# Video911: A Personal Video Security System

Turadg Aleahmad

Dept. of Computer Science
University of California, Berkeley
Berkeley, CA USA
turadg@csua.berkeley.edu

## 1 Abstract

The proliferation of high-bandwidth mobile connectivity, affords many new applications for mobile communications devices. This paper describes an application for mobile phones with video input for personal and community security.

## 2 The problem

A society is nothing without rules and a means to enforce them. From the dawn of man and his club, people have employed myriad technologies to this end. Though these technologies have evolved greatly, their essential function is to deter parties from committing crimes. Let us evaluate two deterrence technologies that enjoy contemporary success.

### 2.1 Guns

Patty is walking through a dangerous quarter of town. The area is poor and the city government has had no interest in installing surveillance cameras. Our poor perambulating Patty is alone and becomes scared as she notices that an ominous man has spotted her and is rushing towards her. She draws her handgun from her purse and unlatches the safety. He reaches for her; she turns and shoots. He just wanted to ask her if he could help her find her way.

This passage is unabashedly melodramatic, but the point is irrefutable: guns cause violence. Whether used for attack or self-defense, their purpose is to cause bodily harm. When a civilian buys a gun, it means one more gun in use. As of April 1999, there were more licensed gun dealers in America than there are McDonald's franchises. [1] A zealous campaign to cast handguns as ideal self-defense weapons "has driven handguns from a mere eight percent of firearms offered for sale in the civilian market in 1946 to 54 percent in 1994."[2]

Studies show that using a handgun in self-defense is more dangerous than nothing at all.[3] Violence, even "justified", is violence and begets more violence. The American gun mythos perpetuates the cult of violence and alternatives must be developed.

## 2.2 CCTV

Closed Circuit Television is marketed as an alternative. It is exceedingly widespread in England despite public opposition. It succeeds because it reduces crime at each installation. Opponents argue that it's Big Brother and point out that it merely moves crime from surveyed spaces to non-surveyed ones[4]. To remain effective, it must continue to expand, until there is non-stop ubiquitous surveillance.

One would think that such a system would never succeed here, but cities around the country are installing cameras at busy intersections and monitoring for fugitive faces. Every attendee of the last Super Bowl had their face captured and analyzed. Government monitoring of its citizens whole lives is antithetic to democracy. A decentralized alternative must be developed.

---

[1] http://www.bradycampaign.org/facts/issuebriefs/preventing.asp
[2] http://www.vpc.org/studies/uninsum.htm
[3] http://www.vpc.org/studies/unincont.htm
4 http://www.privacy.org/pi/issues/cctv/cctv_faq.html

# 3 The Alternative

The goal of CITRIS is to employ technology in the service of society. I argue that a new deterrence technology can be developed as an alternative to guns and CCTV. It is based in two enabling technologies that are posed for rapid growth.

Mobile networks are entering their third generation (3G). European telecom giants have bet billions on this technology so it is guaranteed to be deployed and marketed zealously. Already available in Europe and Japan, it's beginning to arrive in the US. While 2G throughput is 10kb/sec and 2.5G is 64-144kb/sec, 3G is up to 2000kb/sec (384kb/sec on foot and 144kb/sec driving).[5]

Higher throughput enables richer data. Mobile phone manufacturers now offer cameras add-ons. These should soon be standard on high-end models and eventually ubiquitous.

So how do 3G and cameras supplant guns and CCTV? I call it Video911. It offers the personal empowerment of a gun and the accountability of CCTV but minus the violence and authoritarianism.

# 4 Scenario

Patty is again walking alone through a dangerous quarter of town, this time equipped with Video911 on her mobile phone. She notices an equally ominous figure approaching and decides to activate Video911. She does this by flipping a safety cover over the control button and pressing down. The phone initiates a session with the security service to which she subscribes and begins transmitting her GPS location and video camera stream. She lifts her phone so that the camera lens is facing her suspected assailant. As he nears, she informs him that his face is recorded already in a database tied

[5] http://www.3gnewsroom.com/html/what_is_3g/index.shtml

to this location and that all evidence will be used against him. He looks puzzled and offended, then explains that she looked lost and was trying to help. Patty is disappointed in herself for being so suspicious but she is relieved to have employed Video911 instead of a gun or mace. With her finger still firmly on the security button, she types a code into the phone to cancel the session and indicate that she is okay. Because Patty has the opted for the premium tier service, an operator calls her phone to confirm.

Sam the Samaritan, still flustered, demands that she tell the operator to erase the recording. She explains that although the system did record him, it is inaccessible. "I asked about that myself and they told me that the recording is encrypted until I go to emergency level." "What's emergency level?" Sam asks.

At that moment, she sees a third ominous figure running towards them. She presses the activation button to arm her mobile phone and initiates another session. The camera records as Craig the Criminal raises his gun and demands their wallets. Patty informs him that the camera on her mobile phone has recorded his face and that he will be caught if he doesn't leave. He grabs the phone from her, which would have caused the release of the security button, had she not already released it upon sight of the gun. Upon release of the button, the session entered emergency mode and sent the decryption key to the security service. When the session began, they alerted the nearest patrol car to her location. Now they send the officers notice of an altercation and an image of the assailant's face. Craig smashes the phone to the ground and laughs, "There's your recording." Sam replies, "Actually, the recording is safe and sound at the police station." "And they'll be here any second," Patty adds.

Craig turns to flee but runs into the hood of the screeching police car. The officers get out and handcuff him immediately. “We’ve got you now, Gil.” “How do you know who he is?” poses Patty. “The system ran his face through the criminal database. This is Gil Bates. We’ve been trying to prosecute him for a long time but he’s got a brigade of lawyers who keep him free. Thanks to you and Video911, he won’t be monopolizing these streets any longer.” Patty nods accordingly, but thinks to herself, “Funny, I thought his name was Craig for some reason.”

# 5 Architecture

Implementing the design described above requires various components. I will describe how I implemented them in my prototype and how they might be implemented in a production system.

- Client Application
- Messaging
- Media streaming
- Switchboard Application
- Security
- Privacy
- Face Matching
- Face Detection

## 5.1 Client Application


In mobile applications, the choice of runtime environment is crucial. My prototype was written for a laptop and I chose Java 2 Standard Edition (J2SE) with Java Media Framework (JMF). With small modifications, my client code can be ported to Java 2 Mobile Edition (J2ME) and run directly on mobile phones available for sale today. But there is nothing tying the application to Java. It can

run also run on any platform (e.g. Qualcomm's BREW) that provides the necessary media services. Prototyping on a laptop, I was not able to acquire GPS data, but I expect that new mobile phone models provide an API for accessing this information. **NOTE ON MODELING THE UI BUTTONS**

## 5.2 Messaging

For expediency I chose Java Remote Method Invocation (RMI) for the prototype. A production implementation would use a platform-neutral messaging protocol such as XML-RPC or SOAP. The messaging model I designed goes strictly from client to server, but could also go vice-versa if necessary. The methods include:

`createSession()` takes an stream source identifier and returns a session identifier
`setUrgency()` takes a session identifier and a new urgency level
`actionPerformed()` takes a session identifier and an action code (action codes currently include: `CANCELED` and `LOST_CONTROL`)
`ping()` takes nothing and always returns true

The client pings periodically to test the condition of the server link. It can also use this information to pick the best among a list of servers. When the user activates recording, it invokes `createSession()` on the server object with the URI to receive its video stream and stores the session identifier. It sends the `CANCELED` action when the user disarms the device and `LOST_CONTROL` when the button is depressed without first being disarmed.

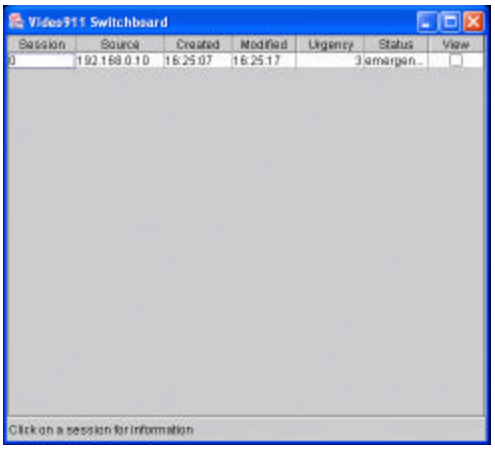
## 5.3 Media Streaming

For the prototype, it was important to stick as close to a possible production implementation in order to evaluate the transmission quality in real-world use. For compression, I chose the h.263 (YUV model) because it is an open standard (RFC2190, RFC2429) and designed for 64-128kbps. For the transport protocol I chose RTP because

it is also an open standard (RFC1889, RFC2198) and is designed specifically for real-time transmission. When the client initiates the session, it provides the server with a URI of the form rtp://client-ip/video. The server detects the codec upon initiation of reception. While I believe H263/RTP is the best streaming combination, the openness of this design allows any server to receive streams with diverse codecs and transport protocols.

## 5.4 Switchboard Application

It receives the session messages from the client, records the stream, and presents the session information to the operator. I wrote my prototype in J2SE with JMF (at right). It presents the operator with the session serial number, the source IP, the time of session initiation, the time of the last message received, the urgency level, the current status, and a checkbox to view the stream.




A production system will require careful research of the user interface to maximize the efficiency and efficacy of the operator. Since efficiency is key, it will also have to be coded in a higher performance language than Java. But because the client-server interface is made exclusively of open standards, it will be easy to write different server implementations and compare how they perform.

## 5.5 Security

A secure communications channel is necessary for several reasons. 1) To prevent modification of the data stream. If it is to be used as evidence in court, it must be reliable and tamper-proof. 2) To authenticate the server. The user's stream should only go be allowed to transmit to a server that they trust. 3) To authenticate the client. This makes

abusers of the system accountable. It also enables accounting, a means for service providers to restrict access only to paying clients. 4) To encrypt the data. This is the least necessary, but easily achievable given 1-3. This will preclude a new generation of (vide) police scanners.

All of the above requirements are provided for by a public-key infrastructure (PKI). A user begins a subscription with a service provider and sends them their public key. The service provider then sends its public key to the user. Now they each have a means of authenticating the other. As a user roams, the client software receives messages from the service provider to switch to closer servers. The certificate of each proposed server is verified against the root certificate authority.

I did not include this functionality in my prototype. My research suggests that the best implementation is the use of secure sockets layer (SSL) in both directions, but this is an open part of the design.

## 5.6 Privacy

Thousands of recordings in a centralized database will surely receive the attention of the ACLU. To address their concerns, the client application could encrypt the stream using private-key encryption (over the already secure channel). The client would prevent the server from viewing the stream by encrypting it with a randomly generated private key. When the client enters the emergency state, it sends the key to the server, allowing it to decrypt the incoming stream and the already-recorded stream. This means that if the client cancels the session before emergency state, it is as if it never sent anything at all. But if it does enter emergency state, the recording preceding the emergency event is still available to law enforcement.

Another privacy feature would be expiration times for records. After a certain amount of time, the server would be required to purge the recording. Whether this would be necessary is a legal question.

Neither of these features was implemented in my prototype. Encrypting the video would require a stream cipher such as RC4. Purging expired records is trivial.

### 5.7 Face Matching

Face matching requires both matching algorithms and a large database of faces to match against. Since many systems exist already to do this, it's merely a matter of connecting to the chosen provider. This feature was not implemented.

### 5.8 Face Acquisition

Since the face is the most useful personal identifier, it would help if the client software were able to focus on it somehow. If the user is too far from the assailant, the face may comprise only a small part of the picture and if it's moving quickly it will be blurry. The client software could be designed to detect the region of the image that is a face and send high-resolution snapshots separate from the main video stream.

This does little to help detect a face in the complete dark. This must be addressed at the hardware level with infrared imaging. Mobile phone manufacturers will surely add this to their models if there is enough demand.

## 6 Criticisms

*There's not enough mobile bandwidth for this.*

The codec chosen (h.263) is designed for 64kbps ISDN throughput. 2.5G promises 64-144kbps and 3G up to 384kbps for pedestrian traffic. AT&T has deployed GPRS, which is classified as 2.5G, and Sprint is deploying 1xRTT, which is classified as

3G. Not to mention the major deployment of 3G in Europe and Japan, nor the use of more efficient codecs.

*There's not enough Internet bandwidth for this.*

This is a bit ridiculous since Internet bandwidth is a commodity. Furthermore, 100 streams at 64kbps require 6.4mbps, which is less than 10base-T. Assuming that at peak 2% of clients are simultaneously in distress, a user base of 5,000 can be served from a dorm room.

*Mobile phones can't run it.*

Mobile phones are increasingly powerful and many provide J2ME. Several models also include video camera options. In fact, Sprint's new Vision service sells[6] this exact combination for use on its 3G network[7].

*It requires too much infrastructure.*

All it requires is already available: a mobile phone with a video camera that can communicate over the Internet. All a service provider needs to do is put up a server and provide the user with the client software. It doesn't need to be involved with law enforcement bureaucracy more than calling them when they've detected a crime. It is akin to home security alarm service providers. Someday it may be integrated with the real 911 system, but this is not necessary to begin.

*It can't do anything in the dark.*

A gun isn't very useful if you can't see either. It can work in the dark at least as well at CCTV and even better with an infrared camera

---

[6] http://www.pcsvision.com/pictures.html
[7] http://industry.java.sun.com/javanews/stories/story2/0,1072,47400,00.html

*It will make us into a nation of paranoid snitches.*

Though some may go around recording everything, isn't Little Brother better than Big Brother? Privacy is important, but documenting crime is indubitably beneficial. With this system, the individual makes the choice to record rather than a face-less government agency. Further privacy measures can be implemented as described in section 5.6.

# 7 Conclusion

The Video911 system could prevent crime. Criminals would be deterred by the specter of an incontrovertible video record of their actions. Guns escalate risk of violence, but Video911 is purely defensive. CCTV actualizes the omnipresent eye of Big Brother, but Video911 lets citizens protect themselves wherever and whenever they choose.

I've presented a detailed design and basic prototype. The design is based in practical, well-tested technologies and will be straightforward to implement. Further research should test the effectiveness of the system in real-world situations. With those results, potential service providers can evaluate the economics and viability of full-scale deployment. Perhaps your next mobile phone will come with Video911™.