The National Network-Centric System and its components in the age of information warfare

Z. Piotrowski
Telecommunications Institute, Faculty of Electronics,
Military University of Technology, Poland

Abstract

This paper describes the new model of information warfare system based on a data hiding platform. This concept is based on the assumption that data hiding technology is the most effective mechanism for creating the new inter-layer, called the stego layer, in the known Internet Protocols. The main idea is to create network botnet bases on standard client-server architecture using specific agents – Hidden Protocol Interpreters (HPI). The HPIs are responsible for bidirectional hidden and secret communication with the administrator's workstation. The HPI serves as a passive monitoring station of the Internet infrastructure, intercepts points for remote control commands and finally plays the role of an active element in order to take control of allocated IP mechanisms (computers, routers, etc.). In this paper, a likely new attack on telecommunications links called a voice spoofing attack, in the form of, for example, a subscriber's voice impersonation, and a method of protection against this attack – Personal Trusted Terminal (PTT) – is also described.

Keywords: NNS, voice spoofing attack, Hidden Protocol Interpreters (HPIs), steganographic router, Personal Trusted Terminal (PTT), stego botnet.

1 Introduction

The National Network-Centric System (NNS) is the concept of the new information platform used for the hidden and secure transmission of additional information in known network environments: Internet, VHF/HF radio, WLAN, PSTN, GSM, etc. This concept assumes that the NNS plays a role as an additional TRANSEC and COMSEC security mechanism in existing well-known standard protocols. We propose to use additional technology – digital

WIT Transactions on the Built Environment, Vol 108, © 2009 WIT Press
www.witpress.com, ISSN 1743-3509 (on-line)
doi:10.2495/SAFE090291

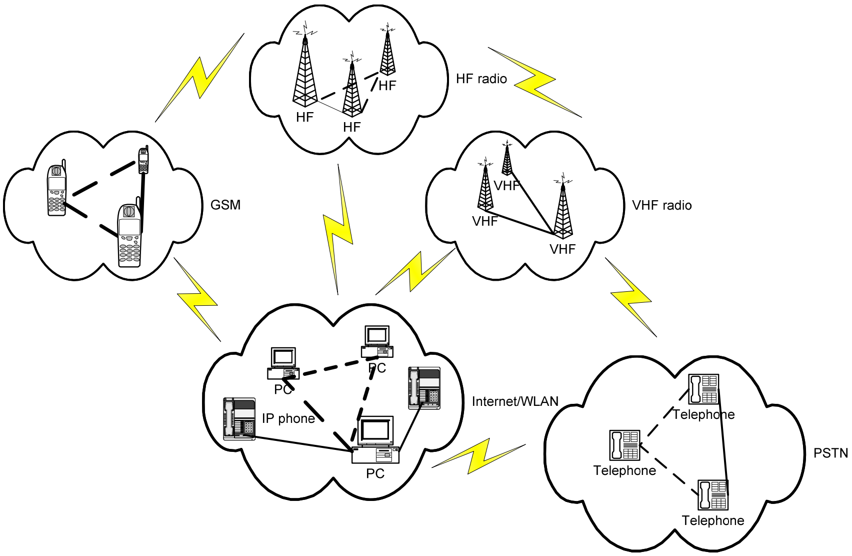

Figure 1: NNS - functional scheme.

watermarking – to hide another layer in the standard protocol. In Figure 1 we propose the base scheme of this system.

The main idea the NNS is as follows: all known network elements can be integrated into one, forming a coherent environment base for the digital watermark technology.

The main features of the NNS are:

- digital watermark perceptual transparency of the additional signal hidden in the original host signal
- statistic similarity
- dedicated data payload
- dedicated crypto power
- robustness against noise and artefacts in the transmission channel
- robustness against intentional attacks
- message integrity verification

A detailed experimental scheme of the proposed base NNS elements is presented in Figure 2.

1.1 Personal Trusted Terminal

The Personal Trusted Terminal (PTT) is the key element of the NNS. This element can be used for remote control of the network elements. The main features of the PTT are as follows:

- hidden and secure data communication in GSM links
- hidden and secure data communication in PSTN telephony

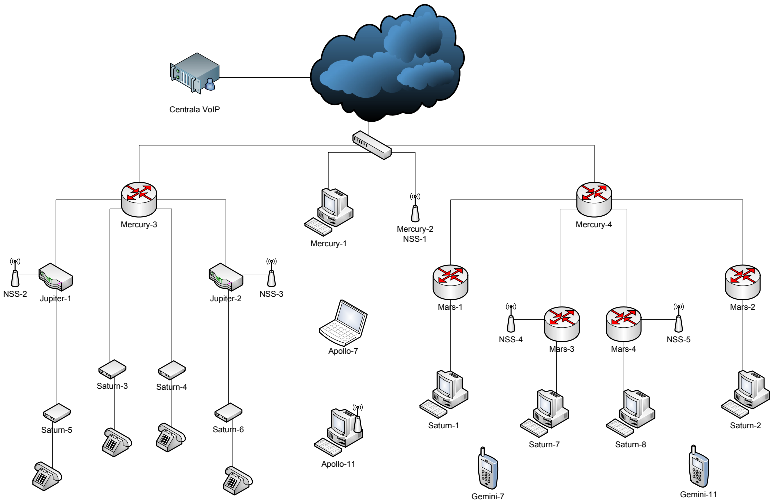

Figure 2: Detailed scheme of the NSS base elements.

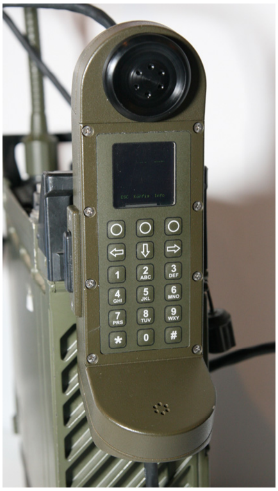

Figure 3: Personal Trusted Terminal (PTT).

- hidden and secure data communication in HF/VHF radio links
- hidden and secure data communication in VoIP telephony
- voice integrity verification

At present, the PTT is used for hidden data communication (subscriber authorisation and voice integrity verification) [1] in HF/VHF radio links. Figure 3 presents the PTT.

1.2 Hidden Protocol Interpreters

It is assumed that the NNS will operate on another key element – Hidden Protocol Interpreters (HPIs). The HPI is a hidden network controller based on an implemented and built-in digital watermark coder and decoder as well as having the main rules of interpretation hidden layer – watermark bits sequences – built-in. The HPI is allocated in the network infrastructure and can be considered as a hidden agent ready for receiving and transmitting hidden commands during the voice transmission.

2 Voice spoofing attack – subscriber's voice impersonation

Until now, we have not seriously considered the potential dangers of artificial speech synthesis in modern voice communication using Public Switched Telephone Networks (PSTNs), the Global System for Mobile Communications (GSM) or Voice over Internet Protocol (VoIP) standard telephony. Homer Dudley, a Bell Labs engineer, patented his solution "voice coder" in 1935. Since then many improvements have been made, completely rebuilding Dudley's vocoder prototype [3], but the idea of the vocoder work is always the same: voice parameterization as well as synthesis. Nowadays, vocoders are able to almost reproduce perfectly the speech signal of any person using predefined specific characteristics of the human voice.

In this section we try to answer a fundamental question concerning the public telephony domain: can a "voice spoofing attack", defined by us as a voice impersonation attack using a fake voice, be treated as a serious problem in telecommunication, and if yes – are we ready to change subscribers' habits to counter voice impersonation in telephone calls?

The two main technologies, speech synthesis and data hiding, both of which are from the Digital Signal Processing domain, have made very fast progress during the last few years. In the voice spoofing attack scheme they play the main roles of:

- means of attack – speech synthesis and
- means of defense – data hiding

In the context of active attacks and defense on telecommunications links they can play more serious roles than at present.

2.1 Speech synthesis technology improvements

Speech synthesis is a very popular technique used in gsm phones and its hybrid vocoder implementations are built into gsm terminals. We can assess everyday quality and intelligibility of synthesized speech using different gsm model terminals and generally we accept speech synthesized by those mobile devices. The European Telecommunications Standards Institute (ETSI www.etsi.com) gsm vocoder standards were changed from the very popular:

- GSM 06.10 RPELTP (Regular-Pulse Excitation-Long Term Prediction-Linear Predictive Coder) with a coding rate of 13.2 kbps (better known as GSM Full Rate (FR) standard) through to the

- GSM 06.20 Half Rate (HR) standard with VSELP (Vector-Sum Excited Linear Prediction) algorithm as well as the
- GSM 06.60 Enhanced Full Rate (EFR) possessing coding a scheme, the so-called Algebraic Code Excited Linear Prediction Coder (ACELP) and finally
- GSM 06.90 Adaptive Multi Rate (AMR) vocoder with the variable coding rates of the bit stream of 4.75, 5.15, 5.90, 6.70, 7.40, 7.95, 10.2 or 12.2 kbps.

We can observe that every time a new standard is issued, it encompasses better quality and intelligibility of speech synthesis using lower bit rates and more advanced algorithms, but consuming much more computational power and of course energy within the terminal.

It is not only the standards that should be considered as an indicator of speech synthesis progress, in fact we can also observe the growing role of "not-certified" synthesizers in popular text-to-speech applications. The famous text-to-speech (TTS) contest Blizzard Challenge [9] organized by Carnegie Mellon University [10] consists of taking publicly available speech databases and building fully synthetic voices. In this contest unknown text sentences from independent sources are generated and each participant synthesizes those sentences with their system. The speech is evaluated and assessed using a Mean Opinion Score (MOS) speech quality measurement test. The contest has taken place annually since 2005 and many new TTS systems have been evaluated. One of the most successful winners was the IVONA TTS system, which can almost "perfectly" synthesize the human voice using a Unit Selection with a Limited Time-scale Modification (USLTM) algorithm [6]. We can find the basic description of this algorithm in the Blizzard Challenge pages [11]. From the normal TTS application user's of the point of view, we are interested in using such algorithms for:

- efficient and human-like fluent reading of text from electronic books,
- speech tracks for automatic telecommunication auto responders,
- Internet text page reading, etc.,

but from the point of view of telecommunication hackers it could be perceived as yet another tool for telephonic pranks and mayhem. When will the boundary between progress in technology of synthesis and voice spoofing be overstepped? We can predict that it depends on the availability of perfect or almost perfect voice synthesis of a specific person, i.e., a real telephone subscriber rather than a virtual, artificial speaker.

Phone hackers have learned how to use the Public Switched Telephone Network (PSTN) signalization protocol named Signaling System #7 (SS7) to simulate or to "spoof" caller ID numbers [12]; thus, we are able to observe a large number of "caller ID spoofing" attacks. Why not to try to spoof a specific subscriber's voice? The changing of the caller ID number is legal according to the SpoofCard, a US company [13] that uses caller ID number changing for business purposes. Thus, when you receive a call from a well-known number assigned to, e.g., a bank listed in your local phone book, you can never be sure if the indicated number and the name are truly connected. Care is especially needed because an additional feature offered is the substitution of the subscriber's voice transformation in real-time. Thus, the subscriber can change the voice gender to male or female as required. Another example of basic “voice spoofing” implementation is the FoneFaker service [14], which offers to change your voice as well as your caller ID number. Thus, we can observe the first attempts of a “voice spoofing” attack in voice call links. Of course, changing your voice to the opposite gender as an unidentified person is a very primitive technique, (in telecommunications terminology – a subscriber’s ‘stunt’), but there will probably be more sophisticated and advanced attempts in the future. To counter such misrepresentation a receiving subscriber should have the possibility of checking the real identity of the calling subscriber.

2.2 Human Auditory System

Using our Human Auditory System (HAS) [4], it is sometimes possible to receive the illusory feedback that the calling subscriber is exactly the person we think it is. It is much easier to appreciate the imperfection of our HAS when listening to music on our mp3 players. To record music in mp3 format, time-frequency psychoacoustic phenomena is used to deceive our HAS. The standard algorithm computes time masking dependencies and the so called Minimum Masking Threshold, which are used to cancel all irrelevant and redundant time and frequency components in the music. The algorithm is used to allocate bits only to the essential music components [15]. When we play music at the rate of 128 kbps we probably do not hear any difference in comparison with the original music, but when the music is played in 16 kbps it will sound like “torture”. It is much easier to authenticate “spoofed” voices over telecommunication links because speech quality transmitted over links is not perfect (the so-called “communications quality”). Even if the impersonated, faked voice is far from the original one, it is still possible to mistakenly recognize the faked voice because of problems with low link quality. There are various tests for assessing the two basic sound features: quality and fidelity, the main representative test being MOS [16, 17]. A low percentage of people have so-called “golden-ears” in that they possess special talents in hearing and they are ideal for subjective assessment tests where hearing sensitivity plays the main role. When two “same” sound stimuli are presented they can easily recognize which one is the original and which is the distorted fake. Of course, in this fidelity test the fixed level of sound quality should be taken into account. When the same test is performed over communications links the probability of difference previewing is almost close the zero. The influence of sound quality on intelligibility are well documented in the [5, 18].

2.3 Voice authorization and message integration ID in NNS using PTT

There are potentially two features in new “voice spoofing attack” using voice calls: voice impersonation by changing distinctive speech features of a specific person and changing the meaning of the statements by expressing words differently.

Could we stand up to a "voice spoofing attack" in our telephone calls right now? When we consider the speaker verification classifications proposed by Markowitz [7] we can use the following methods:

- text dependent based on passwords,
- text prompted based on challenge-response,
- text independent based on free speech.

All of these methods are based on *a priori* knowledge of the subscriber, who often may be personally unknown to us. Sometimes it is impossible to have this *a priori* knowledge, for example during a telephone interview with a Very Important Person. This is one reason why authorized interviews for radio and newspapers are very rarely carried out using telephone voice calls. The PTT proposed method of authorization checking and message integrity verification is based on the data hiding technology named watermarking or steganography, depending on the purposes of use [8]. Generally, a data hiding technique is used to hide additional information represented by a 'watermark' signal in the presence of an original host.

The watermark is perceptually transparent and inaudible in the host signal's presence and is robust against intentional and unintentional attacks. For unintentional attacks we can additionally consider Additive White Gaussian Noise (AWGN), resampling, requantization, all-pass filtering and loss compression, i.e. all the signal processing stages in telecommunication links. For intentional attacks we can counter with erasing, multiple copying and embedding, cropping, stretching, etc. The idea is to send, together with the speech signal, a Personal Identification Number (PIN) assigned to the specific subscriber using encryption based on a single use key so that the PIN number changes during each call session. At the receiver's location the watermark is decoded and the received binary signature (PIN) is compared to the database in the handset. When the received PIN is checked as conforming to the original one, the subscriber is authorized during connection (call). In this case the receiving handset's LCD display indicates the confirmed identity of the sender. In cases when only one subscriber has the handset with the authorization function and the other one has standard handset, conversation is possible but no message about authorization can be displayed. Generally, the PTT is primarily for use in interaction with military battlefield radio stations (VHF and HF radio bands), but it can be of commercial use after changing the handset interface in Internet telephony (Voice Over IP), public switched telephone networks (PSTN) and cellular telephony (GSM).

We propose the idea of the PTT as a trusted handset with a hidden authorization function of the correspondent in telecommunications links. The handset would be connected to a radio station or terminal sending hidden PIN numbers through the radio link together with the conversation signal to authorize the subscriber's communication.

There is another idea when a watermark, representing information about the original, host voice integrity, is transmitted; it can be easily verified, at the receiver side, by comparison of the decoded, extracted watermark as a binary signature from the received signal with another binary signature computed from the received signal. If these two binary signatures are equal it means that the received voice message is not edited. For continued security of conversation the recipient needs to make sure that the voice integrity marker is on and the PIN number is correct (AUTH OK message is displayed), because otherwise it is possible for a substituted conversation to take place with automatic or semi-automatic synthesizers, revealing important information to unauthorized third-party subscribers.

The proposed method can be used for voice secure ciphered links with fixed ciphering power, because the potential voice spoofing attack does not operate on transmission and communication security levels and does not require interaction with internal infrastructure of the link.

2.4 The need for the new security standard for voice calls

Until now, many protocols were developed and established as standards in communications networks focusing on the security and protection aspects of voice calls. The most representative are the SS7 for PSTN, defined in 1987 by ITU-T in recommendations series Q.7xx [19], the well-known H.323 for VoIP, first recommendation defined in 1996 [20] and the relatively new Session Initiation Protocol (SIP) for VoIP, defined by the Internet Engineering Task Force (IETF) [21].

None of the above detailed “mature” standards enable a basic level of security of voice calls in the context of a “voice spoofing attack”. This is because the source information, the voice, is usually allocated by us to the specific person in a subjective manner.

There is an urgent need to devise additional security mechanisms for authorization and message integrity verification at the source – the voice input-output sockets. At the very beginning of this long trip we proposed a mechanism based on watermarking technology – only for dedicated, specific users. When we think about distributing secure key mechanisms they can be commercially available, for example in the form of the well-known pop-up cards; if you did not enter your single use key for the voice session you would be not allowed to authorize your own voice or your own voice messages.

3 Conclusions

The NNS is still in the experimental phase of tests. Until now the main elements of the NNS developed are the PTT and the HPI. The described voice spoofing attack can be ineffective when the PTT is used. The PTT can also be used for remote control in a NNS environment. Hidden data transmutation using the PTT during conversation in unprotected links from hidden communication can be interpreted by HPIs. The proposed system is the new network botnet concept based on the assumption that all local networks (radio, Internet, PSTN) can be integrated into one easily managed digital environment. The NNS idea seems to fulfil all requirements for network-centric mechanisms and even upgrades these into the new security level – the hidden information layer.

References

[1] Piotrowski Z., Gajewski P.: Novel method for watermarking system operating on the HF and VHF radio links, Computational Methods and Experimental Measurements XIII, CMEM XIII, , Southampton, Boston, WIT Press, pp.791-800, 2007

[2] Richard G. Lyons, Understanding Digital Signal Processing, Prentice Hall PTR Publication, 2004

[3] H. W. Dudley, *The vocoder*, Bell Labs Rec., vol. 18, pp. 122-126, 1939, Reprinted in [R. W. Schafer and J. D. Markel, eds., *Speech Analysis*, New York: IEEE Press, 1979, pp. 347-351]

[4] H. Fastl, E. Zwicker, *Psychoacoustic: Facts and Models,* Springer-Verlag, Berlin Heidelberg, 1990

[5] T. Houtgast and H. J. M. Steeneken, A multi-lingual evaluation of the Rasti-method for estimating speech intelligibility in auditoria. Acustica, 54:185--199, 1984.

[6] M. Kaszczuk, L. Osowski, *The IVO Software Blizzard 2007 Entry: Improving Ivona Speech Synthesis System*, 6th ISCA Workshop on Speech Synthesis, Bonn, Germany, August 25, 2007 http://www.festvox.org/blizzard/bc2007/blizzard_2007/blz3_010.html

[7] J. Markowitz, *Anti-Spoofing Techniques for Voice*, September 2005, www.jmarkowitz.com

[8] J. Cox, M. Miller, J. Bloom, J. Fridrich, T. Kalker, *Digital Watermarking and Steganography,* Morgan Kaufman Publishers, 2008

[9] http://www.festvox.org/blizzard/index.html

[10] http://www.lti.cs.cmu.edu/

[11] http://www.festvox.org/blizzard/bc2006/ivo_blizzard2006.pdf

[12] http://en.wikipedia.org/wiki/Caller_ID_spoofing

[13] www.spoofcard.com

[14] http://www.brickhousesecurity.com/telephone-voice-changers.html

[15] http://www.petitcolas.net/fabien/software/mpeg/index.html

[16] ITU-T Recommendation P.830 (02/1996) *Subjective performance assessment of telephone-band and wideband digital codecs*

[17] ITU-R BS1116-1 standard http://www.itu.int/rec/R-REC-BS.1116-1-199710-I/e

[18] http://cslu.cse.ogi.edu/HLTsurvey/ch13node11.html

[19] http://www.itu.int/rec/T-REC-Q/e

[20] http://www.itu.int/rec/T-REC-H.323-200606-I/en

[21] http://tools.ietf.org/html/rfc3261