

# Covert Channel for Improving VoIP Security

Wojciech Mazurczyk, Zbigniew Kotulski

<sup>1</sup> Warsaw University of Technology, Faculty of Electronics and Information Technology, Institute of Telecommunications 15/19 Nowowiejska Str.  
00-665 Warszawa, Poland

{W.Mazurczyk Z.Kotulski}@tele.pw.edu.pl

<sup>2</sup> Polish Academy of Sciences, Institute of Fundamental Technological Research  
zkotulsk@ippt.gov.pl

**Abstract.** In this paper a new way of exchanging data for Voice over Internet Protocol (VoIP) service is presented. With use of audio watermarking and network steganography techniques we achieve a covert channel which can be used for different purposes e.g. to improve IP Telephony signaling protocol's security or to alternate existing protocols like RTCP (Real-Time Control Protocol). In this paper we focus on improving VoIP security. The main advantage of this solution is that it is lightweight (it does not consume any transmission bandwidth) and the data sent is inseparably bound to the voice content.

## 1. Introduction

Nowadays VoIP (Voice over Internet Protocol) is one of the most popular communication technology designed for IP networks. Although there are many standards proposed (SIP, H.323) there are two fields in which IP Telephony is lacking. The first one is providing certain Quality of Service (QoS) parameters (e.g. low end-to-end latency, packet loss) and the second are security considerations as described in [3]. The latter causes to seek for a new approaches. In [16] we proposed to improve signalling protocol's and conversation's security with digital watermarking technique, while in [15] to alternate RTCP (Real-Time Control Protocol) [5] functionality for VoIP RTP digital streams with digital watermarking and network steganography. In this paper we would like to present the possibilities that covert channels in IP Telephony can offer. We would like also to focus more on improving VoIP security. As in [15] we will use two information hiding techniques, mentioned above. In this way we gain important advantages such as: verification of the transmission's source and the content sent (both authentication and integrity services). Additionally, this solution is lightweight, and does not consume transmission bandwidth, because the control bits (a header of the new, proposed protocol) are transmitted in a covert (steganographic) channel and appropriate, the protocol data is inseparably bound to the voice content as a digital watermark.

The paper is organized as follows. In Section 2 both techniques, digital watermarking and steganography, are described. Next, we give details about proposed solution in Section 3. Finally, we sum up with conclusions in Section 4.

## 2. Steganography and Digital Watermarking

Steganography and Digital Watermarking are Information Hiding subdisciplines [9]. The general difference between those two techniques is that the steganography’s aim is to keep the existence of the information secret whereas the watermarking aim is to make it imperceptible.

### 2.1 Steganography: a covert channel

Steganography is a process of hiding secret data inside other, normally transmitted data. Usually it means hiding a secret message within an ordinary message and its extraction at the destination point. In ideal situation, anyone scanning data will fail to know it contains covert data. In modern digital steganography, data is inserted into redundant (provided but often unneeded) data, e.g. fields in communication protocols, graphic image, etc. TCP/IP (or network) steganography utilizes the fact that few headers in the packet are changed during transit ([9], [7], [8], [10]).

In this paper we will exploit a covert channel, which is a method of communication that is not a part of an actual computer system design, but can be used to transfer information to users or system processes that normally would not be allowed to access the information. In TCP/IP stack, there is a number of methods available, whereby covert channels can be established and data can be exchanged secretly between communication parties. An analysis of the headers of typical TCP/IP protocols e.g. IP, UDP, TCP, HTTP, ICMP results in fields that are either unused or optional [8]. This reveals many possibilities where data can be stored and transmitted. As described in [7] the IP header possesses fields that are available to be used as a covert channel. Those fields are marked in Figure 1 with italics. The total capacity of those fields exceeds 60 bits per packet. And there are potentially UDP and RTP protocol’s fields left that can be also used for this purpose.

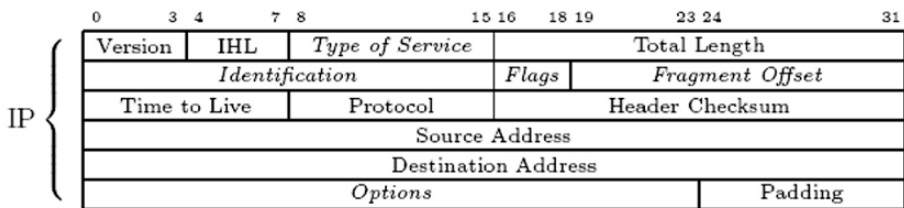


Fig. 1. The IP header with marked fields (italics) available for network steganography [7]

Furthermore, we can distribute those bits that we want to transmit among protocol header’s fields in a predetermined fashion (this pattern can be exchanged during a signalling phase of conversation). In those chosen fields we will transmit only the header (control bits) of our protocol. That is how we will use the network steganography technique. The header consists of 6 bits per packet (as will be described in details in Section 3), so such a type of the transmission is potentially hard to discover.

## 2.2 Watermarking: the imperceptible information

Digital watermarking is a multidisciplinary methodology widely developed in the last decade. It covers a large field of various aspects, from cryptography to signal processing, and is generally used for marking the digital data (images, video, audio or text). There are several applications for the digital watermarks, described in [1] and [2], that include: **fingerprinting** (embedding a distinct watermark into every copy of the author's data), **annotation watermark/content labelling** (embedding information, which describes the digital work that can be later extracted) and **usage control/copy control** (authors can insert a watermark that indicates the number of copies permitted for each user). However, the most important applications for our purposes are: the possibility of embedding the **authentication and integrity watermark** and exchanging additional information inside this watermark.

The audio digital watermark that will be used in the proposed here authentication and integrity solution must possess certain parameters like: robustness, security, transparency, complexity, capacity, verification and invertibility. Those parameters are described in details in [1] and [2]. Their optimization for real-time audio system is crucial. Additionally one has to take into consideration that they are often mutually competitive, so there is always a compromise necessary. That is why the embedded watermark, that we will use, should be characterized by **high robustness, high security** and must be **non-perceptual**. Not every audio watermarking technique is applicable for our solution. IP Telephony is a demanding, real-time service. That is why we can apply the watermarking schemes that really work for the real-time conversations. Such algorithms are described e.g. in [2] and [4].

Generally, watermarking algorithm consist of two phases: first is **embedding** of the watermark into the voice at the source and then its **extraction** at the destination. In IP Telephony we can also distinguish those phases: as soon as the conversation begins, certain information is embedded into the voice samples and sent through the communication channel. Then, the digital watermark is extracted from voice stream before it reaches the callee. After that the retrieved information is verified. If the watermark's data sent is correct, the conversation can be continued.

Most digital watermarking algorithms for the real-time communication are designed to survive the typical non-malicious operations like: low bit rate audio compression, codec changes, DA/AD conversion or packet loss. For example, in [2] the watermarking scheme developed at the Fraunhofer IPSI (Institut Integrierte Publikations und Informationssysteme) was tested for different compression methods. Results revealed that the large simultaneous capacity and robustness depend on the scale of the codec compression. When the compression rate is high (1:53), the watermark is robust only when we embed about 1 bit/s. With a lower compression rate we can obtain about 30 bit/s, whereas the highest data rate was 48 bit/s with good robust, transparent and complexity parameters. Moreover for the monophonic audio signal, which is a default type for the IP Telephony the watermark embedding algorithm appeared around 14 times faster and the watermark detector almost 6 times faster than the real-time one.

The next important thing for proposed here scheme is how much information we can embed into the original voice data. This will influence the speed of the authentication and integrity process throughout the conversation. This parameter is

expected to be high but it is not crucial in our solution. With low compression rates we propose to add a pre-conversation stage. In this stage there will be few seconds of the RTP packets exchange without the conversation. It will delay the setup of the call but then, during the conversation, the time of the watermark verification will be shorter. However, the lowest payload watermarks (about 1 bit/s) cannot be accepted in our solution because in this case the conversation would have to last enormously long to work correctly.

### 3. Possible covert channels in VoIP based on Information Hiding

In Section 1 we presented our previous ideas that used Information Hiding techniques to create covert channels to:

- I. Secure media stream along with the signalling protocol's messages exchanged during the initial phase of the call. This mechanism is described in [16] and uses only audio digital watermarking technique.
- II. Secure conversation and to functionally alternate RTCP protocol with use of the steganography and audio digital watermarking as described in [15].

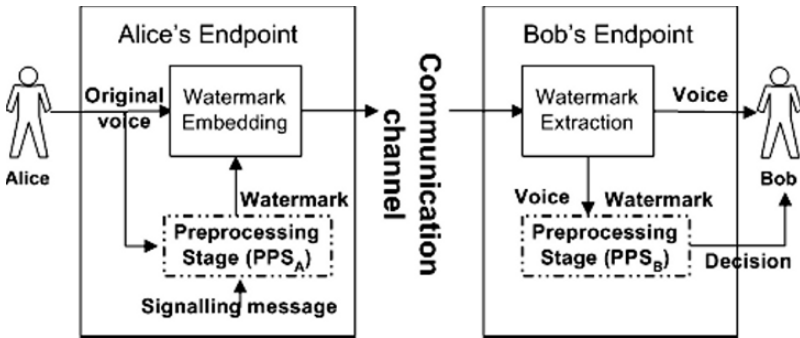
In this paper we will combine those two approaches to achieve conversation and signalling protocol security with use of the both information hiding techniques and a universal, secure channel to exchange additional data (e.g. for RTCP parameters).

So this new, steganographic protocol utilizes covert channel that consists of two subchannels: one created by using digital watermarking, second using network steganography.

For the IP Telephony system the most important security services are: **authentication, integrity and confidentiality**. We must emphasize that the first two can be provided with the use of our protocol. The third should be guaranteed in a different manner, e.g., with the use of the security mechanisms from a classical security model (the cryptographic mechanisms).

#### 3.1 General protocol overview

The solution presented here requires modifications of the general watermarking system presented with the continuous line in Figure 2. We are proposing to add a new functional block called Pre-processing Stage (**PPS**) which is marked in this figure with the dotted line. It will be responsible for preparing (processing) data before the watermark embedding stage.



**Fig. 2.** Modified watermarking scheme with the new Preprocessing Stage (PPS) block for voice transmission from Alice to Bob

The mechanism works as follows: we provide a signalling message and a sample of the caller’s original voice at Alice’s endpoint as an input to the PPS block in the transmitter. The way the PPS block processes information is covered in [16]. After the digital watermark is embedded and sent through the communication channel, the information in the receiver is retrieved and verified in an analogous block in the other endpoint. If the retrieved information is correct, the connection will continue.

But as we circumscribed earlier we will also use the network steganography technique to create an additional covert channel that will be used to transmit header (control bits). The control bits will be used to distinguish the parameters sent. In this case the digital watermarking will be used only to carry the appropriate parameters (user data, security information and additional parameters).

Additionally, to simplify characterization of the presented solution we assume that our solution will be used in the IP protocol version 4 networks [14].

### 3.2 Protocol Data Unit description

The protocol we are proposing here possesses PDU (Protocol Data Unit), the size of which must be kept to minimum. It is important, because as we said in the Section 2, the capacity of the watermark is limited (if we want watermark also to possess simultaneously other parameters like robustness or security).

Every PDU consists of a header (control bits) and a certain number of data bits that are embedded into the sender/receiver voice. The header/control fields are transmitted in a covert channel in unused/optional fields of the IP/UDP/RTP protocol’s headers. The actual value of the data is embedded into voice as a digital watermark. The header bits are organized in fields as shown in the Table 1.

**Table 1.** The header fields and their function

Type of field	No. of bits	Function
P (Parameter)	4	Indicates the parameter that is transmitted inside the watermark
S (Side)	1	Indicates the side of the communication (1 - sender, 0 – receiver)
C (Continuity)	1	Indicates if a packet contains the beginning or continuation of the parameter indicated in the field P (1 – beginning of new parameter, 0 – continuation of the last parameter)

As the capacity of the watermark depends greatly on the codec’s compression rate, so it is possible that one parameter can be distributed into a number of IP packets. The size (number of bits) of each parameter that will be transmitted with described here protocol should be low. We assume that all the parameters should not exceed 32 bits. This is a totally subjective choice based on average watermark’s capacity. However we do not dictate this value. It should depend on the network bandwidth, status and codec’s compression rate. The exemplary values of the field P are shown below:

- 0001 – authentication or integrity parameter (32 bits)
- 0010 – informational parameter 1 (32 bits)
- 0011 – informational parameter 2 (32 bits)
- 0100 – informational parameter 3 (32 bits)
- 0101 – post authentication and integrity parameter (32 bits)
- ...

The number of bits used to indicate the parameters (field P) can be changed. In the above proposition each parameter is identified by 4 bits, which allows to define 16 parameters. If there is need for less number of parameters then the number of bits can be decreased.

If we enclose the information of the side of the communication as well (field S) we can exchange information not only about the data we send but also about the data we receive.

The PDU can have one of the two payload types: **security** or **informational**. The security payload means that the PDU contains certain authentication and/or integrity information that should be verified after its extraction. Two kinds of the security payloads are available, the first is used to provide authentication and integrity of the voice, its source and signalling protocol messages. The role of the second one is to authenticate the protocol’s parameters that were sent earlier (both the security and the informational ones). Details about the security payload and cryptographic operations in this protocol will be covered in Section 3.3.

Another payload type is the informational one. Each PDU carries one parameter’s data (the whole parameter or only part of it). The description of the PDU is also illustrated in Figure 3.

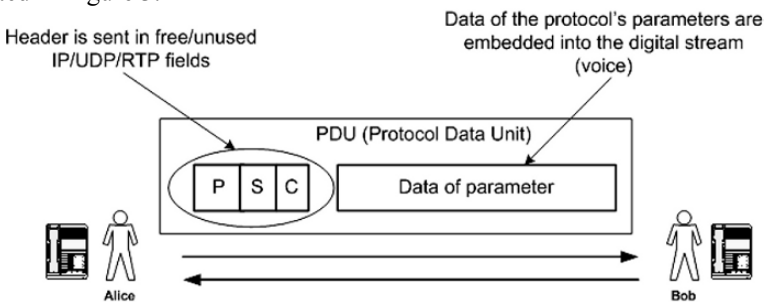


Fig. 3. Description of the Protocol Data Unit

We assume that we want to use suggested solution to improve the conversation and signalling protocol security simultaneously exchange other information too. We will not describe the form of the information parameters here as the exemplary ones based on RTCP protocol are circumscribed in [15]. Generally, we can use this cover channel to exchange any information for the VoIP system that is necessary – it can be

connected with QoS or with other aspects of the call. In this paper we will focus only on the security parameters.

Usually, in one IP/UDP/RTP packet there are about 20-30 milliseconds of voice, which is about 20-30 bytes, depending on the type of the codec used. Supposing that we are able to embed on average about 10 bits/s of the watermark into the voice stream, we must send more than 3 packets to achieve those 10 bits. In this protocol we set parameter's value to 32 bits, so it will be transmitted in about 9-12 packets in more than 3 seconds of the voice signal. In the example scenario in Fig. 4, we see how the exemplary parameter is transmitted (for assumption: 10 bits of watermark per packet).

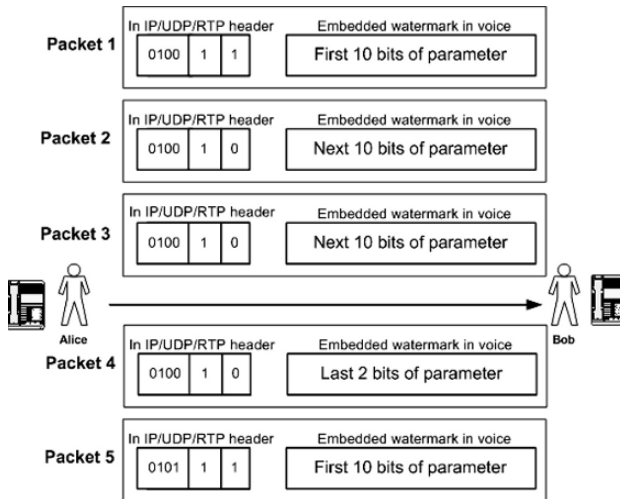


Fig. 4. Exemplary protocol operation (for assumption: 10 bits of watermark per packet)

As we can see in the above figure, the parameter characterized by the code 0100 was sent in four IP/UDP/RTP packets. In the first packet both fields S and C were set to 1. In the next packet field C changed its value to 0 because it is a continuation of the parameter's data that was sent in the last packet. At the destination there must be a buffer to extract all data from each packet. After transmitting all the packets for one parameter, the data is available to be used (if this is informational parameter) or to be verified (for security reasons).

### 3.3 Authentication and integrity parameter calculation and security payload

In Section 3.2 we mentioned that two security payloads are available:

- One is used to provide the authentication and integrity of the voice, its source and signalling protocol that is used in a particular VoIP system
- Second is to authenticate the protocol parameters (both the security and informational ones) that were sent earlier. The authentication and integrity calculation will be performed similarly as described in [6] but with the watermark specific considerations.

The first security parameter is created in a following way: hash function (H) is performed on the certain voice sample (VF) and on signalling protocol messages stored in a special buffer (inside the PPS block marked in Figure 1). Then this value is concatenated with the global identifier of caller (IDX), a preshared password (PASS), and, eventually, the random value (R) and the time stamp (TS). Using the last parameter is optional, because it requires tightly synchronized clocks. However, it is useful since it can protect against the replay attacks. After that, the hash function is performed again. The result, which we will call a **token**, is embedded as a watermark into the voice content. The token for calling party A is shown below:

$$TokenA_N = H \left( H(VF_N) \parallel H(SM_N) \parallel \begin{pmatrix} TS \\ PASS \\ ID_A \end{pmatrix} \parallel R \right) \parallel R \quad (1)$$

On the other side of the communication channel, before the caller’s voice reaches the callee, the token from the watermark must be retrieved and verified. This can be done because the callee computes analogous token locally, and then the two tokens are being compared.

We assumed earlier that each parameter transmitted will consist of 32 bits. So, if the token exceeds this value, there will be additional hash function performed. Then, only the predetermined, chosen bits will be transmitted as a security parameter.

The second security payload is a special purpose parameter that will be used to improve security of the protocol internally and the transmission. The general idea of its calculation is presented in Figure 5.

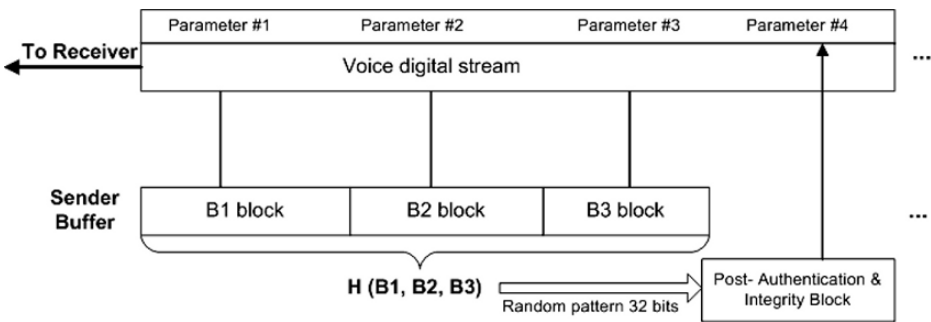


Fig. 5. Example of authentication and integrity mechanism for transmitted parameters

First, we must emphasize that during a conversation (RTP packets flow) there will be constant two-way exchange of a certain sequence of parameters. Those parameters can be susceptible to e.g. modifications or other attacks. To prevent this situation every *n-th* parameter is used to authenticate and provide integrity of *n-1* parameters that were transmitted earlier.



For the situation in Fig. 5.  $n=4$ , three parameters that contain informational or first kind of the security payload are stored in the sender buffer. After they are all in place (B1, B2 and B3 blocks), the hash function (H) can be calculated, if the result value is too long. Since we assumed certain parameter length, we have to choose only 32 bits from the hash to be transmitted. For every conversation this pattern, in which the bits are chosen, should be changed and its determination should be set and sent in a signalling phase of the connection.

Additionally, we assume that we use the mechanism of LoT (Level of Trust) value described in [15] and [16]. This is because the attacker can disrupt the transmission of the header/controls bits or because the poor network conditions can cause the situation in which the receiver is unable to retrieve any parameters that were transmitted by the sender. That is why both parties of the conversation will update special parameter named LoT (Level of Trust) during the conversation. If a parameter (security or informational) is received and verified, the LoT value increases. In any other situation its value decreases. Additionally, the parameters that are exchanged during the conversation influence the LoT value differently. The informational parameters add/subtract to the LoT's value 1, the first kind security parameters 2 and the second kind security parameter 5. The breakage of the call (or notification to the calling parties) will take place if the value of the LoT parameter is equal or below the given threshold or if the certain timer expires.

## 4. Conclusions

The protocol that uses a covert channel for the VoIP transmission was presented. It is based on the information hiding techniques: network steganography which helps to pass the header (control bits) and audio digital watermarking to transmit the actual parameter's data in a voice stream. The most important advantages of our solution are that it does not consume available, transmission bandwidth and that it improves IP Telephony's security.

What we want to emphasize is that the process of sending information for this protocol is continuous in time and, although the bit rate per second offered by the audio digital watermarking and network steganography is usually not very high we are able to exchange quite an amount of the data, if the whole conversation is considered.

The variety of different kinds of the parameters that can be used in our solution is not limited to the security ones. That is why this protocol can be freely extended to other data (informational data) that can be helpful for the VoIP connections. Finally, by using the presented solution one can gain: the signaling protocol and conversation security as well as a secure multipurpose channel for exchanging other parameters.

## References

- [1] J. Dittmann, A. Mukherjee, M. Steinebach: "Media-independent Watermarking Classification and the need for combining digital video and audio watermarking for media authentication", Proceedings of the International Conference on Information Technology: Coding and Computing, IEEE Computer Science Society, Las Vegas, Nevada, USA (2000) 62-67
- [2] M. Steinebach, F. Siebenhaar, C. Neubauer, R. Ackermann, U. Roedig, J. Dittmann: "Intrusion Detection Systems for IP Telephony Networks, Real time intrusion detection symposium, Estoril, Portugal (2002) (17)1-9"
- [3] D. Richard Kuhn, Thomas J. Walsh, Steffen Fries: "Security Considerations for Voice Over IP Systems, Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology (2004)"
- [4] S. Yuan, S. Huss: "Audio Watermarking Algorithm for Real-time Speech Integrity and Authentication, International Multimedia Conference Proceedings of the 2004 Multimedia and security workshop on Multimedia and security, Magdeburg, Germany (2004) 220 - 226"
- [5] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson: "RTP: A Transport Protocol for Real-Time Applications, IETF, RFC 3550, July 2003."
- [6] S. Miner and J. Staddon. "Graph-based authentication of digital streams. In Proceedings of the IEEE Symposium on Research in Security and Privacy (2001) 232-246"
- [7] S. J. Murdoch, S. Lewis: "Embedding Covert Channels into TCP/IP. Information Hiding (2005) 247-26"
- [8] K. Ahsan, D. Kundur: "Practical Data Hiding in TCP/IP. In: Proceedings of Workshop on Multimedia Security at ACM Multimedia '02, Juan-les-Pins (on the French Riviera), December 2002"
- [9] Petitcolas, F., Anderson, R., Kuhn, M.: "Information Hiding – A Survey: IEEE Special Issue on Protection of Multimedia Content, July 1999"
- [10] K. Szczypiorski: "HICCUPS: Hidden Communication System for Corrupted Networks. In Proc. of: The Tenth International Multi-Conference on Advanced Computer Systems ACS'2003, October 22-24, 2003 Międzyzdroje, Poland, pp.31-40, ISBN 83-87362-61-1"
- [11] T. Friedman, R. Caceres, A. Clark: "RTP Control Protocol Extended Reports (RTCP XR), IETF, RFC 3611, November 2003"
- [12] V. Korjik, G. Morales-Luna: "Information Hiding through Noisy Channels, Proceedings of 4th International Information Hiding Workshop, Pittsburgh, PA, USA, (2001) 42-50"
- [13] M. K. Mihcak, R. Venkatesan: "A Perceptual Audio Hashing Algorithm: A Tool For Robust Audio Identification and Information Hiding, Proceedings of 4th International Information Hiding Workshop, Pittsburgh, PA, April 2001."
- [14] Information Sciences Institute University of Southern California: "IP (Internet Protocol), IETF, RFC 791, September 1981."
- [15] W. Mazurczyk, Z. Kotulski: "New security and control protocol for VoIP based on steganography and digital watermarking - Informatyka - Badania i Zastosowania (IBIZA 2006), Kazimierz Dolny 9-11 February 2006"
- [16] W. Mazurczyk, Z. Kotulski: "New VoIP traffic security scheme with digital watermarking - In Proceedings of SafeComp 2006, Lecture Notes in Computer Science 4166, pp. 170 - 181, Springer-Verlag, Heidelberg 2006"