitters. In other words, simultaneous communications of neighboring devices are feasible without complex receivers as required by CDMA.
- The large instantaneous bandwidth enables fine time resolution for accurate position estimation [49] and for network time distribution (synchronization).
- UWB signals have extremely low-power spectral density, with low probability of intercept/detection (LPI/D), which is particularly appealing for military covert operations.

10.1.1. Ranging capabilities of UWB

Particularly appealing for WMSNs are UWB high data rate with low-power consumption, and its positioning capabilities. Positioning capabilities are needed in sensor networks to associate physical meaning to the information gathered by sensors. Moreover, knowledge of the position of each network device allows for scalable routing solutions [84]. While angle-of-arrival techniques and signal strength based techniques do not give particular advantages with respect to other transmission techniques, time-based approaches in UWB allow ranging accuracy in the order of centimeters [49]. This can be intuitively explained by the expression in (1), which gives a lower bound on the best achievable accuracy of a distance estimate $\hat{d}$ [49]:

$$\sqrt{\mathrm{Var}(\hat{d})} \geqslant \frac{c}{2\sqrt{2}\pi\sqrt{\mathrm{SNR}}\beta}, \qquad (1)$$

where c is the speed of light, SNR represents the signal-to-noise ratio, and β is the effective signal bandwidth. As can be seen, the accuracy of the time-based localization technique can be improved by either increasing the effective bandwidth or the SNR. For this reason, the large bandwidth of UWB systems allows extremely accurate location estimations, e.g., within one inch at SNR = 0 dB and with a pulse of 1.5 GHz bandwidth. Excellent comprehensive surveys of the UWB transmission technique, and of localization techniques for UWB systems, are provided in [132,49], respectively.

10.1.2. Standards based on UWB

The IEEE 802.15.3a task group has been discussing for three years an alternate physical layer for its high data rate Wireless Personal Area Networks (WPAN) standard. However, in early 2005 the group has been disbanded after not being able to reach a consensus on a single UWB-based standard between two competing proposal from two leading industry groups, the UWB Forum and the WiMedia Alliance. The UWB Forum proposal was based on a Direct Sequence DS-UWB technology, while the Wimedia alliance was proposing a Multi-band Orthogonal Frequency Division Multiplexing (MB-OFDM). The IEEE 802.15.4a task group is developing an alternate physical layer for low data rate, very low-power consumption sensors, based on impulse radio UWB.

10.1.3. Open research issues

- While the UWB transmission technology is advancing rapidly, many challenges need to be solved to enable multi-hop networks of UWB devices. In particular, although some recent efforts have been undertaken in this direction [39,87], how to efficiently share the medium in UWB multi-hop networks is still an open issue.
- As a step ahead, research is needed aimed at designing a cross-layer communication architecture based on UWB with the objective of reliably and flexibly *delivering QoS to heterogeneous applications in WMSNs*, by carefully leveraging and controlling interactions among layers according to the applications requirements.
- It is necessary to determine how to provide provable latency and throughput bounds to multimedia flows in an UWB environment.
- It is needed to develop analytical models to quantitatively compare different variants of UWB to

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

determine trade-offs in their applicability to high data rate and low-power consumption devices such as multimedia sensors.

- A promising research direction may also be to integrate UWB with advanced cognitive radio [21] techniques to increase the spectrum utilization. For example, UWB pulses could be adaptively shaped to occupy portions of the spectrum that are subject to lower interference.

10.2. Other physical layer technologies

Devices based on the specifications issued by the Zigbee alliance can find applicability in low data rate applications that require simple forms of QoS guarantees. Zigbee [7] is a specification for a suite of high-level communication protocols using small, low-power digital radios based on the IEEE 802.15.4 standard for wireless personal area networks (WPANs). IEEE 802.15.4 can operate at bandwidths of 250 kbit/s at 2.4 GHz, 40 kbit/s at 915 MHz (America) and 20 kbit/s at 868 MHz (Europe). While the data rate is much lower than Bluetooth, energy consumption is much lesser here and recent low-cost commercial versions have demonstrated the viability of this technology for low duty cycle (<0.01) sensor applications. The CSMA-CA MAC allows a large number of devices to be connected simultaneously while provision is also made for guaranteed time slot communication. The IEEE 802.15.4 standard allows three traffic types, namely, periodic data, intermittent data and low frequency data through its contention-based and contention-free channel access methods. In particular, for delay-sensitive applications, slots can be reserved every super-frame that allows contention-free and high-priority access. The standard specifies a reduced functionality device (RFD) and a full functionality device (FFD) in which only the latter can talk with other RFDs or FFDs and assume the role of network coordinators/traffic forwarders. Distinguished on the basis of memory resource availability and communication capability, this standard introduces heterogeneity thus cutting deployment costs further.

10.2.1. Open research issues

Though dedicated communication slots are possible in Zigbee, the low data rate limits its applicability for multimedia applications. The standard describes a self-organizing network but heterogeneous nodes necessitate some form of topology control in order to derive optimum ratios of FFD and RFD devices. Such a ratio will be dependent on the region being monitored, the desired coverage accuracy, amongst others. There is no built-in support in Zigbee that splits up large data into smaller packets and additional code has to be inserted at the application layer. Video and image capturing sensors will hence need a specialized PHY/MAC aware application layer.

11. Cross-layer design

As previously discussed, in multi-hop wireless networks there is a strict interdependence among functions handled at all layers of the communication stack. The physical, MAC, and routing layers together impact the contention for network resources. The physical layer has a direct impact on multiple access of nodes in wireless channels by affecting the interference at the receivers. The MAC layer determines the bandwidth allocated to each transmitter, which naturally affects the performance of the physical layer in terms of successfully detecting the desired signals. On the other hand, as a result of transmission schedules, high packet delays and/or low bandwidth can occur, forcing the routing layer to change its route decisions. Different routing decisions alter the set of links to be scheduled, and thereby influence the performance of the MAC layer. Furthermore, congestion control and power control are also inherently coupled [32], as the capacity available on each link depends on the transmission power. Moreover, specifically to multimedia transmissions, the application layer does not require full insulation from lower layers, but needs instead to perform source coding based on information from the lower layers to maximize the multimedia performance. Existing solutions often do not provide adequate support for multimedia applications since the resource management, adaptation, and protection strategies available in the lower layers of the stack are optimized without explicitly considering the specific characteristics of multimedia applications. Similarly, multimedia compression and streaming algorithms do not consider the mechanisms provided by the lower layers for error protection and resource allocation [112].

The additional challenges brought about by real-time streaming of multimedia content in WMSNs call for new research on cross-layer optimization and cross-layer design methodologies, to leverage potential improvements of exchanging information between different layers of the communication stack. However, the increased complexity of functionalities needed to deliver QoS to multimedia applications needs to be managed as well. In particular, it is important to keep some form of logical separation of these functionalities to preserve upgradability and ease of design and testing. To this aim, it is needed to specify standardized interfaces that will allow leveraging these interactions.

Although a consistent amount of recent papers have focused on cross-layer design and improvement of protocols for WSNs, a systematic methodology to accurately model and leverage cross-layer interactions is still largely missing. Most of the existing studies decompose the resource allocation problem at different layers, and consider allocation of the resources at each layer separately. In most cases, resource allocation problems are treated either heuristically, or without considering cross-layer interdependencies, or by considering pairwise interactions between isolated pairs of layers.

In [112], the cross-layer transmission of multimedia content over wireless networks is formalized as an optimization problem. Several different approaches for cross-layer design of multimedia communications are discussed, including *bottom-up approach*, where the lower layers try to insulate the higher layers from losses and channel capacity variations, and *top-down*, where the higher layer protocols optimize their parameters at the next lower layer. However, only single-hop networks are considered.

In [116], several techniques that provide significant performance gains through cross-layer optimizations are surveyed. In particular, the improvements of adaptive link layer techniques such as adaptive modulation and packet size optimization, joint allocation of capacity and flows (i.e., MAC and routing), joint scheduling and rate allocation, are discussed. While still maintaining a strict layered architecture, it is shown how these cross-layer optimizations help improve the spectral efficiency at the physical layer, and the peak signal-to-noise ratio (PSNR) of the video stream perceived by the user. Clearly, energy-constrained multimedia sensors may need to leverage cross-layer interactions one step further. At the same time, optimization metrics in the energy domain need to be considered as well.

We are currently developing a new cross-layer communication architecture [83] whose objective is to reliably and flexibly *deliver QoS to heterogeneous applications in WMSNs*, by carefully leveraging and controlling interactions among layers according to the applications requirements. Its design is based on the following principles:

- *Network layer QoS support enforced by a cross-layer controller*. The proposed system provides QoS support at the network layer, i.e., it provides packet-level service differentiation in terms of throughput, end-to-end packet error rate, and delay. This is achieved by controlling operations and interactions of functionalities at the physical, MAC, and network layers, based on a unified logic that resides on a cross-layer controller that manages resource allocation, adaptation, and protection strategies based on the state of each functional block, as shown in Fig. 10. The objective of the controller is to optimize some objective function, i.e., minimize energy consumption, while guaranteeing QoS requirements to application flows. While all decisions are jointly taken at the controller, implementation of different functionalities is kept separate for ease of design and upgradeability.
- *UWB physical/MAC layer*. The communication architecture is based on an integrated TH-IR-UWB MAC and physical layer. Similarly to CDMA, TH-IR-UWB allows several transmissions in parallel. Conversely, typical MAC protocols for sensor networks, such as contention-based protocols based on CSMA/CA, require *mutual temporal exclusion* between neighboring transmitters. This allows devising MAC protocols with minimal coordination. While CDMA usually entails complex transceivers and cumber-

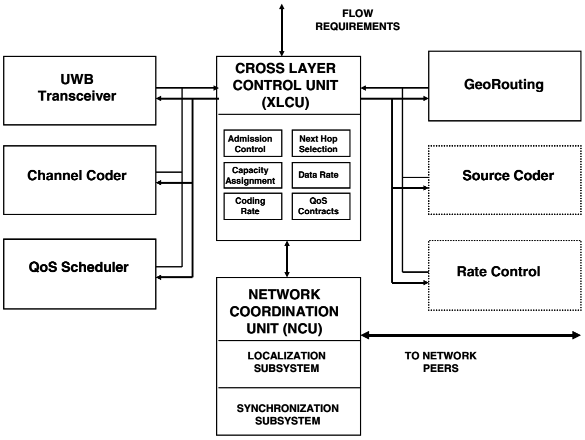

Fig. 10. Cross-layer communication architecture.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

some code assignment protocols, this is achievable with simple transceivers in TH-IR-UWB.

- *Receiver-centric scheduling for QoS traffic*. One of the major problems in multi-hop wireless environments is that channel and interference vary with the physical location of devices. For this reason, we believe that QoS provisioning should be based on receiver-centric scheduling of packets. This way, the receiver can easily estimate the state of the medium at its side. Thus, it can optimally handle loss recovery and rate adaptation, thereby avoiding feedback overheads and latency, and be responsive to the dynamics of the wireless link using the information obtained locally.
- *Dynamic channel coding*. Adaptation to interference at the receiver is achieved through dynamic channel coding, which can be seen as an alternative form of power control, as it modulates the energy per bit according to the interference perceived at the receiver.
- *Geographical forwarding*. We leverage UWB's positioning capabilities, to allow scalable geographical routing. The routing paths are selected by the cross-layer controller by applying an admission control procedure that verifies that each node on the path be able to provide the required service level. The required packet error rate and maximum allowed delay are calculated at each step based on the relative advance of each hop towards the destination.
- *Hop-by-hop QoS contracts*. End-to-end QoS requirements are guaranteed by means of local decision. Each single device that participates in the communication process is responsible for locally guaranteeing given performance objectives. The global, end-to-end requirement is enforced by the joint local behaviors of the participating devices.
- *Multi-rate transmission*. TH-IR-UWB allows varying the data rate at the physical layer, by modifying the *pulse repetition period*. While this functionality has not been fully explored so far, it is possible to devise adaptive systems that modify the achievable data rate at the physical layer based on the perceived interference and on the required power consumption.

12. Other research issues

While most of the challenges in realizing a practical implementation of WMSNs can be classified into layer-specific considerations, there are also additional areas that need to be addressed. This section discusses the impact of recent advances in sensor-actuation, synchronization issues, spatial localization techniques, security and management tools in the context of multimedia transmission.

12.1. Convergence of sensing and actuation

The challenges brought about by WMSNs are not to be limited to resource allocation problems. In fact, WMSNs enable new application scenarios in synergy with other research areas. For example, *Distributed Robotics* [30] has been a hot research topic since the mid-1990s. In distributed robotics, a task is not completed by a single robot but by a *team of collaborating robots*. Information about the surrounding environment is usually gathered by *onboard sensors*, and team members exchange sensor information to move or perform actions (e.g., collaborate to manipulate heavy objects). As opposed to a single robot, a team of robots can perceive the environment from *multiple disparate viewpoints*. In the recently proposed Wireless Sensor and Actor Networks (WSANs) [20] paradigm, the ability of the actors to perceive the environment can be pushed one step further: a dense spatio-temporal sampling of the environment, provided by a pre-deployed sensor network, can be exploited by the whole team of actors, thus increasing the ability of the team to accurately interact with the physical environment. Furthermore, multimedia content gathered by sensors can be used to provide the team of actors with accurate vision from multiple perspectives, while as of today collaborating actors mostly rely on expensive onboard cameras.

Coordination and communication algorithms for static [86] and mobile [85] WSANs have been the focus of our research in recent years. In [86], we introduced a framework for communication and coordination problems in WSANs. The notions of sensor–actor coordination and actor–actor coordination were introduced. The process of establishing data paths between sensors and actors is referred to as sensor–actor coordination. Once an event has been detected, actors coordinate to reconstruct it, to estimate its characteristics, and make a collaborative decision on how to perform the action. This process is referred to as actor–actor coordination. In [85], we introduced a location management scheme to handle the mobility of actors with minimal energy consumption for resource-constrained sensors. The proposed scheme is designed to reduce the energy consumption with respect to existing localization services for ad hoc and sensor networks. This is achieved through a combination of location updating and location prediction, where actors broadcast location updates limiting their scope based on Voronoi diagrams, while sensors predict the movement of actors based on Kalman filtering of previously received updates.

Clearly, further research is needed to fully leverage the opportunities offered by the integration of actors and multimedia sensors in a wireless network.

12.2. Network synchronization

Time synchronization is difficult to achieve on a network-wide basis due to slow clock drift over time, effect of temperature and humidity on clock frequencies, coordination and correction amongst thousands of deployed nodes with low messaging overhead, amongst others [121]. The need for accurate timing in WMSNs is stressed mainly for the following two scenarios:

- In-network processing schemes are often used in WMSNs in order to reduce traffic load in the network. However, flows can only be aggregated at an intermediate node if the difference in the packet generation times is within allowed bounds. We believe that limited synchronization could serve the requirements of WMSNs, in which, it needs to be enforced only along the route chosen to the sink and for a specified length of time. Algorithms specially developed for sensor networks [82] can be easily adapted and large-scale coordination avoided.
- Real-time streaming needs clock synchronization at the sender and receiver to prevent buffer underflows and overflows. Phase Locked Loops (PLL) help in maintaining clock frequencies at the two ends but an efficient digital implementation of the Voltage Controlled Oscillator (VCO) is an important area of research.

12.3. Inter-media synchronization

Multimedia data returned by sensors can be heterogeneous, and possibly correlated. As an example, video and audio samples could be collected over a common geographical region, or still pictures may complement text data containing field measurements. In such cases, the flow of two separate multimedia streams needs to be synchronized as the final data available at the end user will comprise of both of these, played back under a joint timing constraint. The task of coordinating such sequences is called multimedia synchronization. The problem of multimedia synchronization is cross-layer in nature and influences primarily the physical layer and the application layer. At the physical layer, data from different media are multiplexed over shared wireless connections, or are stored in common physical storage. The application layer is concerned with intermedia synchronization necessary for presentation or playout, i.e., with the interactions between the multimedia application and the various media.

The performance metrics for such inter-media synchronization differ from the cases in which only voice or video streaming is desired. Synchronization can be applied to the playout of concurrent or sequential streams of data, and also to the external events generated by a human user. In [75], it is argued that the average instantaneous delay variation (skew) best measures inter-media synchronization for continuous media. The maximum and minimum delay, used in typical real-time scheduling, can be effectively applied for discrete events associated with timed playout of text, graphics, and still images. Minimizing the delay variation at each hop is a challenge yet unaddressed in the context of sensor networks, though the effects of intermedia skew can be mitigated to an extent by dropping and duplicating frames of the different media streams so that they play back in unison at the receiver.

12.4. Localization

Determining the location of the sensor nodes with respect to a common reference, or in the context of the object being monitored by them, is an important aspect in multimedia applications. Cameras and microphones have limited field of operation and hence *reachability* and *coverage* are two important considerations that go into efficient sharing of monitoring tasks. Localization techniques help in allocating resources to events, deciding sensing precision and ensuring complete monitoring of the area under study.

In [104], this common space is provided by automatically determining the relative 3D positions of audio sensors and actors through a closed form approximate solution based on time of flight and time difference of flight. This approach however, requires

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

that audio signals emitted by each sensor does not interfere during the localization process which implies a network-wide ordering or code assignment.

SensEye [69] uses a two-tier localization in which cheap, low resolution cameras are used for object detection and higher resolution web-cameras are used for tracking. It uses two cameras with overlapping coverage to localize an object and compute its Cartesian (x, y, z) coordinates which, in turn, is used to intelligently wake up other nodes or to determine the trajectory of the object. This work assumes that the orientations of the cameras are known relative to a global reference frame and achieved through a GPS like or distributed technique.

12.5. Network security

Security in WMSNs has recently caught the attention of the research community with increasing applications of sensors in military use. While the use of stronger codes, watermarking techniques, encryption algorithms, amongst others, have resulted in secured wireless communication, there are altogether different considerations in WMSNs.

As outlined in [71], a video sensor surveillance system may require in-network processing techniques to reduce the amount of information flowing in the network. At the aggregation point of the incoming streams, the packets would have to be completely decoded and thus the computational complexity of the security algorithms must be low enough to allow real-time processing. There is hence a trade-off between providing enhanced security to the data flow by adopting a higher order code at the source video sensor and permissible multimedia delay requirements. Apart from devising effective light-weight coding techniques, we believe that efforts in this area must be directed to leverage physical layer strategies, as processing power on the battery-powered nodes is likely to be limited.

The delta–sigma ($\Delta\Sigma$) modulator for high-speed speech processing is modified in [71] for simultaneously digitizing and authenticating sensor readings. By exchanging simple keys, filter parameters can be decided that are used to encode the generated stream, thus proving to be an computationally inexpensive scheme. However, this technique has several practical difficulties including modulator matching between the sender and receiver and precision tracking of the signal for accurate demodulation. Other areas that need to be explored are watermarking for heterogeneous streams of voice and video applications. Scalar or voice data may be rendered *invisible* by embedding it in frames of video images thus making eavesdropping difficult.

12.6. Network management

Multimedia network management, when applied to sensor networks, can be considered as a functionality that encompasses resources, theories, tools and techniques to manipulate data provided by different and possibly mixed media with the goal to extract relevant information [74]. Taking into account the concerns of WMSNs, we believe that the design of such systems should be influenced by the following factors:

- *Reduced hardware/application requirement*. Tools for WMSNs may comprise of hand-held devices that are used by the network manager to conduct on-site surveys. Light-weight application environments like the Java Platform, Micro Edition (Java ME) [8] or AJAX [98] could be used for local record manipulation as they considerably reduce traffic between the source and the distant server (or sink). Java ME provides support for networked and offline applications that can be downloaded dynamically. For distant web-based monitoring, AJAX builds upon XML and can be used to perform simple, localized search and modification queries with short length messages exchanged between the source and sink. These tools may allow dynamic reassignment of goals based on perceived QoS or events of interest.
- *Independence of platform/programming environment*. Established proprietary tools like LabView cannot be easily integrated with other languages and there is a clear need of platform independent, general purpose monitoring tools. As a solution, BeanWatcher [74], specially devised for WMSNs, can work with several languages, such as Java, Java ME and C++. Besides, it provides some visual components (e.g., thermometer, speedometer, gauge, and valued maps) to cover different types of sensory data and accommodates the fact that different streams from different types of applications need to be treated differently.

13. Conclusions

We discussed the state of the art of research on Wireless Multimedia Sensor Networks (WMSNs), and outlined the main research challenges. Algorithms, protocols, and hardware for the development of WMSNs were surveyed, and open research issues discussed in detail. We classified currently off-the-shelf hardware as well as available research prototypes for WMSNs. Furthermore, we discussed existing solutions and open research issues at the application, transport, network, link, and physical layers of the communication stack, along with possible cross-layer synergies and optimizations. We pointed out how recent work undertaken in Wyner–Ziv coding at the application layer, specialized spatio-temporal transport layer solutions, delay bounded routing, multi-channel MAC protocols, and UWB technology, amongst others, seem most promising research directions in developing practical WMSNs. We believe that this research area will attract the attention of many researchers and that it will push one step further our ability to observe the physical environment and interact with it.

Acknowledgement

The authors would like to thank Vehbi C. Gungor, Dario Pompili, and Mehmet C. Vuran for their valuable comments.

References

[1] Acroname GARCIA Robotic Platform. <http://www.acroname.com/garcia/garcia.html>.

[2] Advanced Video Coding for Generic Audiovisual Services, ITU-T Recommendation H.264.

[3] Crossbow. <http://www.xbow.com>.

[4] Crossbow MICA2 Mote Specifications. <http://www.xbow.com>.

[5] Crossbow MICAz Mote Specifications. <http://www.xbow.com>.

[6] Crossbow TelosB Mote Specifications. <http://www.xbow.com>.

[7] IEEE 802.15 wpan task group 4 (tg4). <http://grouper.ieee.org/groups/802/15/pub/TG4.html>.

[8] Java Platform, Micro Edition. <http://java.sun.com/javame/index.jsp>.

[9] Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems. First note and Order, Federal Communications Commission, ET-Docket 98-153, adopted February 14, 2002, released April 22, 2002.

[10] The Stargate Platform. <http://www.xbow.com/Products/Xscale.htm>.

[11] Video Coding for Low Bit Rate Communication, ITU-T Recommendation H.263.

[12] JPEG2000 Requirements and Profiles, ISO/IEC JTC1/SC29/WG1 N1271, March 1999.

[13] A. Aaron, S. Rane, E. Setton, B. Girod, Transform-domain Wyner–Ziv codec for video, in: Proc. of Society of Photo-Optical Instrumentation Engineers – Visual Communications and Image Processing, San Jose, CA, USA, January 2004.

[14] A. Aaron, S. Rane, R. Zhang, B. Girod, Wyner–Ziv coding for video: applications to compression and error resilience, in: Proc. of IEEE Data Compression Conf. (DCC), Snowbird, UT, March 2003, pp. 93–102.

[15] A. Aaron, E. Setton, B. Girod, Towards practical Wyner–Ziv coding of video, in: Proc. of IEEE Intl. Conf. on Image Processing (ICIP), Barcelona, Spain, September 2003.

[16] R. Agre, L.P. Clare, G.J. Pottie, N.P. Romanov, Development platform for self-organizing wireless sensor networks, in: Proc. of Society of Photo-Optical Instrumentation Engineers – Aerosense, Orlando, FL, USA, March 1999.

[17] O. Akan, I.F. Akyildiz, Event-to-sink reliable transport in wireless sensor networks, IEEE/ACM Trans. Network. 13 (5) (2005) 1003–1017.

[18] K. Akkaya, M. Younis, An energy-aware QoS routing protocol for wireless sensor networks, in: Proc. of Intl. Conf. on Distributed Computing Systems Workshops (ICSDSW), Washington, DC, 2003.

[19] K. Akkaya, M. Younis, A survey of routing protocols in wireless sensor networks, Ad Hoc Network (Elsevier) 3 (3) (2005) 325–349.

[20] I.F. Akyildiz, I.H. Kasimoglu, Wireless sensor and actor networks: Research challenges, Ad Hoc Networks (Elsevier) 2 (4) (2004) 351–367.

[21] I.F. Akyildiz, W.-Y. Lee, M.C. Vuran, S. Mohanty, Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey, Comput. Networks (Elsevier) 50 (13) (2006) 2127–2159.

[22] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, Comput. Networks (Elsevier) 38 (4) (2002) 393–422.

[23] J.N. Al-Karaki, A.E. Kamal, Routing techniques in wireless sensor networks: a survey, IEEE Wireless Commun. 11 (6) (2004) 6–28.

[24] M.I. Alghamdi, T. Xie, X. Qin, PARM: A power-aware message scheduling algorithm for real-time wireless networks, in: Proc. of ACM Workshop on Wireless Multimedia Networking and Performance Modeling (WMuNeP), Montreal, Que., Canada, 2005, pp. 86–92.

[25] Anuj Batra et al., Multi-band OFDM physical layer proposal for IEEE 802.15 Task Group 3a, IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), March 2004.

[26] R. Biswas, K.R. Chowdhury, D.P. Agrawal, Optimal data-centric attribute allocation and retrieval (DCAAR) scheme for wireless sensor networks, in: Proc. of IEEE Intl. Conf. on Mobile Ad-hoc and Sensor Systems (MASS), Washington, DC, USA, November 2005.

[27] J.-Y.L. Boudec, P. Thiran, Network CalculusLNCS, vol. 2050, Springer-Verlag, New York, 2001.

[28] A. Boulis, M. Srivastava, Node-level energy management for sensor networks in the presence of multiple applications, in: Proc. of IEEE Intl. Conf. on Pervasive Computing and Communications (PerCom), Dallas – Forth Worth, TX, USA, 2003, pp. 41–49.

[29] J. Campbell, P.B. Gibbons, S. Nath, P. Pillai, S. Seshan, R. Sukthankar, IrisNet: an Internet-scale architecture for multimedia sensors, in: Proc. of the ACM Multimedia Conference, 2005.

[30] Y.U. Cao, A.S. Fukunaga, A.B. Kahng, Cooperative mobile robotics: antecedents and directions, Autonom. Robots 4 (1997) 1–23.

[31] J. Capone, I. Stavrakakis, Delivering QoS requirements to traffic with diverse delay tolerances in a TDMA environment, IEEE/ACM Trans. Network. 7 (1) (1999) 75–87.

[32] M. Chiang, Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control, IEEE J. Select. Areas Commun. 23 (1) (2005) 104–116.

[33] K.R. Chowdhury, N. Nandiraju, D. Cavalcanti, D.P. Agrawal, CMAC-A multi-channel energy efficient MAC for wireless sensor networks, in: Proc. of IEEE Wireless Communications and Networking Conference (WCNC), Las Vegas, NV, April 2006.

[34] D. Chu, K. Lin, A. Linares, G. Nguyen, J.M. Hellerstein, sdlib: a sensor network data and communications library for rapid and robust application development, in: Proc. of Information Processing in Sensor Networks (IPSN), Nashville, TN, USA, April 2006.

[35] M. Chu, J.E. Reich, F. Zhao, Distributed attention for large video sensor networks, in: Proc. of the Institute of Defence and Strategic Studies (IDSS), London, UK, February 2004.

[36] R. Cruz, A calculus for network delay. I. Network elements in isolation, IEEE Trans. Informat. Theory 37 (1) (1991) 114–131.

[37] R. Cucchiara, Multimedia surveillance systems, in: Proc. of ACM Intl. Workshop on Video Surveillance and Sensor Networks, Singapore, November 2005.

[38] S. Cui, R. Madan, A. Goldsmith, S. Lall, Energy-delay tradeoffs for data collection in TDMA-based sensor networks, in: Proc. of IEEE Intl. Conf. on Communications (ICC), Seoul, Korea, May 2005, pp. 3278–3284.

[39] F. Cuomo, C. Martello, A. Baiocchi, F. Capriotti, Radio resource sharing for ad-hoc networking with UWB, IEEE J. Select. Areas Commun. 20 (9) (2002) 1722–1732.

[40] T.V. Dam, K. Langendoen, An adaptive energy-efficient MAC protocol for wireless sensor networks, in: Proc. of the ACM Conf. on Embedded Networked Sensor Systems (SenSys), Los Angeles, CA, USA, November 2003.

[41] I. Downes, L.B. Rad, H. Aghajan, Development of a mote for wireless image sensor networks, in: Proc. of COGnitive systems with Interactive Sensors (COGIS), Paris, France, March 2006.

[42] A. Dunkels, T. Voigt, J. Alonso, Making TCP/IP viable for wireless sensor networks, in: Proc. of European Workshop on Wireless Sensor Networks (EWSN), Berlin, Germany, January 2004.

[43] A. Dunkels, T. Voigt, J. Alonso, H. Ritter, Distributed TCP caching for wireless sensor networks, in: Proc. of the Mediterranean Ad Hoc Networking Workshop (MedHocNet), June 2004.

[44] Y. Eisenberg, C.E. Luna, T.N. Pappas, R. Berry, A.K. Katsaggelos, Joint source coding and transmission power management for energy efficient wireless video communications, IEEE Trans. Circ. Syst. Video Technol. 12 (6) (2002) 411–424.

[45] E. Felemban, C.-G. Lee, E. Ekici, MMSPEED: Multipath multi-SPEED protocol for QoS guarantee of reliability and timeliness in wireless sensor networks, IEEE Trans. Mobile Comput. 5 (6) (2006) 738–754.

[46] W. Feng, B. Code, E. Kaiser, M. Shea, W. Feng, L. Bavoil, Panoptes: scalable low-power video sensor networking technologies, in: Proc. of ACM Multimedia, Berkeley, CA, USA, November 2003.

[47] W. Feng, N.B.W. Feng, Dissecting the video sensing landscape, in: Proc. of the ACM Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV), Stevenson, WA, USA, June 2005.

[48] D. Gay, P. Levis, R. von Behren, M. Welsh, E. Brewer, D. Culler, The nesC language: a holistic approach to network embedded systems, in: Proc. of the ACM SIGPLAN 2003 Conf. on Programming Language Design and Implementation (PLDI), San Diego, CA, USA, June 2003.

[49] S. Gezici, Z. Tian, G.B. Giannakis, H. Kobayashi, A.F. Molisch, H.V. Poor, Z. Sahinoglu, Localization via ultra-wideband radios, IEEE Signal Process. Mag. 22 (4) (2005) 70–84.

[50] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, Distributed video coding, Proc. IEEE 93 (1) (2005) 71–83.

[51] B. Girod, M. Kalman, Y. Liang, R. Zhang, Advances in channel-adaptive video streaming, Wireless Commun. Mobile Comput. 2 (6) (2002) 549–552.

[52] V. Gungor, O. Akan, I. Akyildiz, A real-time and reliable transport protocol for wireless sensor and actor networks, submitted for publication.

[53] C. Guo, L.C. Zhong, J.M. Rabaey, Low power distributed MAC for Ad Hoc sensor radio networks, in: Proc. of the IEEE Global Communications Conference (GLOBECOM), San Antonio, TX, November 2001.

[54] E. Gurses, O.B. Akan, Multimedia communication in wireless sensor networks, Ann. Telecommun. 60 (7–8) (2005) 799–827.

[55] S. Haykin, Cognitive radio: Brain-empowered wireless communications, IEEE J. Select. Areas Commun. 23 (2) (2005) 201–220.

[56] T. He, J.A. Stankovic, C. Lu, T.F. Abdelzaher, A spatio-temporal communication protocol for wireless sensor networks, IEEE Trans. Parallel Distr. Syst. 16 (10) (2005) 995–1006.

[57] T. Herman, S. Tixeuil, A distributed TDMA slot assignment algorithm for wireless sensor networks, in: Proc. of the Workshop on Algorithmic Aspects of Wireless Sensor Networks (ALGOSENSORS), Finland, 2004.

[58] R. Holman, J. Stanley, T. Ozkan-Haller, Applying video sensor networks to nearshore environment monitoring, IEEE Perv. Comput. 2 (4) (2003) 14–21.

[59] F. Hu, S. Kumar, Multimedia query with QoS considerations for wireless sensor networks in telemedicine, in: Proc. of Society of Photo-Optical Instrumentation Engineers – Intl. Conf. on Internet Multimedia Management Systems, Orlando, FL, September 2003.

[60] Y.G. Iyer, S. Gandham, S. Venkatesan, STCP: a Generic Transport Layer Protocol for Wireless Sensor Networks, in: Proc. of IEEE Intl. Conf. on Computer Communications and Networks (ICCCN), San Diego, CA, USA, October 2005, pp. 449–454.

[61] D. James, L. Klibanov, G. Tompkins, S.J. Dixon-Warren, Inside CMOS Image Sensor Technology. Chipworks White Paper. <http://www.chipworks.com/resources/whitepapers/Inside-CMOS.pdf>.

[62] S.K. Jayaweera, An energy-efficient virtual MIMO communications architecture based on V-BLAST processing for

distributed wireless sensor networks, in: Proc. of IEEE Intl. Conf. on Sensor and Ad-hoc Communications and Networks (SECON), Santa Clara, CA, USA, October 2004.

[63] S.K. Jayaweera, M.L. Chebolu, Virtual MIMO and distributed signal processing for sensor networks – an integrated approach, in: Proc. of IEEE Intl. Conf. on Communications (ICC), Seoul, Korea, May 2005.

[64] X. Jiang, J. Polastre, D. Culler, Perpetual environmentally powered sensor networks, in: Proc. of IEEE Workshop on Sensor Platform, Tools and Design Methods for Networked Embedded Systems (SPOTS), Los Angeles, CA, USA, April 2005.

[65] S. Kim, R. Fonseca, D. Culler, Reliable transfer in wireless sensor networks, in: Proc. of IEEE Intl. Conf. on Sensor and Ad-hoc Communications and Networks (SECON), Santa Clara, CA, October 2004.

[66] J. Koshy, R. Pandey, VM*: Synthesizing scalable runtime environments for sensor networks, in: Proc. of the ACM Conf. on Embedded Networked Sensor Systems (SenSys), San Diego, CA, November 2005.

[67] K. Kredo, P. Mohapatra, Medium access control in wireless sensor networks, Comput. Networks (Elsevier), in press, doi:10.1016/j.comnet.2006.06.012.

[68] P. Kulkarni, D. Ganesan, P. Shenoy, The case for multi-tier camera sensor network, in: Proc. of the ACM Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV), Stevenson, WA, USA, June 2005.

[69] P. Kulkarni, D. Ganesan, P. Shenoy, Q. Lu, SensEye: a multi-tier camera sensor network, in: Proc. of ACM Multimedia, Singapore, November 2005.

[70] S.S. Kulkarni, M. Arumugam, TDMA service for sensor networks, in: Proc. of the Intl. Conf. on Distributed Computing Systems Workshops (ICDCSW), Washington, DC, USA, 2004, pp. 604–609.

[71] D. Kundur, T. Zourntos, N.J. Mathai, Lightweight security principles for distributed multimedia based sensor networks, in: Proc. of Asilomar Conf. on Signals, Systems and Computers, November 2004, pp. 368–372.

[72] P. Kyasanur, N.H. Vaidya, Capacity of multi-channel wireless networks: Impact of number of channels and interfaces, in: Proc. of ACM Intl. Conf. on Mobile Computing and Networking (MobiCom), Cologne, Germany, August 2005.

[73] Q. Li, D. Rus, Global clock synchronization in sensor networks, IEEE Trans. Comput. 55 (2) (2006) 214–226.

[74] A. Lins, E.F. Nakamura, A.A.F. Loureiro, C.J.N. Coelho Jr., BeanWatcher: a tool to generate multimedia monitoring applications for wireless sensor networks, in: Proc. of IFIP/IEEE Intl. Conf. on Management of Multimedia Networks and Services, 2003, pp. 128–141.

[75] T.D.C. Little, A framework for synchronous delivery of time-dependent multimedia data, J. Multimedia Syst. 1 (2) (1993) 87–94.

[76] S. Lu, V. Bharghavan, R. Srikant, Fair scheduling in wireless packet networks, IEEE/ACM Trans. Network. 7 (4) (1999) 473–489.

[77] X. Lu, E. Erkip, Y. Wang, D. Goodman, Power efficient multimedia communication over wireless channels, IEEE J. Select. Areas Commun. 21 (10) (2003) 1738–1751.

[78] X. Luo, K. Zheng, Y. Pan, Z. Wu, A TCP/IP implementation for wireless sensor networks, in: Proc. of IEEE Conf. on Systems, Man and Cybernetics, October 2004, pp. 6081–6086.

[79] S. Mao, D. Bushmitch, S. Narayanan, S.S. Panwar, MRTP: a multiflow real-time transport protocol for Ad Hoc networks, IEEE Trans. Multimedia 8 (2) (2006) 356–369.

[80] C.B. Margi, V. Petkov, K. Obraczka, R. Manduchi, Characterizing energy consumption in a visual sensor network testbed, in: Proc. of IEEE/Create-Net Intl. Conf. on Testbeds and Research Infrastructures for the Development of Networks and Communities (TridentCom), Barcelona, Spain, March 2006.

[81] D. McIntire, Energy Benefits of 32-bit Microprocessor Wireless Sensing Systems, Sensoria Corporation White Paper.

[82] L. Meier, P. Blum, L. Thiele, Interval synchronization of drift-constraint clocks in Ad-Hoc sensor networks, in: Proc. of ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), Tokyo, Japan, May 2004.

[83] T. Melodia, I.F. Akyildiz, Cross-layer QoS support for UWB wireless multimedia sensor networks, in preparation.

[84] T. Melodia, D. Pompili, I.F. Akyildiz, On the interdependence of distributed topology control and geographical routing in ad hoc and sensor networks, IEEE J. Select. Areas Commun. 23 (3) (2005) 520–532.

[85] T. Melodia, D. Pompili, I.F. Akyildiz, A communication architecture for mobile wireless sensor and actor networks, in: Proc. of IEEE Intl. Conf. on Sensor and Ad-hoc Communications and Networks (SECON), Reston, VA, USA, September 2006.

[86] T. Melodia, D. Pompili, V.C. Gungor, I.F. Akyildiz, A distributed coordination framework for wireless sensor and actor networks, in: Proc. of ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), Urbana-Champaign, IL, May 2005.

[87] R. Merz, J. Widmer, J.-Y.L. Boudec, B. Radunovic, A joint PHY/MAC architecture for low-radiated power TH-UWB wireless ad-hoc networks, Wireless Commun. Mobile Comput. J. 5 (5) (2005) 567–580.

[88] M.J. Miller, N.H. Vaidya, A MAC protocol to reduce sensor network energy consumption using a wakeup radio, IEEE Trans. Mobile Comput. 4 (3) (2005) 228–242.

[89] A. Mishra, S. Banerjee, W. Arbaugh, Weighted coloring based channel assignment for WLANs, Mobile Comput. Commun. Rev. 9 (3) (2005) 19–31.

[90] S. Misra, M. Reisslein, G. Xue, A survey of multimedia streaming in wireless sensor networks, submitted for publication.

[91] P.D. Mitcheson, T.C. Green, E.M. Yeatman, A.S. Holmes, Architectures for vibration-driven micropower generators, J. Microelectromech. Syst. 13 (3) (2004) 429–440.

[92] M. Miyabayashi, N. Wakamiya, M. Murata, H. Miyahara, MPEG-TFRCP: Video transfer with TCP-friendly rate control protocol, in: Proc. of IEEE Intl. Conf. on Communications (ICC), Helsinki, June 2001.

[93] S. Nath, Y. Ke, P.B. Gibbons, B. Karp, S. Seshan, A distributed filtering architecture for multimedia sensors, Intel Research Technical Report IRP-TR-04-16, August 2004.

[94] T. Nguyen, S. Cheung, Multimedia streaming with multiple TCP connections, in: Proc. of the IEEE Intl. Performance Computing and Communications Conference (IPCCC), Arizona, April 2005.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

[95] E. Ould-Ahmed-Vall, D. Blough, B.S. Heck, G.F. Riley, Distributed global identification for sensor networks, in: Proc. of IEEE Intl. Conf. on Mobile Ad-hoc and Sensor Systems (MASS), Washington, DC, USA, November 2005.

[96] J. Paradiso, T. Starner, Energy scavenging for mobile and wireless electronics, IEEE Perv. Comput. 4 (1) (2005) 18–27.

[97] M. Perillo, W. Heinzelman, Sensor management policies to provide application QoS, Ad Hoc Networks (Elsevier) 1 (2–3) (2003) 235–246.

[98] B.W. Perry, Ajax Hacks, O'Reilly Publications, 2006.

[99] G.J. Pottie, W.J. Kaiser, Wireless integrated network sensors, Commun. ACM 43 (May) (2000) 51–58.

[100] R. Puri, K. Ramchandran, PRISM: a new robust video coding architecture based on distributed compression principles, in: Proc. of the Allerton Conf. on Communication, Control, and Computing, Allerton, IL, USA, October 2002.

[101] V. Raghunathan, S. Ganeriwal, M. Srivastava, C. Schurgers, Energy efficient wireless packet scheduling and fair queuing, ACM Trans. Embedded Comput. Syst. 3 (1) (2004) 3–23.

[102] M. Rahimi, S. Ahmadian, D. Zats, J. Garcia, M. Srivastava, D. Estrin, Deep vision: Experiments in exploiting vision in wireless sensor networks, submitted for publication.

[103] M. Rahimi, R. Baer, O. Iroezi, J. Garcia, J. Warrior, D. Estrin, M. Srivastava, Cyclops: in situ image sensing and interpretation in wireless sensor networks, in: Proc. of the ACM Conf. on Embedded Networked Sensor Systems (SenSys), San Diego, CA, November 2005.

[104] V.C. Raykar, I. Kozintsev, R. Lienhart, Position calibration of microphones and loudspeakers in distributed computing platforms, IEEE Trans. Speech Audio Process. 13 (1) (2005) 70–83.

[105] J. Reed, Introduction to Ultra Wideband Communication Systems, Prentice Hall, Englewood Cliffs, NJ, 2005.

[106] A.A. Reeves, Remote Monitoring of patients suffering from early symptoms of Dementia, in: Intl. Workshop on Wearable and Implantable Body Sensor Networks, London, UK, April 2005.

[107] A. Rowe, C. Rosenberg, I. Nourbakhsh, A low cost embedded color vision system, in: Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS), Lausanne, Switzerland, October 2002.

[108] Y. Sankarasubramaniam, I.F. Akyildiz, S.W. McLaughlin, Energy efficiency based packet size optimization in wireless sensor networks, in: Proc. of IEEE Sensor Network Protocols and Applications (SNPA), Anchorage, Alaska, USA, April 2003, pp. 1–8.

[109] C. Santivanez, I. Stavrakakis, Study of various TDMA schemes for wireless networks in the presence of deadlines and overhead, IEEE J. Select. Areas Commun. 17 (7) (1999) 1284–1304.

[110] L. Savidge, H. Lee, H. Aghajan, A. Goldsmith, QoS-based geographic routing for event-driven image sensor networks, in: Proc. of IEEE/CreateNet Intl. Workshop on Broadband Advanced Sensor Networks (BaseNets), Boston, MA, October 2005.

[111] A. Savvides, L. Girod, M. Srivastava, D. Estrin, Localization in sensor networks, in: T.Z.C.S. Raghavendra, K.M. Sivalingam (Eds.), Wireless Sensor Networks, Springer, New York, 2005.

[112] M.V.D. Schaar, S. Shankar, Cross-layer wireless multimedia transmission: Challenges, principles and new paradigms, IEEE Wireless Commun. 12 (4) (2005) 50–58.

[113] R.A. Scholtz, N.Z. Win, Impulse radio: How it works, IEEE Commun. Lett. 2 (2) (1998) 36–38.

[114] C. Schurgers, V. Tsiatsis, S. Ganeriwal, M. Srivastava, Optimizing sensor networks in the energy-latency-density design space, IEEE Trans. Mobile Comput. 1 (1) (2002) 70–80.

[115] C. Schurgers, V. Tsiatsis, S. Ganeriwal, and M. Srivastava, Topology management for sensor networks: Exploiting latency and density, in: Proc. of ACM Intl. Conf. on Mobile Computing and Networking (MobiCom), Atlanta, GA, USA, September 2002.

[116] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, B. Girod, Cross-layer design of ad hoc networks for real-time video streaming, IEEE Wireless Commun. 12 (4) (2005) 59–65.

[117] D. Slepian, J. Wolf, Noiseless coding of correlated information sources, IEEE Trans. Informat. Theory 19 (4) (1973) 471–480.

[118] S. Soro, W.B. Heinzelman, On the coverage problem in video-based wireless sensor networks, in: Proc. of the IEEE Intl. Conf. on Broadband Communications, Networks, and Systems (BroadNets), Boston, MA, USA, October 2005.

[119] F. Stann, J. Heidemann, RMST: Reliable data transport in sensor networks, in: Proc. of IEEE Sensor Network Protocols and Applications (SNPA), Anchorage, Alaska, USA, April 2003, pp. 102–112.

[120] H. Stockdon, R. Holman, Estimation of wave phase speed and nearshore bathymetry from video imagery, J. Geophys. Res. 105 (C9) (2000) 22,015–22,033.

[121] B. Sundararaman, U. Buy, A. Kshemkalyani, Clock synchronization for wireless sensor networks: a survey, Ad Hoc Networks (Elsevier) 3 (3) (2005) 281–323.

[122] K. Sundaresan, V. Anantharaman, H.-Y. Hsieh, R. Sivakumar, ATP: a reliable transport protocol for Ad Hoc networks, IEEE Trans. Mobile Comput. 4 (6) (2005) 588–603.

[123] J. Tourrilhes, Packet frame grouping: Improving IP multimedia performance over CSMA/CA, in: Proc. of IEEE Conf. on Universal Personal Communications, Florence, Italy, October 1998.

[124] T. Vu, D. Reschke, W. Horn, Dynamic packet size mechanism (DPSM) for multimedia in wireless networks, in: Proc. of Multimediale Informations- und Kommunikations systeme (MIK), Erfurt, Germany, September 2002.

[125] M.C. Vuran, O.B. Akan, I.F. Akyildiz, Spatio-temporal correlation: Theory and applications for wireless sensor networks, Comput. Networks (Elsevier) 45 (3) (2004) 245–259.

[126] M.C. Vuran, I.F. Akyildiz, Cross-layer analysis of error control in wireless sensor networks, in: Proc. of IEEE Intl. Conf. on Sensor and Ad-hoc Communications and Networks (SECON), Reston, VA, USA, September 2006.

[127] C.Y. Wan, A.T. Campbell, L. Krishnamurthy, PSFQ: a reliable transport protocol for wireless sensor networks, in: Proc. of ACM Workshop on Wireless Sensor Networks and Applications (WSNA), Atlanta, GA, September 2002.

[128] C.Y. Wan, S.B. Eisenman, A.T. Campbell, CODA: Congestion detection and avoidance in sensor networks, in:

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002

Proc. of the ACM Conf. on Embedded Networked Sensor Systems (SenSys), Los Angeles, CA, November 2003.

[129] M.Z. Win, R.A. Scholtz, Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communication, IEEE Trans. Commun. 48 (4) (2000) 679–689.

[130] A. Wyner, J. Ziv, The rate-distortion function for source coding with side information at the decoder, IEEE Trans. Informat. Theory 22 (January) (1976) 1–10.

[131] Z. Xiong, A.D. Liveris, S. Cheng, Distributed source coding for sensor networks, IEEE Signal Process. Mag. 21 (September) (2004) 80–94.

[132] L. Yang, G.B. Giannakis, Ultra-wideband communications: an idea whose time has come, IEEE Signal Process. Mag. 3 (6) (2004) 26–54.

[133] W. Ye, J. Heidemann, D. Estrin, Medium access control with coordinated, adaptive sleeping for wireless sensor networks, IEEE Trans. Network. 12 (3) (2004) 493–506.

[134] K. Yeung, 802.11a Modeling and MAC enhancements for high speed rate adaptive networks, Technical Report UCLA, 2002.

[135] W. Yu, Z. Sahinoglu, A. Vetro, Energy-efficient JPEG 2000 image transmission over wireless sensor networks, in: Proc. of IEEE Global Communications Conference (GLOBECOM), Dallas, TX, USA, January 2004, pp. 2738–2743.

[136] X. Zhu, B. Girod, Distributed rate allocation for multi-stream video transmission over ad hoc networks, in: Proc. of IEEE Intl. Conf. on Image Processing (ICIP), Genoa, Italy, September 2005, pp. 157–160.

[137] B. Zitová, J. Flusser, Image registration methods: a survey, Image Vis. Comput. 21 (11) (2003) 977–1000.

[138] M. Zuniga, B. Krishnamachari, Integrating future large-scale sensor networks with the Internet, USC Computer Science Technical Report CS 03-792, 2003.

Ian F. Akyildiz received the B.S., M.S., and Ph.D. degrees in Computer Engineering from the University of Erlangen-Nuernberg, Germany, in 1978, 1981 and 1984, respectively.

Currently, he is the Ken Byers Distinguished Chair Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, and Director of Broadband and Wireless Networking Laboratory. He is an Editor-in-Chief of *Computer Networks Journal (Elsevier)* as well as the founding Editor-in-Chief of the *Ad Hoc Network Journal (Elsevier)*. His current research interests are in next generation wireless networks, sensor networks, and wireless mesh networks.

He received the "Don Federico Santa Maria Medal" for his services to the Universidad of Federico Santa Maria, in 1986. From 1989 to 1998, he served as a National Lecturer for ACM and received the ACM Outstanding Distinguished Lecturer Award in 1994. He received the 1997 IEEE Leonard G. Abraham Prize Award (IEEE Communications Society) for his paper entitled "Multimedia Group Synchronization Protocols for Integrated Services Architectures" published in the IEEE Journal of Selected Areas in Communications (JSAC) in January 1996. He received the 2002 IEEE Harry M. Goode Memorial Award (IEEE Computer Society) with the citation "for significant and pioneering contributions to advanced architectures and protocols for wireless and satellite networking". He received the 2003 IEEE Best Tutorial Award (IEEE Communication Society) for his paper entitled "A Survey on Sensor Networks," published in IEEE Communications Magazine, in August 2002. He also received the 2003 ACM Sigmobile Outstanding Contribution Award with the citation "for pioneering contributions in the area of mobility and resource management for wireless communication networks". He received the 2004 Georgia Tech Faculty Research Author Award for his "outstanding record of publications of papers between 1999–2003". He also received the 2005 Distinguished Faculty Achievement Award from School of ECE, Georgia Tech. He has been a Fellow of the Association for Computing Machinery (ACM) and IEEE Fellow since 1996.

Tommaso Melodia received his Laurea and Doctorate degrees in Telecommunications Engineering from the University of Rome "La Sapienza", Rome, Italy, in 2001 and 2005, respectively. He is currently pursuing his Ph.D. in Electrical and Computer Engineering while working as a research assistant at the Broadband and Wireless Networking Laboratory (BWN-Lab), Georgia Institute of Technology, Atlanta, under the guidance of Dr. Ian F. Akyildiz. His main research interests are in wireless sensor and actor networks, wireless multimedia sensor networks, underwater acoustic sensor networks, and in wireless networking in general. He is the recipient of the BWN-Lab Researcher of the Year 2004 award.

Kaushik R. Chowdhury received his B.E. degree in Electronics Engineering with distinction from VJTI, Mumbai University, India, in 2003. He received his M.S. degree in computer science from the University of Cincinnati, OH, in 2006, graduating with the best thesis award. He is currently a Research Assistant in the Broadband and Wireless Networking Laboratory and pursuing his Ph.D. degree at the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA. His current research interests include multichannel medium access protocols, dynamic spectrum management, and resource allocation in wireless multimedia sensor networks. He is a student member of IEEE.

Please cite this article in press as: I.F. Akyildiz et al., A survey on wireless multimedia sensor networks, Comput. Netw. (2006), doi:10.1016/j.comnet.2006.10.002