AnySense: a Video Communication Architecture for Urban Sensing Applications

Weijia Jia, Guoliang Xing, Posco Tso, and Bin Dai
Department of Computer Science
City University of Hong Kong
{itjia,glxing,poscotso,daibin}@cityu.edu.hk

ABSTRACT
Real-time video capture and communication is expected to be an important function in many urban sensor networks composed of camera-equipped mobile phones. In this paper, we present AnySense, a network architecture that supports video communication between 3G phones and Internet hosts in urban sensing applications. AnySense implements transcoding of video streams between the Internet and circuit-switched 3G cellular networks, and is transparent to 3G service providers. A prototype of AnySense has been built and a video clip of system demos is available at http://www.anyserver.org/.

1. INTRODUCTION
The advances of mobile communications and embedded sensing have made it possible to build urban sensor networks composed of mobile phones that collect information in daily urban life [4, 5, 6]. We envision the popularity of a class of video-based urban sensing applications in which camera-equipped mobile phones are used to capture, send and receive *real-time* videos. For instance, a mobile phone user may record events occurring in his proximity and send to friends or broadcast through the Internet in real time. For another example, voluntary mobile phone users may collaborate to collect and broadcast real-time video of city-wide traffic condition during rush hour. In addition to video capture, mobile phones can also be used for receiving videos from cameras connected to the Internet. For instance, a mobile phone user may want to receive real-time video sent from the surveillance cameras installed in his/her home once an abnormal event is detected. In the aforementioned applications, real-time video provides much richer information about the events of interest than other types of data formats such as text messages or static images.

Compared with existing applications on hand-held devices like video downloading and viewing, video-based urban sensing introduces several new challenges. First, in order to capture and publish the video of unpredictable events, mobile phones must have ubiquitous broadband network access, which puts the solutions that rely on short-range wireless networks into question. For instance, although wireless LANs can provide high bandwidth for the smart phones equipped with WiFi interfaces, they usually only cover a small portion of a metropolitan area due to the short range. WiMAX and wireless mesh networks are two emerging technologies that have the potential to provide wide-area high-speed Internet access. However, their deployments are still in a nascent stage. Second, many video-based sensing applications (e.g., instant news coverage) are expected to involve unplanned information collection and substantial numbers of participants. As a result, video capture and communication should be implemented by "out-of-the-box" functions (e.g., video calls) on off-the-shelf mobile phones. For instance, it is undesirable to require all the mobile phones in an application to install additional software or support a special functionality that is only available on the mobile phones from particular manufacturers.

We develop a video communication architecture based on 3rd-generation (3G) mobile technology, called AnySense, to support video-based urban sensing applications. 3G networks are capable of providing both high-speed data and voice/video call services. The number of 3G cellular network deployments has been rapidly increasing worldwide. Video communication between 3G phones and Internet hosts can be achieved by all IP links. However, real-time video streams often suffer from high jitters when transmitted over 3G IP links due to poor link quality and variable bandwidth [7]. As a result, complex software QoS control often needs to be implemented on 3G phones to support IP video calls. The solution of AnySense to this issue is to seamlessly bridge circuit-switched (CS) 3G networks with the Internet. Video data from mobile phones will be carried by CS wireless links with guaranteed bandwidth and built-in QoS control, and are transcoded into IP video streams at a gateway before being sent to Internet hosts. This approach has several advantages compared to the all IP solution. First, due to the built-in CS video call support, no change or installation of new software is needed for a 3G phone to communicate with Internet hosts. This feature is particularly important for in-situ video collection and communication, e.g., instant news coverage, which may involve a large number of participants that carry off-the-shelf 3G phones. Second, connections over CS have built-in support for bandwidth reservation and QoS control, which is key to achieving optimized video quality on resource-limited mobile phones.

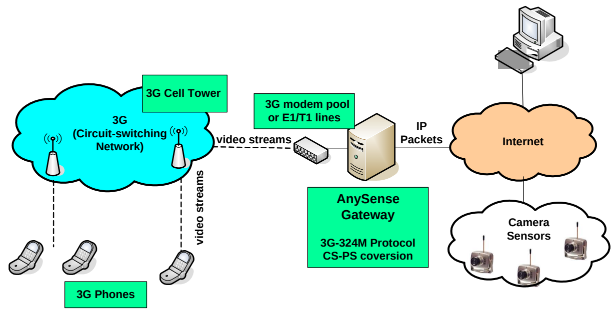

Figure 1: Architecture of AnySense

2. SYSTEM DESCRIPTION

Fig. 1 illustrates the architecture of AnySense. The core component of AnySense is the gateway that interconnects the 3G cellular network and the Internet. When a 3G phone sends a video stream to an Internet host, the 3G phone initiates a video call to the gateway and the video stream is then transmitted over the 3G circuit-switched network. The gateway then converts the stream to IP packets and sends to the host using TCP, UDP or RTP. With an IP address and a phone number, the gateway essentially serves as the proxy for both the 3G phone and the Internet host. We note that the operation of the gateway is transparent to 3G service providers. Furthermore, a 3G phone handles video streams from an Internet host in the same way it handles video calls from 3G phones. Therefore, our architecture supports the video communication between Internet hosts and any off-the-shelf commercial 3G phones.

The gateway implements two important functionality. First, it implements 3G video circuit transmission and internet signaling protocols including 3G-324M [2] and Session Initiation Protocol (SIP) [3] stacks. Second, it performs conversion between 3G circuit switching and IP packet switching. 3G-324M is the International Telecommunication Union (ITU) standard for real-time multimedia services over circuit-switched cellular networks. 3G-324M includes an umbrella protocol, H.324M [1], and several core protocols such as control protocol H.245 and (de)multiplexing protocol H.223. Several key features are defined by 3G-324M to support real-time streaming over circuit-switched networks, including fixed communication delays and low-overhead codecs. H.245 is used by 3G-324M to perform call connection setup and tearing down, capability and codec negotiation, etc. Data (de)multiplexing is performed by H.223. At the multiplexing stage, control and multimedia data (e.g., voice and video) are multiplexed into a bit stream with appropriate delimiters and output to the physical layer (e.g., air interface). At the de-multiplexing stage, multimedia data are extracted from the bit stream received from the physical layer.

We now briefly describe the process of converting circuit switching video streams to IP packets. First, an incoming video call is accepted by the modem pool or T1/E1 interfaces on the gateway. A T1/E1 link is a fiber optic line that can handle 24 call connections. Second, after the connection is established, data breakdown and re-encapsulation are performed on the video streams sent from a 3G phone by H.245 and H.223 protocols. Third, voice and video data are transcoded through efficient media codec such as H.263 and H.264. Finally, Session initialization protocol (SIP) manages the multimedia sessions among the gateway and Internet hosts. SIP has been widely used as signaling protocol for Voice over IP (VoIP). Therefore, AnySense gateway supports the communication between 3G phones and any SIP based VoIP softphones. Depending on network environment, TCP, UDP or RTP is used to transmit data between the gateway and Internet hosts.

3. APPLICATIONS

AnySense supports a number of video applications that involve 3G phone users. We now describe two possible application scenarios.

Real-time video surveillance. In this scenario, camera sensors are installed in the home/office of a 3G user for security surveillance. We assume that a camera sensor is capable of analyzing the videos captured (e.g., detecting human motions). When a camera sensor detects an event, it transmits the real-time video (through a wired/wireless LAN) to a base station, e.g., a PC connected to the Internet. The PC then initiates a UDP connection with the AnySense gateway and sends the real-time video recorded by the camera. The gateway dials the number of the 3G user, establishes a video call connection with the 3G phone, and sends the video stream through 3G circuit-switched cellular network.

Video blogging. In this scenario, 3G users record the events of interest and publish the recorded videos on an Internet webblog server. AnySense allows real-time broadcast of the video being recorded by 3G phones on the blog server. With the support of AnySense, a user's task of video blogging is as simple as making a video call. This key feature encourages the participation of a large number of phone users and the creation of communication-based web blogging.

4. CONCLUSION

In this paper, we present AnySense, a communication architecture that supports video communication between 3G phones and the Internet hosts. AnySense implements transcoding of video streams between the Internet and circuit-switched 3G cellular networks and is transparent to 3G service providers. A class of video-based urban sensing applications can be built with the support of AnySense. We have built a prototype of AnySense. The details of the AnySense project and a video clip of system demos are available at http://www.anyserver.org/.

5. ACKNOWLEDGEMENT

The work described in this paper was partially supported by the City University of Hong Kong under a grant ARD 9668009 and the Research Grants Council of Hong Kong under grants RGC 9041129 and 9041266.

6. REFERENCES

[1] ITU-T Recommendation H.324: Terminal for low bitrate multimedia communication.

[2] 3GPP TR 26.111 V5.0.0, Codec for circuits witched multimedia telephony service; Modifications to H.324. 2002

[3] SIP: Session Initiation Protocol, RFC 3261. June 2002

[4] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory Sensing. In *ACM Sensys World Sensor Web Workshop*, 2006.
[5] Sensorplanet, http://www.sensorplanet.org/.
[6] Mobile landscapes: Graz in real time. http://senseable.mit.edu/projects/graz/.
[7] Eli Orr. Understanding the 3G-324M Spec, Online technical tutorial.