

Propuesta de una Arquitectura en Capas para el Desarrollo de Sistemas Distribuidos de Video-vigilancia

Ignacio Huitzil Velasco¹, Eduardo López Domínguez², Jorge de la Calleja³

^{1,3}Universidad Politécnica de Puebla, Tercer Carril del Ejido "Serrano" s/n San Mateo Cuanalá. Juan C. Bonilla, Puebla, Puebla, México,

²Laboratorio Nacional de Informática Avanzada A.C, Rébsamen 80, esq. Circuito Presidentes Col. Centro Xalapa, Veracruz, México

¹ignacio.h.velasco@gmail.com, ²elopez@lania.mx, ³jdelacalleja@up Puebla.edu.mx

(Paper received on August 10, 2012, accepted on August 24, 2012)

Abstract. In general, the security can be defined as saving the fundamental interests of the human life. In this context, distributed surveillance systems have become an effective tool for providing various security services to people; however the development of these kind of systems include implement various elemental services of computer networks, computer vision, system management, information management, among others. In this work we propose a layer architecture that permits to identify, classify and sort the various necessities services for realizing the design and development of distributed intelligent system of surveillance. In our case, the architecture proposed includes four layers; interconnection, protection, analysis and identification of abnormal events and response. Each of these layers defines the necessary services that should be implemented for the development of third generation systems (intelligent distributed system of surveillance).

Keywords: distributed systems, surveillance system, layer software architecture.

1 Introducción

El término seguridad puede definirse desde distintos puntos de vista y contextos, sin embargo, considerando sólo al que compete al ser humano, se concibe como el salvaguardar sus intereses fundamentales y hasta su propia vida. En la actualidad se han propuesto y desarrollado diversos sistemas de seguridad que emplean alarmas, luz infrarroja, cámaras, sensores, entre otros. En general, los sistemas de seguridad deben considerar y llevar a cabo tres aspectos propuestos por Maloof [1]: el primer es la *protección* que consiste en prevenir hechos indeseables que puedan ocurrir. En segundo lugar la *detección*, ésta se refiere a determinar el momento exacto en el que ocurrió el hecho anormal y que es auxiliado por los mecanismos de protección. Finalmente, después de la detección, la *respuesta* que se refiere a la toma acciones tales como activar alarmas.

Debido a la gran cantidad de información útil que puede ser obtenida y analizada de una secuencia de video, los sistemas de video-vigilancia se han convertido en una herramienta efectiva para brindar servicios de seguridad a las personas en diversos contextos [2, 3, 4]. A los sistemas de video-vigilancia se les puede definir como un sistema de cámaras colocadas en una área (pública o privada en el interior o exterior) con el fin de monitorear y vigilar la actividad que se lleva a cabo [5]. Actualmente, la demanda de este tipo de sistemas es visible en la seguridad de supermercados, edificios inteligentes, en el hogar, el uso militar y gubernamental, entre otros [6, 7].

Por lo tanto, los sistemas de video-vigilancia son un tópico de investigación que permiten su estudio a partir de tres generaciones [4, 8, 9]: la primera generación o *sistemas analógicos* (se basan en Circuitos Cerrados de Televisión CCTV), la segunda generación conocida como *sistemas semiautomáticos* que proveen una mejora tecnológica y de desempeño que permite identificar hechos anormales sobre el video ayudando al operador a realizar esta tarea y finalmente los sistemas de tercera generación los cuales trataremos en este trabajo, también llamados *sistemas inteligentes o automáticos*, que tienen como meta incorporar el enfoque distribuido e incrementar la heterogeneidad. En general, un *sistema inteligente o automático* debe estar compuesto por elementos de: visión por computadora, poseer un control, contar con módulos de almacenamiento y recuperación de evidencias, ser diseñados de manera secuencial y síncrona además de adaptarse en diferentes contextos [8].

Diversos trabajos han propuesto sistemas de video-vigilancia de tercera generación [3, 8,10-13]. En estos trabajos se han propuesto diversas arquitecturas basadas en capas para el desarrollo de sistemas de video-vigilancia de tercera generación; sin embargo, estas propuestas carecen de una estructura y organización adecuada de cada uno de los elementos y servicios que componen a un sistema de tercera generación.

En este artículo se presenta una propuesta de una arquitectura en capas que permite identificar, clasificar, ordenar e integrar los distintos servicios necesarios para llevar a cabo el diseño y desarrollo de sistemas distribuidos de video-vigilancia de tercera generación. La arquitectura propuesta, con base en los aspectos definidos por Maloof [1], está organizada en cuatro capas: *interconexión, protección, análisis e identificación de sucesos anormales y respuesta*. 1) La *capa de interconexión* suministra los servicios necesarios para la comunicación entre los diversos dispositivos y los servicios de seguridad, 2) La *capa de protección* provee el servicio de calibración de cámaras y la transmisión del flujo de video a los clientes finales con el fin de que éstos puedan visualizar lo que sucede en el área vigilada, 3) La *capa de análisis e identificación de sucesos anormales* proporciona los servicios de sensado y análisis del video mediante algoritmos o hardware especial con el fin de detectar, reconocer y rastrear situaciones anormales, además de almacenar y recuperar evidencias, y finalmente 4) La *capa de respuesta* ofrece el servicio de alarmas una vez identificado el hecho anormal.

El artículo se encuentra organizado de la siguiente manera. En la sección 2 se describe el estado del arte. Se define a detalle la propuesta de este trabajo en la sección 3. En la sección 4 se describe de manera breve la implementación de un sistema distribuido móvil que cumple con las primeras dos capas que se proponen y por último, la sección 5 presenta las conclusiones y el trabajo a futuro.

2 Estado del Arte

Diversos trabajos [8, 10-13] proponen una arquitectura en capas para la implementación de sistemas distribuidos de video-vigilancia de tercera generación. A continuación se describen estos trabajos:

Un sistema de video-vigilancia móvil en tiempo real es propuesto por Wang C. et al. (2003) en [10], donde el usuario final puede monitorear una región y controlar las cámaras desde un asistente personal digital (PDA). Este trabajo propone dos capas para su desarrollo:

- 1- *Capa de arquitectura general*: Constituida por los módulos de cámaras, servidor principal de control, terminales de monitoreo y componentes de una LAN inalámbrica. La tarea principal de esta capa consiste en captar el video del sistema de cámaras y transmitirlo a un servidor central donde se visualiza a través de monitores, además de codificar, almacenar y transmitirlo a las terminales móviles.
- 2- *Capa de video-vigilancia móvil*: Integrada por los módulos de: recepción de video para un móvil PDA, transmisión y codificación de video, protocolos de transporte de red, multicast IP, transmisión y recepción de señales de control de cámaras remotas y funciones de la terminal móvil PDA.

La arquitectura propuesta por [10] no cuenta con el aspecto de *protección* (identificar hechos anormales y realizar el almacenamiento de las evidencias) y el aspecto de *respuesta* (alarmas) de acuerdo al ciclo de seguridad de [1].

Por otra parte, MASCOT es un método de diseño para implementar sistemas concurrentes en tiempo real de gran escala que proponen Valera M. y Velatin S.A. (2004) en [8]. Este método incluye el diseño de un sistema de video-vigilancia distribuido e inteligente formado por dos módulos: Módulo de Procesamiento de Datos (MPD) y el Módulo de Control (MC). El módulo MPD se encarga de las funciones de adquisición de video, procesamiento de imágenes y el almacenamiento de evidencias (base de datos). El módulo MC tiene dos tareas importantes: controlar las funciones de MPD y mostrar lo monitoreado al usuario final. Este método omite los aspectos de la seguridad, alarmas y el control de las cámaras.

Mientras tanto, una arquitectura de red inteligente llamada FACET es definida por Bolliger P., Köhler M. y Römer K. (2007) en [13]. La base de esta propuesta es el uso de teléfonos celulares con cámaras. Esta red inteligente tiene la capacidad de detectar movimiento de las zonas monitoreadas mediante el análisis de eventos (sucesos) que indican la entrada y salida de intrusos. Estos eventos son analizados y procesados en cada uno de los teléfonos celulares (llamados eventos distribuidos). La arquitectura FACET se encuentra estructurada por los módulos: cámaras, captura y análisis de imágenes, tiempo de sincronización, sistema de eventos distribuidos, red de comunicación y calibración. El módulo de análisis de imágenes, utiliza el método de substracción de *background* una técnica de segmentación para detección de movimiento. Sin embargo, esta arquitectura carece de los servicios que permiten almacenar los flujos de videos, control de las cámaras y de alarmas para advertir sobre la existencia de movimiento a los usuarios finales.

Por último, el desarrollo de un sistema de video-vigilancia a bajo costo en

hardware es propuesto por Xu L. et al. (2008) y Aswin S. et al. (2009) en [11, 12], integrando teléfonos celulares con cámaras que procesan un algoritmo de substracción de *background* para detectar movimiento. Ambos trabajos incluyen un modelo de tres capas: 1) la adquisición de video a partir de los teléfonos celulares, 2) la detección de movimiento y 3) el envío de alarmas (llamada telefónica o mensaje de texto SMS) que consiste en notificar al usuario final la activación de la capa dos. Este sistema carece de los servicios de monitoreo en tiempo real para usuarios finales, del almacenamiento de evidencias, la calibración y control de cámaras.

3 Arquitectura en Capas

El presente trabajo propone una arquitectura en capas para el desarrollo de sistemas distribuidos de video-vigilancia inteligentes. La arquitectura propuesta está estructurada en capas. Cada capa proporciona ciertos servicios a las capas superiores, ocultando los detalles de los servicios ofrecidos por las capas inferiores. Para el diseño de la arquitectura se consideraron principalmente dos aspectos: las fases del ciclo de seguridad definidas en [1] y los servicios que un sistema de video-vigilancia de tercera generación debe cumplir [3,6-8].

La arquitectura en capas de este trabajo tomó como referencia de diseño a los trabajos de [3, 8, 11, 12].

La arquitectura propuesta en este estudio está formada por cuatro capas: interconexión, protección, análisis e identificación de sucesos anormales y respuesta, que se encuentran colocadas de manera ascendente, ver Figura 1. A continuación se describe cada una de estas capas y los módulos que las integran.

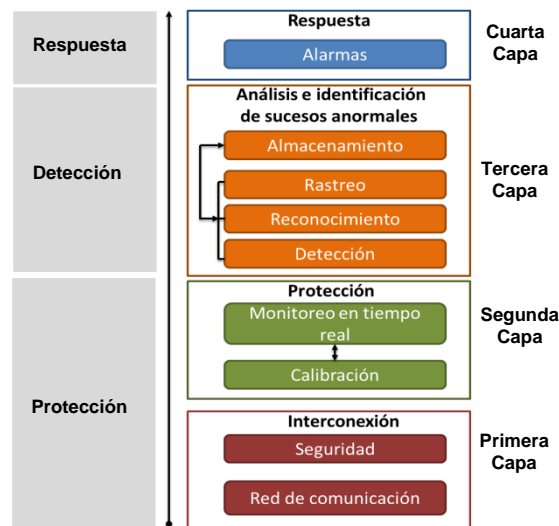


Fig. 1. Arquitectura de software en capas para sistemas de video-vigilancia de tercera generación, (los rectángulos en gris representan las fases del ciclo de seguridad [1]).

3.1 Capa de Interconexión

Es la capa base debido a que el servicio que provee consiste en lograr la comunicación de todos los elementos que intervienen con el sistema de video-vigilancia. Esta capa está integrada por los módulos de red de comunicación y de seguridad, los cuales son descritos enseguida.

Módulo Red de Comunicación.- El módulo considera aquellos elementos físicos y lógicos que intervienen en la comunicación para satisfacer los objetivos que propone el sistema de video-vigilancia y es necesario definir los elementos y características de cada uno de ellos. Por ejemplo en los elementos físicos el número y tipo de cámaras (IP, *Webcam*, celulares, etc.), los dispositivos de red (*routers*, *switches*, servidores), la red (cableada o inalámbrica), los dispositivos o terminales de usuario final (computadoras, móviles, televisores). Por otra parte, en los elementos lógicos, considerar el modelo de conexión y comunicación entre los elementos físicos (modelo cliente-servidor, peer to peer, entre otros) y los protocolos de comunicación tales como TCP/IP o UDP, RTSP, entre otros.

Módulo Seguridad.- El propósito de este módulo se refiere a la seguridad del sistema en su totalidad comprendiendo los aspectos de: confidencialidad, integridad, autenticidad, disponibilidad de los recursos y de la información, consistencia, control de acceso a los recursos.

3.2 Capa Protección

El objetivo de la capa de protección es brindar el servicio de monitoreo (visualizar lo que ocurre actualmente en un área) y el control remoto de las cámaras en tiempo real, a un número determinado de usuarios. En esta capa es necesario definir en primera instancia el tipo de ambiente donde se colocarán las cámaras, en un ambiente controlado (debe cumplir con ciertas características) o dinámico (no hay restricciones) o combinación de éstos. El definir el tipo de ambiente ayudará a implementar con éxito los dos módulos que conforman esta capa: el módulo de calibración y el módulo de monitoreo que son descritos enseguida:

Módulo de Calibración.- Consiste en determinar los parámetros *internos* (distancia focal, factores de distorsión y puntos centrales del plano imagen) y los parámetros *externos* (posición y orientación) de una cámara con el fin de realizar las tareas de monitoreo y posteriormente el análisis de video [14, 15].

Para la calibración de la cámara se determina la geometría, los parámetros *internos* y *externos* que normalmente son calculados de dos formas: 1) mediante un patrón de calibración (marco de referencia del mundo) y 2) mediante un método de autocalibración, el cual se basa en el movimiento de la cámara observado una escena estática y a partir de su desplazamiento, utiliza la información de la imagen para obtener los parámetros. Existen diferentes técnicas y algoritmos de calibración que pueden emplearse en este módulo, para más detalles consulte [15].

Módulo Monitoreo en Tiempo Real.- Este módulo ofrece a los usuarios remotos la capacidad de visualizar el video captado por las cámaras (un número determinado de cámaras y de diferentes tipos) en tiempo real. Para lograr este servicio, se debe establecer la manera en que las cámaras capturen el flujo de video y lo codifiquen mediante un *encoder* (convierte y lo comprime a un nuevo formato de video), para

que posteriormente sea transmitido y visualizado por el cliente. El servicio de monitoreo le permite al usuario acceder en cualquier instante de tiempo que él disponga para supervisar las zonas.

Dentro de este módulo también se incluye el aspecto de manipular las cámaras de manera remota, de tal forma que el usuario tiene el control de las características propias de las cámaras como ampliar o reducir la visualización de la zona (*zoom*), encender luces, sirenas, sistema de voz, cambiar la posición y orientación.

3.3 Capa de Análisis e Identificación de Sucesos Anormales

Consiste en ofrecer el servicio de identificación de sucesos anormales ocurridos en el área monitoreada y del almacenamiento de evidencias (fotos, videos) de estos hechos y de las grabaciones realizadas por el cliente. Esta capa consta de cuatro módulos: detección, reconocimiento, seguimiento de objetos y el almacenamiento.

Los módulos detección, reconocimiento y seguimiento de objetos se encargan de medir las variaciones en el ambiente e identificar situaciones anormales. Dos formas en las que pueden desempeñar estas tareas son: 1) mediante el análisis o tratamiento de imágenes del flujo de video captada por las cámaras (ejemplo algoritmos) y 2) mediante dispositivos electrónicos como los sensores de movimiento. Cada uno de estos módulos tiene un objetivo en particular que se describe a continuación.

Módulo de Detección.- Consiste en localizar situaciones anormales tales como: detectar humo, cambios de temperatura, movimiento (algoritmo *motion template*, entre otros [14]).

Módulo de Reconocimiento.- La función de este módulo es exclusivamente definir o identificar de qué situación anormal se trata, por ejemplo el reconocimiento de objetos animados e inanimados en particular (objetos olvidados, personas) así también, de comportamientos sospechosos de los humanos como hechos violentos y conglomeraciones.

Módulo de Rastreo.- El objetivo de este módulo consiste en que una vez identificado el foco del suceso anormal éste no debe perderse de vista, ejemplo un objeto sospechoso en movimiento.

Módulo de Almacenamiento.- La tarea de este módulo consiste en realizar la grabación de evidencias (fotografía, video y variaciones en el ambiente) y el registro en una base de datos de tipo relacional o distribuida, con el fin de hacer un futuro análisis de las evidencias.

Concluida la grabación y el registro en la base de datos, el sistema tiene la capacidad de proveer a los usuarios finales, los servicios de uso y administración del material multimedia (recuperación de evidencias).

3.4 Capa de Respuesta

Es la última capa de esta arquitectura, la cual proporciona el servicio de informar o advertir al usuario final sobre la existencia de un hecho anormal, que fue identificado sobre el área vigilada. La capa de respuesta esta integrada por el módulo de alarmas que se describe enseguida.

Módulo de Alarmas.- En este módulo se incluyen aquellos elementos o dispositivos que permitan advertir sobre la situación anormal al usuario final. Algunos ejemplos de

alarmas son: envío de correos electrónicos, llamadas telefónicas, mensaje de texto (SMS) y multimedia (MMS) que hacen mención de lo identificado, sirenas, luces, entre otros. Dentro del módulo de alarmas pueden ser incluidos opcionalmente elementos que permitan realizar acciones para tratar de resolver la situación anormal por ejemplo, un detector de humo advierte sobre un incendio y activa un sistema contraincendios.

4 Resultados Preliminares

En esta sección se presenta un prototipo inicial de un sistema distribuido de video-vigilancia que implementa las capas de interconexión y protección. La implementación e integración de las capas de análisis e identificación de sucesos anormales y respuesta se encuentran en desarrollo.

Capa de interconexión

Los resultados del módulo de red de comunicación comprenden a los elementos físicos y lógicos. Los elementos definidos para el sistema distribuido móvil son:

Elementos físicos: 3 cámaras (Dos IP con resolución de 320 x 240 y un dispositivo móvil con cámara de 5MP resolución de 2560 x 1920), 1 servidor de *Streaming*, Red LAN (alámbrica e inalámbrica) y un dispositivo móvil Smartphone con sistema operativo Android 2.1 o superior.

Elementos lógicos: Protocolo de comunicación RTSP (modelo cliente servidor) y Software libre para codificar el flujo de video (*encoder*).

El módulo de seguridad fue implementado del lado del servidor de *Streaming*, que consiste en un módulo de autenticación de usuario (en Java) que permite el acceso al sistema y se definen permisos correspondientes al cliente o al administrador.

Capa de protección

Para el módulo de calibración se estipularon solo los parámetros externos de la cámara definiendo la posición, orientación y el ángulo de visión, colocadas en el interior de un edificio con un ambiente controlado con una luminosidad de 30 a 60 watts.

El módulo de monitoreo en tiempo real se realiza conforme a lo siguiente: Primero, las *cámaras* captan el video de la zona de vigilancia y transmiten este flujo de video al *encoder*. Posteriormente, el *encoder* realiza la conversión y compresión del video a formato H.264 además de cumplir con las características para dispositivos móviles. Una vez tratado el flujo de video por el *encoder* se envía al servidor de *Streaming* que permite al usuario autenticado consultar o visualizar el video en tiempo real desde un dispositivo móvil Android 2.1. Sobre el dispositivo móvil se desarrollo una aplicación que permite la autenticación del usuario, la selección de la cámara y el monitoreo del área vigilada de manera remota en tiempo real.

5 Conclusiones y Trabajo Futuro

El diseño de una arquitectura en capas es indispensable para el desarrollo de sistemas de video-vigilancia en especial de la tercera generación. En este trabajo se ha

propuesto una arquitectura de cuatro capas: interconexión, protección, análisis e identificación de sucesos anormales y respuesta. La arquitectura propuesta permite identificar, clasificar, ordenar e integrar los distintos servicios necesarios para llevar a cabo el diseño y desarrollo de sistemas distribuidos de video-vigilancia de tercera generación. Una de las ventajas de esta arquitectura es la flexibilidad de integrar o extender nuevos módulos con nuevos servicios de acuerdo las necesidades y contextos de vigilancia que se requieran. Como trabajo futuro se implementarán las capas de análisis e identificación de sucesos anormales y respuesta al sistema distribuido de video-vigilancia móvil, así también las pruebas correspondientes.

Referencias

1. Maloof A.M. Machine Learning and Data Mining for Computer Security. Springer, United States of America, (2006)
2. Gonzalez G. Sistema de vigilancia ip vía teléfono móvil, (2001). www.rnds.com.ar, www.flexisoft.com
3. San Miguel J.C, Bescós J., Martínez J., y García A. Diva: A distributed video analysis framework applied to video-surveillance systems. IEEE computer society. Ninth International Workshop on Image Analysis for Multimedia Interactive Services, pages 207–210, (2003).
4. Kurman P., Miltan A., y Kurman P. Study of robust and intelligence in visible and multi-modal framework. Informatics (Slovenia), pages 63–77, (2008).
5. Davis K., Kelsey J., Langellier D., Mapes M., y Rosendahl J. Security cameras, (2003)]. CTER Program at UIUC. <http://students.ed.uiuc.edu/jkelsey/surveillance>
6. Song H., Feng X., Zhao Q., Chi Y., y Yeng H. Mobil video surveillance system of 3g network based on arm9. IEEE e ICCASM (International Conference on Computer Application and System Modeling), pages 400–403,(2010)
7. Zang Q. y Klette R. Object classification and tracking in video surveillance. In Lecture Notes in Computer Science, volume 2756, pages 198–205. Springer Berlin, (2003)
8. Valera M. y Velatin S.A. Real-time architecture for a large distributed surveillance system. IEEE Intelligent Distributed Surveillance Systems, pages 41–45, (2004)
9. Valera M. and Valastin S. A. Intelligent distributed surveillance systems: a review. IEEE Processing Vision, image and signal. Image signal Process, 152(2), 192-204 , (2005)
10. Wang C., Mao Z., Zhnag Y. y Luo H. The mobile video surveillance system based on wireless LAN [ol], (2003)
11. Xu L., Wang Z., Wang H., Shi A. y Li C. A J2ME based intelligent video surveillance system using moving object recognition technology. Congress on Image and Signal Processing CISP. 2008, 2:281–285, (2008)
12. Aswin S., Snthiyan Sethuram A., Varun A. y Vasanth P. A 2ME based wireless automated video surveillance system using motion detection method. Reporte técnico, Department of Electronics and Communication Engineering, AMRITA School of Engineering, Coimbatore, India, (2009)
13. Bolliger P., Köhler M. y Römer K. Facet: Towards a smart camera network of mobile phones. Proceedings of the 1st ACM International Conference on Autonomic Computing and Communication Systems (Autonomics '07), (2007).
14. Bradski G. and Kaehler A. Learning OpenCV. O'reilly, United States of America, (2008)
15. Aracena P. D., Campos P. y Tozzi L. C. Comparación de técnicas de calibración de cámaras digitales Revista Facultad de Ingeniería Universidad de Tarapacá, Arica Chile, vol.13, n°1, pp. 57-67, (2005)